Higher-order stochastic integration through cubic stratification

how to get a $\mathcal{O}(N^{-5})$ error when you compute your favourite integral

Nicolas Chopin

ENSAE, IPP (Institut Polytechnique de Paris)

Joint work with Mathieu Gerber, University of Bristol, UK

This talk

- Generalities on random algorithms for quadrature and best possible rates;
- Two new classes of estimators with optimal rate.

Section 1

Generalities on random quadrature

Formal definition of the considered problem

- We wish to obtain the best possible approximation of

$$\mathcal{I}(f) := \int_{[0,1]^s} f(u)du$$

based on n evaluations of function $f : [0,1]^s \to \mathbb{R}$.

Formal definition of the considered problem

- We wish to obtain the best possible approximation of

$$\mathcal{I}(f) := \int_{[0,1]^s} f(u)du$$

based on n evaluations of function $f : [0,1]^s \to \mathbb{R}$.

- For a random algorithm, our optimality criterion is simply the RMSE (root mean square error), i.e. the root square of:

$$\mathbb{E}\left[\left(\hat{\mathcal{I}}(f) - \mathcal{I}(f)\right)^2\right]$$

Formal definition of the considered problem

- We wish to obtain the best possible approximation of

$$\mathcal{I}(f) := \int_{[0,1]^s} f(u)du$$

based on n evaluations of function $f : [0,1]^s \to \mathbb{R}$.

- For a random algorithm, our optimality criterion is simply the RMSE (root mean square error), i.e. the root square of:

$$\mathbb{E}\left[\left(\hat{\mathcal{I}}(f) - \mathcal{I}(f)\right)^2\right]$$

- We focus on unbiased estimation, $\mathbb{E}\left[\hat{\mathcal{I}}(f)\right] = \mathcal{I}(f)$. Then the RMSE equals the square root of the variance.

Advantages of unbiased algorithms

It is possible to generate k independent unbiased estimates

- in parallel;

Advantages of unbiased algorithms

It is possible to generate k independent unbiased estimates

- in parallel;
- to use them assess the numerical error, through the empirical variance $\hat{\sigma}^2$;

Advantages of unbiased algorithms

It is possible to generate k independent unbiased estimates

- in parallel;
- to use them assess the numerical error, through the empirical variance $\hat{\sigma}^2$;
- and compute their average (the variance of which is $\approx \hat{\sigma}^2/k$)

Advantages of unbiased algorithms

It is possible to generate k independent unbiased estimates

- in parallel;
- to use them assess the numerical error, through the empirical variance $\hat{\sigma}^2$;
- and compute their average (the variance of which is $\approx \hat{\sigma}^2/k$)

Moreover, random estimates tends to have better error rates than deterministic ones, see below.

Most common unbiased algorithm: Monte Carlo

$$\hat{\mathcal{I}}(f) = \frac{1}{n}\sum_{i=1}^{n} f(U_i), \quad U_i \sim \mathcal{U}([0,1]^s)$$

- Unbiased, with RMSE $\mathcal{O}(n^{-1/2})$.

Most common unbiased algorithm: Monte Carlo

$$\hat{\mathcal{I}}(f) = \frac{1}{n}\sum_{i=1}^{n} f(U_i), \quad U_i \sim \mathcal{U}([0,1]^s)$$

- Unbiased, with RMSE $\mathcal{O}(n^{-1/2})$.
- Minimal assumption on f: $\int f^2 < \infty$.

Most common unbiased algorithm: Monte Carlo

$$\hat{\mathcal{I}}(f) = \frac{1}{n}\sum_{i=1}^{n} f(U_i), \quad U_i \sim \mathcal{U}([0,1]^s)$$

- Unbiased, with RMSE $\mathcal{O}(n^{-1/2})$.
- Minimal assumption on f: $\int f^2 < \infty$.
- We can construct estimators with better rates if we focus on a smaller class of functions.

Most common unbiased algorithm: Monte Carlo

$$\hat{\mathcal{I}}(f) = \frac{1}{n}\sum_{i=1}^{n} f(U_i), \quad U_i \sim \mathcal{U}([0,1]^s)$$

- Unbiased, with RMSE $\mathcal{O}(n^{-1/2})$.
- Minimal assumption on f: $\int f^2 < \infty$.
- We can construct estimators with better rates if we focus on a smaller class of functions.

Note: I will **not** talk about (randomised) quasi-Monte Carlo today.

Optimality results

Assuming $f \in \mathcal{C}^r([0,1]^s)$, the best RMSE one may achieve for a random algorithm is (Bakhvalov, 1959)

$$\mathcal{O}(n^{-1/2-r/s})$$

while the best (absolute) error for a deterministic algorithm is

$$\mathcal{O}(n^{-r/s}).$$

Optimality results

Assuming $f \in \mathcal{C}^r([0,1]^s)$, the best RMSE one may achieve for a random algorithm is (Bakhvalov, 1959)

$$\mathcal{O}(n^{-1/2-r/s})$$

while the best (absolute) error for a deterministic algorithm is

$$\mathcal{O}(n^{-r/s}).$$

Statement above is slightly sloppy; for something more formal see e.g. Novak (2015).

Connexion with function approximation

Optimal function approximation: based on n evaluations of f, provide approximation f_n such that $\|f - f_n\|_\infty = \mathcal{O}(n^{-r/s})$.

Connexion with function approximation

Optimal function approximation: based on n evaluations of f, provide approximation f_n such that $\|f - f_n\|_\infty = \mathcal{O}(n^{-r/s})$.

The following estimate (based on $2n$ evaluations of f)

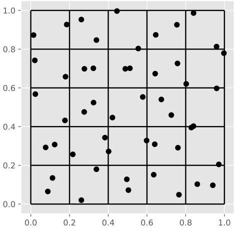
$$\hat{\mathcal{I}}(f) = \mathcal{I}(f_n) + \frac{1}{n}\sum_{i=1}^{n}\{f - f_n\}(U_i), \quad U_i \sim \mathcal{U}\left([0,1]^s\right).$$

is then unbiased and with optimal rate, since:

$$\begin{aligned}\mathrm{Var}\left[\hat{\mathcal{I}}(f)\right] &= \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left[\{f - f_n\}(U_i)\right] \\ &= \mathcal{O}(n^{-1-2r/s})\end{aligned}$$

Stratification

Our approach relies on splitting $[0,1]^s$ into k^s "sub-cubes", and performing l evaluations of f inside each: $n = l \times k^s$, so $k = \mathcal{O}(n^{1/s})$.



Let $\mathfrak{C}_k$ denote the set of centres of the sub-cubes, and $B_k(c)$ the sub-cube with center $c \in \mathfrak{C}_k$.

Haber (1966)'s estimator (optimal for $r = 1$)

$$\hat{\mathcal{I}}(f) := \frac{1}{n} \sum_{c \in \mathfrak{C}_k} f(c + U_c), \qquad U_c \sim \mathcal{U}(\left[-\frac{1}{2k}, \frac{1}{2k}\right]^s)$$

note that $c + U_c \in B_k(c)$, the sub-cube with centre c.

Haber (1966)'s estimator (optimal for $r = 1$)

$$\hat{\mathcal{I}}(f) := \frac{1}{n} \sum_{c \in \mathfrak{C}_k} f(c + U_c), \qquad U_c \sim \mathcal{U}\left(\left[-\frac{1}{2k}, \frac{1}{2k}\right]^s\right)$$

note that $c + U_c \in B_k(c)$, the sub-cube with centre c.

For $f \in \mathcal{C}^1([0,1]^s)$, each term $f(c + U_c)$:

- has expectation $n \int_{B_k(c)} f(u) du$;

Haber (1966)'s estimator (optimal for $r = 1$)

$$\hat{\mathcal{I}}(f) := \frac{1}{n} \sum_{c \in \mathfrak{C}_k} f(c + U_c), \qquad U_c \sim \mathcal{U}(\left[-\frac{1}{2k}, \frac{1}{2k}\right]^s)$$

note that $c + U_c \in B_k(c)$, the sub-cube with centre c.

For $f \in \mathcal{C}^1([0,1]^s)$, each term $f(c + U_c)$:

- has expectation $n \int_{B_k(c)} f(u) du$;
- has variance $n^{-2/s}$, since for $u, v \in B_k(c)$,

$$|f(u) - f(v)| \leq \|\nabla f\|_\infty \frac{1}{k}$$

Haber (1966)'s estimator (optimal for $r = 1$)

$$\hat{\mathcal{I}}(f) := \frac{1}{n} \sum_{c \in \mathfrak{C}_k} f(c + U_c), \qquad U_c \sim \mathcal{U}\left(\left[-\frac{1}{2k}, \frac{1}{2k}\right]^s\right)$$

note that $c + U_c \in B_k(c)$, the sub-cube with centre c.

For $f \in \mathcal{C}^1([0,1]^s)$, each term $f(c + U_c)$:

- has expectation $n \int_{B_k(c)} f(u) du$;
- has variance $n^{-2/s}$, since for $u, v \in B_k(c)$,

$$|f(u) - f(v)| \leq \|\nabla f\|_\infty \frac{1}{k}$$

$\Rightarrow$ the RMSE of $\hat{\mathcal{I}}(f)$ is $\mathcal{O}(n^{-1/2-1/s})$ provided $f \in \mathcal{C}^1$.

In terms of function approximation

Alternatively, we could approximate f by a piecewise constant function:

$$f_n(x) = \sum_{c \in \mathfrak{C}_k} f(c) \times \mathbf{1}_{B_k(c)}(x)$$

and we would recover a very similar estimator.

In terms of function approximation

Alternatively, we could approximate f by a piecewise constant function:

$$f_n(x) = \sum_{c \in \mathfrak{C}_k} f(c) \times \mathbf{1}_{B_k(c)}(x)$$

and we would recover a very similar estimator.

Only difference: we use stratified random variables (rather than IID variables). Variance reduction, but does not change the rate.

Haber (1967)'s estimator (optimal for $r = 2$)

$$\hat{\mathcal{I}}(f) := \frac{1}{k^s} \sum_{c \in \mathfrak{C}_k} g_c(U_c), \qquad U_c \sim \mathcal{U}(\left[-\frac{1}{2k}, \frac{1}{2k}\right]^s)$$

where

$$g_c(u) = \frac{f(c+u) + f(c-u)}{2}$$

so $n = 2k^s$ evaluations; local antithetic effects.

Haber (1967)'s estimator (optimal for $r = 2$)

$$\hat{\mathcal{I}}(f) := \frac{1}{k^s} \sum_{c \in \mathfrak{C}_k} g_c(U_c), \qquad U_c \sim \mathcal{U}\left(\left[-\frac{1}{2k}, \frac{1}{2k}\right]^s\right)$$

where

$$g_c(u) = \frac{f(c+u) + f(c-u)}{2}$$

so $n = 2k^s$ evaluations; local antithetic effects.

Each term $g_c(U_c)$:

- has expectation $n \int_{B(c)} f(u) du$
- has variance $n^{-4/s}$ since $g_c(u) = f(c) + \mathcal{O}(\|u\|^2)$

Haber (1967)'s estimator (optimal for $r = 2$)

$$\hat{\mathcal{I}}(f) := \frac{1}{k^s} \sum_{c \in \mathfrak{C}_k} g_c(U_c), \qquad U_c \sim \mathcal{U}\left(\left[-\frac{1}{2k}, \frac{1}{2k}\right]^s\right)$$

where

$$g_c(u) = \frac{f(c+u) + f(c-u)}{2}$$

so $n = 2k^s$ evaluations; local antithetic effects.

Each term $g_c(U_c)$:

- has expectation $n \int_{B(c)} f(u) du$
- has variance $n^{-4/s}$ since $g_c(u) = f(c) + \mathcal{O}(\|u\|^2)$

$\Rightarrow$ the RMSE of $\hat{\mathcal{I}}(f)$ is $\mathcal{O}(n^{-1/2-2/s})$ provided $f \in \mathcal{C}^2$.

We generalise Haber's estimators

Two approaches:

1. Higher-order cancellations by combining 3 or more terms
2. Control variates

Section 2

Vanishing functions: combining 3 or more terms

Cancellation: combining 4 terms

for $f \in \mathcal{C}^4([0,1]^s)$, since

$$g_c(u) := \frac{f(c+u) + f(c-u)}{2} = f(c) + \frac{1}{2} u^T H(c) u + \mathcal{O}(k^{-4})$$

one has

$$g_c(\lambda u) - \lambda^2 g_c(u) = (1-\lambda^2) f(c) + \mathcal{O}(k^{-4})$$

Cancellation: combining 4 terms

for $f \in \mathcal{C}^4([0,1]^s)$, since

$$g_c(u) := \frac{f(c+u) + f(c-u)}{2} = f(c) + \frac{1}{2} u^T H(c) u + \mathcal{O}(k^{-4})$$

one has

$$g_c(\lambda u) - \lambda^2 g_c(u) = (1-\lambda^2) f(c) + \mathcal{O}(k^{-4})$$

Problem: if $|\lambda| \neq 1$, $\sum_c f(c + \lambda U_c)$ does not have the right expectation, since the support of $c + \lambda U_c$ is now a hyper-cube with centre c and side length $|\lambda|/k$.

Vanishing function

Assume that f may be extended to $\bar{f} : \mathbb{R}^s \to \mathbb{R}$, with $\bar{f} \in \mathcal{C}^r(\mathbb{R}^s)$ and $\bar{f}(x) = 0$ for $x \notin [0,1]^s$. Then

$$\int_{[0,1]^s} f(u)du = \int_{\mathbb{R}^s} \bar{f}(u)du.$$

Vanishing function

Assume that f may be extended to $\bar{f} : \mathbb{R}^s \to \mathbb{R}$, with $\bar{f} \in \mathcal{C}^r(\mathbb{R}^s)$ and $\bar{f}(x) = 0$ for $x \notin [0,1]^s$. Then

$$\int_{[0,1]^s} f(u)du = \int_{\mathbb{R}^s} \bar{f}(u)du.$$

Take $|\lambda| = 1, 3, 5, \ldots$ and $m \geq (|\lambda| - 1)/2$, so that

$$\frac{1}{k^s} \sum_{c \in \mathfrak{C}_{m,k}} f(c + \lambda U_c)$$

remains an unbiased estimator of $\mathcal{I}(f)$. (Here, $\mathfrak{C}_{m,k}$ is $\mathfrak{C}_k$ plus the centers of a few cubes immediately outside of $[0,1]^s$.)

Vanishing function

Assume that f may be extended to $\bar{f} : \mathbb{R}^s \to \mathbb{R}$, with $\bar{f} \in \mathcal{C}^r(\mathbb{R}^s)$ and $\bar{f}(x) = 0$ for $x \notin [0,1]^s$. Then

$$\int_{[0,1]^s} f(u) du = \int_{\mathbb{R}^s} \bar{f}(u) du.$$

Take $|\lambda| = 1, 3, 5, \ldots$ and $m \geq (|\lambda| - 1)/2$, so that

$$\frac{1}{k^s} \sum_{c \in \mathfrak{C}_{m,k}} f(c + \lambda U_c)$$

remains an unbiased estimator of $\mathcal{I}(f)$. (Here, $\mathfrak{C}_{m,k}$ is $\mathfrak{C}_k$ plus the centers of a few cubes immediately outside of $[0,1]^s$.)

For instance, for $\lambda = 3$, the expectation of a given term is the integral of f over 3^s sub-cubes. In return, each sub-cube is "visited" exactly 3^s times.

Vanishing estimator (any $r \geq 1$)

Based on $n = rk^s$ evaluations of f:

$$\hat{\mathcal{I}}_{r,k}(f) = \frac{1}{k^s} \sum_{c \in \mathfrak{C}_{m,k}} \sum_{i=1}^{r} \gamma_i^r f(c + \lambda_i U_c)$$

where $(\lambda_1, \lambda_2, \lambda_3, \cdots) = (1, -1, 3, -3, 5, -5, \cdots)$, and the γ_i^rs are chosen so that

$$\sum_{i=1}^{r} \gamma_i^r f(c + \lambda_i u) = f(c) + \mathcal{O}(\|u\|^r)$$

(Vandermonde system)

Conclusion

We get an unbiased estimator, with (optimal) RMSE $\mathcal{O}(n^{-1/2-r/s})$ (provided f is $\mathcal{C}^r$), which is easy to compute.

Conclusion

We get an unbiased estimator, with (optimal) RMSE $\mathcal{O}(n^{-1/2-r/s})$ (provided f is $\mathcal{C}^r$), which is easy to compute.

However, this is restricted to vanishing functions.

Practical use of the vanishing estimator

An integral over $\mathbb{R}^s$ may be rewritten as:

$$\begin{aligned}\int_{\mathbb{R}^s} h(x)dx &= \int_{\mathbb{R}^s} q(x)\frac{h(x)}{q(x)}dx \\ &= \int_{[0,1]^s} \frac{h(T(u))}{q(T(u))}du\end{aligned}$$

where T is the map such that $T(U)$ is a r.v. with probability density q (Rosenblatt transformation, a.k.a. multivariate inverse CDF).

Practical use of the vanishing estimator

An integral over $\mathbb{R}^s$ may be rewritten as:

$$\int_{\mathbb{R}^s} h(x)dx = \int_{\mathbb{R}^s} q(x)\frac{h(x)}{q(x)}dx$$
$$= \int_{[0,1]^s} \frac{h(T(u))}{q(T(u))}du$$

where T is the map such that $T(U)$ is a r.v. with probability density q (Rosenblatt transformation, a.k.a. multivariate inverse CDF).

The so-obtained integrand is vanishing as soon as $q \ll h$ in the tails; take q as e.g. a product of independent Student variables.

Practical use of the vanishing estimator

An integral over $\mathbb{R}^s$ may be rewritten as:

$$\begin{aligned}\int_{\mathbb{R}^s} h(x)dx &= \int_{\mathbb{R}^s} q(x)\frac{h(x)}{q(x)}dx \\ &= \int_{[0,1]^s} \frac{h(T(u))}{q(T(u))}du\end{aligned}$$

where T is the map such that $T(U)$ is a r.v. with probability density q (Rosenblatt transformation, a.k.a. multivariate inverse CDF).

The so-obtained integrand is vanishing as soon as $q \ll h$ in the tails; take q as e.g. a product of independent Student variables.

See numerical experiments.

Section 3

General case: control variates and numerical derivatives

Control variates

General recipe to reduce the variance of a Monte Carlo estimator:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Find variables Z_i such that:

- $\mathbb{E}[Z_i] = 0$
- $\mathrm{Corr}(Y_i, Z_i) \gg 0$

Then replace the above estimator by:

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - Z_i)$$

Back to Haber's second estimator

$$\hat{\mathcal{I}}(f) = \frac{1}{k^s} \sum_{c \in \mathfrak{C}_k} g_c(U_c)$$

where

$$g_c(u) = \frac{1}{2} f(c+u) + \frac{1}{2} f(c-u) = f(c) + \frac{1}{2} u^T H_f(c) u + \mathcal{O}(\|u\|^4)$$

Back to Haber's second estimator

$$\hat{\mathcal{I}}(f) = \frac{1}{k^s} \sum_{c \in \mathfrak{C}_k} g_c(U_c)$$

where

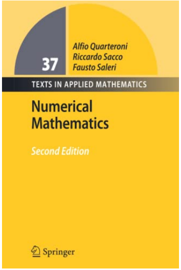
$$g_c(u) = \frac{1}{2} f(c+u) + \frac{1}{2} f(c-u) = f(c) + \frac{1}{2} u^T H_f(c) u + \mathcal{O}(\|u\|^4)$$

Obvious control variate:

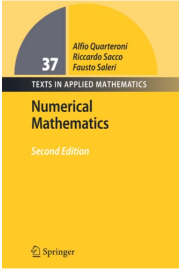
$$\frac{1}{2} U_c^T H_f(c) U_c - \mathbb{E}\left[\frac{1}{2} U_c^T H_f(c) U_c\right]$$

Drawback: requires to compute the Hessian.

Numerical derivatives, a hot topic



Numerical derivatives, a hot topic



Cover of A 660+ page book on numerical mathematics with **0** page on numerical derivatives.

Things to know (about numerical derivatives)

Based on finite differences, e.g. for $f : \mathbb{R} \to \mathbb{R}$

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h^2)$$

$$\frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)}{12h} = f'(x) + \mathcal{O}(h^4)$$

Things to know (about numerical derivatives)

Based on finite differences, e.g. for $f : \mathbb{R} \to \mathbb{R}$

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \mathcal{O}(h^2)$$

$$\frac{-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)}{12h} = f'(x) + \mathcal{O}(h^4)$$

Choice of h: non-trivial trade-off between formula error and finite precision error.

One thing to know about *multivariate* numerical derivatives

It is actually easy (and fast) to approximate any partial derivative **on a grid**: that is, if you have already computed $f(c)$ for each $c \in \mathfrak{C}_k$, then you can approximate $D^\alpha f(c)$ by combining the neighbour terms, $f(c \pm \lambda h)$, with $h = 1/k$, $\lambda \in \mathbb{N}$.

Back to our problem

We compute $f(c)$ for each centre $c \in \mathfrak{C}_k$ of our stratified sub-cubes.

Back to our problem

We compute $f(c)$ for each centre $c \in \mathfrak{C}_k$ of our stratified sub-cubes.

We obtain all the numerical derivatives of order 2, 4, ... by doing linear combinations of these $f(c)$.

Back to our problem

We compute $f(c)$ for each centre $c \in \mathfrak{C}_k$ of our stratified sub-cubes.

We obtain all the numerical derivatives of order 2, 4, ... by doing linear combinations of these $f(c)$.

We replace in our control variates the true derivatives with these numerical derivatives.

Back to our problem

We compute $f(c)$ for each centre $c \in \mathfrak{C}_k$ of our stratified sub-cubes.

We obtain all the numerical derivatives of order 2, 4, ... by doing linear combinations of these $f(c)$.

We replace in our control variates the true derivatives with these numerical derivatives.

This means that $h = 1/k$, so our numerical derivatives are not very accurate, but they are just accurate enough to avoid changing the order of convergence.

Formal definition of our estimator

For a given $r \geq 2$

$$\frac{1}{k^s} \sum_{c \in \mathfrak{C}_{0,k}} \left\{ g_c(U_c) + \sum_{l=2,4,\ldots,r} \sum_{|\alpha|=l} \frac{1}{\alpha!} \hat{D}^\alpha f(c)(U_c^\alpha - \mathbb{E} U_c^\alpha) \right\}$$

with $g_c(u) = \{f(c+u) + f(c-u)\}/2$.

Formal definition of our estimator

For a given $r \geq 2$

$$\frac{1}{k^s} \sum_{c \in \mathfrak{C}_{0,k}} \left\{ g_c(U_c) + \sum_{l=2,4,\ldots,r} \sum_{|\alpha|=l} \frac{1}{\alpha!} \hat{D}^\alpha f(c)(U_c^\alpha - \mathbb{E} U_c^\alpha) \right\}$$

with $g_c(u) = \{f(c+u) + f(c-u)\}/2$.

Based on $n = 3k^s$ evaluations: 2 third at random places, one third at the centres c.

Formal definition of our estimator

For a given $r \geq 2$

$$\frac{1}{k^s} \sum_{c \in \mathfrak{C}_{0,k}} \left\{ g_c(U_c) + \sum_{l=2,4,\ldots,r} \sum_{|\alpha|=l} \frac{1}{\alpha!} \hat{D}^\alpha f(c)(U_c^\alpha - \mathbb{E} U_c^\alpha) \right\}$$

with $g_c(u) = \{f(c+u) + f(c-u)\}/2$.

Based on $n = 3k^s$ evaluations: 2 third at random places, one third at the centres c.

When $r = 1, 2$, we recover Haber's estimators.

Properties of our general estimator

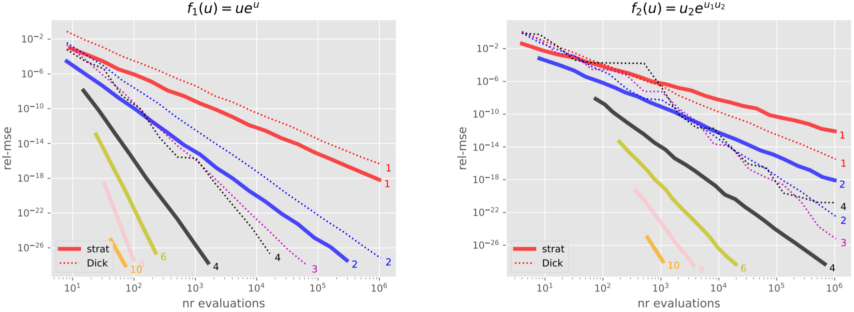
Assuming $f \in \mathcal{C}^r([0,1]^s)$:

- Optimal RMSE $\mathcal{O}(n^{-1/2-r/s})$.
- Error is $\mathcal{O}(n^{-r/s})$ with probability one.
- Error is zero if f is a polynomial of order $p < r$.

Section 4

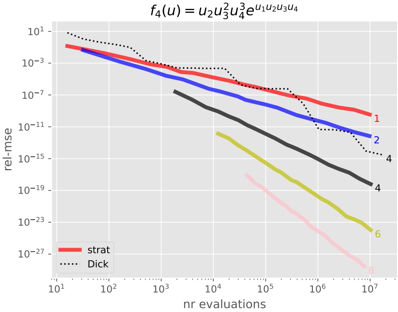
Numerical results: general estimator

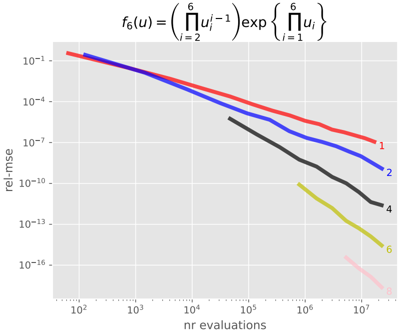
Dick (2011)'s example ($s = 1, 2$)



Relative MSE (mean squared error) vs number of evaluations for the vanishing estimator (thick lines) and Dick's estimator (dotted line). The value of r (stratified) or α (Dick's) are printed next to each curve. Left: f_1; Right: f_2.

Dick (2011)'s example ($s = 4, 6$)





Section 5

Numerical results: vanishing estimator

Setup

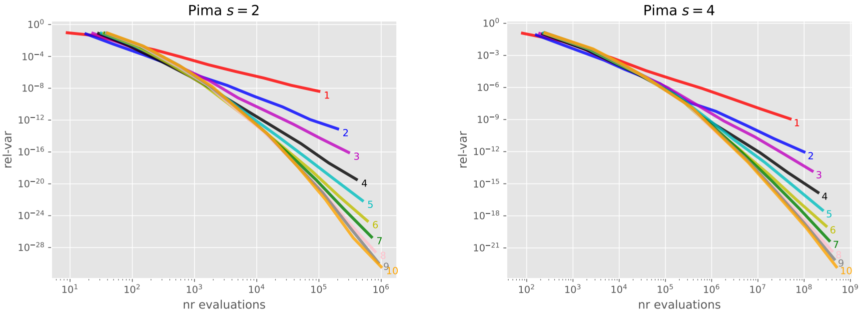
Marginal likelihood (evidence) of a Bayesian logistic model, Pima dataset; we take the s first predictors for $s = 1, \ldots, 8$.

Setup

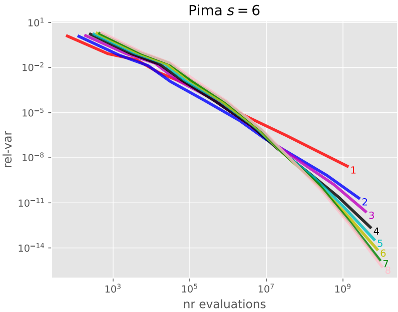
Marginal likelihood (evidence) of a Bayesian logistic model, Pima dataset; we take the s first predictors for $s = 1, \ldots, 8$.

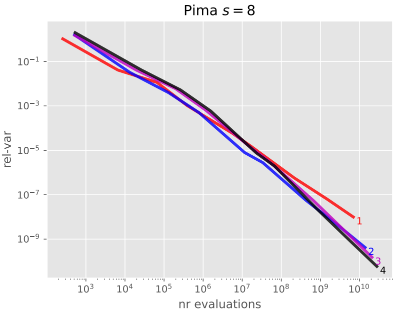
We rewrite this quantity as an integral over $[0, 1]^s$ using importance sampling (and a heavy-tailed proposal).

Plots $s = 2, 4$


Pima $s = 2$
rel-var
nr evaluations
10^0
10^{-4}
10^{-8}
10^{-12}
10^{-16}
10^{-20}
10^{-24}
10^{-28}
10^1
10^2
10^3
10^4
10^5
10^6
1
2
3
4
5
6
7
8
9
10
Pima $s = 4$
rel-var
nr evaluations
10^0
10^{-3}
10^{-6}
10^{-9}
10^{-12}
10^{-15}
10^{-18}
10^{-21}
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
1
2
3
4
5
6
7
8
9
10

Plots $s = 6, 8$


Pima $s = 6$
rel-var
nr evaluations
1
2
3
4
5
6
7
8


Pima $s = 8$
rel-var
nr evaluations
1
2
3
4

Section 6

Conclusion

Concluding remarks

- Optimality results for random quadrature were known for a long time, but practical estimators were more or less lacking.

Concluding remarks

- Optimality results for random quadrature were known for a long time, but practical estimators were more or less lacking.
- Variance estimates (from a few runs) available.

Concluding remarks

- Optimality results for random quadrature were known for a long time, but practical estimators were more or less lacking.
- Variance estimates (from a few runs) available.
- Of course, they are not so practical when $s \gg 10$. Consider the scrambling QMC strategy of Owen (1998) instead.