Bounds on the Variance of Randomly Shifted Lattice Rules

Joint work with Pierre L'Ecuyer.

Bounds on the Variance of Randomly Shifted Lattice Rules

Outline

- Expressing the variance of randomly shifted lattice rules
- Establishing bounds on the variance
- Approach and theoretical results
- Examples of problems
- Construction algorithms
- Conclusions

Multivariate integrals

High dimensional integrals of the form

$$I_d(f) = \int_{[0,1]^d} f(\mathbf{x})\mathrm{d}\mathbf{x},$$

can be approximated by quadrature rules given by

$$Q_{n,d}(f) = \frac{1}{n}\sum_{k=0}^{n-1} f\left(\mathbf{x}_k\right).$$

Here we consider randomly shifted lattice rules.

Shifted rank-1 lattice rules

Are quadrature rules given by

$$Q_{n,d}(f) = \frac{1}{n} \sum_{k=0}^{n-1} f\left(\left\{ \frac{k\mathbf{z}}{n} + \boldsymbol{\Delta} \right\}\right),$$

where the generating vector $\mathbf{z} \in \mathcal{Z}_n^d$, where
$\mathcal{Z}_n := \{z : z \in \{1, 2, \dots, n-1\},\ \gcd(z, n) = 1\}$ and $\boldsymbol{\Delta}$ is a random shift.
The number of elements of the set $\mathcal{Z}_n$ is given by $|\mathcal{Z}_n| = \varphi(n)$, where φ is Euler's totient function.

These shifted lattice rules preserve the geometrical structure of the point set and allow randomisation.

Variance of randomly shifted lattice rules

Theorem

If f is square integrable, then the variance of a randomly shifted lattice rule is given by

$$\mathrm{Var}[Q_{n,d}(f)] = {\sum_{\mathbf{h}\in L^\perp}}' |\hat{f}(\mathbf{h})|^2,$$

where $L^\perp$ denotes the dual of a lattice and the $'$ at the sum indicates that we omit the $\mathbf{h} = \mathbf{0}$ term.

For rank-1 lattice rules the dual is given by $L^\perp = \{\mathbf{h} : \mathbf{h} \cdot \mathbf{z} \equiv 0 \pmod{n}\}$.

ANOVA decomposition

The variance of randomly shifted lattice rules can be analysed through the ANOVA decomposition. The term "ANOVA" stands for *analysis of variance* and the technique is widely used in statistics or finance. The expansion

$$f(\mathbf{x}) = \sum_{\mathfrak{u} \subseteq \mathcal{D} := \{1,2,\ldots,d\}} f_{\mathfrak{u}}(\mathbf{x})$$

is an ANOVA decomposition of f, where each term is given by

$$f_{\mathfrak{u}}(\mathbf{x}) := \int_{[0,1)^{d-|\mathfrak{u}|}} f(\mathbf{x})\,\mathrm{d}\mathbf{x}_{\mathcal{D}\setminus\mathfrak{u}} - \sum_{\mathfrak{g} \subset \mathfrak{u}} f_{\mathfrak{g}}(\mathbf{x}),$$

with the last sum taken over strict subsets of $\mathfrak{u}$. It can then be checked that $f_{\emptyset} = \int_{[0,1)^d} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$ and that for each $f_{\mathfrak{u}}$, we have $\int_0^1 f_{\mathfrak{u}}(\mathbf{x})\,\mathrm{d}x_j = 0$ if $j \in \mathfrak{u}$.

Effective dimension

Since the ANOVA decomposition $f(\mathbf{x}) = \sum_{\mathfrak{u} \subseteq \mathcal{D}} f_{\mathfrak{u}}(\mathbf{x})$ is an orthogonal one, the variance can be written as

$$\sigma^2(f) = \sum_{\mathfrak{u} \subseteq \mathcal{D}} \sigma^2(f_{\mathfrak{u}}(\mathbf{x})).$$

We can then see how much of the total variance is accounted for each projection $\mathfrak{u} \subseteq \mathcal{D}$.

This leads to the idea of *effective dimension*, concept first introduced by Caflisch, Morokoff and Owen.

An example

Consider the 3-dimensional function $f(x_1, x_2, x_3) = 3x_1x_2 + 3x_3^2 + x_2$.

The ANOVA terms of this function are
$f_\emptyset = 2, f_{\{1\}}(\mathbf{x}) = x_1 - 1/2, f_{\{2\}}(\mathbf{x}) = 2x_2 - 1, f_{\{3\}}(\mathbf{x}) = 3x_3^2 - 1, f_{\{1,2\}}(\mathbf{x}) = 2x_1x_2 - x_1 - x_2 + 1/2$, while all the other terms are zero.

The Fourier coefficients of the ANOVA components are
$\hat{f}_{\{1\}}(\mathbf{h}) = \mathrm{i}/(2\pi h_1), \hat{f}_{\{2\}}(\mathbf{h}) = \mathrm{i}/(\pi h_2), \hat{f}_{\{3\}}(\mathbf{h}) = 3/(2\pi^2 h_3^2) + 3\mathrm{i}/(2\pi h_3), \hat{f}_{\{1,2\}}(\mathbf{h}) = -1/(2\pi^2 h_1 h_2)$.

An example

Consider the 3-dimensional function $f(x_1, x_2, x_3) = 3x_1x_2 + 3x_3^2 + x_2$.

The variance can be written as the sum of the variances over the ANOVA terms. Hence $\sigma^2 = \sum_{\mathfrak{u} \subseteq \mathcal{D}} \sigma_{\mathfrak{u}}^2 = \sum_{\mathfrak{u} \subseteq \mathcal{D}} \sum'_{\mathbf{h} \in \mathbb{Z}^d} |\hat{f}_{\mathfrak{u}}(\mathbf{h})|^2$.

For our example, we have $\sigma^2 = 56/45$ and it can be broken down as $\sigma_{\{1\}}^2 = 1/12, \sigma_{\{2\}}^2 = 1/3, \sigma_{\{3\}}^2 = 4/5, \sigma_{\{1,2\}}^2 = 1/36$. So the unidimensional ANOVA components $f_{\{3\}}$ and $f_{\{2\}}$ account for about 64% and 27% of the total variance, respectively.

Observation

It has been observed that functions deriving from practical fields (e.g. finance) are often of a low effective dimension in the sense that they can be well approximated by their low-order ANOVA terms.

In the ANOVA expansion of f, each term $f_{\mathfrak{u}}$ describes the interaction between variables that belong to $\mathfrak{u}$, since it only depends on these variables.

Hence, the importance of each dimension is naturally weighted. For instance, it might be the case that only interactions between two variables are important, but those involving more than two variables can be ignored.

Weights

The role of the weights is to account for the importance of variables or group of variables.

We denote by $\gamma_{\mathfrak{u}} \geq 0$, the weight associated with the subset $\mathfrak{u} \subseteq \mathcal{D} := \{1, 2, \ldots, d\}$.

It is taken bigger/smaller if the importance of variables belonging to $\mathfrak{u}$ is bigger/smaller.

Bounds on the variance

Bounds on the variance will follow by assuming a certain rate of decay of the Fourier coefficients. Here we assume that

$$|\hat{f}_{\mathfrak{u}}(\mathbf{h})| \leq \gamma_{\mathfrak{u}}^{1/2} \prod_{j \in \mathfrak{u}} |h_j|^{-a}, \quad \forall a > 1/2.$$

An interesting situation is when $1/2 < a \leq 1$. In this case the associated Fourier series is not absolutely convergent. However the variance of randomly shifted lattice rules will converge!

Bounds on the variance

With the assumption above, it will follow that the variance is bounded by

$$\mathrm{Var}[Q_{n,d}(f)] \leq D_{n,d,\gamma}(\mathbf{z},\alpha) := \sum_{\mathfrak{u}\subseteq\mathcal{D}} \sum_{\mathbf{h}\in L^{\perp}\cap(\mathbb{Z}\setminus\{0\})^d} \gamma_{\mathfrak{u}} \prod_{j\in\mathfrak{u}} |h_j|^{-\alpha}, \quad \forall\alpha > 1.$$

It follows that the variance cannot exceed the discrepancy $D_{n,d,\gamma}$ defined above.

Bounds on the variance

This discrepancy can be written as

$$D_{n,d,\gamma}(\mathbf{z},\alpha) = \frac{1}{n}\sum_{k=0}^{n-1}\sum_{\mathfrak{u}\subseteq\mathcal{D}}\gamma_{\mathfrak{u}}\prod_{j\in\mathfrak{u}}\left(\sum_{h\in\mathbb{Z}}{}' \frac{e^{2\pi\mathrm{i}hkz_j/n}}{|h|^{\alpha}}\right).$$

We recognise that this expression is actually the square worst-case error in Korobov spaces of functions, which is well studied (Dick, Joe, Kuo, Sloan, Wang, Woźniakowski etc.). Then, what is new?

Observations

1. We assume functions are only square integrable (minimal assumption made also by Monte Carlo). The associated Fourier series may not be absolutely convergent.
2. No need for periodicity assumption or assumptions on the derivatives; no need for a RKHS.
3. **The main benefit of using this discrepancy is that it can be used for a much larger class of integrands than the usual weighted Korobov spaces.**

New theoretical results

There are known results on the existence and construction of lattice rules with a low discrepancy (as defined above) under various assumptions, but there is still room left for extensions.

Here we consider the combination of general weights and composite n, which hasn't been studied before. Also we do not need n to have a fixed number of prime terms in its decomposition.

We will also obtain a refined convergence rate on our discrepancy (explained later).

Our approach

We establish existence and construction results for good shifted lattice rules using the usual averaging argument (details to follow).

We also consider a few construction schemes and compare empirically the performances of these algorithms.

Construction schemes

The most reliable algorithm is currently the well known component-by-component construction.

Component-by-component (CBC) algorithm:
Let $z_1 := 1$.
For $m = 2, 3, \ldots, d$:
find $z_m \in \mathcal{Z}_n$ *such that* $D_{n,m,\gamma}(\mathbf{z}, \alpha)$ *is minimised.*

A mathematical proof that this will produce indeed a good generating vector is usually done by induction over d (and it may be technical).

Construction schemes

Randomised CBC construction algorithm (R-CBC):

Let $z_1 := 1$*;*

For $m = 2, 3, \ldots, d$*:*

choose r *integers* z_m *at random in* $\mathcal{Z}_n$*, and select the one that minimises* $D_{n,m,\gamma}(\mathbf{z}, \alpha)$*.*

One may ask for a comparison on the figure of merit with the mean and repeat the algorithm until the figure of merit is smaller than the mean (an "acceptance-rejection" algorithm).

Construction schemes

An even simpler algorithm is a uniform random search in $(\mathcal{Z}_n)^d$:

Uniform random search algorithm:
Choose r *vectors* $\mathbf{z}$ *at random in* $(\mathcal{Z}_n)^d$*, and select the one that minimises* $D_{n,d,\gamma}(\mathbf{z},\alpha)$*.*

Existence results

Are established by analysing a mean of the quantities $D_{n,d,\gamma}(\mathbf{z},\alpha)$ over all possible generating vectors. Such a mean is given by

$$M_{n,d,\gamma}(\alpha) := \frac{1}{\varphi(n)^d} \sum_{\mathbf{z}\in\mathcal{Z}_n^d} D_{n,d,\gamma}(\mathbf{z},\alpha).$$

Recall that $\mathcal{Z}_n := \{z : z \in \{1,2,\ldots,n-1\}, \gcd(z,n) = 1\}$.

Closed form formulae for the mean are usually available only for prime n. In the non-prime case, we establish a bound for the mean.

Results

We established that there exists a generating vector $\mathbf{z}^*$ such that

$$D_{n,d,\boldsymbol{\gamma}}(\mathbf{z}^*, \alpha) \leq M_{n,d,\boldsymbol{\gamma}}(\alpha) \leq \frac{n + \varphi(n)}{n\varphi(n)} \sum_{\mathfrak{u} \subseteq \mathcal{D}} \gamma_{\mathfrak{u}} (2\zeta(\alpha))^{|\mathfrak{u}|},$$

where ζ denotes the Riemann-zeta function.

We also proved that the CBC construction yields a vector that satisfies the same inequality as above.

Convergence

It is known that the square worst-case error in Korobov spaces converges as $O(n^{-\alpha}(\log n)^{d\alpha})$. So the same convergence rate will follow for our discrepancy. Under appropriate conditions on the weights, the convergence rate is almost $O(n^{-\alpha})$ with the involved constant independent of the dimension (strong tractability).

Here, we establish a refined convergence rate as follows.

Convergence

Theorem

For any integer $n \geq 3$, *any dimension* $d \geq 2$, *and any weights, there exists a vector* $\mathbf{z}^* \in \mathcal{Z}_n^d$ *such that for any* β *satisfying* $1 \leq \beta < \alpha$, *we have*

$$D_{n,d,\gamma}(\mathbf{z}^*, \alpha) \leq \left(\frac{C \log\log n}{n} \sum_{\mathfrak{u} \subseteq \mathcal{D}} \gamma_{\mathfrak{u}}^{1/\beta} (2\zeta(\alpha/\beta))^{|\mathfrak{u}|} \right)^{\beta},$$

where $C > 0$ *is an absolute constant.*

This allows us to conclude that there exists a generating vector such that the discrepancy converges as $O(n^{-\beta}(\log\log n)^{\beta})$. Although $\beta < \alpha$, the merit of such a convergence rate resides in the fact that it is independent of the dimension, under no additional conditions on the weights.

Strong tractability

If in addition, the weights are chosen such that

$$\left(\sum_{\mathfrak{u} \subseteq \mathcal{D}} \gamma_{\mathfrak{u}}^{1/\beta} (2\zeta(\alpha/\beta))^{|\mathfrak{u}|} \right)^{\beta} \leq C(\alpha, \beta, \boldsymbol{\gamma}),$$

where the constant $C(\alpha, \beta, \boldsymbol{\gamma})$ is independent of d, then

$$D_{n,d,\boldsymbol{\gamma}}(\mathbf{z}^*, \alpha) \leq C(\alpha, \beta, \boldsymbol{\gamma}) n^{-\alpha+\varepsilon},$$

which shows also that the variance is almost $O(n^{-\alpha})$ (with $\alpha > 1$), so better than Monte Carlo.

Practical problems

One example is the **mixed logit model** (Train, Bastin, Cirillo etc.), which, in simple terms, is a discrete choice model that provides a representation on how random individuals make choices from a finite number of alternatives.

The problem requires evaluation of the probability that a random individual selects a given alternative – this involves evaluation of a multidimensional integral.

We computed the variance and the fraction of the variance for each projection in the ANOVA decomposition for several individuals and several dimensions.

Mixed logit model

Empirical observations showed that:

- The variance decreases as the cardinality of the projection increases and becomes negligible for projections larger than 4 (about 75% for 1-dimensional projections, 15% for 2-dimensional projections and about 5% for 3-dimensional projections).
- There is little fluctuation among variances of projections having the same cardinality.

This suggests that so-called "order-dependent" weights might be the most appropriate. Weights are named **order-dependent** if sets having the same cardinality have equal weights associated.

Barrier option

We did the same experiments for a barrier option in finance. This involves calculation of a high-dimensional integral with the integrand being a discontinuous function.

$$f(\mathbf{x}) = e^{-r(t_d - t_0)} \max\left(S_d(\mathbf{x}) - K\right) \times \mathbf{1}_{\min\{1 \leq j \leq d\}}[S_j(\mathbf{x}) \geq B],$$

with the usual notations (r for the interest rate, S_d the value of the asset at time t_d, K the strike price and B the barrier).

Barrier option

Our empirical observations showed that the variances corresponding to individual variables (1-dimensional projections) tend to be in decreasing order. Overall, as in the previous example, 1-dimensional projections account for most of the variance, followed by 2 and 3-dimensional projections, while higher dimensional projections account a negligible variance to the total. According to our empirical observations, successive coordinates should have decreasing importance so the sequence of corresponding weights should decrease.

Product weights are more appropriate in this case, that is $\gamma_{\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_j$.

Experiments

If the Fourier coefficients decay as $\hat{f}_{\mathfrak{u}}(\mathbf{h}) = \gamma_{\mathfrak{u}} \prod_{j \in \mathfrak{u}} |h_j|^{-1}$, then the associated Fourier series is not absolutely convergent. The figure of merit is then given by

$$D_{n,d,\gamma}(\mathbf{z}, 2) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{\mathfrak{u} \subseteq \mathcal{D}} \gamma_{\mathfrak{u}} \prod_{j \in \mathfrak{u}} \left({\sum_{h \in \mathbb{Z}}}' \frac{e^{2\pi \mathrm{i} h k z_j / n}}{|h|^2} \right).$$

We plotted the values of the figure of merit for different choices of n, d and weights, for the construction algorithms presented earlier.

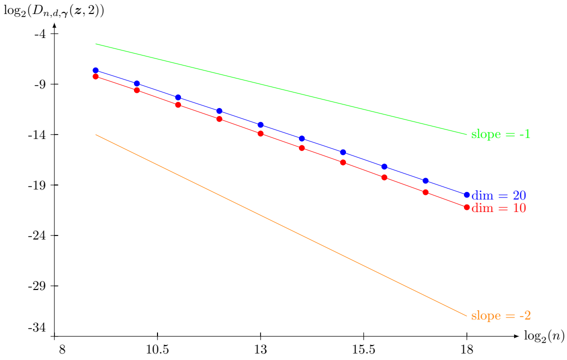
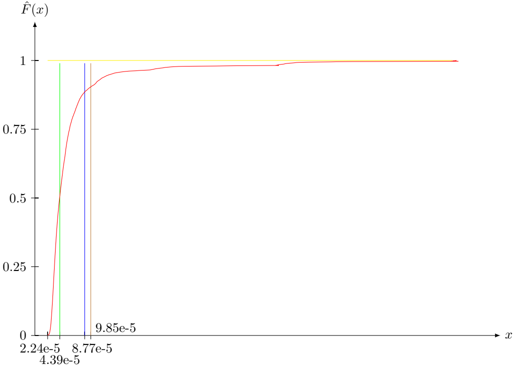
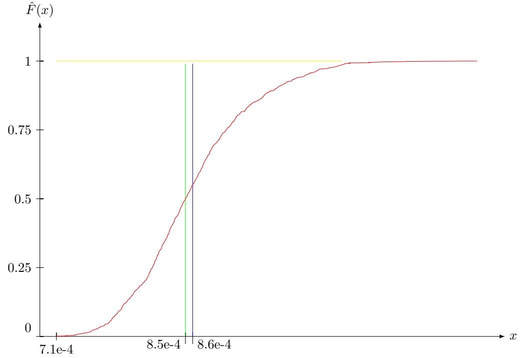


Figure 1: Values of the figure of merit produced by the systematic CBC construction for product weights $\gamma_j = 1/j^2$.



Empirical distribution of $D_{n,d,\gamma}(\boldsymbol{z}, 2)$ for 10^4 random vectors when $n = 2621444$, $d = 10$ and order-dependent weights $\Gamma_\ell = (d(d-1)\cdots(d-\ell+1))^{-1}$, for all $\ell = 1, \ldots, d$. The green, blue and brown lines indicate the empirical median, the empirical mean and the theoretical bound for the mean repsectively.



Empirical distribution of $D_{n,d,\boldsymbol{\gamma}}(\boldsymbol{z}, 2)$ out of 1000 realisations with the R-CBC algorithm when $r = 10, n = 2^{16}, d = 10$ and order-dependent weights $\Gamma_\ell = (d(d-1)\cdots(d-\ell+1))^{-1}$, for all $\ell = 1, \ldots, d$. Green and blue vertical lines indicate the empirical median and empirical mean. All values are smaller than the theoretical bound for the mean!!

Comparison of algorithms

1. CBC produces usually the best results.
2. The randomised algorithms produce better results on rare occasions (by hazard).
3. R-CBC is surprisingly good even for small number of values tested at each step. And it's easy to implement.

Conclusions

1. We bounded the variance of randomly shifted lattice rules by a quantity that is the same with the square worst-case error in Korobov spaces, but applicable for a much larger class of integrands.
2. Extended known results on the construction of lattice rules in Korobov spaces to the combination of non-prime number of points and general weights.
3. Obtained a refined convergence for the figure of merit that is independent of the dimension under no additional conditions on the weights. Also sufficient conditions for strong tractability were given.
4. Analysed a few practical problems and provided some (partial) answers to the eternal question "how to choose the weights".
5. Compared the performances of different construction schemes.

Thank you !