On adaptive stratification

Pierre Etore · **Gersende Fort** · **Benjamin Jourdain** ·
Eric Moulines

Abstract This paper investigates the use of stratified sampling as a variance reduction technique for approximating integrals over large dimensional spaces. The accuracy of this method critically depends on the choice of the space partition, the *strata*, which should be ideally fitted to the subsets where the functions to integrate is nearly constant, and on the allocation of the number of samples within each strata. When the dimension is large and the function to integrate is complex, finding such partitions and allocating the sample is a highly non-trivial problem. In this work, we investigate a novel method to improve the efficiency of the estimator "on the fly", by jointly sampling and adapting the strata which are hyperrectangles and the allocation within the strata. The accuracy of estimators when this method is used is examined in detail, in the so-called asymptotic regime (*i.e.* when both the number of samples and the number of strata are large). It turns out that the limiting variance depends on the directions defining the hyperrectangles but not on the precise abscissae of their boundaries along these directions, which gives a mathematical justification to the common choice of equiprobable strata. So, only the directions are adaptively modified by our algorithm. We illustrate the use of the method for the computation of the price of path-dependent options in models with both constant and stochastic volatility. The use of this adaptive technique yields variance reduction by factors sometimes larger than 1000 compared to classical Monte Carlo estimators.

P. Etore
CMAP, École Polytechnique,
Route de Saclay, 91128 Palaiseau Cedex
Tel.: +33 (0)1 69 33 45 67
E-mail: etore@cmap.polytechnique.fr

G. Fort and E. Moulines
Institut des Télécoms, Télécom ParisTech,
46 Rue Barrault, 75634 Paris Cedex 13, France
Tel.: +33 (0)1 45 81 77 82
E-mail: surname.name@telecom-paristech.fr

B. Jourdain
Université Paris-Est, CERMICS, Projet MathFi ENPC-INRIA-UMLV,
6 et 8 avenue Blaise Pascal, 77455 Marne La Vallée, Cedex 2, France
Tel.: +33 (0)1 64 15 35 67 E-mail: benjamin.jourdain@enpc.fr

1 Introduction

A number of problems in statistics, operation research and mathematical finance boils down to the evaluation of the expectation (or higher order moments) of a random variable $\phi(Y)$, known to be a complicated real valued function of a vector $Y = (Y_1, \ldots, Y_d)$ of independent random variables. In our applications, we will mainly focus on simulations driven by a sequence of independent standard normal random variables, in situations where the dimension d is very large. Such problems arise in particular in computational finance for the pricing of path-dependent options, either when the number of underlying assets is large, or when additional source of randomness is present such as in stochastic volatility models.

The stratification approach consists in dissecting $\mathbb{R}^d$ into mutually exclusive *strata* and ensuring that ϕ is evaluated for a prescribed and appropriate number of points in each stratum (see [7], [3], [16]).

The main purpose of this paper is to discuss a way of dissecting the space into strata and sampling from the strata, adapted to the case where Y is a standard Gaussian vector. We also address the accuracy of estimators when this method of sampling is used, and give conditions upon which the variance reduction is most effective.

Our method makes use of orthogonal directions, to induce a dissection of $\mathbb{R}^d$ with the right property. These directions and the associated allocation are learnt adaptively, while the simulations are performed. The advantage of the adaptive method, similar to those introduced for importance sampling by [15] is that information is collected as the simulations are done, and computations of means and variance of $\phi(Y)$ in strata are used to update the choice of these strata and of the allocation. We investigate in some details the asymptotic regime *i.e.* where the number of simulations and the number of the strata both go to infinity.

The method is illustrated for pricing path-dependent options driven by high-dimensional Gaussian vectors, combining adaptive importance sampling based on a change of drift together with the suggested adaptive stratification. The combination of these two methods, already advocated in an earlier work by [8], is very effective; nevertheless, these examples show that, contrary to what is suggested in this work, the asymptotical optimal drift vector is not always the most effective direction of stratification.

The paper is organized as follows. In section 2, an introduction to the main ideas of the stratification is presented. Section 3 addresses the behavior of the stratified estimator in the asymptotic regime (*i.e.* when both the number of samples and the number of strata go to infinity). The roles of the stratification directions, the strata in each direction of stratification and of the allocation within each strata are evidenced. In section 4, an algorithm is proposed to adapt the directions of stratifications and the allocation of simulations within each stratum. In Section 5, the proposed adaptive stratification procedure is illustrated using applications for the pricing of path-dependent options.

2 An introduction to stratification

Suppose we want to compute an expectation of the form $\mathrm{E}\left[\phi(Y)\right]$ where $\phi : \mathbb{R}^d \to \mathbb{R}$ is a measurable function and Y is a $\mathbb{R}^d$-valued random variable. We assume hereafter that

$$\mathrm{E}\left[\phi^2(Y)\right] < +\infty \,. \tag{1}$$

We consider a *stratification variable* of the form $\mu^T Y$ where μ is an orthonormal $(d \times m)$ matrix with $m \leq d$. In all our examples, m is equal to one or two. Given a finite partition $\{\mathsf{S_i}, \mathbf{i} \in \mathcal{I}\}$ of $\mathbb{R}^m$, the sample space $\mathbb{R}^d$ of Y is divided into *strata* defined by

$$\mathsf{S}_{\mu,\mathbf{i}} \stackrel{\text{def}}{=} \left\{ x \in \mathbb{R}^d, \mu^T x \in \mathsf{S_i} \right\} \,, \quad \mathbf{i} \in \mathcal{I} \,. \tag{2}$$

It is assumed in the sequel that the probability of the strata $\{p_\mathbf{i}, \mathbf{i} \in \mathcal{I}\}$

$$p_\mathbf{i}(\mu) \stackrel{\text{def}}{=} \mathbb{P}\left(Y \in \mathsf{S}_{\mu,\mathbf{i}}\right) = \mathbb{P}\left(\mu^T Y \in \mathsf{S_i}\right) \,, \tag{3}$$

are known which is the case when Y is a standard Gaussian vector and the $S_\mathbf{i}$ are hyperrectangles. Up to removing some strata, we may assume without loss of generality that $p_\mathbf{i}(\mu) > 0$, for any $\mathbf{i} \in \mathcal{I}$.

Let M be the total number of draws and $\mathcal{Q} = \{q_\mathbf{i}, \mathbf{i} \in \mathcal{I}\}$ be an allocation vector (i.e. $q_\mathbf{i} \geq 0$ and $\sum_{\mathbf{i} \in \mathcal{I}} q_\mathbf{i} = 1$) : the number $M_\mathbf{i}$ of samples allocated to the $\mathbf{i}$-th stratum is given by

$$M_\mathbf{i} \stackrel{\text{def}}{=} \left\lfloor M \sum_{\mathbf{j} \leq \mathbf{i}} q_\mathbf{j} \right\rfloor - \left\lfloor M \sum_{\mathbf{j} < \mathbf{i}} q_\mathbf{j} \right\rfloor \,, \quad \mathbf{i} \in \mathcal{I} \,, \tag{4}$$

where $\lfloor \cdot \rfloor$ denotes the lower integer part and by convention, $\sum_{\emptyset} q_{\mathbf{j}} = 0$ (it is assumed that the set of indices $\mathcal{I}$ is totally ordered). If the number of points in each stratum is chosen to be proportional to the probability of the strata, the allocation is said to be *proportional*. Given the strata $\{\mathsf{S}_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}\}$ and the allocation $\mathcal{Q}$, the *stratified estimator* with M draws is defined by

$$
\sum_{\mathbf{i} \in \mathcal{I}: M_{\mathbf{i}} > 0} p_{\mathbf{i}}(\mu) \left\{ \frac{1}{M_{\mathbf{i}}} \sum_{j=1}^{M_{\mathbf{i}}} \phi(Y_{\mathbf{i},j}) \right\} , \tag{5}
$$

where $\{Y_{\mathbf{i},j}, j \leq M_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}\}$ are independent random variables with $Y_{\mathbf{i},j}$ distributed according to the conditional distribution $\mathrm{P}\left[Y \in \cdot \,|\, \mu^T Y \in \mathsf{S}_{\mathbf{i}} \right]$ for $j \leq M_{\mathbf{i}}$.

The stratified estimator is an unbiased estimator of $\mathrm{E}[\phi(Y)]$ if the $M_{\mathbf{i}}$'s are all positive (a sufficient condition is $M \geq 1/\min_{\mathbf{i}} q_{\mathbf{i}}$). Its variance is given by $\sum_{\mathbf{i} \in \mathcal{I}: M_{\mathbf{i}} > 0} M_{\mathbf{i}}^{-1} p_{\mathbf{i}}^2(\mu) \sigma_{\mathbf{i}}^2(\mu)$ where $\sigma_{\mathbf{i}}^2(\mu)$ is the conditional variance of the random vector $\phi(Y)$ given $\mu^T Y \in \mathsf{S}_{\mathbf{i}}$,

$$
\sigma_{\mathbf{i}}^2(\mu) \stackrel{\mathrm{def}}{=} \mathrm{E}\left[\phi^2(Y) \,\middle|\, \mu^T Y \in \mathsf{S}_{\mathbf{i}} \right] - \left(\mathrm{E}\left[\phi(Y) \,\middle|\, \mu^T Y \in \mathsf{S}_{\mathbf{i}} \right] \right)^2 . \tag{6}
$$

When M goes to infinity and the number of strata is either fixed or goes to infinity slowly enough, the variance of the stratified estimator is equivalent to $M^{-1} \sum_{\mathbf{i} \in \mathcal{I}: q_{\mathbf{i}} > 0} q_{\mathbf{i}}^{-1} p_{\mathbf{i}}^2(\mu) \sigma_{\mathbf{i}}^2(\mu)$ (see Lemma 1). The two key questions that arise in every application of the stratified sampling method are (i) the choice of the dissection of the space and (ii) for a fixed M, the determination of the number of samples $M_{\mathbf{i}}$ to be generated in each stratum $\mathbf{i}$. The optimal allocation minimizing the above asymptotic variance subject to the constraint $\sum_{\mathbf{i} \in \mathcal{I}} q_{\mathbf{i}} = 1$ is given by :

$$
q_{\mathbf{i}}^\star(\mu) \stackrel{\mathrm{def}}{=} \frac{p_{\mathbf{i}}(\mu)\, \sigma_{\mathbf{i}}(\mu)}{\sum_{\mathbf{j} \in \mathcal{I}} p_{\mathbf{j}}(\mu) \sigma_{\mathbf{j}}(\mu)} . \tag{7}
$$

For a given stratification matrix μ, we refer to $\mathcal{Q}^\star(\mu) = \{q_{\mathbf{i}}^\star(\mu), \mathbf{i} \in \mathcal{I}\}$ as the *optimal stratification vector*. Of course, contrary to the proportions $p_{\mathbf{i}}(\mu)$, the conditional expectations $\mathrm{E}\left[\phi(Y) \,|\, Y \in \mathsf{S}_{\mu,\mathbf{i}} \right]$ are unknown and so are the conditional variances $\sigma_{\mathbf{i}}^2(\mu)$. Because μ is also unknown, an adaptive procedure is required.

The simplest approach would be to estimate these conditional variance in a pilot run, to determine the optimal allocation vector from these estimates, and then to use this allocation vector in a second stage to determine the stratified estimator. Such a procedure is clearly suboptimal, since the results obtained in the pilot step are not fully exploited. This calls for a more sophisticated procedure, in the spirit of those used for adaptive importance sampling; see for example, [15] and [16]. In these algorithms, the estimate of conditional variance and the stratification directions is gradually improved while computing the stratified estimator and estimating its variance. Such algorithm extends the procedure by [6], who proposed to adaptively learn the optimal allocation vector for a set of given strata and derived a central limit theorem for the adaptive estimator (with the optimal asymptotic variance).

3 Asymptotic analysis of the stratification performance

We derive in this Section the asymptotic variance of the stratified estimator when both the total number of draws M and the number of strata (possibly depending upon M) tend to $+\infty$. The variance of the estimator depends on the stratification matrix μ, on the partition $\{\mathsf{S}_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}\}$ of the sample space of $\mu^T Y$ and on the allocation $\mathcal{Q}$.

For any integer k, we denote by λ the Lebesgue measure on $\mathbb{R}^k$, equipped with its Borel sigma-field (the dependence in the dimension k is implicit). For a probability density h w.r.t the Lebesgue measure on $\mathbb{R}$, we denote by H the cumulative distribution function, and H^{-1} the *quantile* function, defined as the generalized inverse of H,

$$
H^{-1}(u) = \inf\{x \in \{H > 0\} : H(x) \geq u\} , \quad \text{for any } u \in [0, 1] ,
$$

where, by convention, $\inf \emptyset = +\infty$. Let I be a positive integer. The choice of the strata boundaries is parameterized by an m-uplet $(g_1, \ldots, g_m)$ of probability densities on $\mathbb{R}$ in the following sense: for all m-uplet $\mathbf{i} = (i_1, \ldots, i_m) \in \{1, \cdots, I\}^m$,

$$
\mathsf{S}_{\mathbf{i}} \stackrel{\mathrm{def}}{=} \left\{ (x_1, \ldots, x_m) \in \mathbb{R}^m : G_k^{-1}\left(\frac{i_k - 1}{I} \right) < x_k \leq G_k^{-1}\left(\frac{i_k}{I} \right) \right\} . \tag{8}
$$

We denote by $g(x_1, \ldots, x_m) \stackrel{\text{def}}{=} \prod_{k=1}^m g_k(x_k)$ the associated joint density. Let μ be a $d \times m$ orthonormal matrix. We consider the stratification $\mathcal{S}(\mu) = \{\mathsf{S}_{\mu,\mathbf{i}}, \mathbf{i} \in \{1, \ldots, I\}^m\}$ of the space $\mathbb{R}^d$. Denote by $\varsigma_{I,M}^2(\mu, g, \mathcal{Q})$ the asymptotic variance of the stratified estimator, given by

$$\varsigma_{I,M}^2(\mu, g, \mathcal{Q}) \stackrel{\text{def}}{=} \sum_{\mathbf{i} \in \{1, \ldots, I\}^m : M_{\mathbf{i}} > 0} M_{\mathbf{i}}^{-1} \, p_{\mathbf{i}}^2(\mu) \sigma_{\mathbf{i}}^2(\mu) \,, \tag{9}$$

where the number of draws $M_{\mathbf{i}}$ is given by (4) and $p_{\mathbf{i}}(\mu), \sigma_{\mathbf{i}}^2(\mu)$, the probability and the conditional variance are given by (3), and (6), respectively. The dependence w.r.t. g and $\mathcal{Q}$ of $M_{\mathbf{i}}, p_{\mathbf{i}}(\mu)$ and $\sigma_{\mathbf{i}}^2(\mu)$ is implicit.

We consider allocations $\mathcal{Q}_\chi = \left\{ q_{\mathbf{i}}(\chi) \stackrel{\text{def}}{=} \int_{\mathsf{S}_{\mathbf{i}}} \chi \, d\lambda \,, \mathbf{i} \in \{1, \ldots, I\}^m \right\}$ parameterized by a probability density $\chi : \mathbb{R}^m \to \mathbb{R}_+$. We assume that the random variable $\mu^T Y$ possesses a density f_μ w.r.t. the Lebesgue measure (on $\mathbb{R}^m$). We consider the functions

$$\psi_\mu(x) \stackrel{\text{def}}{=} \mathrm{E}\left[\phi\left(Y\right) \Big| \mu^T Y = x\right], \quad \text{and} \quad \zeta_\mu(x) \stackrel{\text{def}}{=} \mathrm{E}\left[\phi^2(Y) \Big| \mu^T Y = x\right].$$

Using these notations, the asymptotic variance of the stratified estimator may be rewritten as

$$\varsigma_{I,M}^2(\mu, g, \mathcal{Q}_\chi) = \sum_{\mathbf{i} \in \{1, \ldots, I\}^m : M_{\mathbf{i}} > 0} M_{\mathbf{i}}^{-1} \left\{ \left(\int_{\mathsf{S}_{\mathbf{i}}} f_\mu \, d\lambda\right) \left(\int_{\mathsf{S}_{\mathbf{i}}} \zeta_\mu f_\mu \, d\lambda\right) - \left(\int_{\mathsf{S}_{\mathbf{i}}} \psi_\mu f_\mu \, d\lambda\right)^2 \right\}.$$

We will investigate the limiting behavior of asymptotic variance $\varsigma_{I,M}^2(\mu, g, \mathcal{Q}_\chi)$ when the total number of samples M and the number of strata I both tend to $+\infty$. For that purpose, some technical conditions are required. For ν a measure on $\mathbb{R}^n$ and h a real-valued measurable function on $\mathbb{R}^n$, we denote by $\operatorname{essinf}_\nu(h)$ and $\operatorname{esssup}_\nu(h)$ the essential infimum and supremum w.r.t. the measure ν. From now on we use the following convention : $z/0$ is equal to $+\infty$ if $z > 0$ and to 0 if $z = 0$.

A1 $\int_{\mathbb{R}^m} \chi^2/g \, d\lambda < +\infty$ and $\operatorname{essinf}_{g \cdot \lambda}(\chi/g) > 0$.

A2 for $h \in \{f_\mu, \zeta_\mu f_\mu, \psi_\mu f_\mu\}$, $\int_{\mathbb{R}^m} h^2/g \, d\lambda < +\infty$.

Under A2, λ-a.e. , $g = 0$ implies that $f_\mu = 0$. Finally, a reinforced integrability condition is needed

A3 $\int_{\mathbb{R}^m} f_\mu^4 (\zeta_\mu - \psi_\mu^2)^2 / [\chi^2 g] \, d\lambda < +\infty$.

When $m < d$, we establish the expression of the limit as the number of strata I goes to $+\infty$ of the limiting variance (as the number of simulations M goes to $+\infty$) of the stratified estimator. Define

$$\varsigma_\infty^2(\mu, \chi) \stackrel{\text{def}}{=} \int_{\mathbb{R}^m} f_\mu^2 (\zeta_\mu - \psi_\mu^2)/\chi \, d\lambda \,. \tag{10}$$

Proposition 1 *Let m be an integer such that $m < d$, $g_1, \cdots, g_m$ be probability density functions (pdf) w.r.t. to the Lebesgue measure of $\mathbb{R}$, μ be a $d \times m$ orthonormal matrix, and χ be a pdf w.r.t. the Lebesgue measure on $\mathbb{R}^m$. Assume that g and χ satisfy assumptions A1-A3. Then*

$$\lim_{I \to +\infty} \lim_{M \to +\infty} M \varsigma_{I,M}^2(\mu, g, \mathcal{Q}_\chi) = \varsigma_\infty^2(\mu, \chi).$$

Assume in addition one of the following conditions

(i) *$\operatorname{esssup}_{\chi \cdot \lambda}(f_\mu/\chi) < +\infty$ and $\{I_M, M \geq 1\}$ is an integer-valued sequence such that $I_M^{-1} + I_M^m M^{-1} \to 0$ as M goes to infinity.*

(ii) *$\{I_M, M \geq 1\}$ is an integer-valued sequence such that $I_M^{-1} + I_M^{2m} M^{-1} \to 0$ as M goes to infinity.*

Then,

$$\lim_{M \to +\infty} M \varsigma_{I_M, M}^2(\mu, g, \mathcal{Q}_\chi) = \varsigma_\infty^2(\mu, \chi).$$

The proof is given in Section 6.1. It is worthwhile to note that the limiting variance of the stratified estimator $\varsigma_\infty^2(\mu, \chi)$ does not depend on the densities $(g_1, \ldots, g_m)$ that define the strata : only the directions of stratification μ and the allocation distribution $\mathcal{Q}_\chi$ enters in the limit. The contribution to the variance of the randomness in the directions orthogonal to the rows of μ dominates at the first order. In practice, this means that asymptotically, once the directions of stratification are chosen, the choice of the strata is irrelevant; the usual choice of g_i as the distribution of the i-th component of the random vector $\mu^T Y$, $i \in \{1, \ldots, m\}$ is asymptotically optimal.

On the contrary, the limiting variance $\varsigma_\infty^2(\mu,\chi)$ depends on the allocation density χ. For a given value of the stratification directions μ, it is possible to minimize the function $\chi \mapsto \varsigma_\infty^2(\mu,\chi)$. Assume that $\int_{\mathbb{R}^m} f_\mu \sqrt{\zeta_\mu - \psi_\mu^2}\, d\lambda > 0$. Since $\int_{\mathbb{R}^m} f_\mu \sqrt{\zeta_\mu - \psi_\mu^2}\, d\lambda = \mathrm{E}\left[\sqrt{\mathrm{Var}\left[\phi(Y)|\mu^T Y\right]}\right] \leq \sqrt{\mathrm{Var}(\phi(Y))}$, the integral is finite by (1) and it is possible to define a density $\chi_\mu^\star$ by

$$\chi_\mu^\star \stackrel{\text{def}}{=} f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \Big/ \int_{\mathbb{R}^m} f_\mu \sqrt{\zeta_\mu - \psi_\mu^2}\, d\lambda\,. \tag{11}$$

Then $\chi_\mu^\star$ is the minimum of $\chi \mapsto \varsigma_\infty^2(\mu,\chi)$ and the minimal limiting variance is

$$\varsigma_\infty^2(\mu,\chi_\mu^\star) = \left(\int_{\mathbb{R}^m} f_\mu \sqrt{\zeta_\mu - \psi_\mu^2}\, d\lambda\right)^2 = \left(\mathrm{E}\left[\sqrt{\mathrm{Var}\left[\phi(Y)|\mu^T Y\right]}\right]\right)^2\,.$$

Provided $\chi_\mu^\star$ satisfies assumptions A1-2 (note that in that case, A3 is automatically satisfied), the choice $\chi = \chi_\mu^\star$ for the allocation of the drawings in the strata is asymptotically optimal.

Remark 1 *An expression of the limiting variance $\varsigma_\infty^2(\mu,\chi)$ has been obtained in [8, Lemma 4.1] in the case $m = 1$ and for the proportional allocation rule which corresponds to $\chi = f_\mu$. It is shown by these authors that the limiting variance is $\mathrm{E}\left(\mathrm{Var}\left[\phi(Y)\,|\,\mu^T Y\right]\right)$ which is equal to $\varsigma_\infty^2(\mu,f_\mu)$ (note that in this case the stratification density $g = f_\mu$, satisfies the assumptions A1-3 provided that $\mathrm{E}[\phi^4(Y)] < \infty$). Unless $\mathrm{Var}\left[\phi(Y)\,|\,\mu^T Y\right]$ is a.s. constant, the asymptotic variance is strictly smaller for the optimal choice of the allocation density.*

The optimal allocation density $\chi_\mu^\star$ cannot in general be computed explicitly but, as shown in the following Proposition, can be approximated by computing the optimal allocation within each stratum.

Proposition 2 *Let $m < d$ be an integer and μ be an $(d \times m)$ orthonormal matrix. Assume that A2 is satisfied. Then,*

$$\lim_{I \to +\infty} \sum_{\mathbf{i} \in \{1,\ldots,I\}^m} \left| q_{\mathbf{i}}^\star(\mu) - \int_{\mathsf{S}_{\mathbf{i}}} \chi_\mu^\star\, d\lambda \right| = 0\,,$$

where $\mathcal{Q}^\star(\mu) \stackrel{\text{def}}{=} \{q_{\mathbf{i}}^\star(\mu), \mathbf{i} \in \{1,\ldots,I\}^m\}$ is given by (7). Let $\{I_M, M \geq 1\}$ be an integer-valued sequence such that $I_M^{-1} + I_M^m M^{-1} \to 0$ as M goes to infinity. Then,

$$\lim_{M \to +\infty} M \varsigma_{I_M,M}^2(\mu,g,\mathcal{Q}^\star(\mu)) = \varsigma_\infty^2(\mu,\chi_\mu^\star)\,.$$

The proof is given in Section 6.1. As the number of strata goes to infinity, the stratified estimator run with the optimal allocation $\mathcal{Q}^\star(\mu)$ has the same asymptotic variance as the stratified estimator run with the allocation deduced from the optimal density $\chi_\mu^\star$. In practice, of course, the optimal allocation $\mathcal{Q}^\star(\mu)$ is unknown, but it is possible to construct an estimator of this quantity by estimating the conditional variance of $\mathrm{Var}[\phi(Y)|\mu^T Y \in \mathsf{S}_{\mathbf{i}}]$ within each stratum [6].

Remark 1 When $m = d$, the results obtained are markedly different since the accuracy of the stratified estimator now depends on the definition of the strata along each direction. Let $\phi_\mu(x) \stackrel{\text{def}}{=} \phi(\mu^T x)$, $\partial_k \phi_\mu$ be the partial derivative of ϕ_μ w.r.t. its k-th coordinate for $k \in \{1,\ldots,d\}$. Let g_k still denote the function $x = (x_1,\ldots,x_d) \in \mathbb{R}^d \mapsto g_k(x_k)$. Assuming A1, $\mathrm{esssup}_\lambda (f_\mu/g) < +\infty$ and $\phi \in C^1$ satisfies $\mathrm{esssup}_\lambda\left(\sum_{k=1}^d |\partial_k \phi_\mu|/g_k\right) < +\infty$, one checks in [5] that for any integer-valued sequence $\{I_M, M \geq 1\}$ such that $\lim_{M \to \infty}\left(I_M^{-1} + I_M^{d+2} M^{-1}\right) = 0$,

$$\lim_{M \to +\infty} M I_M^2 \varsigma_{I_M,M}^2(\mu,g,\mathcal{Q}_\chi) = \varsigma_\infty^2(\mu,g,\chi) \stackrel{\text{def}}{=} \frac{1}{12} \int_{\mathbb{R}^d} \frac{f_\mu^2}{\chi} \sum_{k=1}^d \left(\frac{\partial_k \phi_\mu}{g_k}\right)^2 d\lambda.$$

In addition, $\lim_{M \to +\infty} M I_M^2 \varsigma_{I_M,M}^2(\mu,g,\mathcal{Q}^\star(\mu)) = \varsigma_\infty^2(\mu,g,\chi_{\mu,g}^\star)$ with $\chi_{\mu,g}^\star \propto f_\mu \sqrt{\sum_{k=1}^d \left(\frac{\partial_k \phi_\mu}{g_k}\right)^2}$.

4 An adaptive stratification algorithm

As shown in the asymptotic theory presented above, under optimal allocation, it is more important to optimize the stratification matrix μ than the strata boundaries along each stratification direction [1]. Propo-

---

[1] Of course, this is an asymptotic result, but our numerical experiments suggest that optimizing the strata boundaries along each stratification direction does not lead to a significant reduction of the variance. This is why we concentrate on the optimization of the stratification matrix

sition 1 suggests the following strategy: the "optimal" matrix $\mu_\star$ is defined as a minimizer of the limiting variance $\mu \mapsto \varsigma_\infty^2(\mu, \chi_\mu^\star)$. Of course, this optimization problem does not have a closed form expression because the functions $x \mapsto \psi_\mu(x)$, $x \mapsto \zeta_\mu(x)$ are not available.

We rather use the optimal limiting variance of the stratified estimator given in Proposition 2. The problem boils down to search for a minimizer μ of the variance $\varsigma_{I,M}^2(\mu, g, \mathcal{Q}^\star(\mu))$. In our applications, Y is a d-dimensional standard normal vector, and $\mu^T Y$ is a m-dimensional standard Gaussian vector. In this case, we set g_i, $i = \{1, \ldots, m\}$ to be the standard Gaussian distribution so that the strata boundaries in each directions are the quantiles of the standard normal variable. This choice leads to equiprobable strata for the vector $\mu^T Y$.

Of course, the optimization of $\varsigma_{I,M}^2(\mu, g, \mathcal{Q}^\star(\mu))$ is a difficult task because in particular the definition of this function involves multidimensional integrals, which cannot be computed with high accuracy. Note also that, in most situations, the optimization should be done in parallel to the main objective, namely, the estimation of the quantity of interest $\mathrm{E}[\phi(Y)]$, which is obtained using a stratified estimator based on the adaptively defined directions of stratification μ. The adaptive stratification is analog to the popular adaptive importance sampling; see for example [15], [2], [10], and [16].

When the function to minimize is an expectation, the classical approaches to tackle this problem are based on Monte Carlo approximations for the integrals appearing in the expression of the objective function and its gradients. There are typically two approaches to Monte Carlo methods, the stochastic approximation procedure and the sample average approximation method; see [9]. In the adaptive stratification context, these Monte Carlo estimators are based on the current fit of the stratification matrix and of the conditional variances within each stratum, the underlying idea being that the algorithm is able to progressively learn the optimal stratification, while the stratified estimator is constructed.

The algorithm described here is closely related to the sample average approximation method, the main difference with the classical approach being that, at every time a new search direction is computed, a new Monte Carlo sample (using the current fit of the strata and of the allocation) is drawn.

Denote by f the density of Y w.r.t. the Lebesgue measure. Define for $\mathbf{i} \in \{1, \cdots, I\}^m$, a function $h \in \{f, \phi f, \phi^2 f\}$, and an orthonormal $d \times m$ matrix μ,

$$\nu_\mathbf{i}(h, \mu) \overset{\text{def}}{=} \int_{\mathsf{S}_{\mu,\mathbf{i}}} h \, d\lambda = \int \prod_{k=1}^m \mathbb{1}_{\{y, G_k^{-1}((\mathbf{i}_k-1)/I) \leq \langle \mu_k, y\rangle \leq G_k^{-1}(\mathbf{i}_k/I)\}} h \, d\lambda \,, \tag{12}$$

where $\langle x, y\rangle$ denotes the scalar product of the vectors x and y. Using the definition of $\nu_\mathbf{i}$, the proportions $p_\mathbf{i}(\mu)$ and the conditional variances with each stratum $\sigma_\mathbf{i}^2(\mu)$ respectively given by (3) and (6) may be expressed as, when $\nu_\mathbf{i}(f, \mu) > 0$,

$$p_\mathbf{i}(\mu) = \nu_\mathbf{i}(f, \mu) \,, \quad \text{and} \quad \sigma_\mathbf{i}^2(\mu) = \frac{\nu_\mathbf{i}(f\phi^2, \mu)}{\nu_\mathbf{i}(f, \mu)} - \left(\frac{\nu_\mathbf{i}(f\phi, \mu)}{\nu_\mathbf{i}(f, \mu)}\right)^2 \,. \tag{13}$$

When M is large and I is fixed, minimizing the asymptotic variance of the stratified estimate with optimal allocation is equivalent to minimize $V(\mu)$ w.r.t. the stratification matrix μ where (see Lemma 1)

$$V(\mu) \overset{\text{def}}{=} \sum_{\mathbf{i}=1}^{\mathcal{I}} p_\mathbf{i}(\mu)\sigma_\mathbf{i}(\mu) = \sum_{\mathbf{i}=1}^{\mathcal{I}} \left(\nu_\mathbf{i}(f, \mu)\nu_\mathbf{i}(f\phi^2, \mu) - \nu_\mathbf{i}^2(f\phi, \mu)\right)^{1/2} \,.$$

Assuming that the functions $\mu \mapsto \nu_\mathbf{i}(h, \mu)$ are differentiable at μ for $h \in \{f, f\phi, f\phi^2\}$, the gradient may be expressed as

$$\nabla_\mu V(\mu) = \sum_{\mathbf{i}=1}^{\mathcal{I}} \frac{\nabla_\mu \nu_\mathbf{i}(f, \mu) \, \nu_\mathbf{i}(f\phi^2, \mu) + p_\mathbf{i}(\mu) \, \nabla_\mu \nu_\mathbf{i}(f\phi^2, \mu) - 2\nu_\mathbf{i}(f\phi, \mu) \, \nabla_\mu \nu_\mathbf{i}(f\phi, \mu)}{2 \, p_\mathbf{i}(\mu)\sigma_\mathbf{i}(\mu)} \mathbb{1}_{\{p_\mathbf{i}(\mu)\sigma_\mathbf{i}(\mu) \neq 0\}} \,. \tag{14}$$

The computation of this gradient thus requires to calculate $\nabla_\mu \nu_\mathbf{i}(h, \mu)$ for $h \in \{f, f\phi, f\phi^2\}$. For a vector $\nu \in \mathbb{R}^d$, $\nu \neq 0$, and $z \in \mathbb{R}$, define λ_z^ν, the restriction of the Lebesgue measure on the hyperplane $\{y \in \mathbb{R}^d, \langle \nu, y\rangle = z\}$.

Proposition 3 *Let $z \in \mathbb{R}$, $h : \mathbb{R}^d \to \mathbb{R}$ be a locally bounded integrable real function, $g_z : \mathbb{R}^d \ni \nu \mapsto \int \mathbb{1}_{\{y, \langle \nu, y\rangle \leq z\}} h(y) \, d\lambda(y)$ and $\mu \in \mathbb{R}^d$ be a non-zero vector. Assume that h is continuous λ_z^μ almost everywhere and that there exists $\varepsilon > 0$ such that*

$$\lim_{M \to +\infty} \sup_{|\nu - \mu| \leq \varepsilon} \int |y| \mathbb{1}_{\{|y| \geq M\}} |h(y)| \, d\lambda_z^\nu(y) = 0 \,. \tag{15}$$

Then, the function $\nu \mapsto g_z(\nu)$ is differentiable at μ and $\nabla_\mu g_z(\mu) = -\int \frac{y}{|\mu|} h(y) \, d\lambda_z^\mu(y)$.

Corollary 1 *Assume that h is a real locally bounded integrable function. Let m be an integer, $z = (z_1, \ldots, z_m) \in \mathbb{R}^m$, $g_z : \mathbb{R}^{d \times m} \ni (\nu_1, \ldots, \nu_m) \mapsto \int \prod_{k=1}^{m} \mathbb{1}_{\{y, \langle \nu_k, y \rangle \leq z_k\}} h(y) \, d\lambda(y)$ and $\mu = [\mu_1, \ldots, \mu_m] \in \mathbb{R}^{d \times m}$ be a full rank matrix. Assume that h is continuous $\sum_{k=1}^{m} \lambda_z^{\mu_k}$ almost everywhere and that there exists $\varepsilon > 0$ such that, for any $k \in \{1, \ldots, m\}$, $\lim_{M \to +\infty} \sup_{|\nu - \mu_k| \leq \varepsilon} \int |y| \mathbb{1}_{\{|y| \geq M\}} |h(y)| \, d\lambda_z^{\nu}(y) = 0$. Then, g_z is differentiable at μ and the differential $\nabla_\mu g_z$ is given by $\nabla_\mu g_z = [\nabla_{\mu_1} g_z, \ldots, \nabla_{\mu_m} g_z]$, where*

$$\nabla_{\mu_i} g_z(\mu) = - \int \frac{y}{|\mu_i|} \prod_{k \neq i} \mathbb{1}_{\{y, \langle \mu_k, y \rangle \leq z_k\}} h(y) \, d\lambda_{z_i}^{\mu_i}(y) \, .$$

The algorithm goes as follows. Denote by $\{\gamma_t\}$ a sequence of stepsizes. Consider the strata $\{\mathsf{S_i}, \mathbf{i} \in \{1, \cdots, I\}^m\}$ given by (8) for some product density g.

1. Initialization. Choose initial stratification directions $\mu^{(0)}$ and an initial number of draws in each statum $M^{(0)} \stackrel{\text{def}}{=} \{M_{\mathbf{i}}^{(0)}, \mathbf{i} \in \stackrel{\text{def}}{=} \{1, \ldots, I\}^m\}$ such that $\sum_{\mathbf{i}} M_{\mathbf{i}}^{(0)} = M$. Compute the probabilities $p_{\mathbf{i}}(\mu^{(0)})$ of each stratum.

2. Iteration. At iteration $t+1$, given $\mu^{(t)}$, $M^{(t)}$ and $\{p_{\mathbf{i}}(\mu^{(t)}), \mathbf{i} \in \{1, \cdots, I\}^m\}$,

   (a) *Compute $\widehat{\nabla V}(\mu^{(t)})$:*

   (i) for $\mathbf{i} \in \{1, \cdots, I\}^m$, draw $M_{\mathbf{i}}^{(t)}$ realizations of i.i.d. random variables $\{Y_{\mathbf{i},k}^{(t)}, k \leq M_{\mathbf{i}}^{(t)}\}$ with distribution $\mathbb{P}(Y \in \cdot | Y \in \mathsf{S}_{\mu^{(t)}, \mathbf{i}})$ and evaluate $\hat{\nu}_{\mathbf{i}}^{(t+1)}(h) = \frac{p_{\mathbf{i}}(\mu^{(t)})}{M_{\mathbf{i}}^{(t)}} \sum_{k=1}^{M_{\mathbf{i}}^{(t)}} h\left(Y_{\mathbf{i},k}^{(t)}\right)$ for $h \in \{\phi, \phi^2\}$.

   (ii) for $k \in \{1, \cdots, m\}$, $s \in \{G_k^{-1}(1/I), \cdots, G_k^{-1}((I-1)/I)\}$, draw $\tilde{M}_{k,s}^{(t)}$ realizations of i.i.d. random variables with distribution $\mathbb{P}(Y \in \cdot | [\mu_k^{(t)}]^T Y = s)$. Compute a Monte Carlo estimate of $\nabla_\mu \nu_{\mathbf{i}}(h, \mu^{(t)})$ for $h \in \{f, f\phi, f\phi^2\}$ based on Corollary 1.

   (iii) Compute a Monte Carlo estimate of $\nabla V(\mu^{(t)})$ based on the expression (14).

   (b) *Update the direction of stratification:* Set $\tilde{\mu} = \mu^{(t)} - \gamma_t \widehat{\nabla V}(\mu^{(t)})$; define $\mu^{(t+1)}$ as the orthonormal matrix found by computing the singular value decomposition of $\tilde{\mu}$ and keeping the m left singular vectors.

   (c) *Update the allocation policy:*

   (i) compute an estimate $\hat{\sigma}_{\mathbf{i}}^{(t+1)}$ of the standard deviation within stratum $\mathbf{i}$

   $$\hat{\sigma}_{\mathbf{i}}^{(t+1)} = \left(\frac{\hat{\nu}_{\mathbf{i}}^{(t+1)}(\phi^2)}{p_{\mathbf{i}}(\mu^{(t)})} - \left(\frac{\hat{\nu}_{\mathbf{i}}^{(t+1)}(\phi)}{p_{\mathbf{i}}(\mu^{(t)})} \right)^2 \right)^{1/2} \, .$$

   (ii) Update the allocation vector

   $$q_{\mathbf{i}}^{(t+1)} = \frac{p_{\mathbf{i}}(\mu^{(t)}) \, \hat{\sigma}_{\mathbf{i}}^{(t+1)}}{\sum_{\mathbf{j} \in \{1, \ldots, I\}^m} p_{\mathbf{j}}(\mu^{(t)}) \, \hat{\sigma}_{\mathbf{j}}^{(t+1)}} \, ,$$

   and the number of draws $\{M_{\mathbf{i}}^{(t+1)}, \mathbf{i} \in \{1, \ldots, I\}^m\}$ by applying the formula (4) with a total number of draws equal to M.

   (d) *Update the probabilities $p_{\mathbf{i}}(\mu^{(t+1)})$, $\mathbf{i} \in \{1, \cdots, I\}^m$.*

   (e) *Compute an averaged stratified estimate of the quantity of interest:* Estimate the Monte Carlo variance of the stratified estimator for the current fit of the strata and the optimal allocation

   $$[\varsigma^2]^{(t+1)} = \frac{1}{M} \left(\sum_{\mathbf{i} \in \{1, \cdots, I\}^m} p_{\mathbf{i}}(\mu^{(t)}) \, \hat{\sigma}_{\mathbf{i}}^{(t+1)} \right)^2 \, .$$

   Compute the current fit of the stratified estimator by the following weighted average

   $$\mathcal{E}^{(t+1)} = \left(\sum_{\tau=1}^{t+1} \frac{1}{[\varsigma^2]^{(\tau)}} \right)^{-1} \sum_{\tau=1}^{t+1} \frac{1}{[\varsigma^2]^{(\tau)}} \sum_{\mathbf{i} \in \{1, \cdots, I\}^m} \hat{\nu}_{\mathbf{i}}^{(\tau)}(\phi) \, . \tag{16}$$

There are two options to choose the stepsizes $\{\gamma_t, t \geq 0\}$. The traditional approach consists in taking a decreasing sequence satisfying the following conditions (see for example [14, 11])

$$\sum_{t \geq 0} \gamma_t = +\infty , \qquad \sum_{t \geq 0} \gamma_t^2 < +\infty .$$

If the number of simulations is fixed in advance, say equal to N, then one can use a constant stepsize strategy, *i.e.* choose $\gamma_t = \gamma$ for all $t \in \{1, \ldots, N\}$. As advocated in [9], a sensible choice in this setting is to take γ_t proportional to $N^{-1/2}$. The convergence of this crude gradient algorithm proved to be quite fast in all our applications, so it is not required to resort to computationally involved alternatives.

Step 2(a)ii is specific to the optimization problem to solve and is not related to the stratification itself. The number of draws for the computation of the surface integral (see Corollary 1) can be chosen independently of the allocation $M^{(t)}$. When the samples in steps 2(a)i and 2(a)ii can be obtained by transforming the same set of variables (see Section 5 for such a situation), it is natural to choose $\tilde{M}^{(t)} = \{\tilde{M}_{k,s}^{(t)}, k \in \{1, \cdots, m\}, s \in \{G_k^{-1}(1/I), \cdots, G_k^{-1}((I-1)/I)\}\}$ such that $\sum_{k,s} \tilde{M}_{k,s}^{(t)} = M$.

When f_μ has a product form (which is the case e.g. when Y is a standard d-dimensional Gaussian distribution), we can set $g = f_\mu$. Then, the strata are equiprobable and $p_{\mathbf{i}}(\mu) = 1/I^m$ for any $(\mathbf{i}, \mu)$.

It is out of the scope of this paper to prove the convergence of this algorithm and we refer the reader to classical treatises on this subject. The above algorithm provides, at convergence, both *(i)* "optimal" directions of stratification and an estimate of the associated optimal allocation; *(ii)* an averaged stratified estimate $\mathcal{E}$. By omitting the step 2e, the algorithm might be seen as a mean for computing the stratification directions and the associated optimal allocation, and these quantities can then be plugged in a "usual" stratification procedure.

5 Applications in Financial Engineering

The pricing of an option amounts to compute the expectation $\mathrm{E}[\Xi(Y)]$ for some measurable non-negative function Ξ on $\mathbb{R}^d$, where Y is a standard d-multivariate Gaussian variable. The Cameron-Martin formula implies that for any $\nu \in \mathbb{R}^d$,

$$\mathrm{E}[\Xi(Y)] = \mathrm{E}\left[\Xi(Y + \nu) \exp(-\nu^T Y - 0.5\nu^T \nu)\right] , \qquad (17)$$

The variance of the plain Monte Carlo estimator depends on the choice of ν. In all the experiments below (except for the [8] estimator), we use either $\phi(y) = \Xi(y)$ (case $\nu = 0$) or $\phi(y) = \Xi(y + \nu_\star) \exp(-\nu_\star^T y - 0.5\nu_\star^T \nu_\star)$ where $\nu_\star$ is the solution of the optimization problem

$$\mathrm{argmax}_{\{\nu \in \mathbb{R}^d, \Xi(\nu) > 0\}} \left\{\ln \Xi(\nu) - 0.5\nu^T \nu\right\} , \qquad (18)$$

(case $\nu = \nu_\star$). The motivations for this particular choice of the drift vector ν and procedures to solve this optimization problem are discussed in [8].

We apply the adaptive stratification procedure introduced in Section 4 (hereafter referred to as "`AdaptStr`") in the case $m = 1$. For comparison purposes, we also run the stratification procedure proposed in [8] (hereafter referred to as "`GHS`"), combining *(i)* importance sampling with the drift $\nu_\star$ defined in (18), and *(ii)* stratification with proportional allocation and direction $\mu_{\mathbf{g}}$ defined in [8, Section 4.2].

We also run stratification algorithms with three different directions of stratification: the vector $\mu_\star \propto \nu_\star$, the vector μ_{reg} proportional to the vector of linear regression of the function $\phi(Y)$ on Y (these regression coefficients are obtained in a pilot run), and a vector μ_l which is specific to each application. For these three directions, we run stratification with proportional allocation (case "q_i" set to "*prop*") and with optimal allocation (case "q_i" set to "*opt*"). We also run the plain Monte Carlo estimator (column "`MC`"); when used with $\nu = \nu_\star$, "MC" corresponds to an importance sampling estimator with a drift function $\nu_\star$.

Finally, we compare these stratified estimators to Latin Hypercube (LH) estimators (see [7], [12] for a description of this method). For Y a standard normal vector in $\mathbb{R}^d$, the expectation of interest $\mathrm{E}[\phi(Y)]$ is also equal to $\mathrm{E}[\phi(OY)]$ for any orthogonal matrix $O \in \mathbb{R}^{d \times d}$ but the variance of the LH estimator associated with the variable $\phi(OY)$ depends on the choice of O. Unfortunately, it is very difficult to compute explicitly the asymptotic variance of LH estimators and therefore to adapt the matrix O; see [13]. A possible seemingly sensible choice for O consists in letting O be an unitary matrix with the first column equal to the stratification direction μ. We consider simple guesses, associated to the stratification directions $\mu_\star$, μ_{reg} and the adaptive stratification direction obtained by our algorithm `AdaptStr`.

5.1 Practical implementations of the adaptive stratification procedure

The numerical results have been obtained by running `Matlab` codes available from the authors [2] In the numerical applications below, $m = 1$. We choose $g = f_\mu$ so that the strata are equiprobable ($p_{\mathbf{i}}(\mu) = 1/I$). We choose $I = 100$ strata and $M = 20\,000$ draws per iterations.

The drift vector ν that solves (18) is obtained by running `solnp`, a nonlinear optimization program in Matlab freely available at `http://www.stanford.edu/~yyye/matlab/`. The direction $\mu^{(0)}$ is set to the unitary constant vector $(1, \cdots, 1)/\sqrt{d}$; the initial allocation $M^{(0)}$ is proportional. Exact sampling under the conditional distributions $\mathbb{P}(Y \in \cdot | Y \in \mathsf{S}_{\mu^{(t)}, \mathbf{i}})$ and $\mathbb{P}(Y \in \cdot | [\mu^{(t)}]^T Y = s)$ can be done by linear transformation of standard Gaussian vectors (see [7, section 4.3, p. 223]). The draws in step 2(a)i and 2(a)ii can be obtained by transforming the same set of $M^{(t)}$ Gaussian random variables $\{V_j^{\mathbf{i}}, j \le M_{\mathbf{i}}^{(t)}, \mathbf{i} \in \{1, \cdots, I\}\}$. Therefore, the total number of d-dimensional Gaussian draws by iteration is M (the estimates of $\nu_{\mathbf{i}}(h, \mu)$ and $\nabla_\mu \nu_{\mathbf{i}}(h, \mu)$ are not independent); M uniform draws in $(0, 1)$ are also required to sample under the conditional distribution $\mathbb{P}(Y \in \cdot | Y \in \mathsf{S}_{\mu^{(t)}, \mathbf{i}})$. The criterion is optimized using a fixed stepsize steepest descent algorithm (the stepsize is determined using a limited set of pilot runs).

5.2 Assessing efficiency of the adaptive stratification procedure

We compare the averaged stratified estimate $\mathcal{E}^N$ obtained after $N = 200$ iterations, with different stratification procedures and with the crude Monte Carlo estimate. We report in the tables below the estimate of the option prices and the estimates of the first and third quartiles of the variance of the estimator obtained from 50 independent replications.

The comparison of the procedures relies on the variance of the estimators. The column "`MC`" is an estimate of the variance of $\phi(Y)$ computed with MN i.i.d. samples of a d-multivariate gaussian distribution. In the case $\nu = 0$, this is an estimation of the variance of the plain Monte Carlo estimator; when $\nu = \nu_\star$, this corresponds to an estimation of the Importance Sampling estimator (with importance sampling distribution equal to a standard Gaussian distribution centered at $\nu_\star$). The column "`AdaptStr`" is the limiting variance per sample of $\mathcal{E}^N$ which is equal to

$$ N \left\{ \sum_{t=1}^N \left(\left[\sum_{\mathbf{i}} p_{\mathbf{i}} \, \hat{\sigma}_{\mathbf{i}}^{(t)} \right]^2 \right)^{-1} \right\}^{-1} \sim \left(\sum_{\mathbf{i}} p_{\mathbf{i}} \, \sigma_{\mathbf{i}}(\mu^{(+\infty)}) \right)^2 , $$

when each iteration $t \in \{1, \cdots, N\}$ implies M draws (see the algorithm in Section 4); note that by definition of our procedure, the allocation is optimal. The column "`GHS`" is an estimate of $\sum_{\mathbf{i}} p_{\mathbf{i}} \sigma_{\mathbf{i}}^2(\mu_{\mathbf{g}})$ computed with MN samples; note also that by definition of this procedure, only the case $\nu = \nu_\star$ and the proportional allocation has been considered. The columns "μ_{reg}", "$\mu_\star$", and "μ_l" report the results for the stratification procedures with these directions of stratification: the rows 'proportional allocation' report an estimation of $\sum_{\mathbf{i}} p_{\mathbf{i}} \sigma_{\mathbf{i}}^2(\mu)$ computed with MN samples (for $\mu \in \{\mu_{\mathrm{reg}}, \mu_\star, \mu_l\}$ and $\nu \in \{0, \nu_\star\}$). We also consider the results for the optimal allocation, and to that goal we estimate the standard deviation within each stratum by an iterative algorithm - with N iterations - : the rows 'optimal allocation' report an estimation of $N \left\{ \sum_{t=1}^N \left([\sum_{\mathbf{i}} p_{\mathbf{i}} \, \hat{\sigma}_{\mathbf{i}}^{(t)}(\mu)]^2 \right)^{-1} \right\}^{-1}$ where $\{\hat{\sigma}_{\mathbf{i}}^{(t)}(\mu), \mathbf{i} \le I\}$ is an estimation of the standard deviation of the strata computed with a total number of M draws allocated to each stratum according to the optimal allocation computed at the previous iteration ($t - 1$) (the allocation at iteration 0 is the proportional one). For Latin Hypercube samplers, because we want to report not only a number but also confidence intervals, the total number of draws (MN) are allocated to generate N i.i.d. estimators each based on Latin Hypercube sample of size M.

5.3 Asian options

Consider the pricing of an arithmetic Asian option on a single underlying asset under standard Black-Scholes assumptions. The price of the asset is described by the stochastic differential equation $\frac{dS_t}{S_t} = r \, dt + \upsilon \, dW_t$, $S_0 = s_0$, where $\{W_t, t \ge 0\}$ is a standard Brownian motion, r is the risk-free mean rate of return, υ is the volatility and s_0 is the initial value. The asset price is discretized on a regular grid

---

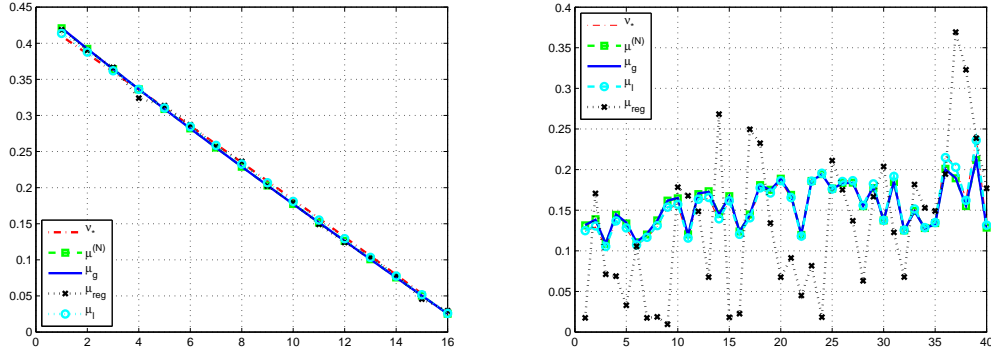[2] These codes are freely available from the url `http://www.tsi.enst.fr/~gfort/`

Fig. 1 [left panel] Asian Option when $(v, K, \nu) = (0.1, 45, \nu_\star)$: drift vector $\nu_\star$ and directions $\mu^{(N)}$, μ_g, μ_reg and μ_l. $\nu_\star$ has been scaled to have norm 1 ($\nu_\star \leftarrow \nu_\star/0.42$). [right panel] Basket Option when $(c, K, \nu) = (0.1, 45, \nu_\star)$: drift vector $\nu_\star$ and directions $\mu^{(N)}$, μ_g, μ_l, μ_reg and μ_l. $\nu_\star$ has been scaled to have norm 1 ($\nu_\star \leftarrow \nu_\star/0.41$).

$0 = t_0 < t_1 < \cdots < t_d = T$, with $t_i \overset{\text{def}}{=} iT/d$. The increment of the Brownian motion on $[t_{i-1}, t_i)$ is simulated as $\sqrt{T/d}\,Y_i$ for $i \in \{1, \cdots, d\}$ where $Y = (Y_1, \cdots, Y_d) \sim \mathcal{N}_d(0, \text{Id})$. The discounted payoff of a discretely monitored arithmetic average Asian option with strike price K is given by $\Xi(Y)$,

$$\Xi(y) = \exp(-rT)\left(\frac{s_0}{d}\sum_{k=1}^{d}\exp\left((r - 0.5v^2)\frac{kT}{d} + v\sqrt{\frac{T}{d}}\sum_{j=1}^{k}y_j\right) - K\right)_+ \quad , \quad y = (y_1, \cdots, y_d) \in \mathbb{R}^d \,,$$

where for $x \in \mathbb{R}$, $x_+ = \max(x, 0)$. In the numerical applications, we take $s_0 = 50$, $r = 0.05$, $T = 1$, $(v, K) \in \{(0.1, 45), (0.5, 45), (0.5, 65), (1, 45), (1, 65)\}$ and $d = 16$. We choose $\mu_l \propto (d, d-1, \cdots, 1)$.

We run `AdaptStr` when $(v, K) = (0.1, 45)$: on Figure 1, the optimal drift vector $\nu_\star$, the direction $\mu^{(N)}$ obtained after N iterations of `AdaptStr`, and the directions of stratification $\mu_\text{g}, \mu_\text{reg}, \mu_l$ are plotted. In Figure 2, the successive directions $t \mapsto \mu^{(t)}$, the successive estimations of the quantity of interest $t \mapsto \mathcal{E}^{(t)}$ and of the variance $t \mapsto (\sum_\mathbf{i} p_\mathbf{i}\hat\sigma_\mathbf{i}^{(t)})^2$ are displayed. We observe that $\{\mu^{(t)}, t \geq 0\}$ converges to the direction μ_g, and the convergence takes place after about 30 iterations. We find the same pattern for a wide range of parameter values. The choice of the stratification direction has a major impact on the variance of the estimate $\mathcal{E}^{(t)}$ as shown on Figure 2 [bottom right]. Along the iterations of the algorithm, the variance decreases from 0.1862 to 0.0016. We also observed that the convergence of the algorithm and the limiting values were independent of the initial values $(\mu^{(0)}, M^{(0)})$ (these results are not reported for brevity). These initial values (and the choice of the sequence $\{\gamma^{(t)}, t \geq 1\}$) only influence the number of iterations required to converge. `AdaptStr` can also be seen as a procedure that computes a stratification direction and provides the associated optimal allocation. These quantities can then be used for running a (usual) stratification procedure with M draws and for the optimal allocation. By doing such with $M = 20\,000$, we obtain an estimate of the quantity $\text{E}[\phi(Y)]$ equal to 6.05 and of the variance equal to $0.0016/M$. We can compare these results to the output of `GHS`: this yields the same estimator of $\text{E}[\phi(Y)]$ and a larger standard deviation equal to $0.0071/M$. Observe that since $\mu^{(N)} = \mu_\text{g}$, the two algorithms differ from the allocations in the strata.

We conclude this study of `AdaptStr` by illustrating the role of the drift vector ν (see Eq. 17). We report on Figure 3 the limiting direction $\mu^{(N)}$, the estimates $t \mapsto \mathcal{E}^{(t)}$ and the variance $t \mapsto (\sum_\mathbf{i} p_\mathbf{i}\hat\sigma_\mathbf{i}^{(t)})^2$ when $\nu = 0$. The limiting direction $\mu^{(N)}$ slightly differs from μ_g and is close to $\nu_\star$. Moreover, the variance reduction is weaker: the limiting value of $t \mapsto (\sum_\mathbf{i} p_\mathbf{i}\hat\sigma_\mathbf{i}^{(t)})^2$ is 0.0035. The efficiency of the adaptive stratification procedure `AdaptStr` is thus related to the drift vector ν in (17); similar conclusions are reached in [8] (see also [7]).

We report in Tables 1 and 2 the variance of different estimators, as described in Section 5.2.

Insert Tables 1 and 2 about here

Consider first the case $\nu = 0$. When the volatility of the asset is low $v = 0.1$ and the strike is in-the-money, the performance of the adaptive stratification estimator "`AdaptStr`" and of the stratification estimator with fixed direction μ_reg, $\mu_\star$ and μ_l and with optimal allocations are equivalent. We observe indeed that the directions $\mu^{(N)}, \mu_\text{reg}, \mu_\star$ and μ_l are almost colinear. Compared to the plain Monte Carlo, the variance
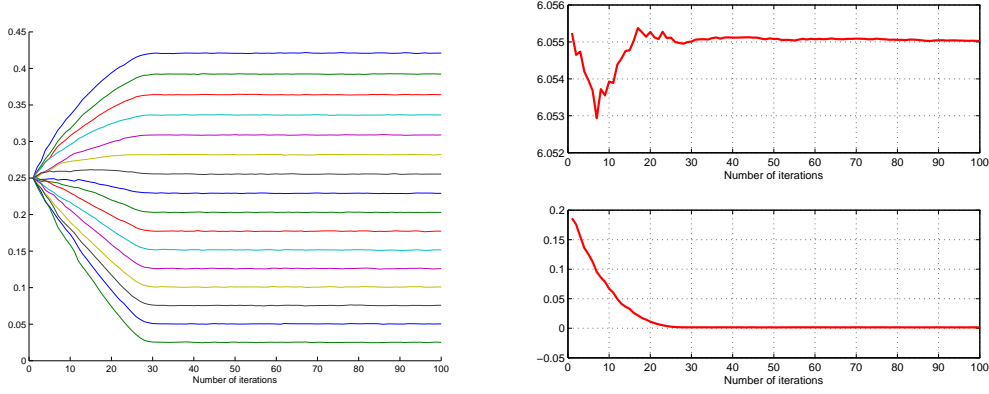
Fig. 2 Asian Option when $(v, K, \nu) = (0.1, 45, \nu_\star)$. [left panel] successive directions of stratification $t \mapsto \mu^{(t)}$. $\mu^{(0)}$ is proportional to the vector $(1, \cdots, 1)$ so that the d curves start from the same point $1/\sqrt{d}$. By convention, the first component of $\mu^{(t)}$ is positive. [top right] successive estimations of the quantity of interest $t \mapsto \mathcal{E}^{(t)}$. [bottom right] successive values of the variance $t \mapsto (\sum_{\mathbf{i}} p_{\mathbf{i}} \hat{\sigma}_{\mathbf{i}}^{(t)})^2$; the limiting value is 0.0016.

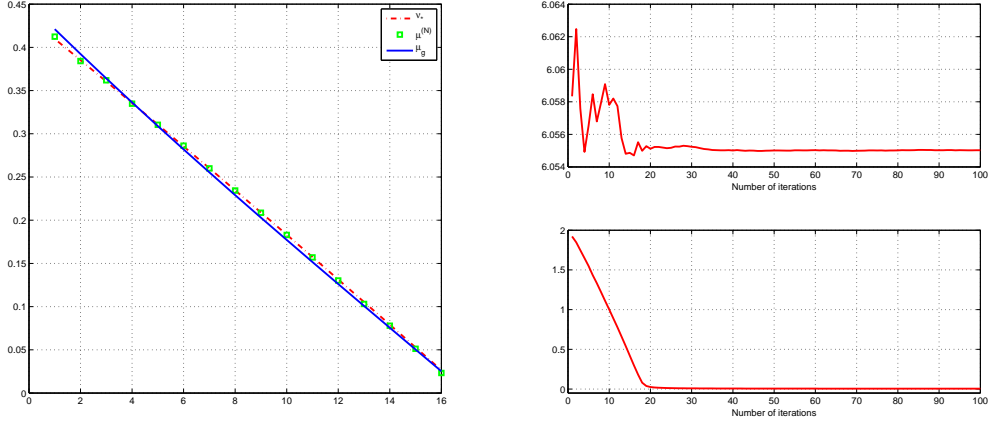**Fig. 3** Asian Option with $(v, K, \nu) = (0.1, 45, 0)$: [left panel] the limiting direction $\mu^{(N)}$ and for comparison, μ_g and $\nu_\star$ normalised to have norm 1 ($\nu_\star \to \nu_\star/0.42$). [top right] successive estimations of the quantity of interest $t \mapsto \mathcal{E}^{(t)}$. [bottom right] successive values of the variance $t \mapsto (\sum_{\mathbf{i}} p_{\mathbf{i}} \hat{\sigma}_{\mathbf{i}}^{(t)})^2$: the limiting value is 0.0035.

reduction is in both case equal to 2500. The LH estimator with a rotation along any of the directions $\mu^{(N)}, \mu_\star$ and μ_{reg} outperforms all the stratified estimators: the variance reduction is by a factor 10500. This reduction in the variance strongly depends upon the choice of the rotation: the LH estimator with no rotation implies a variance reduction by a factor 150.

When the volatility of the asset is high $v = 1$, the conclusions are markedly different. Consider e.g. the case when the strike is out of the money. The adaptive stratification estimator "`AdaptStr`" provides a reduction of variance by a factor 150, which is again similar to the variance reduction afforded by the stratification with fixed directions $\mu_{\mathrm{reg}}, \mu_\star$ and μ_l, and optimal allocation; `AdaptStr` outperforms stratified estimators with any of the fixed direction $\mu_{\mathrm{reg}}, \mu_\star$ or μ_l by a factor 13 when allocation is proportional. The LH estimator with no rotation only provides a reduction in variance by a factor 1.7; when the rotation along $\mu^{(N)}$ is applied, the reduction is by a factor 65. Here again, the LH estimator is very sensitive to the choice of the orthogonal matrix O.

The use of the drift $\nu = \nu^\star$ improves the variance of all the stratified estimators by a factor 2 to 10, depending on the choice of the stratification direction; and by a factor 10 to 25 for "`MC`". Here again, "`AdaptStr`" is the best stratified estimator; its performance can be approached by stratification estimators with fixed directions, but the choice of this fixed direction depends crucially upon the choice of the volatility and the value of the strike. $\mu^{(N)}$ and μ_{reg} are almost colinear in many cases, but not always as observed

e.g. in the cases $(v, K, \nu) = (0.1; 45; \nu_\star)$ and $(v, K, \nu) = (1; 65; \nu_\star)$. It is interesting to note that the use of the drift $\nu_\star$ does not always improve the variance of the LH estimator.

These experiments show that the choice of the stratification direction and of the allocation is crucial. For example, in the case $(v, K, \nu) = (0.5, 65, 0)$, the adaptive stratification estimator improves upon the stratification estimator with fixed direction $(1, \cdots, 1)/\sqrt{d}$ and optimal allocation by a factor 60 (and by a factor 190 when proportional allocation is used) - these results are not reported in the tables for brevity since this direction is rarely optimal - . Even if simple guesses for the direction reduce the variance, this reduction can be improved (by a factor 20) when optimal allocation is used; this allocation is unknown and has to be learnt. In these examples, LH outperforms in many cases stratification procedures provided it is applied with a rotation O: the rotation along $\mu^{(N)}$ outperforms LH with no rotation and when compared with other simple guess rotations, it provides similar or better variance reduction. All these remarks strongly support the use of adaptive procedures.

5.4 Options with knock-out at expiration

A knock-out barrier option is a path-dependent option that expires worthless if the underlier reaches a specified barrier level. The payoff of this option is given by

$$\Xi(y) = \exp(-rT) \left(\frac{s_0}{d} \sum_{k=1}^{d} \exp\left((r - 0.5\sigma^2)\frac{kT}{d} + \sigma\sqrt{\frac{T}{d}} \sum_{j=1}^{k} y_j \right) - K \right)_+ \mathbb{1}_{\{S_T(y) \leq B\}} ,$$

where K is the strike price, B is the barrier and $S_T(y)$ is the underlier price modeled as

$$S_T(y) = s_0 \exp\left((r - 0.5\sigma^2)T + \sigma\sqrt{\frac{T}{d}} \sum_{j=1}^{d} y_j \right) .$$

In the numerical applications, we set $s_0 = 50$, $r = 0.05$, $T = 1$, $\sigma = 0.1$, $d = 16$ and $(K, B) \in \{(50, 60), (50, 80)\}$. We choose $\mu_l \propto (d, d-1, \cdots, 1)$. On Figure 4[left panel], we plot $\mu^{(N)}$ in the case $(K, B, \nu) = (50, 60, 0)$; the limiting direction is μ_l. On Figure 4[right panel], we plot $\mu^{(N)}$ in the case $(K, B, \nu) = (50, 60, \nu_\star)$; for comparison, we also plot $\nu_\star$, μ_g, μ_{reg} and μ_l. This is an example where the optimal stratification direction $\mu^{(N)}$ does not coincide with the different direction of stratification ($\mu_{\mathbf{g}}$, μ_{reg} and μ_l); in this example, $\mu_g \sim \mu_l$ but the optimal direction of stratification is close to $(1, \cdots, 1)/\sqrt{d}$. The direction associated to the regression estimator is far from being optimal.

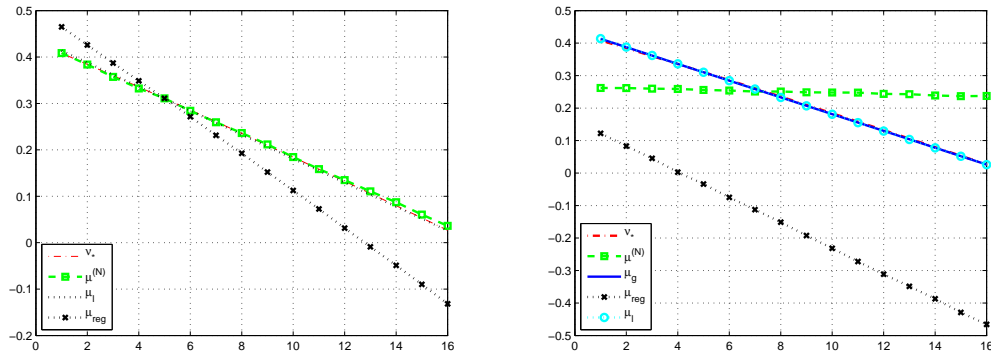**Fig. 4** Barrier Option when $(K, B) = (50, 60)$ and [left panel] $\nu = 0$ or [right panel] $\nu = \nu_\star$: directions $\mu^{(N)}$, μ_{reg}, μ_l, $\nu_\star$ and μ_g. $\nu_\star$ has been scaled to have norm 1 ($\nu_\star \leftarrow \nu_\star/0.84$)

We report in Tables 3 and 4 the variance of the different estimators, as described in Section 5.2.

Insert Table 3 and 4 about here

Consider first the case where the drift is set to 0. When $(K, B) = (50, 60)$ (the option is in the money, and the barrier is close to the money), the adaptive stratification provides a variance reduction by a factor 10 with respect to the plain Monte Carlo estimator. In this case, the stratification directions μ_l and $\mu_\star$ (with an optimal allocation) perform almost as well (and μ_l and $\mu_\star$ are close to $\mu^{(N)}$ at the convergence), while μ_{reg} provides a higher variance than what we have with $\mu^{(N)}$. For the LH estimator, the variance reduction is only by a factor 1.5; it is worthwhile to note that the best choice for the rotation is $\mu^{(N)}$. The use of the drift vector $\nu_\star$ improves the variance of the adaptive stratification by a factor 1.8; the optimal stratification vector is now $\mu^{(N)}$ which surpasses μ_{reg}. The variance of the LH estimator is also reduced, but the variance of the best LH estimator is more than twice higher than the variance of "AdapStr" in the case $\nu = 0$.

When $(K, B) = (50, 80)$ (the barrier is out of the money), a factor reduction 2800 is obtained by the adaptive stratification estimator "AdaptStr"; a similar variance reduction is achieved using stratified estimator with direction μ_{reg} and with optimal allocation. For the LH estimator, the variance reduction is by a factor 2000, when the rotation is μ_{reg}. The use of the drift vector $\nu_\star$ improves the behavior of all the algorithms: the variance of "AdaptStr" is reduced by a factor 3.8. Finally, LH with rotation $\mu^{(N)}$ reduces the variance of LH with no rotation by a factor 1200.

To conclude, this example shows again the interest of adaptive procedures in order to find a stratification direction, the optimal allocation or a rotation in LH.

5.5 Basket options

Consider a portfolio consisting of d assets. The portfolio contains a proportion α_k of asset k, $k \in \{1, \ldots, d\}$. The price of each asset is described by a geometric Brownian motion (under the risk neutral probability measure)

$$\frac{dS_t^{(k)}}{S_t^{(k)}} = r\, dt + \upsilon_k\, dW_t^{(k)}$$

but the standard Brownian motions $\{W^{(k)}, k \in \{1, \ldots, d\}\}$ are not necessarily independent. For any $t \geq s$ and $k \in \{1, \ldots, d\}$

$$\ln S_t^{(k)} = \ln S_s^{(k)} + \left(r - 0.5\upsilon_k^2\right)(t - s) + \upsilon_k\sqrt{t - s}\tilde{Y}_k$$

where $\tilde{Y} = (\tilde{Y}_1, \ldots, \tilde{Y}_d) \sim \mathcal{N}_d(0, \Sigma)$. The $d \times d$ matrix Σ is a positive semidefinite matrix with diagonal coefficients equal to 1. Therefore, the variance of the log-return on asset k in the time interval $[s, t]$ is $(t - s)\upsilon_k^2$, and the covariance between the log-returns i, j is $(t - s)\upsilon_i\upsilon_j\Sigma_{i,j}$. It follows that $\Sigma_{i,j}$ is the correlation between the log-returns. The price at time 0 of a European call option with strike price K and exercise time T is given by $\mathrm{E}[\Xi(Y)]$ where

$$\Xi(y) = \exp(-rT)\left(\sum_{k=1}^{d} \alpha_k s_0^{(k)} \exp\left((r - 0.5\upsilon_k^2)T + \upsilon_k\sqrt{T}\tilde{y}_k\right) - K\right)_+$$

and $\tilde{y} = \sqrt{\Sigma}y$ ($\sqrt{\Sigma}$ denotes a square root of the matrix Σ i.e. solves $MM^T = \Sigma$). In the numerical applications, Σ is chosen to be $\Sigma_{i,i} = 1$, $\Sigma_{i,j} = c$, $\alpha_k = 1/d$, $r = 0.05$, $T = 1$, and $d = 40$. We consider $(c, K) \in \{(0.1, 45), (0.5, 45), (0.9, 45)\}$. The initial values $\{s_0^k, k \leq d\}$ are drawn from the uniform distribution in the range $[20, 80]$; the volatilities $\{\upsilon_k, k \leq d\}$ are chosen linearly equally spaced in the set $[0.1, 0.4]$. The assets are sorted so that $\upsilon_1 \leq \cdots \leq \upsilon_d$. We choose

$$\mu_l \propto \left(\alpha_1 s_0^{(1)} \exp((r - 0.5\upsilon_1^2)T)\upsilon_1, \cdots, \alpha_d s_0^{(d)} \exp((r - 0.5\upsilon_d^2)T)\upsilon_d\right)\Sigma.$$

In the case $(c, K, \nu) = (0.1, 45, \nu_\star)$, we plot on Figure 1[right panel] the limiting direction $\mu^{(N)}$ and for comparison, the directions $\nu_\star$, μ_g, μ_{reg}, μ_l. We report in Tables 5 and 6 the variance of the different estimators, as described in Section 5.2.

Insert Tables 5 and 6 about here

In this example again the adaptive stratification estimator improves upon the best stratified estimator with (non-adaptive) stratification direction and optimal allocation. Here again, the knowledge of the optimal allocation improves the variance reduction, by a factor 15 for example in the case $(c, K, \nu) = (0.9, 45, 0)$ for the fixed direction μ_{reg}, $\mu_\star$ or μ_l. It is interesting to note that the variance reduction with respect to

the plain Monte Carlo using the optimal direction of stratification ranges from 100 ($c = 0.1, B = 45$) to 2500 ($c = 0.9, B = 45$) whereas the use of the drift $\nu^\star$ allows only a reduction from a factor 10. The choice of the stratification direction plays a more important role than the choice of drift direction.

The comparison with the LH estimator is more difficult, because this estimator behaves totally differently from the stratified estimator. First, the use of a drift increases the variance: whereas the effect of the drift is always markedly beneficial for the Monte Carlo estimators, the drift $\nu_\star$ may increase the variance by a factor than can be as large as 15 (when $(c, K) = (0.1, 45)$). Second, LH with rotation always outperforms (adaptive) stratification: the main difficulty stems from the choice of the rotation but the obtained results show that rotation along $\mu^{(N)}$ provides the maximal variance reduction, or is similar to the maximal one (maximal among the three considered rotations). Finally, the performance of the estimator is extremely sensitive to the choice of the simulation setting: when the correlation among the assets is large $c = 0.9$, the choice of the first vector of the orthogonal matrix O becomes crucial. Rotation along $\mu^{(N)}$ (resp. μ_{reg}) improves LH with no rotation by a factor 22500 (resp. 3000).

5.6 Heston model of stochastic volatility

We consider a last example which is not covered by the methodology presented in [8]. We price an Asian option in the Heston model, specified as follows

$$dS_t = rS_t dt + \sqrt{\xi_t}\, S_t \left(\sqrt{1-\rho^2}dW_t^1 + \rho dW_t^2\right), \ \ dX_t = S_t dt, \ \ d\xi_t = k(\theta - \xi_t)dt + \sigma\sqrt{\xi_t}\, dW_t^2$$

where $\{W_t^1, t \geq 0\}$ and $\{W_t^2, t \geq 0\}$ are two independent Brownian motions, r is the risk free rate, $\sigma > 0$ the volatility, $k \geq 0$ the mean reversion rate, $\theta \geq 0$ the long run average volatility, and $\rho \in [-1, 1]$ the correlation rate. The price of an Asian Call option with strike K and maturity T is

$$\mathrm{E}\left[\exp(-rT)\left(\frac{1}{T}X_T - K\right)_+\right]. \tag{19}$$

In our tests, we have chosen the parameters so that $\sigma^2 \leq 4k\theta$. This enabled us to replace by Gaussian increments the finitely-valued random variables used to discretize W^1 in the scheme proposed in [1] to approximate the SDE satisfied by (S_t, X_t, ξ_t). We refer to [1] for a precise description of the scheme that we used. The resulting approximation $\hat{X}_d$ of X_T is generated from a vector $Y = (Y_1, \ldots, Y_d, Y_{d+1}, \ldots, Y_{2d}) \sim \mathcal{N}_{2d}(0, \mathrm{Id})$ corresponding to the increments of (W^1, W^2) and a vector $B = (B_1, \ldots, B_d)$ of independent Bernoulli random variables with parameter 0.5. The price (19) is then approximated by

$$\mathrm{E}[\exp(-rT)\left(\hat{X}_d - K\right)_+].$$

In the following tests we keep ν in (17) equal to zero and do not stratify the random vector B.

We choose $S_0 = 100$, $\theta = 0.01$, $k = 2$, $\sigma = 0.2$, $T = 1$, $r = 0.095$, $\rho = -0.5$ and $(\xi_0, K) \in \{(0.01, 120), (0.01, 100), (0.01, 80), (0.04, 130), (0.04, 100), (0.04, 70)\}$. The discretization step of the scheme is $d = 50$.

On Figure 5 we plot the successive estimations of the variance $t \mapsto (\sum_{\mathbf{i}} p_{\mathbf{i}} \hat{\sigma}_{\mathbf{i}}^{(t)})^2$, when $(\xi_0, K) = (0.01; 120)$ and $(\xi_0, K) = (0.04; 70)$.

We plot on Figure 6 the components of $\mu^{(N)}$ with respect to the component index in the cases $(\xi_0, K) = (0.01; 120)$ and $(\xi_0, K) = (0.04; 70)$.

We report in Tables 7 and 8 the variance of somes estimators described in Section 5.2.

Insert Table 7 and 8 about here

The first observation is that even in this case, `AdaptStr` still works and provides variance reduction when compared to Monte Carlo. It is all the more efficient than the option is out of the money : when $(\xi_0, K, \nu) = (0.01, 120, 0)$, the variance reduction is by a factor 85; when $(\xi_0, K, \nu) = (0.04, 130, 0)$, the variance reduction is by a factor 105. `AdaptStr` and stratification with fixed direction μ_{reg} are equivalent, provided the last one is applied with optimal allocation, just necessitating again iterative procedures.

We can wonder on the effect of the moneyness and the volatility of the model on the variance reduction. As shown in Table 7, in general the achieved variance reduction is larger when the option is out of the money (for $K = 130$ and $\xi_0 = 0.04$ the variance is divided by nearly 105 when using `AdaptStr`). We also observe that stratification procedures outperform LH samplers when the option is out of the money, but LH is equivalent to stratification when the option is in the money.
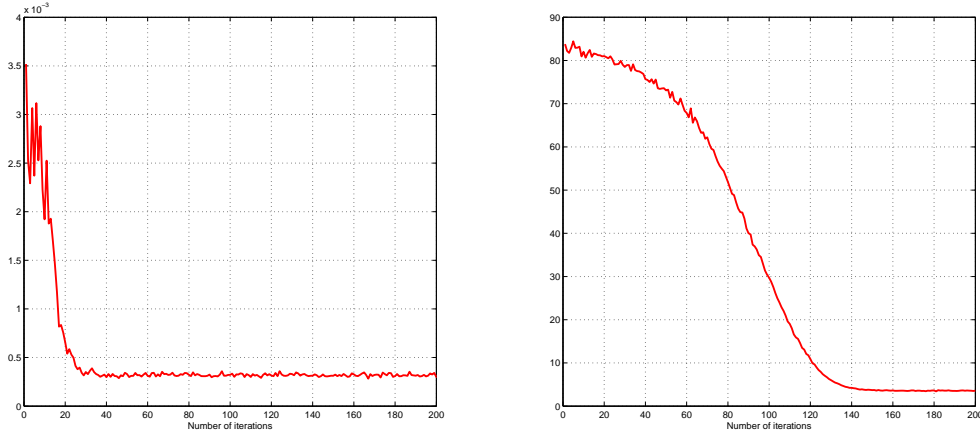
Fig. 5 Asian Call Option in Heston model: successive estimations of the variance $t \mapsto (\sum_{\mathbf{i}} p_{\mathbf{i}} \, \hat{\sigma}_{\mathbf{i}}^{(t)})^2$ [left panel] when $(\xi_0, K) = (0.01; 120)$; [right panel] when $(\xi_0, K) = (0.04; 70)$
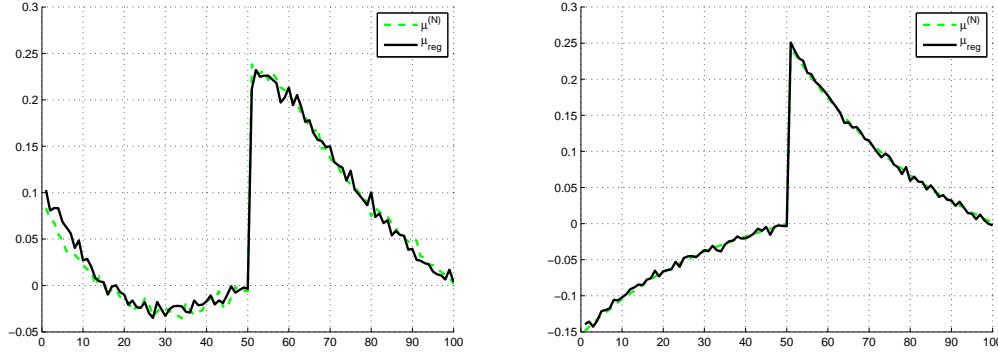

Fig. 6 Asian Call Option in Heston model: vectors $\mu^{(N)}$ and μ_{reg} in the cases [left panel] $(\xi_0, K) = (0.01; 120)$; [right panel] $(\xi_0, K) = (0.04; 70)$

6 Proofs

6.1 Proofs of Section 3

In the sequel, we denote $\mathcal{I}_m \overset{\mathrm{def}}{=} \{1, \ldots, I\}^m$.

Lemma 1 *Let $\{\mathsf{S}_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}_m\}$ be given by (8) and $p_{\mathbf{i}}\sigma_{\mathbf{i}} \overset{\mathrm{def}}{=} \sqrt{\left(\int_{\mathsf{S}_{\mathbf{i}}} f_\mu \, d\lambda\right) \left(\int_{\mathsf{S}_{\mathbf{i}}} \zeta_\mu f_\mu \, d\lambda\right) - \left(\int_{\mathsf{S}_{\mathbf{i}}} \psi_\mu f_\mu \, d\lambda\right)^2}$.*

   (i) $\forall \epsilon > 0$, $\forall M > \epsilon^{-1}$, $\sup_{\mathcal{Q}:\inf_{\mathbf{i} \in \mathcal{I}_m} q_{\mathbf{i}} \geq \epsilon} \left| M\varsigma_{I,M}^2(\mu, g, \mathcal{Q}) - \sum_{\mathbf{i} \in \mathcal{I}_m} \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}} \right| \leq \frac{1}{M\epsilon(\epsilon - M^{-1})} \, \mathrm{Var}[\phi(Y)]$.

   (ii) *Assume that $\mathrm{essinf}_{g\cdot\lambda}(\chi/g) > 0$ and $\mathrm{esssup}_{\chi\cdot\lambda}(f_\mu/\chi) < +\infty$. Let $\epsilon > 0$. For any (I, M) such that $MI^{-m} \, \mathrm{essinf}_{g\cdot\lambda}(\chi/g) \geq 1 + \epsilon$*

$$\left| M\varsigma_{I,M}^2(\mu, g, \mathcal{Q}_\chi) - \sum_{\mathbf{i} \in \mathcal{I}_m} \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}(\chi)} \right| \leq \frac{(1 + \epsilon^{-1})\mathrm{Var}[\phi(Y)]}{\mathrm{essinf}_{g\cdot\lambda}(\chi/g)} \, \frac{I^m}{M} \, \left(\mathrm{esssup}_{\chi\cdot\lambda}\left(\frac{f_\mu}{\chi}\right) \wedge \frac{I^m}{\mathrm{essinf}_{g\cdot\lambda}(\chi/g)} \right).$$

   (iii) *For any positive integers M, I and real $\epsilon > 1$,*

$$\left| M\varsigma_{I,M}^2(\mu, g, \mathcal{Q}^\star(\mu)) - \sum_{\mathbf{i} \in \mathcal{I}_m} \left(q_{\mathbf{i}}^\star(\mu)\right)^{-1} \, p_{\mathbf{i}}^2 \, \sigma_{\mathbf{i}}^2 \right| \leq \mathrm{Var}[\phi(Y)] \, \left((1 + \epsilon)\frac{I^m}{M} + \frac{1}{\epsilon - 1} \right),$$

   where $\mathcal{Q}^\star(\mu) = \{q_{\mathbf{i}}^\star(\mu), \mathbf{i} \in \mathcal{I}_m\}$ is the optimal allocation defined by (7).

Proof By definition of $M_{\mathbf{i}}$ (see Eq. 4), $M_{\mathbf{i}} = 0$ when $q_{\mathbf{i}} = 0$ and $M_{\mathbf{i}} \geq 1$ when $q_{\mathbf{i}} \geq M^{-1}$. One may have $M_{\mathbf{i}} = 1$ when $q_{\mathbf{i}} \in (0, M^{-1})$ but then $M M_{\mathbf{i}}^{-1} \leq q_{\mathbf{i}}^{-1}$. Hence,

$$\left| M \, \varsigma_{I,M}^2(\mu, g, \mathcal{Q}) - \sum_{\mathbf{i} \in \mathcal{I}_m : q_{\mathbf{i}} > 0} \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}} \right| \leq \sum_{\mathbf{i} \in \mathcal{I}_m, q_{\mathbf{i}} \geq 1/M} \left| \frac{M q_{\mathbf{i}} - M_{\mathbf{i}}}{M_{\mathbf{i}}} \right| \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}} + \sum_{\mathbf{i} \in \mathcal{I}_m : 0 < q_{\mathbf{i}} < 1/M} \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}} \quad (20)$$

(i) When $\inf_{\mathbf{i} \in \mathcal{I}_m} q_{\mathbf{i}} \geq \epsilon > M^{-1}$, the second term in the rhs is null and since by (4), $M q_{\mathbf{i}} - 1 < M_{\mathbf{i}} < M q_{\mathbf{i}} + 1$,

$$\sum_{\mathbf{i} \in \mathcal{I}_m, q_{\mathbf{i}} \geq 1/M} \left| \frac{M q_{\mathbf{i}} - M_{\mathbf{i}}}{M_{\mathbf{i}}} \right| \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}} \leq M^{-1} \left(\sup_{\mathbf{i} \in \mathcal{I}_m, q_{\mathbf{i}} \geq 1/M} p_{\mathbf{i}} q_{\mathbf{i}}^{-1} \right) \sum_{\mathbf{i} \in \mathcal{I}_m, q_{\mathbf{i}} \geq 1/M} (q_{\mathbf{i}} - M^{-1})^{-1} \, p_{\mathbf{i}} \sigma_{\mathbf{i}}^2, \quad (21)$$

which yields the desired result upon noting that $p_{\mathbf{i}} q_{\mathbf{i}}^{-1} \leq q_{\mathbf{i}}^{-1} \leq \epsilon^{-1}$ and $\sum_{\mathbf{i}} p_{\mathbf{i}} \sigma_{\mathbf{i}}^2 \leq \mathrm{Var}[\phi(Y)]$.

(ii) Under the stated assumptions, $q_{\mathbf{i}}(\chi) = \int_{\mathsf{S}_{\mathbf{i}}} \chi d\lambda \geq \mathrm{essinf}_{g \cdot \lambda} (\chi/g) \; I^{-m}$. Hence $M q_{\mathbf{i}} \geq 1 + \epsilon$ which implies that the second term in the rhs of (20) is null. This also implies that $q_{\mathbf{i}} - M^{-1} \geq \left(1 - \frac{1}{1+\epsilon} \right) \mathrm{essinf}_{g \cdot \lambda} (\chi/g) \; I^{-m}$. We conclude the proof by combining this bound with (21) and the following one :

$$\frac{p_{\mathbf{i}}}{q_{\mathbf{i}}(\chi)} = \frac{\int_{\mathsf{S}_{\mathbf{i}}} f_\mu \, d\lambda}{\int_{\mathsf{S}_{\mathbf{i}}} \chi \, d\lambda} \leq \mathrm{esssup}_{\chi \cdot \lambda} (f_\mu/\chi) \wedge \frac{1}{q_{\mathbf{i}}(\chi)} \leq \mathrm{esssup}_{\chi \cdot \lambda} (f_\mu/\chi) \wedge \frac{I^m}{\mathrm{essinf}_{g \cdot \lambda} (\chi/g)} \; .$$

(iii) Note that by convention, $p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2 / q_{\mathbf{i}}^\star(\mu) = 0$ when $q_{\mathbf{i}}^\star(\mu) = 0$. By definition of the optimal allocation (see Eq. 7),

$$\left(q_{\mathbf{i}}^\star(\mu) \right)^{-1} \, p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2 \leq q_{\mathbf{i}}^\star(\mu) \left(\sum_{\mathbf{j}} p_{\mathbf{j}} \sigma_{\mathbf{j}} \right)^2 \leq q_{\mathbf{i}}^\star(\mu) \, \mathrm{Var}[\phi(Y)] \; .$$

The second term in the rhs of (20) is upper bounded by $I^m M^{-1} \, \mathrm{Var}[\phi(Y)]$. For the first term,

$$[\mathrm{Var}[\phi(Y)]]^{-1} \sum_{\mathbf{i} \in \mathcal{I}_n, q_{\mathbf{i}}^\star(\mu) \geq 1/M} \left| \frac{M q_{\mathbf{i}}^\star(\mu) - M_{\mathbf{i}}}{M_{\mathbf{i}}} \right| \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}^\star(\mu)}$$

$$\leq \sum_{\mathbf{i} \in \mathcal{I}_m, 1/M \leq q_{\mathbf{i}}^\star(\mu) \leq \epsilon/M} \left| \frac{M q_{\mathbf{i}}^\star(\mu) - M_{\mathbf{i}}}{M_{\mathbf{i}}} \right| q_{\mathbf{i}}^\star(\mu) + \sum_{\mathbf{i} \in \mathcal{I}_m, q_{\mathbf{i}}^\star(\mu) \geq \epsilon/M} \left| \frac{M q_{\mathbf{i}}^\star(\mu) - M_{\mathbf{i}}}{M_{\mathbf{i}}} \right| q_{\mathbf{i}}^\star(\mu) \; .$$

For all $\mathbf{i}$ such that $q_{\mathbf{i}}^\star(\mu) \geq 1/M$, $M_{\mathbf{i}}^{-1} |M q_{\mathbf{i}}^\star(\mu) - M_{\mathbf{i}}| \leq 1$ which implies that

$$\sum_{\mathbf{i} \in \mathcal{I}_m, 1/M \leq q_{\mathbf{i}}^\star(\mu) \leq \epsilon/M} \left| \frac{M q_{\mathbf{i}}^\star(\mu) - M_{\mathbf{i}}}{M_{\mathbf{i}}} \right| q_{\mathbf{i}}^\star(\mu) \leq \frac{\epsilon I^m}{M} \; .$$

For all $\mathbf{i}$ such that $q_{\mathbf{i}}^\star(\mu) \geq \epsilon/M$, $M_{\mathbf{i}}^{-1} |M q_{\mathbf{i}}^\star(\mu) - M_{\mathbf{i}}| \leq M_{\mathbf{i}}^{-1} \leq (M q_{\mathbf{i}}^\star(\mu) - 1)^{-1} \leq (\epsilon - 1)^{-1}$ which implies that

$$\sum_{\mathbf{i} \in \mathcal{I}_m, q_{\mathbf{i}}^\star(\mu) \geq \epsilon/M} \left| \frac{M q_{\mathbf{i}}^\star(\mu) - M_{\mathbf{i}}}{M_{\mathbf{i}}} \right| q_{\mathbf{i}}^\star(\mu) \leq (\epsilon - 1)^{-1} \; .$$

Proof of Proposition 1 To prove the Proposition 1, we need the two following Lemmas. The first is a standard change of variables formula (see for example, [4, Theorem 4.1.11]). Define $G^{-1}(x_1, \ldots, x_m) \overset{\mathrm{def}}{=} (G_1^{-1}(x_1), \ldots, G_m^{-1}(x_m))$ where G_k is the c.d.f. associated to the density g_k on $\mathbb{R}$.

Lemma 2 *Let $h : \mathbb{R}^m \to \mathbb{R}$ be a measurable function. Assume that h is nonnegative or is such that $\int_{\mathbb{R}^m} |h| 1_{\{g > 0\}} \, d\lambda < +\infty$. Then, for all $0 \leq v_k \leq w_k \leq 1$, $k \in \{1, \ldots, I\}$*

$$\int_{\prod_{k=1}^m [G_k^{-1}(v_k), G_k^{-1}(w_k)]} h 1_{\{g > 0\}} \, d\lambda = \int_{\prod_{k=1}^m [v_k, w_k]} \frac{h}{g} \circ G^{-1} \, d\lambda \; . \quad (22)$$

The second technical Lemma is our key approximation result.

Lemma 3 *Let $h, \gamma : \mathbb{R}^m \to \mathbb{R}$ be functions such that $\int_{\mathbb{R}^m} \left(h^2 + \gamma^2 \right) / g \, d\lambda < +\infty$. Define for $\mathbf{i} \in \mathcal{I}_m$,*

$$R_{\mathbf{i}}[h, \gamma] \overset{\mathrm{def}}{=} \int_{\mathsf{S}_{\mathbf{i}}} h \gamma / g \, d\lambda - I^m \left(\int_{\mathsf{S}_{\mathbf{i}}} h \, d\lambda \right) \left(\int_{\mathsf{S}_{\mathbf{i}}} \gamma \, d\lambda \right) \; . \quad (23)$$

Then $\lim_{I \to +\infty} \sum_{\mathbf{i} \in \mathcal{I}_m} |R_{\mathbf{i}}[h, \gamma]| = 0$.

Proof By polarization, it is enough to prove the result when $\gamma = h$ with $\int_{\mathbb{R}^m} h^2/g \; d\lambda < +\infty$. This integrability condition ensures that λ-a.e. , $g = 0$ implies $h = 0$ and by (22), one has

$$R_{\mathbf{i}}[h,h] = \int_{\prod_{k=1}^{m}[(i_k-1)/I, i_k/I]} \frac{h^2}{g^2} \circ G^{-1} \; d\lambda - I^m \left(\int_{\prod_{k=1}^{m}[(i_k-1)/I, i_k/I]} \frac{h}{g} \circ G^{-1} \; d\lambda \right)^2 ,$$

where the right-hand-side is non-negative by Cauchy-Schwarz inequality. Set $\tilde{h}(u) \overset{\text{def}}{=} \frac{h}{g}(G^{-1}(u))$ if $u \in (0,1)^m$ and 0 otherwise. By (22) and the integrability assumption made on h, the function $\tilde{h}$ is square integrable on $\mathbb{R}^m$. Using the definition of $\tilde{h}$ for the first equality and symmetry for the second one, one has

$$\sum_{\mathbf{i} \in \mathcal{I}_m} R_{\mathbf{i}}[h,h] = I^m \sum_{\mathbf{i} \in \mathcal{I}_m} \int_{\mathcal{J}_{\mathbf{i}}^2} \tilde{h}(u)\{\tilde{h}(u) - \tilde{h}(v)\} du dv = \frac{I^m}{2} \sum_{\mathbf{i} \in \mathcal{I}_m} \int_{\mathcal{J}_{\mathbf{i}}^2} \{\tilde{h}(u) - \tilde{h}(v)\}^2 \; du dv$$

$$= \frac{I^m}{2} \sum_{\mathbf{i} \in \mathcal{I}_m} \int_{\mathcal{J}_{\mathbf{i}}} \int_{\mathcal{J}_{\mathbf{i}} - u_k} \{\tilde{h}(u) - \tilde{h}(u+w)\}^2 \; du dw \leq \frac{1}{2} \int_{[0,1]^m} \int_{[-1,1]^m} (\tilde{h}(u) - \tilde{h}(u+z/I))^2 \; du dz .$$

where we have set, for $\mathbf{i} \in \{1,\ldots,m\}$, $\mathcal{J}_{\mathbf{i}} = \prod_{k=1}^{m}[(i_k-1)/I, i_k/I]$. By continuity of the translations in $L^2(\mathbb{R}^m, du)$ and Lebesgue's Theorem, one obtains that the right-hand-side converges to 0 as $I \to \infty$.

We now proceed to the proof of Proposition 1. Under A1, it holds that

$$q_{\mathbf{i}}(\chi) \geq \left(\underset{g \cdot \lambda}{\operatorname{essinf}} \, (\chi/g) \right) \int_{\mathsf{S}_{\mathbf{i}}} g \; d\lambda = I^{-m} \underset{g \cdot \lambda}{\operatorname{essinf}} \, (\chi/g) . \tag{24}$$

Hence, by Lemma 1(i), to prove the first assertion, it is enough to check that $\lim_{I \to +\infty} \sum_{\mathbf{i} \in \{1,\ldots,I\}^m} \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}(\chi)} = \varsigma_\infty^2(\mu, \chi)$. By definition of $R_{\mathbf{i}}$ (see Eq. (23)),

$$\frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}(\chi)} = \frac{\int_{\mathsf{S}_{\mathbf{i}}} f_\mu^2 (\zeta_\mu - \psi_\mu^2)/g \; d\lambda - R_{\mathbf{i}}[f_\mu, \zeta_\mu f_\mu] + R_{\mathbf{i}}[\psi_\mu f_\mu, \psi_\mu f_\mu]}{I^m \int_{\mathsf{S}_{\mathbf{i}}} \chi \; d\lambda} ,$$

$$\varsigma_\infty^2(\mu, \chi) = \frac{\int_{\mathsf{S}_{\mathbf{i}}} f_\mu^2 (\zeta_\mu - \psi_\mu^2)/g \; d\lambda - R_{\mathbf{i}}[\chi, f_\mu^2(\zeta_\mu - \psi_\mu^2)\chi^{-1}]}{I^m \int_{\mathsf{S}_{\mathbf{i}}} \chi \; d\lambda} .$$

Therefore $\displaystyle \sum_{\mathbf{i} \in \mathcal{I}_m} \frac{p_{\mathbf{i}}^2 \sigma_{\mathbf{i}}^2}{q_{\mathbf{i}}(\chi)} - \varsigma_\infty^2(\mu, \chi) = \sum_{\mathbf{i} \in \mathcal{I}_m} \frac{R_{\mathbf{i}}[\chi, f_\mu^2(\zeta_\mu - \psi_\mu^2)\chi^{-1}] + R_{\mathbf{i}}[\psi_\mu f_\mu, \psi_\mu f_\mu] - R_{\mathbf{i}}[f_\mu, \zeta_\mu f_\mu]}{I^m \int_{\mathsf{S}_{\mathbf{i}}} \chi \; d\lambda}$,

and one easily concludes with (24) and Lemma 3 (which applies under A2 and A3). The second assertion is a consequence of Lemma 1(ii).

Proof of Proposition 2 For ease of notations, in this proof, the dependence upon μ and the strata $\{\mathsf{S}_{\mathbf{i}}, \mathbf{i} \in \mathcal{I}_m$ is omitted. We denote $p_{\mathbf{i}} \sigma_{\mathbf{i}} \overset{\text{def}}{=} \sqrt{\left(\int_{\mathsf{S}_{\mathbf{i}}} f_\mu \; d\lambda \right) \left(\int_{\mathsf{S}_{\mathbf{i}}} \zeta_\mu f_\mu \; d\lambda \right) - \left(\int_{\mathsf{S}_{\mathbf{i}}} \psi_\mu f_\mu \; d\lambda \right)^2}$. Since for $a, b \geq 0$, $|\sqrt{a} - \sqrt{b}| \leq \sqrt{|a-b|}$, one has

$$\sum_{\mathbf{i} \in \mathcal{I}_m} \left| p_{\mathbf{i}} \sigma_{\mathbf{i}} - \int_{\mathsf{S}_{\mathbf{i}}} \left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] \; d\lambda \right|$$

$$\leq \sum_{\mathbf{i} \in \mathcal{I}_m} \left| \int_{\mathsf{S}_{\mathbf{i}}} f_\mu \; d\lambda \int_{\mathsf{S}_{\mathbf{i}}} \zeta_\mu f_\mu \; d\lambda - \left(\int_{\mathsf{S}_{\mathbf{i}}} \psi_\mu f_\mu \; d\lambda \right)^2 - \left(\int_{\mathsf{S}_{\mathbf{i}}} \left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] \; d\lambda \right)^2 \right|^{1/2}$$

$$= \sum_{\mathbf{i} \in \mathcal{I}_m} \sqrt{\frac{1}{I^m}} \left| -R_{\mathbf{i}}[f_\mu, \zeta_\mu f_\mu] + R_{\mathbf{i}}[\psi_\mu f_\mu, \psi_\mu f_\mu] + R_{\mathbf{i}}\left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2}, f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] \right|$$

$$\leq \left(\sum_{\mathbf{i} \in \mathcal{I}_m} \left| -R_{\mathbf{i}}[f_\mu, \zeta_\mu f_\mu] + R_{\mathbf{i}}[\psi_\mu f_\mu, \psi_\mu f_\mu] + R_{\mathbf{i}}\left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2}, f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] \right| \right)^{1/2} .$$

Under A2, $\int f_\mu^2 (\zeta_\mu - \psi_\mu^2)/g \; d\lambda < +\infty$, and by Lemma 3, the right-hand-side converges to 0 as $I \to +\infty$. Therefore,

$$\lim_{I \to +\infty} \sum_{\mathbf{i} \in \mathcal{I}_m} \left| p_{\mathbf{i}} \sigma_{\mathbf{i}} - \int_{\mathsf{S}_{\mathbf{i}}} \left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] \; d\lambda \right| = 0 . \tag{25}$$

We write

$$\left(\int \left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] d\lambda \right) \sum_{\mathbf{i} \in \mathcal{I}_m} \left| q_{\mathbf{i}}(\chi_\mu^\star) - q_{\mathbf{i}}^\star(\mu) \right|$$

$$\leq \sum_{\mathbf{i} \in \mathcal{I}_m} q_{\mathbf{i}}^\star(\mu) \left| \sum_{\mathbf{j} \in \mathcal{I}_m} p_{\mathbf{j}} \sigma_{\mathbf{j}} - \int \left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] d\lambda \right| + \sum_{\mathbf{i} \in \mathcal{I}_m} \left| p_{\mathbf{i}} \sigma_{\mathbf{i}} - \int_{\mathsf{S}_{\mathbf{i}}} \left[f_\mu \sqrt{\zeta_\mu - \psi_\mu^2} \right] d\lambda \right| .$$

By Eq.(25), the rhs tends to zero as $I \to +\infty$. The second assertion is a consequence of Lemma 1(iii) applied with $\epsilon = \sqrt{M/I^m}$ and of Eq. (25).

6.2 Proofs of Section 4

We only give the proof of Proposition 3 and refer to [5] for the one of Corollary 1.

Proof of Proposition 3 Let $H \in \mathbb{R}^d$ be such that $|H| < |\mu|$, $e_1 = \frac{\mu}{|\mu|}$, $a = \langle H, e_1 \rangle$, $b = |H - ae_1|$ and e_2 be equal to $\frac{H-ae_1}{b}$ if $b \neq 0$ and to any vector with norm 1 orthogonal to e_1 otherwise. We complete (e_1, e_2) with $(e_3, \ldots, e_d)$ to obtain an orthonormal basis of $\mathbb{R}^d$. For $\alpha \in \mathbb{R}^d$, $\alpha_k = \langle \alpha, e_k \rangle$.

$$g_z(\mu + H) - g_z(\mu) = \int_{\{\alpha, \alpha_1 \leq \frac{z - \alpha_2 b}{|\mu| + a}\}} h(\alpha) \, d\alpha - \int_{\{\alpha, \alpha_1 \leq \frac{z}{|\mu|}\}} h(\alpha) \, d\alpha = \int_{\mathbb{R}^{d-1}} \int_{\frac{z}{|\mu|}}^{\frac{z - \alpha_2 b}{|\mu| + a}} h(\alpha) d\alpha_1 d\alpha_{2:d}$$

$$= - \int_{\mathbb{R}^{d-1}} \int_0^1 h\left(\frac{z - \alpha_2 bs}{|\mu| + as} e_1 + \sum_{k=2}^d \alpha_k e_k \right) \frac{az + \alpha_2 b|\mu|}{(|\mu| + as)^2} ds d\alpha_{2:d}$$

$$= - \int_0^1 \int_{\mathbb{R}^{d-1}} h\left(z \frac{(|\mu| + as)e_1 + bse_2}{(|\mu| + as)^2 + (bs)^2} + \sum_{k=3}^d \alpha_k e_k \right.$$

$$\left. + \left(\alpha_2 - \frac{zbs}{(|\mu| + as)^2 + (bs)^2} \right) \frac{(|\mu| + as)e_2 - bse_1}{|\mu| + as} \right) \frac{az + \alpha_2 b|\mu|}{(|\mu| + as)^2} d\alpha_{2:d} ds$$

$$= - \int_0^1 \int h(y) \frac{\langle y, H \rangle}{|\mu + sH|} d\lambda_z^{\mu + sH} ds , \tag{26}$$

where, for the last equality, we made the change of variable

$$\beta_2 = \frac{\sqrt{(|\mu| + as)^2 + (bs)^2}}{|\mu| + as} \alpha_2 - \frac{zbs}{(|\mu| + as)\sqrt{(|\mu| + as)^2 + (bs)^2}} ,$$

used the equality $(|\mu| + as)e_1 + bse_2 = \mu + sH$ and remarked that $\langle \mu + sH, y \rangle = z$ implies that $az + \langle y, e_2 \rangle b|\mu| = (|\mu| + as) \langle y, H \rangle$. Define, for $\nu \in \mathbb{R}_*^d$, $\gamma(h, \nu) \overset{\text{def}}{=} \int \frac{y}{|\nu|} h(y) d\lambda_z^\nu$. We deduce that

$$g_z(\mu + H) - g_z(\mu) + \left\langle H, \int \frac{y}{|\mu|} h(y) \, d\lambda_z^\mu \right\rangle = \left\langle H, \int_0^1 \{\gamma(h, \mu) - \gamma(h, \mu + sH)\} \, ds \right\rangle .$$

Consider now the following decomposition $\gamma(h, \nu) = \gamma\left(h\mathbb{1}_{\{|\cdot| > M\}}, \nu \right) + \gamma\left(h\mathbb{1}_{\{|\cdot| \leq M\}}, \nu \right)$. Under assumption 15, the first term in the rhs is arbitrarily small as M goes to infinity uniformly in ν close to μ. When $\nu \to \mu$, the measure $\mathbb{1}_{\{|\cdot| \leq M\}} \lambda_z^\nu$ converges weakly to $\mathbb{1}_{\{|\cdot| \leq M\}} \lambda_z^\mu$; hence, the second term converges to $\gamma\left(h\mathbb{1}_{\{|\cdot| \leq M\}}, \mu \right)$. Therefore, the function $\nu \mapsto \gamma(h, \nu)$ is continuous at μ and the conclusion follows easily.

References

1. A. Alfonsi. High order discretization scheme for the CIR process: application to affine term structure and Heston model. *Math. Comp.*, To appear.
2. B. Arouna. Adaptive Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1):1–24, 2004.

3. S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57 of *Stochastic Modelling and Applied Probability*. Springer, New York, 2007.

4. R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

5. P. Etore, G. Fort, B. Jourdain, and E. Moulines. On adaptive stratification, 2008.

6. P. Etore and B. Jourdain. Adaptive optimal allocation in stratified sampling methods. *Methodol. Comput. Appl. Probab.*, 9(2):117–152, to appear (2009).

7. P. Glasserman. *Monte Carlo methods in financial engineering*, volume 53 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 2004. Stochastic Modelling and Applied Probability.

8. P. Glasserman, P. Heidelberger, and P. Shahabuddin. Asymptotically optimal importance sampling and stratification for pricing path-dependent options. *Math. Finance*, 9(2):117–152, 1999.

9. A. Judistsky, G. Lan, A. Nemirovski, and A. Shapiro. Stochastic approximation approach to stochastic programming, 2007.

10. R. Kawai. Adaptive Monte Carlo variance reduction with two-time-scale stochastic approximation. *Monte Carlo Methods Appl.*, 13(3):197–217, 2007.

11. H. J. Kushner and G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.

12. A. Owen. Quasi Monte Carlo sampling. In Monte Carlo Ray Tracing : Siggraph 2003 Course, 2003.

13. Art B. Owen. A central limit theorem for Latin hypercube sampling. *J. Roy. Statist. Soc. Ser. B*, 54(2):541–551, 1992.

14. G. Ch. Pflug. *Optimization of stochastic models*. The Kluwer International Series in Engineering and Computer Science, 373. Kluwer Academic Publishers, Boston, MA, 1996. The interface between simulation and optimization.

15. R. Y. Rubinstein and D. P. Kroese. *The cross-entropy method*. Information Science and Statistics. Springer-Verlag, New York, 2004. A unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning.

16. R. Y. Rubinstein and D. P. Kroese. *Simulation and the Monte Carlo method*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2008.

Model		ν	q_i	Price	Variance					
v	K				MC	AdaptStr	GHS	μ_{reg}	$\mu_\star$	μ_l
0.1	45	0	prop	-	(8.63922; 8.64018)	-	-	(0.01736; 0.01744)	(0.01625; 0.01635)	(0.01672; 0.01673)
			opt	6.05	(8.63922; 8.64018)	(0.00348; 0.00353)	-	(0.00472; 0.00477)	(0.00355; 0.00362)	(0.00373; 0.00379)
		$\nu_\star$	prop	6.05	(0.80272; 0.80311)	-	(0.01375; 0.01376)	(0.01375; 0.01376)	(0.00777; 0.00778)	(0.00739; 0.00740)
			opt	6.05	(0.80272; 0.80311)	(0.00155; 0.00157)	-	(0.00513; 0.00519)	(0.00198;0.00203)	(0.00173; 0.00176)
0.5	45	0	prop	9.00	(158.067; 158.159)	-	-	(2.083; 2.091)	(2.126; 2.130)	(2.239; 2.247)
			opt	9.00	(158.067; 158.159)	(0.348; 0.356)	-	(0.358; 0.366)	(0.365; 0.377)	(0.388; 0.403)
		$\nu_\star$	prop	9.00	(14.9458; 14.9476)	-	(0.2027; 0.2028)	(0.2248; 0.2249)	(0.3244; 0.3246)	(0.2212; 0.2213)
			opt	9.00	(14.9458; 14.9476)	(0.1454; 0.1496)	-	(0.1597; 0.1628)	(0.2204; 0.2260)	(0.1588; 0.1632)
0.5	65	0	prop	2.16	(48.3833; 48.4423)	-	-	(1.8542; 1.8594)	(1.8560; 1.8617)	(2.0956; 2.0994)
			opt	2.16	(48.3833; 48.4423)	(0.0916; 0.0954)	-	(0.0939; 0.0972)	(0.0949; 0.0979)	(0.1437; 0.1490)
		$\nu_\star$	prop	2.16	(2.3203; 2.3204)	-	(0.0387; 0.0388)	(0.0457; 0.0458)	(0.0479; 0.0480)	(0.0488; 0.0489)
			opt	2.16	(2.3203; 2.3204)	(0.0199; 0.0207)	-	(0.0236; 0.0245)	(0.0251;0.0262)	(0.0257; 0.0268)
1	45	0	prop	14.01	(851.670; 852.314)	-	-	(52.011; 52.477)	(54.371; 54.654)	(57.458; 57.916)
			opt	14.01	(851.670; 852.314)	(5.245; 5.540)	-	(5.353; 5.608)	(5.537; 5.830)	(5.911; 6.157)
		$\nu_\star$	prop	14.01	(42.759; 42.763)	-	(3.012; 3.015)	(3.183; 3.187)	(4.359; 4.362)	(3.264; 3.266)
			opt	14.01	(42.759; 42.763)	(2.219; 2.320)	-	(2.365; 2.446)	(3.156; 3.298)	(2.409; 2.506)
1	65	0	prop	7.79	(586.14; 586.94)	-	-	(50.71; 51.05)	(50.51; 50.62)	(55.72; 55.91);
			opt	7.79	(586.14; 586.94)	(3.50; 3.98)	-	(2.96; 3.08)	(3.08;3.25)	(3.86; 4.02)
		$\nu_\star$	prop	7.78	(22.331; 22.337)	-	(1.550; 1.551)	(1.744; 1.746)	(2.003; 2.005)	(1.558; 1.559)
			opt	7.78	(22.331; 22.337)	(0.963; 1.017)	-	(1.109; 1.1562)	(1.272; 1.349)	(0.973; 1.025)

Table 1 Asian Option: Monte Carlo and stratification

Model		ν	Price	Variance			
v	K			Latin	Latin +Rot μ_{reg}	Latin +Rot $\mu_\star$	Latin +Rot $\mu^{(N)}$
0.1	45	0	6.05	(0.05929; 0.05985)	(0.00080; 0.00081)	(0.00075; 0.00076)	(0.00077; 0.00078)
		$\nu_\star$	6.05	(0.59677; 0.60301)	(0.00622; 0.00626)	(0.00093; 0.00094)	(0.00034; 0.00035)
0.5	45	0	9.00	(35.492; 35.607)	(0.373; 0.374)	(0.350; 0.351)	(0.384; 0.385)
		$\nu_\star$	9.00	(11.603: 11.851)	(0.164; 0.168)	(0.239; 0.245)	(0.136; 0.137)
0.5	65	0	2.16	(26.785; 29.454)	(0.148; 0.157)	(0.132; 0.140)	(0.142; 0.153)
		$\nu_\star$	2.16	(1.971; 2.023)	(0.042; 0.044)	(0.036; 0.037)	(0.032; 0.033)
1	45	0	14.00	(357.43; 357.94)	(10.85; 10.87)	(9.83; 9.85)	(12.19; 12.20)
		$\nu_\star$	14.00	(35.90; 36.67)	(2.31; 2.40)	(3.22; 3.29)	(2.09; 2.11)
1	65	0	7.78	(338.59; 339.85)	(7.93; 7.94)	(7.45; 7.94)	(9.13; 9.15)
		$\nu_\star$	7.78	(19.45; 19.83)	(1.47; 1.52)	(1.32; 1.35)	(1.24; 1.26)

Table 2 Asian Option: Latin Hypercube

Variance

Model K	B	Price	ν	q_i	MC	AdaptStr	GHS	μ_{reg}	$\mu_\star$	μ_l
50	60	1.38	0	prop	(2.9932; 2.9935)	-	-	(1.4594; 1.4596)	(1.1323; 1.1326)	(1.1388; 1.1392)
50	60	1.38	0	opt	(2.9932; 2.9935)	(0.3012; 0.3089)	-	(0.8202; 0.8293)	(0.3015; 0.3105)	(0.3059; 0.3134)
50	60	1.38	$\nu_\star$	prop	(1.3393; 1.3394)	-	(0.4968; 0.4972)	(1.1466; 1.1467)	(0.4898; 0.4901)	(0.4982; 0.4986)
50	60	1.38	$\nu_\star$	opt	(1.3393; 1.3394)	(0.1700; 0.1738)	-	(1.1153; 1.1232)	(0.2987; 0.3059)	(0.3050; 0.3129)
50	80	1.92	0	prop	(4.91575; 4.91727)	-	-	(0.01619; 0.01622)	(0.01655; 0.01667)	(0.01637; 0.01643)
50	80	1.92	0	opt	(4.91575; 4.91727)	(0.00174; 0.00178)	-	(0.00178; 0.00182)	(0.00182; 0.00186)	(0.00209; 0.00214)
50	80	1.92	$\nu_\star$	prop	(0.70357; 0.70366)	-	(0.00107; 0.00108)	(0.00124; 0.00125)	(0.00126; 0.00127)	(0.00107; 0.00108)
50	80	1.92	$\nu_\star$	opt	(0.70357; 0.70366)	(0.00046; 0.00047)	-	(0.00055; 0.00056)	(0.00057; 0.00058)	(0.00046; 0.00047)

Table 3 Barrier Option: Monte Carlo and stratification

Variance

Model K	B	Price	ν	Latin	Latin +Rot μ_{reg}	Latin +Rot $\mu_\star$	Latin +Rot $\mu^{(N)}$
50	60	1.38	0	(1.9750; 1.9963)	(1.5074; 1.5416)	(0.9584; 0.9837)	(0.9670; 0.9853)
50	60	1.38	$\nu_\star$	(1.2524; 1.2681)	(1.0048; 1.0209)	(0.3094; 0.3269)	(0.2108; 0.2157)
50	80	1.92	0	(0.72457; 0.73091)	(0.00234; 0.00243)	(0.00241; 0.00246)	(0.00241; 0.00246)
50	80	1.92	$\nu_\star$	(0.44731; 0.45464)	(0.00054; 0.00055)	(0.00055; 0.00056)	(0.00038; 0.00039)

Table 4 Barrier Option: Latin Hypercube

Variance

Model c	K	Price	ν	q_i	MC	AdaptStr	GHS	μ_{reg}	$\mu_\star$	μ_l
0.1	45	11.24	0	prop	(22.1636; 22.1664)	-	-	(0.2483; 0.2488)	(0.2488; 0.2490)	(0.2516; 0.2518)
0.1	45	11.24	0	opt	(22.1636; 22.1664)	(0.2163; 0.2190)	-	(0.2173; 0.2194)	(0.2162; 0.2189)	(0.2185; 0.2211)
0.1	45	11.24	$\nu_\star$	prop	(1.3935; 1.3939)	-	(0.2585; 0.2586)	(0.8709; 0.8711)	(0.2640; 0.2642)	(0.2625; 0.2627)
0.1	45	11.24	$\nu_\star$	opt	(1.3935; 1.3939)	(0.2070; 0.2101)	-	(0.6427; 0.6531)	(0.2122; 0.2145)	(0.2109; 0.2134)
0.5	45	11.56	0	prop	(81.126; 81.135)	-	-	(0.368; 0.371)	(0.367; 0.368)	(0.367; 0.368)
0.5	45	11.56	0	opt	(81.126; 81.135)	(0.099; 0.100)	-	(0.099; 0.100)	(0.099; 0.100)	(0.099; 0.101)
0.5	45	11.56	$\nu_\star$	prop	(8.633; 8.637)	-	(0.078; 0.079)	(0.086; 0.087)	(0.085; 0.086)	(0.080; 0.081)
0.5	45	11.56	$\nu_\star$	opt	(8.633; 8.637)	(0.061; 0.062)	-	(0.067; 0.068)	(0.066; 0.067)	(0.063; 0.064)
0.9	45	12.09	0	prop	(134.1479; 134.1654)	-	-	(0.7460; 0.7496)	(0.7308; 0.7350)	(0.7396; 0.7415)
0.9	45	12.09	0	opt	(134.1479; 134.1654)	(0.0533; 0.0547)	-	(0.0548; 0.0560)	(0.0540; 0.0553)	(0.0534; 0.0549)
0.9	45	12.09	$\nu_\star$	prop	(14.4573; 14.4636)	-	(0.0221; 0.0222)	(0.0294; 0.0295)	(0.0242; 0.0243)	(0.0229; 0.0230)
0.9	45	12.09	$\nu_\star$	opt	(14.4573; 14.4636)	(0.0080; 0.0081)	-	(0.0118; 0.0120)	(0.0094; 0.0095)	(0.0085; 0.0086)

Table 5 Basket Option: Monte Carlo and stratification

Variance

Model c	K	Price	ν	Latin	Latin +Rot μ_{reg}	Latin +Rot $\mu_\star$	Latin +Rot $\mu^{(N)}$
0.1	45	11.24	0	(0.0827; 0.0839)	(0.0664; 0.0680)	(0.0261; 0.0275)	(0.0311; 0.0326)
0.1	45	11.24	$\nu_\star$	(1.1652; 1.2109)	(0.9018; 0.9360)	(0.0476; 0.0505)	(0.0408; 0.0425)
0.5	45	11.56	0	(4.8999; 4.9886)	(0.0187; 0.0189)	(0.0184; 0.0186)	(0.0190; 0.0193)
0.5	45	11.56	$\nu_\star$	(6.8104; 6.9903)	(0.0206; 0.0208)	(0.0230; 0.0240)	(0.0138; 0.0140)
0.9	45	12.09	0	(12.8847; 13.3987)	(0.0072; 0.0074)	(0.0064; 0.0066)	(0.0070; 0.0072)
0.9	45	12.09	$\nu_\star$	(12.3153; 12.6230)	(0.0037; 0.0038)	(0.0025; 0.0026)	(0.0005; 0.0006)

Table 6 Basket Option: Latin Hypercube

Model				Price	Variance		
ξ_0	K	ν	q_i	-	MC	AdaptStr	μ_{reg}
0.01	120	0	prop	0.007	(0.02706; 0.02731)	-	(0.02342; 0.02352)
			opt	0.007	(0.02706; 0.02731)	(0.00031; 0.00033)	(0.00026; 0.00028)
	100	0	prop	5.19	(18.0654; 18.0702)	-	(2.0093; 2.0098)
			opt	5.19	(18.0654; 18.0702)	(1.6846; 1.7122)	(1.7116; 1.7426)
	80	0	prop	22.65	(30.0024; 30.0209)	-	(3.1629; 3.1646)
			opt	22.65	(30.0024; 30.0209)	(2.8168; 2.8737)	(2.9079; 2.9645)
0.04	130	0	prop	0.024	(0.15186; 0.15242)	-	(0.09828; 0.09901)
			opt	0.024	(0.15186; 0.15242)	(0.00142; 0.00148)	(0.00121; 0.00129)
	100	0	prop	6.42	(46.579; 46.591)	-	(2.372; 2.373)
			opt	6.42	(46.579; 46.591)	(1.673; 1.711)	(1.657; 1.693)
	70	0	prop	31.74	(88.358; 88.405)	-	(4.079; 4.080)
			opt	31.74	(88.358; 88.405)	(3.536; 3.610)	(3.673; 3.753)

Table 7 Asian Option in Heston model: Monte Carlo and stratification

Model		ν	Price	Variance		
ξ_0	K			Latin	Latin +Rot μ_{reg}	Latin +Rot $\mu^{(N)}$
0.01	120	0	0.007	(0.0271; 0.0273)	(0.0078; 0.0079)	(0.0087; 0.0088)
	100	0	5.19	(2.741; 2.803)	(2.059; 2.093)	(2.107; 2.186)
	80	0	22.65	(3.618; 3.679)	(2.340; 2.405)	(2.167; 2.236)
0.04	130	0	0.024	(0.1487; 0.1498)	(0.0256; 0.0258)	(0.0292; 0.0294)
	100	0	6.42	(7.513; 7.648)	(2.547; 2.641)	(3.188; 3.264)
	70	0	31.74	(3.096; 3.281)	(3.766; 3.854)	(3.500; 3.573)

Table 8 Asian Option in Heston model: Latin Hypercube