

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

Subhasish Basak^{1,2}, Julien Bect², Emmanuel Vazquez²

JdS 2022, Lyon, 15 Juin

¹ANSES, Maisons-Alfort, France

²Univ. Paris-Saclay, CNRS, CentraleSupélec, L2S, Gif-sur-yvette, France

Problem statement

We are interested in approximating $\mathbb{E}(g(Y)) = \int g(y) P_Y(dy)$

- $g : \mathbb{R} \rightarrow \mathbb{R}$ is **monotonic** and **bounded**
- P_Y is known, CDF F_Y of Y is continuous
- After scaling, the problem reduces to estimating

$$S(f) = \int_0^1 f(x) dx,$$

where $f : [0, 1] \rightarrow [0, 1]$ is a **non-decreasing function**

- We consider a **fixed sample-size** setting where $S(f)$ is estimated from n evaluations of f

Example: Quantitative Risk Assessment (QRA)

- Consider the following example in food safety
 - Y : dose of pathogen STEC (Shiga Toxin producing Escherichia coli)
 - $g(Y)$: conditional risk of food-borne disease (Haemolytic Uremic Syndrome)
 - g is an increasing function (complex numerical model)
- Goal: estimate $\mathbb{E}(g(Y))$

Contents

1 Literature review

2 Non sequential randomized methods

3 A lower bound for the L_p error

4 Uniform i.i.d sampling

5 Stratified sampling

6 Discussion

1 Literature review

- Two classes of methods (sequential or nonsequential):
 - **deterministic**
 - **randomized**
- Kiefer (1957) studied **deterministic** methods
 - Trapezoidal rule is the optimal worst-case **deterministic** algorithm
 - It attains the lower bound for the maximal error: $\frac{1}{2 \cdot (n+1)}$

- Novak (1992) studied unbiased randomized nonsequential methods
 - Unbiasedness comes at the price of a slight loss of performance w.r.t. the trapezoidal rule
- Novak (1992) also studied randomized sequential methods
 - He proposed a rate-optimal algorithm ($n^{-3/2}$ for the L_1 error)

- We focus on nonsequential unbiased Monte Carlo methods
- We provide a lower bound for the maximal L_p error
- We study methods based on uniform i.i.d sampling and stratified sampling
- These **unbiased** methods can be used as building blocks for sequential methods

2 Non sequential randomized methods

- A nonsequential randomized method evaluates the function at n random points X_1, \dots, X_n in $[0, 1]$ and approximates the integral $S(f)$ using an estimator

$$\widehat{S}_n(f) = \varphi(X_1, f(X_1), \dots, X_n, f(X_n)) \quad (1)$$

- The worst-case L^p error of such a method over the class F is

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

F is the class of all non-decreasing functions $f : [0, 1] \rightarrow [0, 1]$

3 A lower bound for the L_p error

- We show that (extending Novak's result)

Theorem

For any nonsequential randomized method with sample size n ,

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

Corollary

$$\sup_{f \in F} \text{Var} \left(\widehat{\mathcal{S}}_n(f) \right) \geq \frac{1}{32n^2}.$$

4 Uniform i.i.d sampling

- We study unbiased nonsequential methods based on uniform i.i.d sampling
- $\hat{S}_n(f)$ uses uniform i.i.d samples $\{X_1, X_2, \dots, X_n\}$ from $[0, 1]$
- We study two unbiased estimators
 - Simple Monte Carlo estimator
 - Control variate estimator

- Simple Monte Carlo estimator $\widehat{S}_n^{\text{MC}}(f)$

- The estimator is defined as

$$\widehat{S}_n^{\text{MC}} = \frac{1}{n} \sum_{i=1}^n f(X_i), \quad X_i \stackrel{iid}{\sim} U_{[0,1]} \quad (3)$$

- Using Popoviciu's inequality:

$$\left(e_2 \left(\widehat{S}_n^{\text{MC}} \right) \right)^2 = \max_{f \in F} \text{Var} \left(\widehat{S}_n^{\text{MC}}(f) \right) = \frac{1}{4n}.$$

- The maximal error is attained when f is a unit step function jumping at $x_0 = 1/2$

- **Control variate estimator** $\widehat{S}_n^{\text{cv}}(f)$

- We consider the control variate $\tilde{f}(X_i) = X_i$ and the estimator

$$\widehat{S}_n^{\text{cv}}(f) = \frac{1}{n} \sum_{i=1}^n (f(X_i) - \tilde{f}(X_i)) + \frac{1}{2}. \quad (4)$$

- Then,

$$\left(e_2\left(\widehat{S}_n^{\text{cv}}\right)\right)^2 = \max_{f \in F} \text{Var}\left(\widehat{S}_n^{\text{cv}}(f)\right) = \frac{1}{12n}.$$

- The maximal error is attained for any unit step function

5 Stratified sampling

- Unbiased randomized nonsequential methods based on stratified sampling
- $[0, 1]$ is divided into K strata, with the k -th stratum denoted by $I_k = [x_{k-1}, x_k]$, $0 = x_0 < x_1 < \dots < x_{K-1} < x_K = 1$
 - stratum size: $w_k = |x_k - x_{k-1}|$
 - stratum budget: n_k
- $\hat{S}_n(f)$ uses n_k uniformly distributed samples from I_k

- Stratified sampling estimator $\widehat{S}_n^{\text{str}}(f)$

- We consider K strata with sizes w_k and budgets n_k

$$\widehat{S}_n^{\text{str}}(f) = \sum_{k=1}^K w_k \cdot \frac{1}{n_k} \sum_{i=1}^{n_k} f(X_{k,i}), \quad (5)$$

- Then,

$$\left(e_2\left(\widehat{S}_n^{\text{str}}\right)\right)^2 = \max_{f \in F} \text{Var}\left(\widehat{S}_n^{\text{str}}(f)\right) = \frac{1}{4} \max_k \frac{w_k^2}{n_k},$$

- The minimal value of the maximal error is $\frac{1}{4 \cdot n^2}$, attained with $K = \sum_k n_k$ and $w_k = n^{-1}$.

6 Discussion

Maximal error of nonsequential randomized methods

lower bound	$\widehat{S}_n^{\text{str}}(f)$	$\widehat{S}_n^{\text{cv}}(f)$	$\widehat{S}_n^{\text{MC}}(f)$
$1/32 \cdot n^2$	$1/4 \cdot n^2$	$1/12 \cdot n$	$1/4 \cdot n$

- The best-known variance upper bound

- $n \leq 2$: Control variate estimator $\widehat{S}_n^{\text{cv}}(f)$

- $n \geq 3$: Stratified sampling estimator $\widehat{S}_n^{\text{str}}(f)$

(The squared error for the **deterministic biased** trapezoidal rule is $\frac{1}{4(n+1)^2}$
 → outperforms randomized unbiased methods.)

Ratio of variance upper bound to lower bound

$n = 1$	$n = 2$	$n \geq 3$
2.67	5.33	8

- We do not know at the moment if these results are optimal in the class of unbiased nonsequential methods
- The lower bound of variance is not known to be sharp
- **Future work:**
 - Multivariate case with partial monotonicity
 - Constructing sequential randomized methods



This work is part of the [ArtiSaneFood](#) project (grant number : [ANR-18-PRIM-0015](#)) which is part of the [PRIMA](#) program supported by the [European Union](#).



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

Subhasish Basak^{1,2}, Julien Bect², Emmanuel Vazquez²

JdS 2022, Lyon, 15 Juin

¹ANSES, Maisons-Alfort, France

²Univ. Paris-Saclay, CNRS, CentraleSupélec, L2S, Gif-sur-yvette, France