# Speeding up Monte Carlo Integration: Control Neighbors for Optimal Convergence

**Rémi LELUC**

*Ecole Polytechnique, Institut Polytechnique de Paris, France*

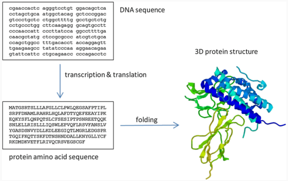
Joint work with François Portier, Johan Segers and Aigerim Zhuman

paper

Published in *Bernoulli, 2024.*

# Motivation: Machine Learning recent advances

AlphaGo (2016)



AlphaFold (2018)

GPT-3/4(2020/2023)

**"Intelligence"**

=

Data + Models + **Algorithms** + Computing Power

# Motivation: need for integral estimators

**Central Question:** *Integration*

Computation of an *integral* through probabilistic objective $\mathcal{F}$

$$\mathcal{F}(\theta) = \mathbb{E}_{\pi_\theta(x)}[f(x)] = \int_{\mathcal{X}} f(x)\pi_\theta(x)\mathrm{d}x. \quad (1)$$

**Main issue:** intractability and computational cost

- **(RL)** Trajectory $\tau = (s_0, a_0, \dots, s_{T-1}, a_{T-1})$ with policy $\pi_\theta$ and cumulative return $\mathcal{R}(\tau) = \sum_{t=0}^{T-1} \gamma^t r(s_t, a_t)$.

$$\mathcal{F}(\theta) = \mathbb{E}_{\pi_\theta(\tau)}[\mathcal{R}(\tau)]$$

*(2016) AlphaGo A.I. beats champion Lee Sedol in Go.*

- **(VI)** $\mathcal{F}$ optimises the log-likelihood $\log p(x|z)$ under a regularization constraint which promotes closeness between the density $q$ and the prior distribution $p(z)$

$$\text{ELBO} = \mathcal{F}(\theta) = \mathbb{E}_{q_\theta(z|x)}[\log p(x|z)] - \text{KL}(q_\theta(z|x)||p(z)).$$

# Advantages of Random estimates

***Easy and Practical***

→ Requires only three steps: sampling, evaluating, averaging

***Randomness as a Strength***

→ Naturally escape local optima

→ Complete exploration of the search space

***Large-Scale learning***

→ simple, scalable, parallelizable

→ in supervised learning, deterministic gradient scales as $O(nd)$, stochastic version reduces to $O(d)$ operations

***Theoretical justifications***[1]

→ deterministic methods $O(n^{-s/d})$

→ optimal random procedure $O(n^{-1/2}n^{-s/d})$

[1](Novak, 2016): Some results on the complexity of numerical integration

# Integration $\mathcal{F}$

## Monte Carlo Integration & Variance Reduction

# Monte Carlo integration

### Underlying integration problem

Let $(\mathcal{X}, \mathcal{A}, \pi)$ be a probability space, $f : \mathcal{X} \to \mathbb{R}$ with $f \in L_2(\pi)$.

- **Goal:**

$$\pi(f) := \int_{\mathcal{X}} f(x)\pi(\mathrm{d}x) = \mathbb{E}_\pi[f(X)].$$

- **Constraints:** $f$ is unknown (black-box) or no approximation is sufficiently accurate, sampling from $\pi$ may be hard.

Let $X_1, ..., X_n \overset{\text{i.i.d.}}{\sim} \pi$, naive Monte Carlo estimator $\hat{\alpha}_n^{\mathrm{mc}}(f)$ of $\pi(f)$ is

$$\hat{\alpha}_n^{\mathrm{mc}}(f) := \frac{1}{n}\sum_{i=1}^{n} f(X_i) \tag{2}$$

### Research Questions

- How to reduce the variance of Monte Carlo estimates?
- How to sample from $\pi$?
- How to achieve optimal convergence rates?

Ref: Metropolis and Ulam (1949); Robert and Casella (1999); Evans and Swartz (2000); Glasserman (2004); Owen (2013); Novak (2016); Chopin and Gerber (2024)

# Complexity rates for integration error

**Definition: Root Mean Squared Error (RMSE)**

The error $\delta_n$ of a procedure $\hat{\alpha}_n(f)$ that approximates $\pi(f)$ is

$$\delta_n = \mathbb{E}\left[|\hat{\alpha}_n(f) - \pi(f)|^2\right]^{1/2}$$

$\rightarrow$ Lipschitz integrands[2], **optimal rate** in $O(n^{-1/2}n^{-1/d})$ (Novak, 2016)

| | |
|---|---|
| **OLS control variates** (Portier and Segers, 2019) | $O(n^{-1/2}m^{-1/d})$ |
| **Determinantal sampling** (Bardenet and Hardy, 2020) | $O(n^{-1/2}n^{-1/2d})$ |
| **Control Functionals** (Oates et al., 2017) | $O(n^{-7/12})$ |
| **Cubic Stratification** (Haber, 1966; Chopin and Gerber, 2024) | $O(n^{-1/2}n^{-1/d})$ |

[2] for integrand with $s$ bounded derivatives, rate in $O(n^{-1/2}n^{-s/d})$

# General view of Control Variates

## Control Functionals

- Build surrogate function $\hat{f}$ with known integral $\pi(\hat{f})$
- Use centered variables $\hat{f}(X_i) - \pi(\hat{f})$ to derive the following enhanced Monte Carlo estimate with control variates

$$\hat{\alpha}_n^{(CV)}(f) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \left(\hat{f}(X_i) - \pi(\hat{f})\right)\right\}$$

## Approximation in $L_2(\pi)$

Let $(X_1, \dots X_n) \sim \pi$. Suppose that $\hat{f}$ depends only on a surrogate sample $\tilde{X}_1, \dots, \tilde{X}_N$ which is independent from $(X_1, \dots X_n)$, then

$$\mathbb{E}\left[|\hat{\alpha}_n^{(CV)}(f) - \pi(f)|^2\right] \leq \frac{1}{n}\mathbb{E}\left[\int (f - \hat{f})^2 \mathrm{d}\pi\right].$$

# Control Functionals examples

- **RKHS approximation:** (Oates, Girolami, and Chopin, 2017)

Ridge regression in Hilbert space $\mathcal{H}$

$$\hat{f} \in \arg\min_{\varphi \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} (f(\tilde{X}_i) - \varphi(\tilde{X}_i))^2 + \lambda \|\varphi\|_{\mathcal{H}}^2$$

- **Basis functions:** (Portier and Segers, 2019; Leluc et al., 2021)

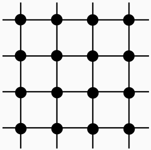
Use $m$ basis functions $h_1, \ldots, h_m$ to fit OLS:

$$\hat{f} = \hat{\beta}_n^\top h, \qquad \left(\hat{\alpha}_n, \hat{\beta}_n\right) = \arg\min_{(\alpha,\beta) \in \mathbb{R} \times \mathbb{R}^m} \|f^{(n)} - \alpha \mathbb{1}_n - H\beta\|_2^2$$

- **Partitioning and Stratification:** (Chopin and Gerber, 2024)

$(\tilde{X}_1, \ldots, \tilde{X}_N)$ is the $(1/\ell)$-equidistant grid of $[0,1]^d$ with $N = \ell^d$, $\ell \geq 1$ and $(R_i)_{i=1,\ldots,N}$ is the partition of $[0,1]^d$ made of the rectangles.

$$\hat{f}(x) = \sum_{i=1}^{N} f(\tilde{X}_i) \mathbb{1}_{R_i}(x)$$



# Nearest Neighbors

## Control Neighbors

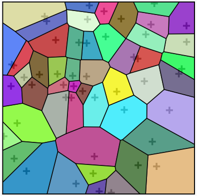
$$\hat{\alpha}_n^{(CVNN)}(f) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \left(\hat{f}_n^{(i)}(X_i) - \pi(\hat{f}_n)\right)\right\}$$

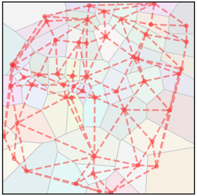
**Leave-one-out Nearest Neighbors:**

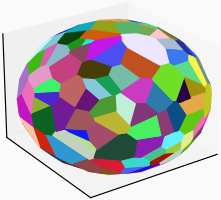
Take same sample $(X_1, \ldots, X_n)$ and define

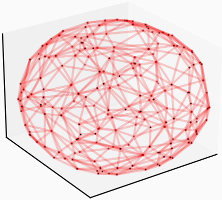
$$\hat{f}_n(x) = \sum_{j=1}^{n} f(X_j)\mathbb{1}_{S_{n,j}}(x), \qquad \hat{f}_n^{(i)}(x) = \sum_{j\neq i} f(X_j)\mathbb{1}_{S_{n,j}^{(i)}}(x)$$

where $S_{n,j}$ are **Voronoï cells**









# Control Neighbors properties

## Control Neighbors

$$\hat{\alpha}_n^{(CVNN)}(f) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \left(\hat{f}_n^{(i)}(X_i) - \pi(\hat{f}_n)\right)\right\}$$

- (a) (Same framework as naive MC) does not require the existence of control variates with known integrals
- (b) (Quadrature Rule) $\hat{\alpha}_n = \sum_{i=1}^{n} w_{n,i} f(X_i)$, for **quadrature weights** $w_{n,i}$ **that do not depend on the function** $f$.
- (c) (Practical tool box) The weights $w_{n,i}$ are built using efficient nearest neighbors estimates (Bentley, 1975; Pedregosa et al., 2011)

## Complexity rate for integration error of Control Neighbors

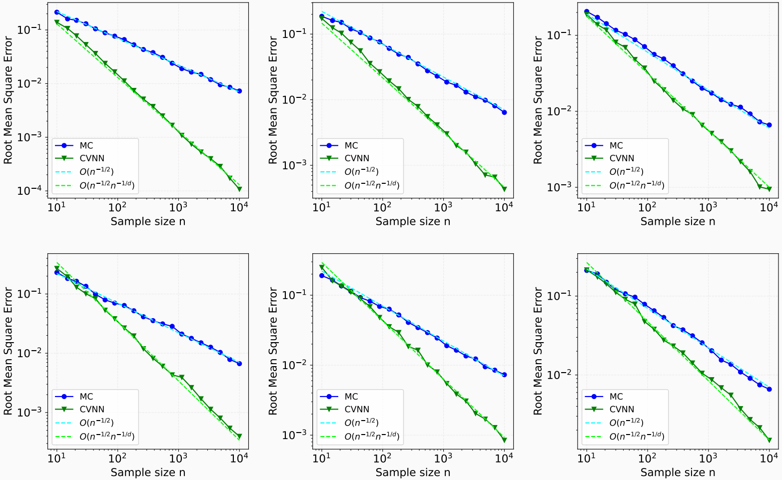
$$\mathbb{E}\left[|\hat{\alpha}_n^{(CVNN)}(f) - \pi(f)|^2\right]^{1/2} \lesssim n^{-1/2} n^{-s/d}$$

$$|\hat{\alpha}_n^{(CVNN)}(f) - \pi(f)| \lesssim \sqrt{\log(1/\varepsilon)}(\log n)^{1+s/d} n^{-1/2} n^{-s/d}$$

(with proba greater than $1-\varepsilon$)

# Control Neighbors on synthetic integrands

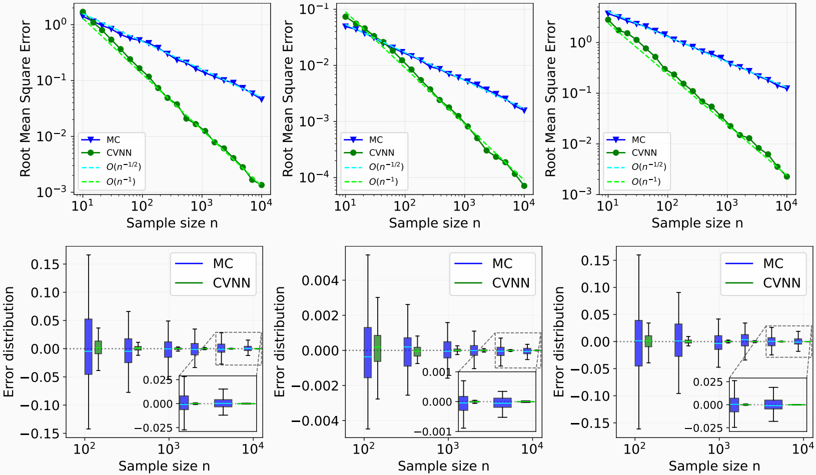
- $f_1(x_1, \ldots, x_d) = \sin(\pi(\frac{2}{d}\sum_{i=1}^{d} x_i - 1))$ with $\pi = \mathbb{1}_{[0,1]^d}$
- $f_2(x_1, \ldots, x_d) = \sin(\frac{\pi}{d}\sum_{i=1}^{d} x_i)$ with $\pi = \mathcal{N}_d(0, I_d)$




Error curves for $f_1$(top) and $f_2$(bottom) with $d \in \{2; 3; 4\}$

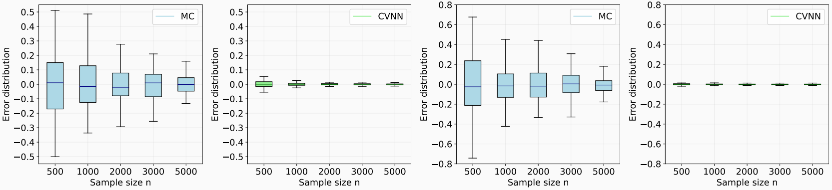
# Control Neighbors on Sphere $\mathbb{S}^2$

$$f_3(x,y,z) = \cos(x+y+z),\ f_4(x,y,z) = \cos(x)\cos(y)\cos(z),\ f_5(x,y,z) = \exp(x-y)$$

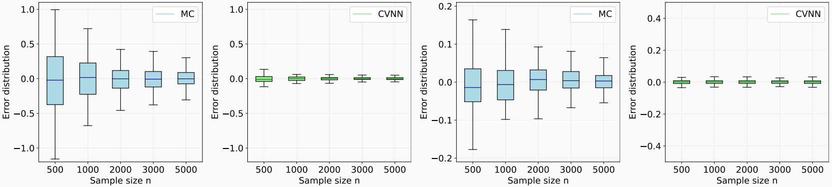



RMSE and boxplots for $f_3$(left), $f_4$(center) and $f_5$(right)

# Control Neighbors for Option Pricing




Black-Scholes model with spot price $S_0 = 100$, strike $K = S_0$, maturity $T = 2$ months, risk-free rate $r = 0.1$, constant volatility $\sigma = 0.3$, barrier price $H = 130$.(left: "Up-In"/right: "Up-Out")

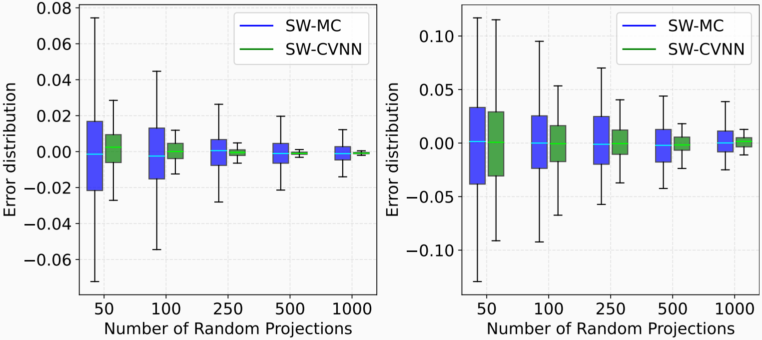



Heston Model with spot ptice $S_0 = 100$, strike $K = S_0$, barrier price $H = 130$, maturity $T = 2$ months, risk-free rate $r = 0.1$, initial volatility $v_0 = 0.1$, long-run average variance $\theta = 0.02$, rate of mean reversion $\kappa = 4$, instanteneous correlation $\rho = 0.8$ and volatility of volatility $\xi = 0.9$.(left: "Up-In"/right: "Up-Out")

# Control Neighbors for Sliced-Wasserstein

• Compare standard Monte Carlo estimate (SW-MC) with the proposed control neighbors estimate (SW-CVNN) when computing the Sliced-Wasserstein distance between two Gaussian distributions $SW_2(P, Q)$.

•$P = \mathcal{N}_q(m_X, \sigma_X^2 \mathrm{I}_q)$ and $Q = \mathcal{N}_q(m_Y, \sigma_Y^2 \mathrm{I}_q)$, $m_X, m_Y \sim \mathcal{N}(0, \mathrm{I}_q)$ and $\sigma_X = 2$ and $\sigma_Y = 5$, empirical distributions $P_m$ and $Q_m$ based on $m = 2\,000$ samples, $n \in \{50; 100; 250; 500; 1000\}$.




Boxplots of Sliced-Wasserstein estimates SW-MC and SW-CVNN for Gaussian distributions on $\mathbb{R}^q$ with $q \in \{3; 6\}$. The boxplots are obtained over $100$ replications.

# Conclusion

- We have explored the use of **nearest neighbors** in the construction of control variates for variance reduction in Monte Carlo integration.

- We have shown that for Hölder integrands of regularity $s \in (0, 1]$ on bounded metric spaces of dimension $d$ as measured by a sufficiently regular probability distribution, a **faster rate of convergence**, in $\mathcal{O}(n^{-1/2} n^{-s/d})$ as $n \to \infty$, is possible through the construction of a control variate via leave-one-out neighbors.

- (**Theory**) Theoretical guarantees are given both in terms of bounds on the root mean squared error and as concentration inequalities (requiring an additional logarithmic factor).

- (**Practice**) In numerical experiments, the method enjoyed a notable error reduction with respect to Monte Carlo integration.

# References

Bardenet, R. and A. Hardy (2020). Monte carlo with determinantal point processes. *The Annals of Applied Probability 30*(1), 368–417.

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM 18*(9), 509–517.

Chopin, N. and M. Gerber (2024). Higher-order monte carlo through cubic stratification. *SIAM Journal on Numerical Analysis 62*(1), 229–247.

Evans, M. and T. Swartz (2000). *Approximating integrals via Monte Carlo and deterministic methods*. Oxford Statistical Science Series. Oxford University Press, Oxford.

Glasserman, P. (2004). *Monte Carlo methods in financial engineering*, Volume 53. New York, NY, USA: Springer.

Haber, S. (1966). A modified monte-carlo quadrature. *Mathematics of Computation 20*(95), 361–368.

# Bibliography ii


Leluc, R., F. Portier, and J. Segers (2021, 07). Control variate selection for Monte Carlo integration. *Statistics and Computing 31*.

Metropolis, N. and S. Ulam (1949). The monte carlo method. *Journal of the American statistical association 44*(247), 335–341.

Novak, E. (2016). Some results on the complexity of numerical integration. In *Monte Carlo and Quasi-Monte Carlo Methods*, pp. 161–183. Springer.

Oates, C. J., M. Girolami, and N. Chopin (2017). Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(3), 695–718.

Owen, A. B. (2013). Monte carlo theory, methods and examples.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research 12*, 2825–2830.

Portier, F. and J. Segers (2019). Monte Carlo integration with a growing number of control variates. *Journal of Applied Probability 56*, 1168–1186.

Robert, C. P. and G. Casella (1999). *Monte Carlo statistical methods* (Second ed.), Volume 2 of *Springer Texts in Statistics*. Springer.