Control Variates Selection for Monte-Carlo Integration

Rémi Leluc
Joint work with François Portier and Johan Segers

June 28th - July 10th 2020

Machine Learning Summer School (MLSS) 2020

Outline

Outline

Introduction

- Solve deterministic problems by a stochastic approach

Introduction

- Solve deterministic problems by a stochastic approach
- Simple, Flexible, Scalable

Introduction

- Solve deterministic problems by a stochastic approach
- Simple, Flexible, Scalable
- Many fields of applications that include: physical science, engineering, climate change, biology, applied statistics, artificial intelligence for games, finance and business.

Outline

Mathematical Background

Mathematical Background

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space and let X be a random variable with distribution P. Let $f \in L^2(P)$ be a square integrable, real-valued function on $\mathcal{X}$ of which we would like to calculate the integral.

Mathematical Background

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space and let X be a random variable with distribution P. Let $f \in L^2(P)$ be a square integrable, real-valued function on $\mathcal{X}$ of which we would like to calculate the integral.

Goal: Estimate

$$P(f) = \int_{\mathcal{X}} f(x)P(dx) = \mathbb{E}_{X\sim P}[f(X)].$$

Mathematical Background

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space and let X be a random variable with distribution P. Let $f \in L^2(P)$ be a square integrable, real-valued function on $\mathcal{X}$ of which we would like to calculate the integral.
Goal: Estimate

$$P(f) = \int_{\mathcal{X}} f(x) P(dx) = \mathbb{E}_{X \sim P} [f(X)] .$$

Monte-Carlo procedure:

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space and let X be a random variable with distribution P. Let $f \in L^2(P)$ be a square integrable, real-valued function on $\mathcal{X}$ of which we would like to calculate the integral.

Goal: Estimate

$$P(f) = \int_{\mathcal{X}} f(x) P(dx) = \mathbb{E}_{X \sim P}\left[f(X)\right].$$

Monte-Carlo procedure:

- Choose *randomly* some points, called *nodes* or *particles*, $X_1, \ldots, X_n$ in S, $n \in \mathbb{N}^\star$.

Mathematical Background

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space and let X be a random variable with distribution P. Let $f \in L^2(P)$ be a square integrable, real-valued function on $\mathcal{X}$ of which we would like to calculate the integral.

Goal: Estimate

$$P(f) = \int_{\mathcal{X}} f(x) P(dx) = \mathbb{E}_{X \sim P}[f(X)].$$

Monte-Carlo procedure:

- Choose *randomly* some points, called *nodes* or *particles*, $X_1, \ldots, X_n$ in $S, n \in \mathbb{N}^\star$.
- Evaluate the function at nodes $f(X_1), \ldots, f(X_n)$.

Let $(\mathcal{X}, \mathcal{A}, P)$ be a probability space and let X be a random variable with distribution P. Let $f \in L^2(P)$ be a square integrable, real-valued function on $\mathcal{X}$ of which we would like to calculate the integral.

Goal: Estimate

$$P(f) = \int_{\mathcal{X}} f(x)P(dx) = \mathbb{E}_{X \sim P}\left[f(X)\right].$$

Monte-Carlo procedure:

- Choose *randomly* some points, called *nodes* or *particles*, $X_1, \ldots, X_n$ in S, $n \in \mathbb{N}^\star$.
- Evaluate the function at nodes $f(X_1), \ldots, f(X_n)$.
- Compute an approximation of $P(f)$ based on $((X_1, f(X_1)), \ldots, (X_n, f(X_n)))$.

Monte-Carlo Estimator

Monte-Carlo Estimator

- Let $X_1, ..., X_n \overset{\text{i.i.d}}{\sim} P$ be an independent random sample from P on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The naive Monte-Carlo estimator of $P(f)$ is given by the empirical mean

$$\boxed{\hat{I}_n^{\mathrm{mc}}(f) = P_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)}$$

Monte-Carlo Estimator

- Let $X_1, ..., X_n \overset{\text{i.i.d}}{\sim} P$ be an independent random sample from P on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The naive Monte-Carlo estimator of $P(f)$ is given by the empirical mean

$$\boxed{\hat{I}_n^{\mathrm{mc}}(f) = P_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_i)}$$

- The Monte-Carlo estimator $P_n(f)$ of $P(f)$ is unbiased and has variance $\sigma^2(f)/n$, where $\sigma^2(f) = P[(f - P(f))^2]$. By the central limit theorem, we have the convergence in distribution

$$\boxed{\sqrt{n}(P_n(f) - P(f)) \xrightarrow[n \to +\infty]{d} \mathcal{N}(0, \sigma^2(f))}$$

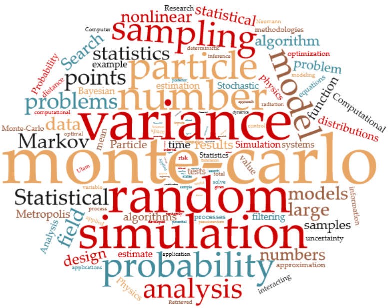
Popular Methods

Popular Methods

- Antithetic variates
- Control variates

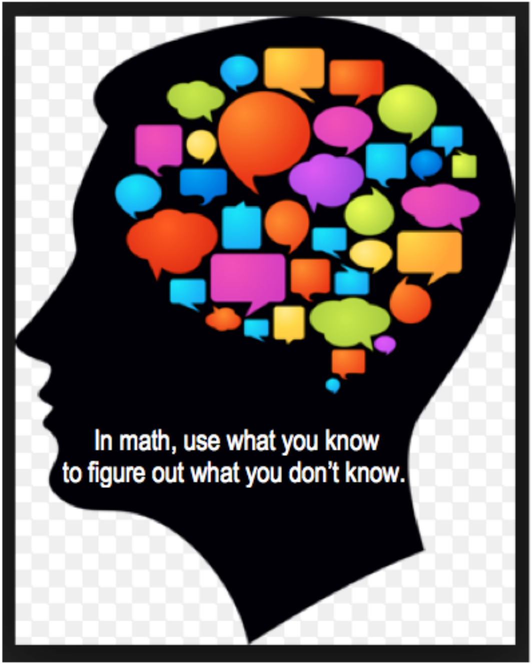
Popular Methods

- Antithetic variates
- Control variates
- Importance sampling
- Sequential Monte-Carlo/MCMC



Books: **Evans** (Approximating integrals via Monte Carlo and deterministic methods,2000), **Robert** (Monte Carlo Statistical Methods,2005), **Glasserman** (Monte Carlo Methods in Financial Engineering,2003)

Outline


In math, use what you know
to figure out what you don't know.

Monte-Carlo Control Variates

Monte-Carlo Control Variates

- If you wish to evaluate the (unknown) integral of a certain function you better use functions of which you know the integral.

Monte-Carlo Control Variates

- If you wish to evaluate the (unknown) integral of a certain function you better use functions of which you know the integral.
- The control variates are functions $h_1, ...h_m \in L^2(P)$ with known expectations. Assume that $P(h_k) = 0$ for all $k = 1, ..., m$. Let $h = (h_1, \ldots, h_m)^T$ denote the $\mathbb{R}^m$-valued function with the m control variates as elements.

Monte-Carlo Control Variates

- If you wish to evaluate the (unknown) integral of a certain function you better use functions of which you know the integral.
- The control variates are functions $h_1, ...h_m \in L^2(P)$ with known expectations. Assume that $P(h_k) = 0$ for all $k = 1, ..., m$. Let $h = (h_1, \dots, h_m)^T$ denote the $\mathbb{R}^m$-valued function with the m control variates as elements.
- For $\beta = (\beta_1, \dots, \beta_m)^T \in \mathbb{R}^m$, $P(f - \beta^T h) = P(f)$, so $P_n(f - \beta^T h)$ is an unbiased estimator of $P(f)$.

Control Variates estimator

Control Variates estimator

- Class of Monte-Carlo estimators

$$\boxed{\hat{I}_n^{(\mathrm{cv})}(f,\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \beta^T h(X_i)\right\}}, (\beta \in \mathbb{R}^m)$$

Control Variates estimator

- Class of Monte-Carlo estimators

$$\boxed{\hat{I}_n^{(\mathrm{cv})}(f,\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \beta^T h(X_i)\right\}}, (\beta \in \mathbb{R}^m)$$

- Minimize the variance to find optimal β

$$\beta^\star(f) \in \underset{\beta\in\mathbb{R}^m}{\arg\min}\ P[(f - P(f) - \beta^T h)^2] = \underset{\beta\in\mathbb{R}^m}{\arg\min}\ ||(f - P(f)) - \beta^T h||_{L^2}^2$$

Control Variates estimator

- Class of Monte-Carlo estimators

$$\boxed{\hat{I}_n^{(\mathrm{cv})}(f,\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \beta^T h(X_i)\right\}}, (\beta \in \mathbb{R}^m)$$

- Minimize the variance to find optimal β

$$\beta^\star(f) \in \underset{\beta\in\mathbb{R}^m}{\arg\min}\; P[(f - P(f) - \beta^T h)^2] = \underset{\beta\in\mathbb{R}^m}{\arg\min}\; ||(f - P(f)) - \beta^T h||_{L^2}^2$$

- Oracle estimator of the integral

$$\hat{\alpha}_n^{\mathrm{or}}(f) = P_n[f - \beta^\star(f)^T h].$$

Control Variates estimator

- Class of Monte-Carlo estimators

$$\boxed{\hat{I}_n^{(\mathrm{cv})}(f,\beta) = \frac{1}{n}\sum_{i=1}^{n}\left\{f(X_i) - \beta^T h(X_i)\right\}}, (\beta \in \mathbb{R}^m)$$

- Minimize the variance to find optimal β

$$\beta^\star(f) \in \underset{\beta\in\mathbb{R}^m}{\arg\min}\ P[(f - P(f) - \beta^T h)^2] = \underset{\beta\in\mathbb{R}^m}{\arg\min}\ ||(f - P(f)) - \beta^T h||_{L^2}^2$$

- Oracle estimator of the integral

$$\hat{\alpha}_n^{\mathrm{or}}(f) = P_n[f - \beta^\star(f)^T h].$$

- If $\beta^*(f)$ would be known, the use of control variates would always reduce the variance of the basic Monte Carlo estimator.

Control Variates estimator

Control Variates estimator

- The integral $P(f)$ thus appears as the intercept of a linear regression model with response f and explanatory variables $h_1, \ldots, h_m$,

$$\boxed{(P(f), \beta^\star(f)) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} P[(f - \alpha - \beta^T h)^2]}$$

Control Variates estimator

- The integral $P(f)$ thus appears as the intercept of a linear regression model with response f and explanatory variables $h_1, \ldots, h_m$,

$$\boxed{(P(f), \beta^\star(f)) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min}\ P[(f - \alpha - \beta^T h)^2]}$$

- (Hilbert projection) Normal equation: $P(hh^T)\beta^\star(f) = P(hf)$

Control Variates estimator

- The integral $P(f)$ thus appears as the intercept of a linear regression model with response f and explanatory variables $h_1, \ldots, h_m$,

$$\boxed{(P(f), \beta^\star(f)) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} \; P[(f - \alpha - \beta^T h)^2]}$$

- (Hilbert projection) Normal equation: $P(hh^T)\beta^\star(f) = P(hf)$
- Need to estimate the Gram matrix $P(hh^T)$ and $P(hf)$

Ordinary Least Squares Monte-Carlo (OLSMC)

Ordinary Least Squares Monte-Carlo (OLSMC)

- Empirical risk minimization paradigm (Replace P by measure P_n)

$$\boxed{\left(\hat{\alpha}_n^{\mathrm{ols}}(f), \hat{\beta}_n^{\mathrm{ols}}(f)\right) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} \|f^{(n)} - \alpha\mathbb{1}_n - H\beta\|_2^2}$$

Ordinary Least Squares Monte-Carlo (OLSMC)

- Empirical risk minimization paradigm (Replace P by measure P_n)

$$\boxed{\left(\hat{\alpha}_n^{\mathrm{ols}}(f), \hat{\beta}_n^{\mathrm{ols}}(f)\right) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} \|f^{(n)} - \alpha\mathbb{1}_n - H\beta\|_2^2}$$

- $f^{(n)} = (f(X_1), \ldots, f(X_n))^T \in \mathbb{R}^n$ is the vector of evaluations of f.

Ordinary Least Squares Monte-Carlo (OLSMC)

- Empirical risk minimization paradigm (Replace P by measure P_n)

$$\boxed{\left(\hat{\alpha}_n^{\mathrm{ols}}(f), \hat{\beta}_n^{\mathrm{ols}}(f)\right) \in \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} \|f^{(n)} - \alpha\mathbb{1}_n - H\beta\|_2^2}$$

- $f^{(n)} = (f(X_1), \ldots, f(X_n))^T \in \mathbb{R}^n$ is the vector of evaluations of f.
- $\mathbb{1}_n = (1, \ldots, 1)^T \in \mathbb{R}^n$, $\|\cdot\|_2$ denotes the Euclidean norm and H is the random $n \times m$ matrix defined by $H = \left(h_j(X_i)\right)_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}$.

$$\boxed{\hat{\alpha}_n^{\mathrm{ols}}(f) = P_n[f - \hat{\beta}_n^{\mathrm{ols}}(f)^T h] = \sum_{i=1}^n w_{n,i} f(X_i)}$$

Control Variates in the literature

Control Variates in the literature

- Stein method to build control functionals with non-parametric extension $\hookrightarrow$ **Control functionals for Monte Carlo integration**, Oates, C. J., M. Girolami, and N. Chopin (2014)

Control Variates in the literature

- Stein method to build control functionals with non-parametric extension $\hookrightarrow$ **Control functionals for Monte Carlo integration**, Oates, C. J., M. Girolami, and N. Chopin (2014)
- Central Limit Theorem $m \to +\infty, n \to +\infty$ $\hookrightarrow$ **Monte Carlo integration with a growing number of control variates**, Portier, F. and J. Segers (2018)

Control Variates in the literature

- Stein method to build control functionals with non-parametric extension $\hookrightarrow$ **Control functionals for Monte Carlo integration**, Oates, C. J., M. Girolami, and N. Chopin (2014)
- Central Limit Theorem $m \to +\infty, n \to +\infty \hookrightarrow$ **Monte Carlo integration with a growing number of control variates**, Portier, F. and J. Segers (2018)
- Langevin diffusion in MCMC methods $\hookrightarrow$ **Diffusion approximations and control variates for MCMC**, Brosse, N., A. Durmus, S. Meyn, É. Moulines, and A. Radhakrishnan(2018)

Control Variates in the literature

- Stein method to build control functionals with non-parametric extension $\hookrightarrow$ **Control functionals for Monte Carlo integration**, Oates, C. J., M. Girolami, and N. Chopin (2014)
- Central Limit Theorem $m \to +\infty, n \to +\infty \hookrightarrow$ **Monte Carlo integration with a growing number of control variates**, Portier, F. and J. Segers (2018)
- Langevin diffusion in MCMC methods $\hookrightarrow$ **Diffusion approximations and control variates for MCMC**, Brosse, N., A. Durmus, S. Meyn, É. Moulines, and A. Radhakrishnan(2018)
- Variance reduction via regularization $\hookrightarrow$ **Regularised Zero-Variance Control Variates for High-Dimensional Variance Reduction**, South, L. F., C. J. Oates, A. Mira, and C. Drovandi (2018)

Limitations of OLSMC

Limitations of OLSMC

- Too many variables or/and few samples (case $m >> n$)

Limitations of OLSMC

- Too many variables or/and few samples (case $m >> n$)
- Dependence among variables $\rightarrow$ very large coefficients

Limitations of OLSMC

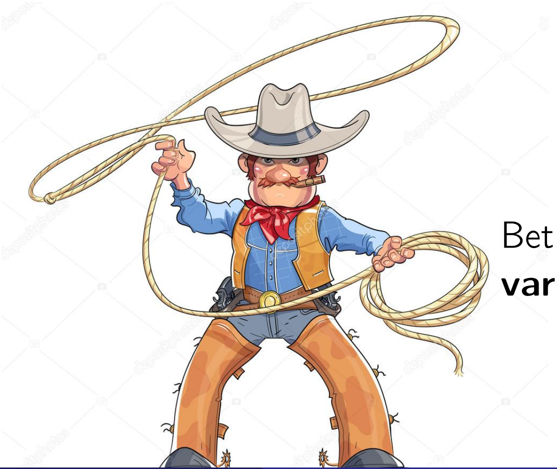
- Too many variables or/and few samples (case $m >> n$)
- Dependence among variables $\rightarrow$ very large coefficients

How to avoid those problems ?

Limitations of OLSMC

- Too many variables or/and few samples (case $m >> n$)
- Dependence among variables $\rightarrow$ very large coefficients

How to avoid those problems ?



Bet on sparsity

variable selection !

Lasso Monte-Carlo (LASSOMC)

Lasso Monte-Carlo (LASSOMC)

- Adding ℓ_1-penalization leads to

$$\boxed{\left(\hat{\alpha}_n^{\mathrm{lasso}}(f), \hat{\beta}_n^{\mathrm{lasso}}(f)\right) = \operatorname*{arg\,min}_{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m} \frac{1}{2n}\|f^{(n)} - \alpha\mathbb{1}_n - H\beta\|_2^2 + \lambda\|\beta\|_1}$$

Lasso Monte-Carlo (LASSOMC)

- Adding ℓ_1-penalization leads to

$$\boxed{\left(\hat{\alpha}_n^{\mathrm{lasso}}(f), \hat{\beta}_n^{\mathrm{lasso}}(f)\right) = \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} \frac{1}{2n}\|f^{(n)} - \alpha\mathbb{1}_n - H\beta\|_2^2 + \lambda\|\beta\|_1}$$

- Choice of the regularization parameter λ.

Lasso Monte-Carlo (LASSOMC)

- Adding ℓ_1-penalization leads to

$$\boxed{\left(\hat{\alpha}_n^{\mathrm{lasso}}(f), \hat{\beta}_n^{\mathrm{lasso}}(f)\right) = \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} \frac{1}{2n}\|f^{(n)} - \alpha\mathbb{1}_n - H\beta\|_2^2 + \lambda\|\beta\|_1}$$

- Choice of the regularization parameter λ.
- Lasso takes advantage of *sparse* regression models.The *active set* associated to the coefficient vector $\beta \in \mathbb{R}^m$ is

$$S(\beta) = \{j = 1, \dots, m \, : \, \beta_j \neq 0\}.$$

Lasso Monte-Carlo (LASSOMC)

- Adding ℓ_1-penalization leads to

$$\boxed{\left(\hat{\alpha}_n^{\text{lasso}}(f), \hat{\beta}_n^{\text{lasso}}(f)\right) = \underset{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^m}{\arg\min} \frac{1}{2n}\|f^{(n)} - \alpha\mathbb{1}_n - H\beta\|_2^2 + \lambda\|\beta\|_1}$$

- Choice of the regularization parameter λ.
- Lasso takes advantage of *sparse* regression models.The *active set* associated to the coefficient vector $\beta \in \mathbb{R}^m$ is

$$S(\beta) = \{j = 1, \ldots, m \,:\, \beta_j \neq 0\}.$$

- The number of elements in $S^\star = S(\beta^\star(f))$, denoted by $\ell^\star := |S^\star|$, quantifies the level of sparsity associated to the regression model.

Least Square Lasso Monte-Carlo (LSLASSOMC)

Least Square Lasso Monte-Carlo (LSLASSOMC)

- Lasso to select the active variables among a large number of control variates, then compute OLSMC using only the variables selected at the previous stage.

Least Square Lasso Monte-Carlo (LSLASSOMC)

- Lasso to select the active variables among a large number of control variates, then compute OLSMC using only the variables selected at the previous stage.
- Let $\hat{S} = \{k \in \{1, \ldots, m\} : \hat{\beta}^{\text{lasso}}_{N,k}(f) > 0\}$ denote the estimated active set of control variates based on the subsample of size N. The LSLASSOMC estimate $\hat{\alpha}^{\text{lslasso}}_n(f)$ of $P(f)$ is defined by

$$\boxed{\left(\hat{\alpha}^{\text{lslasso}}_n(f), \hat{\beta}^{\text{lslasso}}_n(f)\right) \in \operatorname*{arg\,min}_{(\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^{\hat{\ell}}} \|f^{(n)} - \alpha \mathbb{1}_n - H_{\hat{S}}\beta\|_2^2}$$

Outline

Non-asymptotic Error Analysis

Non-asymptotic Error Analysis

Assumptions: **sub-gaussian** residuals with factor τ, **linearly independent** and **bounded** control variates, appropriate λ
$(U_h = \max_{j=1,\dots,m} \sup_{x \in \mathcal{X}} |h_j(x)|, \gamma = \lambda_{\min}(G), \zeta_h = U_h^2/\gamma)$

Concentration inequalities (Leluc,Portier,Segers, 2019)

For $\delta \in (0,1)$ with probability at least $1-\delta$,

Non-asymptotic Error Analysis

Assumptions: **sub-gaussian** residuals with factor τ, **linearly independent** and **bounded** control variates, appropriate λ
$(U_h = \max_{j=1,\ldots,m} \sup_{x\in\mathcal{X}} |h_j(x)|, \gamma = \lambda_{\min}(G), \zeta_h = U_h^2/\gamma)$

Concentration inequalities (Leluc,Portier,Segers, 2019)

For $\delta \in (0,1)$ with probability at least $1-\delta$,

$$|\hat{\alpha}_n^{\mathrm{ols}}(f) - P(f)| \leq \sqrt{2\log(8/\delta)}\frac{\tau}{\sqrt{n}} + 27m\log(8m/\delta)\zeta_h\frac{\tau}{n}$$

Non-asymptotic Error Analysis

Assumptions: **sub-gaussian** residuals with factor τ, **linearly independent** and **bounded** control variates, appropriate λ
$(U_h = \max_{j=1,\ldots,m} \sup_{x\in\mathcal{X}} |h_j(x)|, \gamma = \lambda_{\min}(G), \zeta_h = U_h^2/\gamma)$

Concentration inequalities (Leluc,Portier,Segers, 2019)

For $\delta \in (0,1)$ with probability at least $1-\delta$,

$$|\hat{\alpha}_n^{\mathrm{ols}}(f) - P(f)| \leq \sqrt{2\log(8/\delta)}\frac{\tau}{\sqrt{n}} + 27m\log(8m/\delta)\zeta_h\frac{\tau}{n}$$

$$|\hat{\alpha}_n^{\mathrm{lasso}}(f) - P(f)| \leq \sqrt{2\log(8/\delta)}\frac{\tau}{\sqrt{n}} + 680\ell^\star\log(8m/\delta)(U_h^2/\gamma^\star)\frac{\tau}{n}$$

Non-asymptotic Error Analysis

Assumptions: **sub-gaussian** residuals with factor τ, **linearly independent** and **bounded** control variates, appropriate λ
$(U_h = \max_{j=1,\ldots,m} \sup_{x\in\mathcal{X}} |h_j(x)|, \gamma = \lambda_{\min}(G), \zeta_h = U_h^2/\gamma)$

Concentration inequalities (Leluc,Portier,Segers, 2019)

For $\delta \in (0,1)$ with probability at least $1-\delta$,

$$|\hat{\alpha}_n^{\mathrm{ols}}(f) - P(f)| \leq \sqrt{2\log(8/\delta)}\frac{\tau}{\sqrt{n}} + 27m\log(8m/\delta)\zeta_h\frac{\tau}{n}$$

$$|\hat{\alpha}_n^{\mathrm{lasso}}(f) - P(f)| \leq \sqrt{2\log(8/\delta)}\frac{\tau}{\sqrt{n}} + 680\ell^\star\log(8m/\delta)(U_h^2/\gamma^\star)\frac{\tau}{n}$$

$$|\hat{\alpha}_n^{\mathrm{lslasso}}(f) - P(f)| \leq \sqrt{2\log(16/\delta)}\frac{\tau}{\sqrt{n}} + 27\ell^\star\log(16\ell^\star/\delta)\zeta_h^\star\frac{\tau}{n}$$

Outline

Numerical experiments

- Different integrands on $[0, 1]^d$:

Numerical experiments

- Different integrands on $[0,1]^d$:

$$\varphi(x_1, \ldots, x_d) = 1 + \sin\left(\pi\left(\frac{2}{d}\sum_{i=1}^{d} x_i - 1\right)\right)$$

Numerical experiments

- Different integrands on $[0,1]^d$:

$$\varphi(x_1,\ldots,x_d) = 1 + \sin\left(\pi\left(\frac{2}{d}\sum_{i=1}^{d} x_i - 1\right)\right)$$

$$\forall 1 \leq j \leq d, \quad f_j(x_1,\ldots,x_d) = \prod_{i=1}^{j} (2/\pi)^{1/2} x_i^{-1} \exp\left(-(\log(x_i)^2/2\right),$$

$$g_j(x_1,\ldots,x_d) = \prod_{i=1}^{j} \frac{\log(2)}{2^{x_i-1}} = \log(2)^j 2^{\sum_{i=1}^{j}(1-x_i)}.$$

Numerical experiments

- Different integrands on $[0,1]^d$:

$$\varphi(x_1,\dots,x_d) = 1 + \sin\left(\pi\left(\frac{2}{d}\sum_{i=1}^{d} x_i - 1\right)\right)$$

$$\forall 1 \le j \le d, \quad f_j(x_1,\dots,x_d) = \prod_{i=1}^{j} (2/\pi)^{1/2} x_i^{-1} \exp\left(-(\log(x_i)^2/2\right),$$

$$g_j(x_1,\dots,x_d) = \prod_{i=1}^{j} \frac{\log(2)}{2^{x_i-1}} = \log(2)^j 2^{\sum_{i=1}^{j}(1-x_i)}.$$

- Methods in competition: OLSMC, LassoMC, LSLassoMC(X)

Numerical experiments

Numerical experiments

- $h_j(x) = L_j(2x-1)$ for $x \in [0,1]$, with L_j the univariate Legendre polynomial (Legendre function of the first kind) of degree j.

Numerical experiments

- $h_j(x) = L_j(2x-1)$ for $x \in [0,1]$, with L_j the univariate Legendre polynomial (Legendre function of the first kind) of degree j.
- For a multi-index $\ell = (\ell_1, \ldots, \ell_d)$ in $\{0, 1, \ldots, k\}^d \setminus \{(0, \ldots, 0)\}$, build

$$h_\ell(x_1, \ldots, x_d) = \prod_{j=1}^{d} h_{\ell_j}(x_j) = h_{\ell_1}(x_1) \times \ldots \times h_{\ell_d}(x_d)$$

Numerical experiments

- $h_j(x) = L_j(2x-1)$ for $x \in [0,1]$, with L_j the univariate Legendre polynomial (Legendre function of the first kind) of degree j.
- For a multi-index $\ell = (\ell_1, \ldots, \ell_d)$ in $\{0,1,\ldots,k\}^d \setminus \{(0,\ldots,0)\}$, build

$$h_\ell(x_1,\ldots,x_d) = \prod_{j=1}^{d} h_{\ell_j}(x_j) = h_{\ell_1}(x_1) \times \ldots \times h_{\ell_d}(x_d)$$

- Sort in ascending order according to the total degree $\sum_{j=1}^{d} \ell_j$.

Numerical experiments

- $h_j(x) = L_j(2x-1)$ for $x \in [0,1]$, with L_j the univariate Legendre polynomial (Legendre function of the first kind) of degree j.
- For a multi-index $\ell = (\ell_1, \ldots, \ell_d)$ in $\{0,1,\ldots,k\}^d \setminus \{(0,\ldots,0)\}$, build

$$h_\ell(x_1,\ldots,x_d) = \prod_{j=1}^{d} h_{\ell_j}(x_j) = h_{\ell_1}(x_1) \times \ldots \times h_{\ell_d}(x_d)$$

- Sort in ascending order according to the total degree $\sum_{j=1}^{d} \ell_j$.

d	k	Degree threshold				
		1	3	5	10	12
3	12	3	19	55	285	454
5	10	5	55	251	3 001	6 157
8	3	8	164	1 214	20 993	36 813

Table: Number of control variates by degrees

Numerical experiments

Numerical experiments

- λ is selected by imposing a lower bound and an upper bound on the number of activated random variables $\rightarrow$ dichotomic search.

Numerical experiments

- λ is selected by imposing a lower bound and an upper bound on the number of activated random variables $\rightarrow$ dichotomic search.
- initialize $\lambda = \lambda_{max}$ and decrease it to have more and more control variates until their number lies in the range $[c_1\sqrt{n}, c_2\sqrt{n}]$.

Numerical experiments

- λ is selected by imposing a lower bound and an upper bound on the number of activated random variables $\rightarrow$ dichotomic search.
- initialize $\lambda = \lambda_{max}$ and decrease it to have more and more control variates until their number lies in the range $[c_1\sqrt{n}, c_2\sqrt{n}]$.

n	N	$\lfloor 3\sqrt{n} \rfloor$	$\lfloor 12\sqrt{n} \rfloor$
2 000	700	134	536
5 000	1 000	212	848
10 000	2 000	300	1 200

Table: Parameters setting with range $(c_1\sqrt{n}, c_2\sqrt{n})$ of selected control variates.

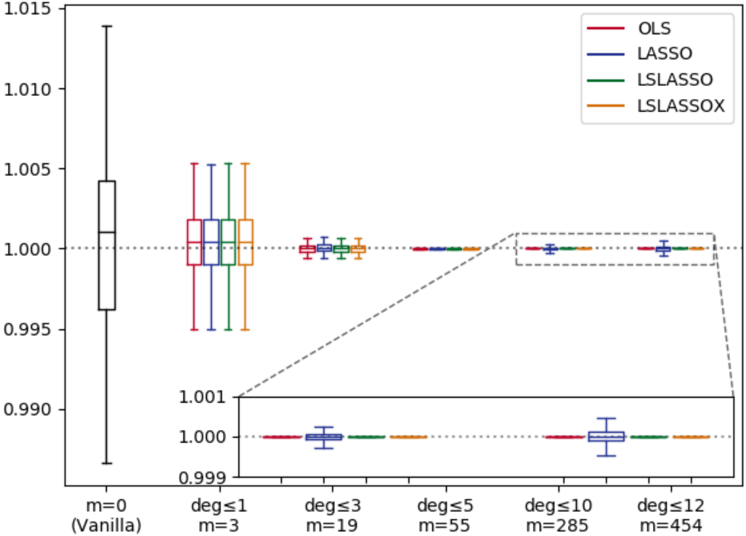


Figure: $\varphi, d = 3, n = 10\,000, N = 2\,000$.

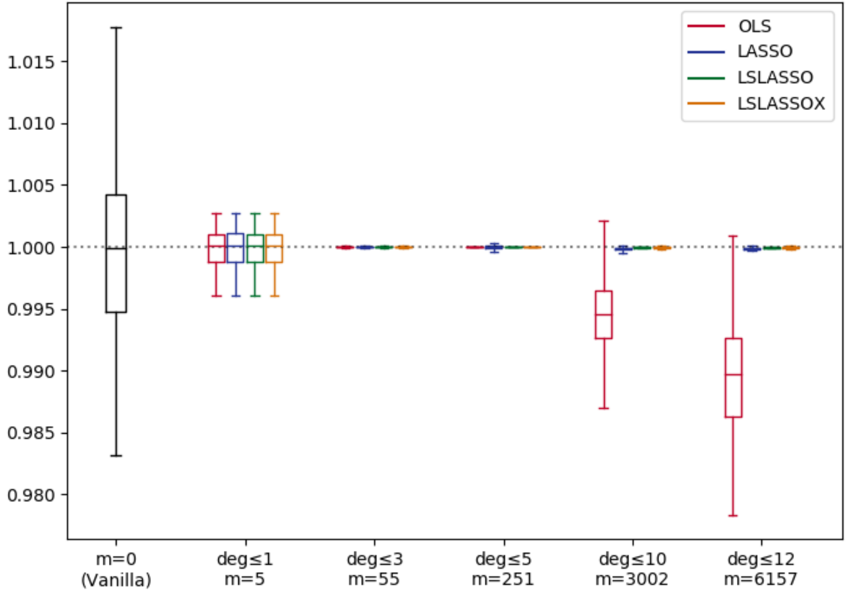


Figure: $g_3, d = 5, n = 2\,000, N = 700.$

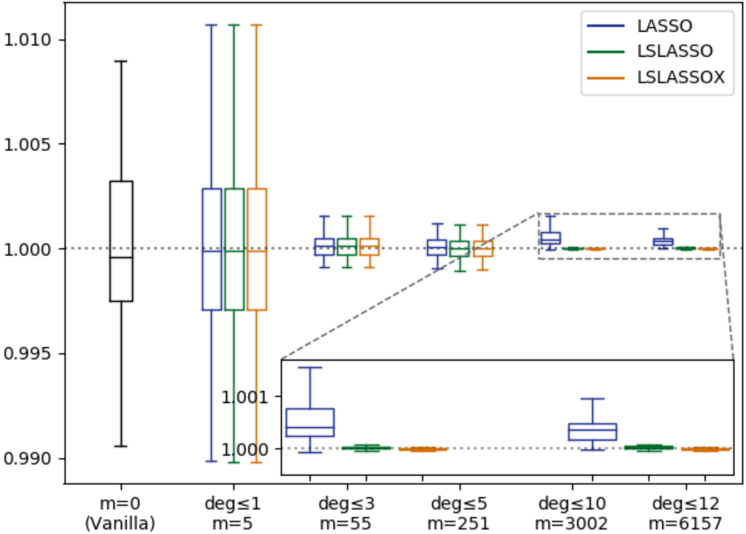


Figure: $f_1, d = 5, n = 5\,000, N = 1\,000.$

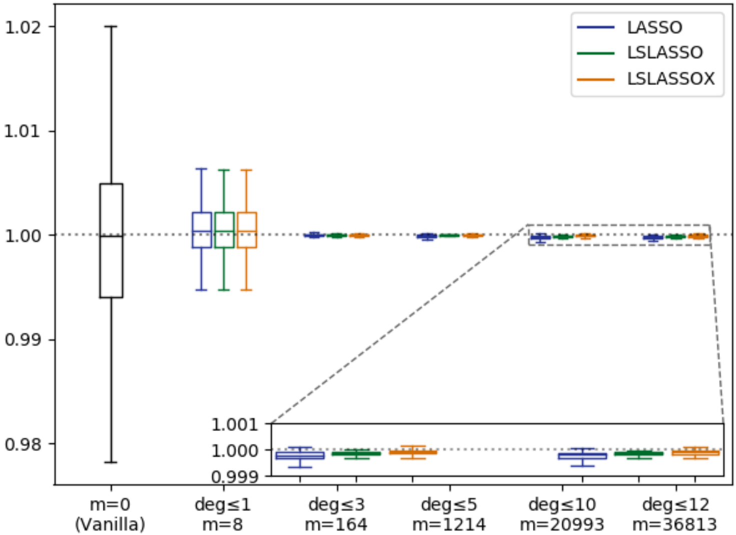


Figure: $g_4, d = 8, n = 2\,000, N = 700.$

Outline

Conclusion

Conclusion

- The particular variance reduction technique of control variates offers many advantages as it relies on a simple and intuitive paradigm.

Conclusion

- The particular variance reduction technique of control variates offers many advantages as it relies on a simple and intuitive paradigm.
- Regularizing the ordinary least squares estimator by preselecting appropriate control variates via the Lasso turns out to increase the accuracy without additional computational cost.

Conclusion

- The particular variance reduction technique of control variates offers many advantages as it relies on a simple and intuitive paradigm.
- Regularizing the ordinary least squares estimator by preselecting appropriate control variates via the Lasso turns out to increase the accuracy without additional computational cost.
- The proposed numerical method performs better than any other state of the art method.

Questions and Answers

Bibliography

Brosse, N., A. Durmus, S. Meyn, É. Moulines, and A. Radhakrishnan (2018). Diffusion approximations and control variates for mcmc. *arXiv preprint arXiv:1808.01665*.

Evans, M. and T. Swartz (2000). *Approximating integrals via Monte Carlo and deterministic methods*, Volume 20. OUP Oxford.

Glasserman, P. (2013). *Monte Carlo Methods in Financial Engineering*, Volume 53. Springer Science & Business Media.

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*.

Portier, F. and J. Segers (2018). Monte carlo integration with a growing number of control variates. *arXiv preprint arXiv:1801.01797*.

Robert, C. P. and G. Casella (2005). *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Berlin, Heidelberg: Springer-Verlag.

South, L. F., C. J. Oates, A. Mira, and C. Drovandi (2018). Regularised zero-variance control variates for high-dimensional variance reduction. *arXiv preprint arXiv:1811.05073*.

Assumptions

- The control variates $h_1, \dots, h_m \in L^2(P)$ are **linearly independent**. As a consequence, the **Gram matrix** $G := P(hh^T) \in \mathbb{R}^m$ is **positive definite** and its smallest eigenvalue $\gamma := \lambda_{\min}(G)$ is positive.
- The residual function $\epsilon = f - P(f) - \beta^\star(f)^T h$ satisfies $\epsilon \in \mathcal{G}(\tau^2)$ for some $\tau > 0$, that is, $\int_{\mathcal{X}} \exp(\lambda x)\, \epsilon(x)\, P(dx) \leq \exp(\lambda^2 \tau^2 / 2), \forall \lambda \in \mathbb{R}$.
- The control variates $h_1, \dots, h_m \in L^2(P)$ are **uniformly bounded**. Put $U_h := \max_{j=1,\dots,m} \sup_{x \in \mathcal{X}} |h_j(x)|$
- We have **orthogonality between active and inactive control variates** $P(h_j h_k) = 0$ for all $j \in \{1, \dots, m\} \setminus S^\star$ and all $k \in S^\star$.

Support recovery

Assumptions: **sub-gaussian** residuals, **linearly independent** and **bounded** control variates, appropriate λ
$(U_h = \max_{j=1,\dots,m} \sup_{x\in\mathcal{X}} |h_j(x)|, \gamma = \lambda_{\min}(G))$

Support recovery LASSOMC (Leluc,Portier,Segers, 2019)

For all $\delta \in (0,1)$, all integer n such that $n \geq 70(\ell^\star U_h^2/\gamma^\star)^2 \log(10\ell^\star m/\delta)$, and all λ such that

$$17 U_h \sqrt{\log(10m/\delta)}\tau/\sqrt{n} \leq \lambda \leq (\gamma^\star/(3\sqrt{\ell^\star})) \min_{k\in S^\star} |\beta_k^\star(f)|,$$

it holds that, with probability at least $1-\delta$, the LASSO solution $\hat{\beta}_n^{\text{lasso}}(f)$ is unique and the true active set is recovered, $\text{supp}(\hat{\beta}_n^{\text{lasso}}(f)) = S^\star$.