Integration of bounded monotone functions: Revisiting the nonsequential case, with a focus on unbiased Monte Carlo (randomized) methods

Subhasish Basak, Julien Bect, Emmanuel Vazquez

Integration of bounded monotone functions: Revisiting the nonsequential case, with a focus on unbiased Monte Carlo (randomized) methods

Subhasish Basak[1,2,⋆] & Julien Bect[2] & Emmanuel Vazquez[2]

[1] *ANSES, Maison-Alfort, France.* ⋆*E-mail :* subhasish.basak@centralesupelec.fr
[2] *Université Paris-Saclay, CNRS, CentraleSupélec,*
Laboratoire des signaux et systèmes, Gif-sur-Yvette, France.

Résumé. Dans cet article, nous revisitons le problème de l'intégration numérique d'une fonction monotone bornée, en nous concentrant sur la classe des méthodes de Monte Carlo non séquentielles. Nous établissons dans un premier une borne inférieure pour l'erreur maximale dans L^p d'un algorithme non séquentiel, qui généralise pour $p > 1$ un théorème de Novak. Nous étudions ensuite, dans le cas $p = 2$, l'erreur maximale de deux méthodes sans biais—une méthode fondée sur l'utilisation d'une variable de contrôle, et la méthode de l'échantillonnage stratifié.

Mots-clés. Intégration monotone, Monte Carlo, échantillonnage stratifié, échantillonnage hypercube latin, variable de contrôle, complexité

Abstract. In this article we revisit the problem of numerical integration for monotone bounded functions, with a focus on the class of nonsequential Monte Carlo methods. We first provide new a lower bound on the maximal L^p error of nonsequential algorithms, improving upon a theorem of Novak when $p > 1$. Then we concentrate on the case $p = 2$ and study the maximal error of two unbiased methods—namely, a method based on the control variate technique, and the stratified sampling method.

Keywords. Monotone integration, Monte Carlo, stratified sampling, Latin hypercube sampling, control variate, information-based complexity

1 Introduction

We address in this article the problem of constructing a numerical approximation of the expectation $\mathbb{E}(g(Y)) = \int g(y)\, \mathrm{P}_Y(\mathrm{d}y)$, where Y is a real random variable with known distribution P_Y, and g is a real function that is bounded and monotone. Such a problem occurs naturally in applications where one is interested in computing a risk using a model that provides an increasing conditional risk $g(Y)$ with respect to some random variable Y. This situation occurs for instance in the field of food safety, with Y a dose of pathogen and $g(Y)$ the corresponding probability of food-borne illness (see, e.g., Perrin et al., 2014).

Assuming that the cumulative distribution function of Y is continuous, the problem reduces after change of variable and scaling to the computation of $S(f) = \mathbb{E}(f(X)) = \int_0^1 f(x)\,\mathrm{d}x$, where X is uniformly distributed on $[0,1]$ and f belongs to the class F of all non-decreasing functions defined on $[0,1]$ and taking values in $[0,1]$. We work in this article in a fixed sample-size setting, where the number n of evaluations of f to be performed is chosen beforehand.

This problem was first studied by Kiefer (1957), who proved that considering regularly-spaced evaluations at $x_i = i/(n+1)$, $1 \le i \le n$, and then using the trapezoidal integration rule assuming $f(0) = 0$ and $f(1) = 1$, is optimal in the worst-case sense among all deterministic—possibly sequential—methods, with maximal error $1/(2(n+1))$. Novak (1992) later studied Monte Carlo (a.k.a. randomized) methods, and established that sequential methods are better in this setting than nonsequential ones, with a minimax rate of $n^{-3/2}$ over F for the L^1 error. Novak's proof relies on the construction of a particular two-stage algorithm, using stratified sampling in the second stage.

This article revisits the nonsequential setting with a focus on unbiased Monte Carlo methods, which are a key building block for the construction of good (rate-optimal) sequential methods—as can be learned from the proof of Theorem 3 in Novak's article. Section 2 derives a lower bound for the maximal L^p-error of nonsequential methods, for any $p \ge 1$, which is a generalization of a result by Novak (1992) concerning the L^1 error. Sections 3 and 4 then study the maximal L^2 error (variance) of two simple unbiased methods, based respectively on the control variate technique and on stratification. Section 5 concludes the article with a discussion.

2 A lower bound for the maximal L^p error

A nonsequential (also called non-adaptive) Monte Carlo method first evaluates the function at n random points $X_1, \ldots, X_n$ in $[0,1]$, and then approximates the integral $S(f)$ using an estimator

$$\hat{S}_n(f) = \varphi\left(X_1,\, f(X_1),\, \ldots,\, X_n,\, f(X_n)\right), \tag{1}$$

where $\varphi : [0,1]^{2n} \to \mathbb{R}$ is a measurable function. A nonsequential method is thus defined by two ingredients: the distribution of $(X_1, \ldots, X_n)$ and the function φ. The worst-case L^p error of such a method over the class F is

$$e_p(\hat{S}_n) \;=\; \sup_{f \in F} \mathbb{E}\left(\left|S(f) - \hat{S}_n(f)\right|^p\right)^{1/p}. \tag{2}$$

Remark 1. The class of nonsequential Monte Carlo methods as usually defined in the literature also allows $\hat{S}_n$ to be randomized (i.e., allows φ to be a random function). We have not considered randomized estimators in our definition, however, since Rao-Blackwell's theorem implies that they do not help in this setting, for any convex loss function.

Novak (1992) proved that for any nonsequential Monte Carlo method with sample size n, the maximal L^1 error $e_1(\hat{S}_n)$ is greater or equal to $1/(8n)$. We generalize this result to the case of the L^p error.

Theorem 2.1. *For any nonsequential Monte Carlo methods with sample size n,*

$$e_p(\hat{S}_n) \;\geq\; \left(\frac{1}{2}\right)^{2+1/p} \frac{1}{n}\,.$$

Observe that Novak's lower bound is recovered for $p = 1$. Using Theorem 2.1 with $p = 2$ we can deduce of lower bound for the variance of unbiased nonsequential methods.

Corollary 2.2. *For any unbiased nonsequential Monte Carlo method with sample size n,*

$$\sup_{f\in F} \mathrm{var}\left(\hat{S}_n(f)\right) \;\geq\; \frac{1}{32n^2}.$$

Proof of Theorem 2.1. Consider a nonsequential Monte Carlo methods with evaluation points $X_1, \ldots, X_n$ and estimator $\hat{S}_n$. Divide the interval $[0,1]$ into $2n$ equal subintervals of length $1/(2n)$: then at least one of the subintervals, call it I, will contain no evaluation point with probability at least $1/2$. Now construct two functions $f_1, f_2 \in F$ that are both equal to zero on the left of I, equal to one on the right, and such that $f_1 = 1$ and $f_2 = 0$ on I. Then $S(f_1) - S(f_2) = 1/(2n)$, and $\hat{S}_n(f_1) = \hat{S}_n(f_2)$ on the event $A = \{\{X_1, X_2, \ldots, X_n\} \cap I = \emptyset\}$, since f_1 and f_2 coincide outside of I. It follows that

$$\begin{aligned}
\left(e_p(\hat{S}_n)\right)^p \;&\geq\; \sup_{f\in\{f_1,f_2\}} \mathbb{E}\Big(|S(f) - \hat{S}_n(f)|^p\Big) \;\geq\; \frac{1}{2}\sum_{j=1}^{2} \mathbb{E}\Big(|S(f_j) - \hat{S}_n(f_j)|^p\Big)\\
&\geq\; \frac{1}{2}\sum_{j=1}^{2} \mathbb{E}(|S(f_j) - \hat{S}_n(f_j)|^p \cdot \mathbb{1}_A) \;=\; \frac{1}{2}\sum_{j=1}^{2} \mathbb{E}(|S(f_j) - T|^p \cdot \mathbb{1}_A),
\end{aligned}$$

where T denotes the common value of $\hat{S}_n(f_1)$ and $\hat{S}_n(f_2)$ on A. We conclude that

$$\left(e_p(\hat{S}_n)\right)^p \;\geq\; \frac{1}{2}\,\frac{|I|^p}{2^{p-1}}\,P(A) \;\geq\; \left(\frac{1}{2}\right)^{2p+1} \cdot \frac{1}{n^p},$$

using the fact that, for any $a, b, x \in \mathbb{R}$ and $p \geq 1$, $|a-x|^p + |b-x|^p \geq |a-b|^p/2^{p-1}$. □

3 Uniform i.i.d. sampling

The simple Monte Carlo method is the most common example of a nonsequential method: the evaluation points $X_1, \ldots, X_n$ are drawn independently, uniformly in $[0,1]$, and then the integral is estimated by $\hat{S}_n^{\mathrm{MC}}(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$. The estimator is clearly unbiased, and it follows from Popoviciu's inequality—i.e., $\mathrm{var}(Z) \le 1/4$ for any random variable Z taking values in $[0,1]$—that

$$\left(e_2\left(\hat{S}_n^{\mathrm{MC}}\right)\right)^2 \;=\; \max_{f\in F} \mathrm{var}\left(\hat{S}_n^{\mathrm{MC}}(f)\right) \;=\; \frac{1}{4n}.$$

The maximal error is attained when f is a unit step function jumping at $x_0 = 1/2$. It turns out that a smaller error can be achieved, for the same (uniform i.i.d.) sampling scheme, using the control variate technique. More specifically, we consider the control variate $\tilde{f}(X_i) = X_i$ and set

$$\hat{S}_n^{\mathrm{cv}}(f) \;=\; \frac{1}{n}\sum_{i=1}^{n}\left(f(X_i) - \tilde{f}(X_i)\right) \,+\, \frac{1}{2}.$$

Theorem 3.1. *The estimator $\hat{S}_n^{\mathrm{cv}}(f)$ is unbiased, and satisfies*

$$\left(e_2\left(\hat{S}_n^{\mathrm{cv}}\right)\right)^2 \;=\; \max_{f\in F} \mathrm{var}\left(\hat{S}_n^{\mathrm{cv}}(f)\right) \;=\; \frac{1}{12n}.$$

The maximal error is attained for any unit step function.

Proof. The estimator is unbiased since $\mathbb{E}(\tilde{f}(X_i)) = 1/2$, and therefore the mean-squared error is equal to $\mathrm{var}(\hat{S}_n^{\mathrm{cv}}(f)) = \frac{1}{n}\,\mathrm{var}(f(X) - X)$. For a unit step function $f = \mathbb{1}_{[x_0,1]}$ with a jump at $x_0 \in [0,1]$, the random variable $f(X) - X$ is uniformly distributed over $[-x_0, 1-x_0]$, which yields $\mathrm{var}(\hat{S}_n^{\mathrm{cv}}(f)) = 1/(12n)$ as claimed. It remains to show that $\mathrm{var}(f(X) - X) \le \frac{1}{12}$ for all $f \in F$.

Let $F_m \subset F$ denote the class of all non-decreasing staircase functions of the form $f = \sum_{k=1}^{m} \alpha_k \cdot \mathbb{1}_{(\frac{k-1}{m},\frac{k}{m}]}$, with $0 \le \alpha_1 \le \alpha_2 \le \ldots \le \alpha_m \le 1$. For any $f \in F$, consider the piecewise-constant approximation $f_m \in F_m$ defined by averaging f over each subinterval of length $1/m$. Then, $\mathbb{E}(f_m(X)) = \mathbb{E}(f(X))$ and $|\mathrm{var}\,(f_m(X) - X) - \mathrm{var}\,(f(X) - X)| \le \frac{1}{m}$. Thus,

$$\sup_{f\in F} \mathrm{var}\,(f(X) - X) \;=\; \lim_{m\to\infty} \sup_{f\in F_m} \mathrm{var}\,(f(X) - X)\,. \tag{3}$$

Let us now show that $\mathrm{var}\,(f(X) - X)$ is maximized over F_m when f is a unit step function. Pick any $f = \sum_{k=1}^{m} \alpha_k \cdot \mathbb{1}_{(\frac{k-1}{m},\frac{k}{m}]} \in F_m$. Set $\alpha_0 = 0$ and $\alpha_{m+1} = 1$. If f is not a unit step function, then there exist $k_1, k_2 \in \{1, \ldots, m\}$ such that $k_1 \le k_2$ and $\alpha_{k_1-1} < \alpha_{k_1} = \ldots = \alpha_{k_2} < \alpha_{k_2+1}$. Denote by $f_u \in F_m$ the function obtained by changing the common value of $\alpha_{k_1}, \ldots, \alpha_{k_2}$ in f to $u \in [\alpha_{k_1-1}, \alpha_{k_2+1}]$. The variance $\mathrm{var}(f_u(X) - X)$ is a convex function of u, since it can be expanded as $au^2 + bu + c$ with $a = \frac{k_2-k_1+1}{m}(1 - \frac{k_2-k_1+1}{m}) > 0$. Consequently, we have $\mathrm{var}(f_u(X) - X) > \mathrm{var}(f(X) - X)$ at one of the two endpoints of $[\alpha_{k_1-1}, \alpha_{k_2+1}]$. Note that the corresponding staircase function f_u has one fewer step than f. Iterating as necessary, we conclude that for any $f \in F_m$ there exists a unit step function $g \in F_m$ such that $\mathrm{var}\,(f(X) - X) \le \mathrm{var}\,(g(X) - X) = \frac{1}{12}$. Therefore $\sup_{f\in F_m} \mathrm{var}\,(f(X) - X) = \frac{1}{12}$, which completes the proof. $\square$

4 Stratified sampling

Consider now a stratified sampling estimator with K strata:

$$\hat{S}_n^{\mathrm{str}}(f) \;=\; \sum_{k=1}^{K} w_k \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} f\left(X_{k,i}\right), \tag{4}$$

where the k-th stratum is $I_k = [x_{k-1}, x_k]$, $0 = x_0 < x_1 < \cdots < x_{K-1} < x_K = 1$, the weight $w_k = |x_{k-1} - x_k|$ is the length of the k-th stratum, the allocation scheme $(n_1, \ldots, n_K)$ is such that $n_k \geq 0$ for all k and $\sum_k n_k = n$, and the random variables $X_{k,i}$ are independent, with the $X_{k,i}$s uniformly distributed in I_k. Note that the sampling points are no longer identically distributed here. The estimator $\hat{S}_n^{\mathrm{str}}(f)$ is unbiased, with variance

$$\operatorname{var}\left(\hat{S}_n^{\mathrm{str}}(f)\right) \;=\; \sum_{k=1}^{K} \frac{w_k^2}{n_k} \operatorname{var}\left(f\left(X_{k,1}\right)\right). \tag{5}$$

Theorem 4.1. *For any $K \leq n$, any choice of strata and any allocation scheme, the stratified sampling estimator* (4) *satisfies*

$$\left(e_2\left(\hat{S}_n^{\mathrm{str}}\right)\right)^2 \;=\; \max_{f \in F} \operatorname{var}\left(\hat{S}_n^{\mathrm{str}}(f)\right) \;=\; \frac{1}{4} \max_k \frac{w_k^2}{n_k}, \tag{6}$$

The maximal error is attained for a unit step function with a jump at the middle of I_{k^*}, where $k^* \in \operatorname{argmax} w_k^2/n_k$. The minimal value of the maximal error (6) *is $\frac{1}{4n^2}$, and is obtained with $K = n$ strata of equal lengths ($w_k = n^{-1}$ and $n_k = 1$ for all k).*

The optimal stratified sampling method can be seen as a one-dimensional special case of the Latin Hypercube Sampling (LHS) method (McKay et al., 1979). (On a related note, McKay et al. (1979) prove that, in any dimension, the LHS method is preferable to the simple Monte Carlo method if the function is monotone in each of its arguments.)

Proof. For all $K \in \mathbb{N}^*$, let $\boldsymbol{\Delta}_K = \left\{(\Delta_1, \ldots, \Delta_K) \in \mathbb{R}_+^K \mid \sum_{k=1}^K \Delta_k \leq 1\right\}$. For a given stratified sampling method with K strata, for all $\underline{\Delta} \in \boldsymbol{\Delta}_K$, define

$$F_{\underline{\Delta}} \;=\; \left\{ f \in F \;\middle|\; \forall k \in \{1, \ldots, K\},\, f(x_k) - f(x_{k-1}) = \Delta_k \right\}.$$

Then it follows from (5) and Popoviciu's inequality that

$$\max_{f \in F_{\underline{\Delta}}} \operatorname{var}\left(\hat{S}_n^{\mathrm{str}}(f)\right) \;=\; \frac{1}{4} \sum_{k=1}^{K} \frac{w_k^2 \Delta_k^2}{n_k}, \tag{7}$$

where the maximum is attained for a non-decreasing staircase function with jumps of height Δ_k at the middle of the strata. Note that $\sum_{k=1}^K \Delta_k^2 \leq \sum_{k=1}^K \Delta_k \leq 1$. Therefore, the right-hand side of (7) is upper-bounded by $\frac{1}{4} \max_k w_k^2/n_k$, which is indeed the value of the variance (5) when f is a unit step function with a jump at the middle of the stratum where w_k^2/n_k is the largest.

In order to prove the second part of the claim, observe that any stratum with $n_k \geq 2$ can be further divided into n_k sub-strata of equal lengths without increasing the upper bound. Considering then the case where $K = n$ and $n_k = 1$ for all k, the upper bound reduces to $\frac{1}{4} \max_k w_k^2$, which is minimal when $w_1 = \cdots = w_n = n^{-1}$ since $\sum_k w_k = 1$. $\square$

5 Discussion

The stratified sampling (LHS) method provides the best-known variance upper bound over the class F for an unbiased nonsequential method as soon as $n \geq 3$, but is outperformed by the control variate method of Section 3 when $n \leq 2$. We do not know at the moment if these results are optimal in the class of unbiased nonsequential methods. (The ratio between the best variance upper bound and the lower bound of Corollary 2.2 is $\frac{8}{3} \approx 2.67$ for $n = 1$, $\frac{16}{3} \approx 5.33$ for $n = 2$ and 8 for $n \geq 3$.)

Relaxing the unbiasedness requirement, it turns out that both methods are outperformed for all n by the (deterministic) trapezoidal method discussed in the introduction, which has a worst-case squared error of $1/(4(n+1)^2)$. The ratio of worst-case mean-squared errors, however, is never very large—at most $\frac{16}{9} \approx 1.78$—and goes to 1 when n goes to infinity.

Directions for future work include constructing better sequential methods using the findings of this article, and extending our results to mutivariate integration with partial monotonicity (see, e.g., McKay et al., 1979, Section 2.1).

References

F. Perrin, F. Tenenhaus-Aziza, V. Michel, S. Miszczycha, N. Bel, and M. Sanaa. Quantitative risk assessment of haemolytic and uremic syndrome linked to O157:H7 and non-O157:H7 shiga-toxin producing escherichia coli strains in raw milk soft cheeses. *Risk Analysis*, 35(1):109–128, 2014.

J. Kiefer. Optimum sequential search and approximation methods under minimum regularity assumptions. *Journal of the Society for Industrial and Applied Mathematics*, 5 (3):105–136, 1957.

E. Novak. Quadrature formulas for monotone functions. *Proceedings of the American Mathematical Society*, 115(1):59–68, 1992.

M. D. McKay, R. J. Beckman, and Conover W. J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

This work is part of the ArtiSaneFood project (ANR-18-PRIM-0015), which is part of the PRIMA program supported by the European Union.