ORTHOGONAL MESH SAMPLING METHOD

—A New Monte Carlo Technique—

HIROTAKA SAKASEGAWA

(Received Sept. 18, 1975; revised Feb. 19, 1976)

1. Introduction

A Monte Carlo sampling method is a very attractive one to evaluate an integral $\theta=\int_D f(\boldsymbol{x})d\boldsymbol{x}$ where D denotes a region in d-dimensional Euclidian space. A simple (crude) Monte Carlo sampling procedure is executed as follows: let $\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N$ be points chosen at random from the region D and calculate function values at these points. Then the unbiased estimator θ is obtained by

$$\hat{\theta}_C=|D|\frac{f(\boldsymbol{x}_1)+\cdots+f(\boldsymbol{x}_N)}{N} \tag{1.1}$$

in which $|D|$ denotes the volume of D. It is well-known that a variance of the above estimator is asymptotically $O(N^{-1})$, so if one desires to estimate more accurately, say one decimal place more, he must increase his sample size hundred times or more. This is an unsatisfactory matter, and many considerations have been done to reduce a sample size or to estimate more accurately.

Haber [1] considered from the point of view of the numerical integration that (let D be a unit hypercube for simplicity) dividing D into congruent subcubes, one can estimate θ more accurate than $\hat{\theta}_C$ if one chooses points at random from each these subcubes one by one, not from D. However, this method, like many formulae in numerical integrations, will be infeasible when d becomes high; let each length of side of subcubes be K^{-1}, then a number of congruent subcubes becomes K^d, i.e. the quantity of computation increases exponentially as d increases linearly. From the idea that the more uniformly one takes sample points from the region, the more accurately one can estimate, we consider in this paper a new technique called orthogonal mesh sampling method. Using this method, the amount of computation increases linearly as d increases, and yet, we can estimate θ more accurately than the crude Monte Carlo method.

Now we explain the outline of this method in the following. Let D be a unit hypercube in d-dimensional Euclidian space, i.e. $D=[0,1]^d$. We divide D into congruent subcubes. All these subcubes, the number of which is, say K^d, have one to one correspondence to d-tuples $(\alpha_1, \cdots, \alpha_d)$ $(\alpha_i=0,1,\cdots,K-1;\ i=1,\cdots,d)$ in a natural manner. Then we choose K subcubes out of these K^d subcubes so that they are distributed uniformly in D. Let $(\alpha_1^1,\cdots,\alpha_d^1),\cdots,(\alpha_1^K,\cdots,\alpha_d^K)$ be K d-tuples corresponding to these selected subcubes. Requirement for uniformity of these K points in D means that each point $(\alpha_i^1,\cdots,\alpha_i^K)$ $(i=1,\cdots,d)$ must be distributed uniformly in $\{0,1,\cdots,K-1\}$. This will be satisfactorily realized by choosing $(\alpha_i^1,\cdots,\alpha_i^K)$ $(i=1,\cdots,d)$ as a random permutation of $\{0,1,\cdots,K-1\}$. Next, we choose a point at random from each selected K subcubes and using these K points $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_K\}$, calculate function values $f(\boldsymbol{x}_1),\cdots,f(\boldsymbol{x}_K)$ and take their average. Repeating this experiment as we desire, we estimate θ by taking an average of whole individual values.

We define the new estimator of θ and derive a variance of this estimator in the next section and in Section 3 some numerical examples are given.

2. The estimator and its variance

The estimator of this method is given as follows:

$$\hat{\theta}_K=\frac{1}{n}\sum_{j=1}^{n}\frac{1}{K}\sum_{k=1}^{K}f\left(\frac{\alpha_1^k+\xi_1^k}{K},\cdots,\frac{\alpha_d^k+\xi_d^k}{K}\right) \tag{2.1}$$

where $\{\alpha_i^1,\cdots,\alpha_i^K\}$ $(i=1,2,\cdots,d)$, same as in Section 1, denotes a random permutation of $\{0,1,\cdots,K-1\}$ (without loss of generality, let $\alpha_1^k=k-1$ for all k), ξ_i^k's are random variables which are distributed uniformly in $[0,1]$ and suffix j is omitted.

That the estimator $\hat{\theta}_K$ is unbiased is proved below.

$$\begin{aligned}\mathrm{E}(\hat{\theta}_K)&=\frac{1}{K}\sum_{k=1}^{K}\mathrm{E}\left(f\left(\frac{\alpha_1^k+\xi_1^k}{K},\cdots,\frac{\alpha_d^k+\xi_d^k}{K}\right)\right)\\&=\frac{1}{K}\sum_{k=0}^{K-1}\frac{1}{K}\sum_{l=0}^{K-1}\cdots\frac{1}{K}\sum_{m=0}^{K-1}\int_0^1\cdots\int_0^1 f\left(\frac{k+\xi_1}{K},\frac{l+\xi_2}{K},\cdots,\right.\\&\qquad\left.\frac{m+\xi_d}{K}\right)d\xi_1\cdots d\xi_d\\&=\sum_{k=1}^{K}\cdots\sum_{m=1}^{K}\int_{(k-1)/K}^{k/K}\cdots\int_{(m-1)/K}^{m/K}f(x_1,\cdots,x_d)dx_1\cdots dx_d=\theta\,.\end{aligned} \tag{2.2}$$

In order to compare this method with that of the crude Monte Carlo, we calculate a variance of $\hat{\theta}_K$.

$$
\begin{aligned}
(2.3)\qquad \operatorname{Var}(\hat{\theta}_K) &= \frac{1}{n}\frac{1}{K^2}\operatorname{Var}\left(\sum_{k=1}^{K} f\left(\frac{\alpha_1^k+\xi_1^k}{K},\cdots,\frac{\alpha_d^k+\xi_d^k}{K}\right)\right) \\
&= \frac{1}{nK}\frac{1}{K}\sum_{k=1}^{K}\operatorname{Var}\left(f\left(\frac{\boldsymbol{\alpha}^k+\boldsymbol{\xi}^k}{K}\right)\right) \\
&\quad + \frac{1}{nK}\frac{1}{K}\sum_{k\neq j}\operatorname{Cov}\left(f\left(\frac{\boldsymbol{\alpha}^k+\boldsymbol{\xi}^k}{K}\right), f\left(\frac{\boldsymbol{\alpha}^j+\boldsymbol{\xi}^j}{K}\right)\right),
\end{aligned}
$$

where we adopted the abbreviation $\boldsymbol{\alpha}^k$ and $\boldsymbol{\xi}^k$ as $(\alpha_1^k,\cdots,\alpha_d^k)$ and $(\xi_1^k,\cdots,\xi_d^k)$, respectively. Now,

$$
\begin{aligned}
&\operatorname{Var}\left(f\left(\frac{\boldsymbol{\alpha}^k+\boldsymbol{\xi}^k}{K}\right)\right) \\
&= \frac{1}{K}\sum_{l=0}^{K-1}\cdots\frac{1}{K}\sum_{m=0}^{K-1}\int_0^1\cdots\int_0^1 f^2\left(\frac{k-1+\xi_1}{K}, \frac{l+\xi_2}{K},\right. \\
&\qquad \left.\cdots, \frac{m+\xi_d}{K}\right)d\xi_1\cdots d\xi_d \\
&\quad -\left\{\frac{1}{K}\sum_{l=0}^{K-1}\cdots\frac{1}{K}\sum_{m=0}^{K-1}\int_0^1\cdots\int_0^1 f\left(\frac{k-1+\xi_1}{K}, \frac{l+\xi_2}{K},\right.\right. \\
&\qquad \left.\left.\cdots, \frac{m+\xi_d}{K}\right)d\xi_1\cdots d\xi_d\right\}^2 \\
&= K\int_{(k-1)/K}^{k/K}\int_0^1\cdots\int_0^1 f^2(\boldsymbol{x})d\boldsymbol{x} - K^2\left\{\int_{(k-1)/K}^{k/K}\int_0^1\cdots\int_0^1 f(\boldsymbol{x})d\boldsymbol{x}\right\}^2,
\end{aligned}
$$

accordingly, the first term in (2.3) becomes

$$
\frac{1}{nK}\int_D f^2(\boldsymbol{x})d\boldsymbol{x} - \frac{1}{n}\sum_{k=1}^{K}\left\{\int_{(k-1)/K}^{k/K}\int_0^1\cdots\int_0^1 f(\boldsymbol{x})d\boldsymbol{x}\right\}^2 .
$$

On the other hand,

$$
\begin{aligned}
&\mathrm{E}\left(f\left(\frac{\boldsymbol{\alpha}^k+\boldsymbol{\xi}^k}{K}\right) f\left(\frac{\boldsymbol{\alpha}^j+\boldsymbol{\xi}^j}{K}\right)\right) \\
&= \frac{1}{K(K-1)}\sum_{k_2\neq j_2}\cdots\frac{1}{K(K-1)}\sum_{k_d\neq j_d}\int_0^1\cdots\int_0^1 f\left(\frac{k-1+\xi_1}{K},\right. \\
&\qquad \left.\frac{k_2+\xi_2}{K},\cdots,\frac{k_d+\xi_d}{K}\right) f\left(\frac{j-1+\eta_1}{K}, \frac{j_2+\eta_2}{K},\cdots,\frac{j_d+\eta_d}{K}\right)d\boldsymbol{\xi}d\boldsymbol{\eta} \\
&= K^2\left(\frac{K}{K-1}\right)^{d-1}\sum_{k_2\neq j_2}\cdots\sum_{k_d\neq j_d}\int_{C_{kk_2\cdots k_d}} f(\boldsymbol{x})d\boldsymbol{x}\int_{C_{jj_2\cdots j_d}} f(\boldsymbol{x})d\boldsymbol{x}
\end{aligned}
$$

where $C_{kl\cdots m}$ denotes the region $[(k-1)/K, k/K]\times[(l-1)/K, l/K]\times\cdots\times[(m-1)/K, m/K]$. Next,

$$
\mathrm{E}\left(f\left(\frac{\boldsymbol{\alpha}^k+\boldsymbol{\xi}^k}{K}\right)\right) = K\int_{(k-1)/K}^{k/K}\int_0^1\cdots\int_0^1 f(\boldsymbol{x})d\boldsymbol{x} .
$$

Using two equations just above, the second term in (2.3) becomes

$$\frac{1}{n}\left(\frac{K}{K-1}\right)^{d-1}\sum_{(k)\neq(j)}\int_{C_{k_1\cdots k_d}} f(\boldsymbol{x})d\boldsymbol{x}\int_{C_{j_1\cdots j_d}} f(\boldsymbol{x})d\boldsymbol{x}$$
$$-\frac{1}{n}\sum_{k\neq j}\int_{C_{k\cdots\cdots}} f(\boldsymbol{x})d\boldsymbol{x}\int_{C_{j\cdots\cdots}} f(\boldsymbol{x})d\boldsymbol{x}$$

where $(k)\neq(j)$ means that $k_i\neq j_i$ for all i and $C_{k\cdots\cdots}=\bigcup_{l=1}^{K}\cdots\bigcup_{m=1}^{K} C_{kl\cdots m}$.

$$\text{(2.4)} \qquad \therefore \quad \mathrm{Var}\,(\hat{\theta}_K)=\frac{1}{nK}\int_D f^2(\boldsymbol{x})d\boldsymbol{x}-\frac{1}{n}\left\{\int_D f(\boldsymbol{x})d\boldsymbol{x}\right\}^2$$
$$+\frac{1}{n}\left(\frac{K}{K-1}\right)^{d-1}\sum_{(k)\neq(j)}\int_{C_{k_1\cdots k_d}} f(\boldsymbol{x})d\boldsymbol{x}\int_{C_{j_1\cdots j_d}} f(\boldsymbol{x})d\boldsymbol{x}\,.$$

Let $A_{kl\cdots m}$ be $\int_{C_{kl\cdots m}} f(\boldsymbol{x})d\boldsymbol{x}$ then the last term in (2.4) becomes

$$\frac{1}{n}\left(\frac{K}{K-1}\right)^{d-1}\{A^2_{\cdot\cdots\cdot}-\sum_k A^2_{k\cdots\cdots}-\sum_l A^2_{\cdot l\cdots\cdots}-\cdots-\sum_m A^2_{\cdots\cdots m}$$
$$+\sum_{k,l} A^2_{kl\cdots\cdots}+\cdots+(-1)^d\sum_{k,l,\cdots,m} A^2_{kl\cdots m}\}\,.$$

Finally,

$$\text{(2.5)} \qquad \mathrm{Var}\,(\hat{\theta}_K)=\frac{\sigma^2}{nK}+\frac{K-1}{nK}\left\{\left(\frac{K}{K-1}\right)^d-1\right\}\theta^2$$
$$-\frac{1}{n}\left(\frac{K}{K-1}\right)^{d-1}\{\sum_k A^2_{k\cdots\cdots}+\sum_l A^2_{\cdot l\cdots\cdots}+\cdots+\sum_m A^2_{\cdots\cdots m}$$
$$-\sum_{k,l} A^2_{kl\cdots\cdots}+\cdots+(-1)^{d-1}\sum_{k,l,\cdots,m} A^2_{kl\cdots m}\}$$

where $\sigma^2=\int_D f^2(\boldsymbol{x})d\boldsymbol{x}-\theta^2$. The first term in (2.5) equals to $\mathrm{Var}\,(\hat{\theta}_C)$ for the sample size $N=nK$, so $\mathrm{Var}\,(\hat{\theta}_K)$ is smaller than $\mathrm{Var}\,(\hat{\theta}_C)$ if and only if the reminder of the right-hand side of (2.5) is negative. Now the second and the third terms are transformed as follows:

$$\text{(2.6)} \qquad -\frac{1}{n}\left(\frac{K}{K-1}\right)^{d-1}\{V_1+V_2+\cdots+V_d-V_{12}-\cdots+(-1)^{d-1}V_{12\cdots d}\}$$

where $V_i=\sum_k A^2_{\cdots\overset{i}{k}\cdots\cdots}-\theta^2/K$, $V_{ij}=\sum_{k,l} A^2_{\cdots\overset{i}{k}\cdots\overset{j}{l}\cdots\cdots}-\theta^2/K^2$ etc. and they are all non-negative. This transformation suggests that (2.6) will be negative if $f(\boldsymbol{x})$ is non-negative in D and $\mathrm{Var}\,(\hat{\theta}_K)$ will be smaller than $\mathrm{Var}\,(\hat{\theta}_C)$, but this is not verified yet.

3. Numerical examples

The last description in the previous section is not accurate but more discussion is not made analytically at present. Making up for analytical difficulties we calculate variances numerically for a few functions in this section.

Example 1. $f(x_1,\cdots,x_d)=\exp(-x_1-\cdots-x_d)$

For large d, the value of this function changes severely in D, and the estimation of the integral by Monte Carlo method is quite difficult. It goes without saying that the integral of this function can be calculated analytically but this is chosen only because the integral value and the variance of $\hat{\theta}_K$ can be expressed explicitly and we can calculate these values exactly for any K and d, e.g.

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_K)=&\frac{\sigma^2}{N}+\frac{K-1}{N}\left\{\left(\frac{K}{K-1}\right)^d-1\right\}(1-e^{-1})^{2d}\\&-\frac{K}{N}\left(\frac{K}{K-1}\right)^{d-1}\sum_{\nu=1}^{d}(-1)^{\nu-1}\binom{d}{\nu}(1-e^{-1})^{2(d-\nu)}\\&\cdot\left\{\sum_{i=1}^{K}(e^{-(i-1)/K}-e^{-i/K})^2\right\}^{\nu}\end{aligned}$$

where $\sigma^2=\int_D f^2(\boldsymbol{x})d\boldsymbol{x}-\left(\int_D f(\boldsymbol{x})d\boldsymbol{x}\right)^2$ and $N=nK$ denotes a sample size. For various d and K, ratios of $\mathrm{Var}(\hat{\theta}_K)$ to $\mathrm{Var}(\hat{\theta}_C)$ $(=\sigma^2/N)$ are tabulated in Table 3.1 together with the ratio of a variance of the antithetic variate method (A.V.M., see Hammersley et al. [2]), where $1-\boldsymbol{x}$ is taken as the antithetic variate of $\boldsymbol{x}$, to $\mathrm{Var}(\hat{\theta}_C)$. Generally speaking,

Table 3.1 Variance ratios in Example 1

$\hat{\theta}_K : K$ \ d	2	3	4	5	6	7	8	9	10
2	0.318	0.365	0.408	0.449	0.487	0.522	0.556	0.586	0.615
3	0.171	0.218	0.263	0.306	0.346	0.384	0.421	0.455	0.487
4	0.116	0.162	0.206	0.248	0.288	0.327	0.363	0.398	0.431
5	0.090	0.135	0.178	0.219	0.259	0.297	0.333	0.368	0.401
8	0.061	0.104	0.145	0.185	0.223	0.260	0.295	0.329	0.362
10	0.054	0.096	0.136	0.175	0.213	0.249	0.284	0.318	0.350
20	0.044	0.084	0.123	0.161	0.197	0.232	0.266	0.299	0.331
50	0.041	0.080	0.118	0.154	0.190	0.224	0.258	0.290	0.322
100	0.040	0.079	0.116	0.153	0.188	0.222	0.255	0.288	0.319
1000	0.039	0.078	0.115	0.151	0.186	0.220	0.253	0.285	0.316
A.V.M.	0.107	0.176	0.240	0.300	0.353	0.403	0.449	0.492	0.531

the antithetic variate method is not much superior to the crude Monte Carlo method, but because of the monotonicity of the function of this example makes this method effctive. According to this table we conclude that the orthogonal mesh sampling method is superior to the crude Monte Carlo method for any K and d, and also is superior to the antithetic variate method for K greater than 4 and for any d.

Example 2. $f(x_1,\cdots,x_d)=\prod_{i=1}^{d}(\sqrt{2\pi})^{-1}\exp\left(-\frac{x_i^2}{2}\right)$

This is a well-known normal density with an identity matrix as a covariance matrix. As we cannot calculate $\mathrm{Var}(\hat{\theta}_K)$ exactly, simulation experiment is undertaken with a sample size $nK\approx 10^4$. For various d and K, ratios of $\mathrm{Var}(\hat{\theta}_K)$ to $\mathrm{Var}(\hat{\theta}_C)$ and the variance of the A.V.M. estimator to $\mathrm{Var}(\hat{\theta}_C)$ are tabulated in Table 3.2. Each experimental result in the table is the average of seven measurements. This method is superior to the crude Monte Carlo method for any d and K, and is also superior to the A.V.M. for any d and for any K greater than 4.

Table 3.2 Variance ratios in Example 2

$\hat{\theta}_K : K$ \ d	2	4	6	8	10
2	0.273	0.300	0.325	0.353	0.368
5	0.055	0.077	0.101	0.120	0.146
10	0.022	0.043	0.063	0.084	0.105
20	0.013	0.034	0.054	0.072	0.094
40	0.011	0.030	0.048	0.071	0.087
50	0.011	0.031	0.051	0.069	0.088
100	0.010	0.029	0.051	0.065	0.090
A.V.M.	0.096	0.128	0.159	0.193	0.226

Example 3. $f(x_1,\cdots,x_d)=(\sum x_i^2)(1+\sum x_i^2)^{-1}$

This function is chosen because it is indecomposable into simple integrals. Results of Monte Carlo simulation experiments are given in

Table 3.3 Variance ratios in Example 3

K \ d	2	4	6	8	10
2	0.292	0.358	0.373	0.370	0.380
5	0.095	0.147	0.161	0.156	0.158
10	0.063	0.110	0.120	0.123	0.116
20	0.054	0.102	0.105	0.110	0.102
50	0.050	0.095	0.099	0.106	0.100

Table 3.3. Each figure in the table represents the ratio of Var $(\hat{\theta}_K)$ to Var $(\hat{\theta}_C)$ for given d and K. Our method is again superior to the crude Monte Carlo method for any d and K.

4. Conclusion

There are many unknown behaviours in the orthogonal mesh sampling method presented here. It is necessary to show the family of functions useful for this method and the quantity of variance of the estimator or the variance ratio to the crude Monte Carlo method in order that our method is effective for general use. It is our regret however that we can only say that the variance of the estimator can be reduced about one tenth if the function value changes moderately such as Example 2 in the previous section.

Acknowledgement

The author wishes to thank the referee for his valuable suggestions and is indebted to him for the revision of Section 3.

THE INSTITUTE OF STATISTICAL MATHEMATICS

REFERENCES

[1] Haber, S. (1969). Stochastic quadrature formulas, *Math. Comp.* **23**, 751-764.
[2] Hammersley, J. M. et al. (1964). *Monte Carlo Method*, Methuen.