

K-ANTITHETIC VARIATES IN MONTE CARLO SIMULATION

ABDELAZIZ NASROALLAH

ABSTRACT. Standard Monte Carlo simulation needs prohibitive time to achieve reasonable estimations. for untractable integrals (i.e. multidimensional integrals and/or intergals with complex integrand forms). Several statistical technique, called variance reduction methods, are used to reduce the simulation time. In this note, we propose a generalization of the well known antithetic variate method. Principally we propose a K –antithetic variate estimator (KAVE) based on the generation of K correlated uniform variates. Some numerical examples are presented to show the improvement of our proposition.

1. INTRODUCTION

The approximation of integrals is one of the major class of problems that arise in statistical inference. There are a lot of situations that can be formulated as integration problems. Generally, in statistic context, integration is associated with the Bayesian approach [9].

Note that it is not always possible to derive explicit probabilistic models and even less possible to analytically compute the estimators associated with a given paradigm (maximum likelihood, Bayes, method of moments, etc ..). Moreover, other statistical methods, such as bootstrap methods [3], although unrelated to the Bayesian approach, may involve the integration of the empirical cumulative distribution function (cdf). Similarly, alternatives to standard likelihood, such as marginal likelihood in [1] for example, may require the integration of the nuisance parameters. Also, several optimization problems in statistical inference use numerical integration steps, such as EM algorithm in [6] and [2]. Here we are concerned with simulation based integration methods that uses the generation of random variables. Such methods are generally known as Monte Carlo simulation methods.

Basically, we interest to numerical approximation of untractable integrals (i.e. integrals in height dimensional case and/or with complex integrand form). It is well known that a standard Monte Carlo simulation of such integrals needs a prohibitive time to achieve reasonable estimations.

To reduce the simulation time, there are many statistical techniques, called variance reduction methods [8], [10], [5], [4], [9], [11], [7]. The *variance* is linked to simulation time via the sample size (the relation can be seen in confidence interval expression).

Our contribution, in numerical integration, is a generalization form of the known antithetic variate estimator [4]. Principally, we propose a K –antithetic variate estimator, based on K negatively correlated uniform variates. To generate such variates, the proposed algorithm ,namely KAVE, uses normal vectors on \mathbb{R}^K .

The improvement and efficiency of KAVE is tested on four numerical examples.

The paper is organized as follows: in section two, we present the standard Monte Carlo integration. In section three we present the antithetic variate method. In section four, we present the proposed estimator which generalizes the antithetic case, in the same section we give a proof of the variance reduction carried with respect to the standard case. We terminate the section by giving our generation algorithm. Finally, in section five, we give simulation results to confirm the improvement of our proposition.

2. STANDARD MONTE CARLO INTEGRATION

Let X be a real random variable with function density $f(x)\mathbb{1}_{[a,b]}(x)$, ($\mathbb{1}_A$ is the indicator function of set A). We denote expectation and variance operators by $\mathbb{E}[\cdot]$ and $Var(\cdot)$ respectively. We consider an integrable function g with respect to the measure $f(x)dx$, and we suppose that g is monotone. A presentation of standard Monte Carlo integration is easily accomplished by looking at the

generic problem of evaluating the integral

$$(1) \quad \theta = \mathbb{E}(g(X)) = \int_a^b g(x)f(x)dx.$$

It is natural to propose, using an independent and identically distributed (i.i.d) sample (X_1, \dots, X_n) generated from the density f , the following empirical average to approximate θ :

$$\theta_n = \frac{1}{n} \sum_{i=1}^n g(X_i),$$

since θ_n converges almost surely to $\mathbb{E}[g(X)]$ by the Strong Law of Large Numbers. Moreover, the speed of convergence of θ_n can be assessed. Construction of convergence test and confidence bounds on the approximation of θ can be made.

Remarks 1. . – Standard Monte Carlo integration in the multivariate case can be accomplished in a similar way.

– The use of Monte Carlo method for approximating any quantity on integral form is possible since the integral can be transformed to an expectation of a random variable.

– For simplicity and without loss of generality, we interest to the basic example $\theta = \int_0^1 g(x)f(x)dx$, where f is a density function of an absolute continuous real random X .

3. STANDARD ANTITHETIC VARIATES

Although a standard Monte Carlo simulation lead to i.i.d samples, it may be preferable to generate samples from correlated variables when estimating an integral, as they may reduce the variance of the corresponding estimator.

A first setting where the generation of independent sample is less desirable corresponds to the estimation of an integral θ written as sum (resp. difference) of two quantities which are close in value: let $\theta_1 = \int_0^1 g_1(x)f_1(x)dx$ and $\theta_2 = \int_0^1 g_2(x)f_2(x)dx$ such quantities, and θ_{1n} and θ_{2n} are independent estimators of θ_1 and θ_2 respectively, the variance of $\theta_n = \theta_{1n} + \theta_{2n}$ (resp. $\theta_{1n} - \theta_{2n}$)

is $var(\theta_{1n}) + var(\theta_{2n})$, which may be too large to support a fine enough analysis on the sum (resp. difference). However, if θ_{1n} and θ_{2n} are negatively (resp. positively) correlated, the variance is reduced by a factor $+2cov(\theta_{1n}, \theta_{2n})$ (resp. $-2cov(\theta_{1n}, \theta_{2n})$), which may greatly improve the analysis of the sum (resp. difference). The philosophy of the antithetic variates algorithm, as generally presented in the literature, is as follows:

$$\theta = \int_0^1 g(x)dx = \frac{1}{2} \int_0^1 g(x)dx + \frac{1}{2} \int_0^1 g(1-x)dx.$$

If (U_1, \dots, U_n) is an i.i.d sample of uniform variables in $[0, 1]$, then a standard estimator of θ , related to the antithetic variates is $\theta_n^{av} = \theta_{1n} + \theta_{2n}$, where

$$\theta_{1n} = \frac{1}{2n} \sum_{i=1}^n g(U_i) \quad \text{and} \quad \theta_{2n} = \frac{1}{2n} \sum_{i=1}^n g(1 - U_i).$$

We have

$$var(\theta_n^{av}) = var(\theta_{1n}) + var(\theta_{2n}) + 2cov(\theta_{1n}, \theta_{2n})$$

Since $cov(U_i, 1-U_i) < 0$, g is monotone and $var(\theta_{1n}) = var(\theta_{2n}) = \frac{1}{4}var(\theta_n)$, then

$$var(\theta_n^{av}) < var(\theta_n).$$

So the antithetic variates technique reduces the variance of the standard estimator of θ .

Remark 1. . θ_n^{av} and θ_n are both unbiased and converge in probability to θ .

In the following section, we propose a generalization of the variable antithetic method.

4. K-ANTITHETIC VARIATES

Let n and K be positif integers and let $S_n(K) := \cup_{i=1}^n \{U_1^i, \dots, U_K^i\}$ be a sample of nK uniform random on $]0, 1[$ such that $\forall i \in$

$\{1, 2, \dots, n\}$, $cov(U_k^i, U_j^i) < 0$ for $j \neq k$ and $cov(U_k^i, U_j^l) = 0$ for $i \neq l$. Now, for a fixed n , let $(S_n(K))_{K \in \mathbb{N}^*}$ be the sequence of samples defined by $S_n(K + 1) = S_n(K) \cup \{U_{K+1}^i, i = 1, \dots, n\}$, and consider the estimator $\theta_n(K)$ defined on the sample $S_n(K)$ by

$$(2) \quad \theta_n(K) = \frac{1}{nK} \sum_{i=1}^n \sum_{j=1}^K g(U_j^i).$$

This estimator is a generalization form of the standard estimator θ_n^{av} by writing θ in the form $\theta = \int_0^1 g(x)dx = \frac{1}{K} \left[\sum_{k=1}^K \int_0^1 g(x)dx \right]$.

Proposition 4.1. *For fixed n , we have*

$$var(\theta_n(K + 1)) \leq var(\theta_n(K)) \quad , \quad K = 1, 2, 3, \dots$$

Where $K = 1$ and $K = 2$ correspond to the standard Monte Carlo and standard antithetic variates case respectively.

Proof

Let's define $a = var(g(U_1^1))$ and $\alpha_K = 2 \sum_{i=1}^n \sum_{j=1}^K cov(g(U_j^i), g(U_{K+1}^i))$. By straightforward calculation, we obtain

$$var(\theta_n(K)) = \frac{a}{nK} + \frac{1}{(nK)^2} \sum_{k=1}^{K-1} \alpha_k, \quad K \geq 2$$

Let's note $\Delta V_n(K) = var(\theta_n(K + 1)) - var(\theta_n(K))$. We have

$$(nK(K + 1))^2 \Delta V_n(K) = -nK(K + 1)a + K^2 \alpha_K - (2K + 1) \sum_{k=1}^{K-1} \alpha_k$$

Using the Schwarz's inequality, we have $|\alpha_k| \leq 2nka$. So

$$-2(K + 1) \leq \frac{nK(K + 1)^2 \Delta V_n(K)}{a} \leq 0 \quad \square$$

To simulate $\theta_n(K)$, we need to generate K correlated uniform random variable. Such a procedure is based on the following lemma.

Lemma 4.2. *Let Z_1 and Z_2 be two standard gaussian random variables. Then*

$$(3) \quad cov(\phi(Z_1), \phi(Z_2)) = \frac{1}{2\pi} \arcsin\left(\frac{cov(Z_1, Z_2)}{2}\right),$$

where ϕ is the cdf of the standard gaussian law $N(0, 1)$.

The proof of this lemma is given in annexe.

Now, for $i = 1, \dots, n$, let $(X_1^{(i)}, \dots, X_K^{(i)})$ be a gaussian random vector $N_K(0, \Sigma^{(i)})$, where $\Sigma^{(i)} = (\sigma_{kj}^{(i)})_{1 \leq k, j \leq K}$ is a covariance matrix such that $\sigma_{kj}^{(i)} < 0$ for all $1 \leq k \neq j \leq K$.

For $1 \leq k \leq K$, $Z_k^{(i)} := \frac{X_k^{(i)}}{(\sigma_{kk}^{(i)})^{1/2}}$ is a standard gaussian variable $N(0, 1)$, for all $i = 1, \dots, n$.

Now, for each $i = 1, \dots, n$, applying the lemma 4.2 to each $(Z_k^{(i)}, Z_j^{(i)})$, $1 \leq k, j \leq K$, we get a sample $(\phi(Z_1^{(i)}), \dots, \phi(Z_K^{(i)}))$ of K negatively correlated uniform random.

4.1. K-Antithetic Variables Algorithm (KAVE). step 0 :
 $i = 0; \theta = 0$

step 1 : $i := i + 1$; give $\Sigma^{(i)}$ such that $\sigma_{kj}^{(i)} < 0$ for all $1 \leq k \neq j \leq K$

step 2 : generate a random vector $(X_1^{(i)}, \dots, X_K^{(i)})$ from $N_K(0, \Sigma^{(i)})$ and take

$$(U_1^{(i)}, \dots, U_K^{(i)}) := (\phi(Z_1^{(i)}), \dots, \phi(Z_K^{(i)})), \text{ where } Z_k^{(i)} = \frac{X_k^{(i)}}{(\sigma_{kk}^{(i)})^{1/2}}$$

step 3 : compute $a = \sum_{k=1}^K g(U_k^i)$

step 4 : $\theta = \theta + a$, if $i < n$ then goto **step 1**

step 5 : $\theta_n(K) = \frac{\theta}{nK}$ is the estimator.

Remark 2. . the statistical table of $N(0, 1)$ is used to get $\phi(\cdot)$.

5. NUMERICAL EXAMPLES

In the present section, we present four numerical examples to show the effectiveness of our algorithm. Since the number of arithmetic operations and the number of generated random is different from a situation to the other, we compare KAVE algorithm and

standard Antithetic variables Algorithm using CPU time.

For each example, we give estimation accompanied, with standard deviation estimation between brackets, for different CPU times and different values of K . For each example the Monte Carlo simulation result is summarized in a table. We denote $t_n(K)$ the estimation of $\theta_n(K)$ where n is fixed by CPU time and s_K the standard deviation corresponding to $t_n(K)$. For all examples, we consider $K = 1, 2, 3, 5$ and 7 . The case $K = 1$ correspond to the standard Monte Carlo simulation and $K = 2$ is similar to the standard antithetic variates. Simulations are carried on a Pinthium 4 computer.

In all these examples, KAVE works well and outperform the standard cases. It gives good estimations with reduced standard deviation. Its performances increase with K for the same CPU time. So the proposed generalization of the standard antithetic variates method, KAVE, can be a competitive algorithm in Monte Carlo integration context.



Example 1 : $\theta = \int_0^1 du$ (the exact value is 0.5)

CPU	$t_n(1)(s_1)$	$t_n(2)(s_2)$	$t_n(3)(s_3)$	$t_n(5)(s_5)$	$t_n(7)(s_7)$
8	0.5003(0.2889)	0.5000(0.2037)	0.5007(0.1661)	0.4999(0.1280)	0.5006(0.1081)
16	0.5005(0.2886)	0.5000(0.2046)	0.4999(0.1643)	0.4999(0.1276)	0.5005(0.1077)
24	0.4971(0.2882)	0.5000(0.2039)	0.5000(0.1661)	0.5004(0.1300)	0.5000(0.1077)

Table 1 : estimation of $\theta_n(K)$ with it's standard deviation between brackets for different values of K and different CPU times.

Example 2 : $\theta = \int_0^1 \log |0.5 - y| dy$ (the exact value is $\theta = -1.6931$)

CPU	$t_n(1)(s_1)$	$t_n(2)(s_2)$	$t_n(3)(s_3)$	$t_n(5)(s_5)$	$t_n(7)(s_7)$
8	-1.6952(1.0119)	-1.6931(0.7056)	-1.6996(0.5842)	-1.6962(0.4401)	-1.7016(0.3784)
16	-1.6901(0.9982)	-1.6919(0.7080)	-1.7018(0.5878)	-1.7015(0.4500)	-1.6928(0.3698)
24	-1.6957(1.0062)	-1.6932(0.7165)	-1.6837(0.5851)	-1.6912(0.4511)	-1.6953(0.3754)

Table 2 : estimation of $\theta_n(K)$ with it's standard deviation between brackets for different values of K and different CPU times.

Example 3 : $\theta = \int_0^1 \int_0^1 \int_0^1 u_1 u_2 u_3 du_1 du_2 du_3$ (the exact value is $\theta = 0.125$)

CPU	$t_n(1)(s_1)$	$t_n(2)(s_2)$	$t_n(3)(s_3)$	$t_n(5)(s_5)$	$t_n(7)(s_7)$
30	0.1260(0.1479)	0.1250(0.0509)	0.1251(0.0266)	0.1249(0.0146)	0.1251(0.0132)
40	0.1241(0.1452)	0.1250(0.0509)	0.1257(0.0264)	0.1252(0.0135)	0.1291(0.0109)
50	0.1256(0.1473)	0.1250(0.0509)	0.1255(0.0293)	0.1270(0.0125)	0.1276(0.0056)

Table 3 : estimation of $\theta_n(K)$ with it's standard deviation between brackets for different values of K and different CPU times.

Example 4 : $\theta = \int_0^1 \int_0^1 \int_0^1 \int_0^1 u_1 u_2 u_3 u_4 u_5 du_1 du_2 du_3 du_4 du_5$ (the exact value is $\theta = 0.0312$)

CPU	$t_n(1)(s_1 \times 10)$	$t_n(2)(s_2 \times 10^2)$	$t_n(3)(s_3 \times 10^2)$	$t_n(5)(s_5 \times 10^3)$	$t_n(7)(s_7 \times 10^3)$
40	0.0314(0.56)	0.0313(0.98)	0.0314(0.39)	0.0313(0.71)	0.0319(0.45)
50	0.0315(0.57)	0.0313(0.99)	0.0311(0.31)	0.0312(1.22)	0.0308(0.54)
60	0.0313(0.56)	0.0313(0.98)	0.0313(0.35)	0.0314(0.97)	0.0311(0.36)

Table 4 : estimation of $\theta_n(K)$ with it's standard deviation between brackets for different values of K and different CPU times.

REFERENCES

- [1] Barndorff-Nielsen O. and Cox D. R. (1994) *Inference and Asymptotics*. Chapman and Hall, London.
- [2] Dempster A. P., Laird N. M. and Rubin D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.
- [3] Efron B., Tibshirani R. J. (1994) *An Introduction to the bootstrap*. Chapman and Hall, London.
- [4] Hammersley J. M. and Handscomb D. C. (1964) *Monte Carlo Methods*. John Wiley, New York.
- [5] Law A. M. and Kelton W. D. (1991) *Simulation Modeling and Analysis*. McGraw-Hill, New York.
- [6] McLachlan G. J. and Krishnam T. (1997) *The EM algorithm and extensions*. John Wiley and sons, New York-London, Sydney-Toronto.
- [7] M. S. (2001) On the use of antithetic variates in particle transport problems. *Annals of Nuclear Energy*, Vol. 28, Num. 4, pp. 297-332(36).
- [8] Nasroallah A. (1991) *Simulation de chaînes de Markov et techniques de réduction de la variance*. Thèse de Doctorat IRISA, Rennes.
- [9] Robert C. P., Casella G. (1999) *Monte Carlo Statistical Methods*. Springer-Verlag, New York.
- [10] Rubinstein R. Y. (1981) *Simulation and the Monte Carlo Method*. John Wiley and Sons, New York.
- [11] Tuffin B. (1996) Variance Reductions Applied to Product-Form Multi-Class Networks; Antithetic Variates and Low Discrepancy Sequences. *Pub. Int. 1005 (Model)*, IRISA, Rennes.



APPENDIX A. PROOF OF LEMMA 4.2

Let X, Y and Z be three independent d -dimensional centered gaussian vectors with covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq d}$. The proof need the following auxillary result:

If $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) are i.i.d gaussian couples, then (X_1, Y_1) and (X_2, Y_3) are independent.

Now return to the proof of our principal lemma: we denote ϕ_i the cdf of X_i , we interest to the computation of $cov(\phi_i(X_i), \phi_j(X_j))$ for $i \neq j$.

Let

$$\alpha_{ij} = \mathbb{E}[\phi_i(X_i)\phi_j(X_j)]$$

we have,

$$\alpha_{ij} = \int \int \phi_i(s)\phi_j(t)dH_{X_iX_j}(s, t),$$

where $H_{X_iX_j}(s, t)$ is the cdf of (X_i, X_j) .

Now X_i, Y_i and Z_j are identically distributed for $i, j = 1, \dots, d$. So

$$\alpha_{ij} = \int \int \mathbb{P}(Y_i < s)\mathbb{P}(Z_j < t)dH_{X_iX_j}(s, t).$$

Since Y and Z are independent, one can easily proof that Y_i and Z_j for $i \neq j$ are independent. So

$$\alpha_{ij} = \int \int \mathbb{P}(Y_i < s, Z_j < t)dH_{X_iX_j}(s, t).$$

Now by the above auxiliary result, we get

$$\begin{aligned} \alpha_{ij} &= \int \int \mathbb{P}(Y_i < x, Z_j < y / X_i = x, X_j = y)dH_{X_iX_j}(x, y) \\ &= \mathbb{P}(Y_i < X_i, Z_j < X_j). \end{aligned}$$

Let $V = X - Y$ and $W = X - Z$, one can easily verify that (V_i, W_j) is $N_2(0, \Sigma^{(ij)})$, where

$$\Sigma^{(ij)} = \begin{pmatrix} 2\sigma_{ii} & \sqrt{\sigma_{ii}\sigma_{jj}}\rho_{ij} \\ \sqrt{\sigma_{ii}\sigma_{jj}}\rho_{ij} & 2\sigma_{jj} \end{pmatrix} \text{ and } \rho_{ij} = cov(X_i, X_j).$$

So

$$\alpha_{ij} = P(V_i > 0, W_j > 0).$$

Now, let A and B be two independent and centered gaussian random variables with the same variance 1, then the vector (A, B) is $N_2(0, I_2)$. Let

$$N = \begin{pmatrix} \sqrt{2\sigma_{ii}}\sqrt{1 - \frac{\rho_{ij}^2}{4}} & \sqrt{\frac{\sigma_{ii}}{2}}\rho_{ij} \\ 0 & \sqrt{2\sigma_{jj}} \end{pmatrix}.$$

We have $NN' = \Sigma^{(ij)}$, where prime stand for transpose. It is easy to see that (V_i, W_j) and $N(A, B)'$ have the same law. So

$$\begin{aligned} \alpha_{ij} &= \mathbb{P} \left(\sqrt{2\sigma_{ii}}\sqrt{1 - \frac{\rho_{ij}^2}{4}}A + \sqrt{\sigma_{ii}}\frac{\rho_{ij}}{\sqrt{2}}B > 0 \ , \ \sqrt{2\sigma_{jj}}B > 0 \right) \\ &= \mathbb{P} \left(\sqrt{1 - \frac{\rho_{ij}^2}{4}}A + \frac{\rho_{ij}}{2}B > 0 \ , \ B > 0 \right). \end{aligned}$$

Now consider the transformation T given by $T : (A, B) \rightarrow (R \cos(\Theta) \ , \ R \sin(\Theta))$, where $R = \sqrt{A^2 + B^2}$ and Θ is uniform on $[0, 2\pi]$. Using this transformation, we get So

$$\alpha_{ij} = \mathbb{P} \left(\sqrt{1 - \frac{\rho_{ij}^2}{4}}R \cos(\Theta) + \frac{\rho_{ij}}{2}R \sin(\Theta) > 0 \ , \ R \sin(\Theta) > 0 \right).$$

Since $R > 0$, then

$$\begin{aligned} \alpha_{ij} &= \mathbb{P} \left(\sqrt{1 - \frac{\rho_{ij}^2}{4}}\cos(\Theta) + \frac{\rho_{ij}}{2}\sin(\Theta) > 0 \ , \ \sin(\Theta) > 0 \right) \\ &= \mathbb{P} \left(\cos \left[\arcsin \left(\frac{\rho_{ij}}{2} \right) - \Theta \right] > 0 \ , \ \sin(\Theta) > 0 \right) \\ &= \mathbb{P} \left(\Theta \in \left] \arcsin \left(\frac{\rho_{ij}}{2} \right) - \frac{\pi}{2} \ , \ \arcsin \left(\frac{\rho_{ij}}{2} \right) + \frac{\pi}{2} \right[\cap]0, \pi[\right) \\ &= \mathbb{P} \left(\Theta \in \left] 0, \arcsin \left(\frac{\rho_{ij}}{2} \right) + \frac{\pi}{2} \right[\right). \end{aligned}$$

Since Θ is uniform in $[0, 2\pi]$, then

$$\alpha_{ij} = \frac{\arcsin\left(\frac{\rho_{ij}}{2}\right) + \frac{\pi}{2}}{2\pi}.$$

Finally, since $\phi_i(X_i)$ $i = 1, 2, \dots, d$ are uniform, then

$$\text{cov}(\phi_i(X_i), \phi_j(X_j)) = \alpha_{ij} - \frac{1}{4} = \frac{\arcsin\left(\frac{\rho_{ij}}{2}\right)}{2\pi}.$$

E-mail address: Email : nasroallah@ucam.ac.ma

CADI AYYAD UNIVERSITY, FACULTY OF SCIENCES SEMLALIA, MARRAKESH

THIS RESEARCH IS SUPPORTED BY LABORATOIRE IBN-AL-BANNA DE MATHÉMATIQUES
ET APPLICATIONS (LIBMA)

