Repulsiveness for a better integration (not my social program)

Workshop on kernel and sampling methods for design and quantization, GDR Mascott-Num

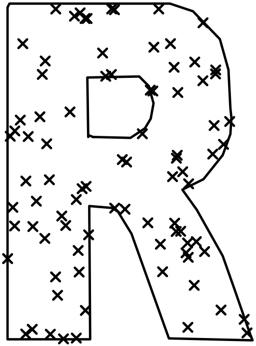
Jean-François Coeurjolly

joint works with A Mazoyer (IMT, Toulouse), PO Amblard (CNRS, U Grenoble Alpes)

November 2021

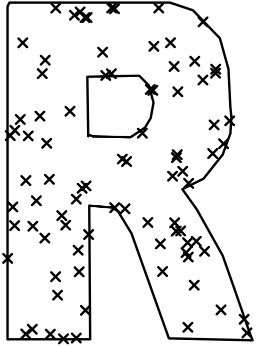
Are you sure you you're not in favor of repulsion ?

- Here are 100 points drawn randomly in . . .

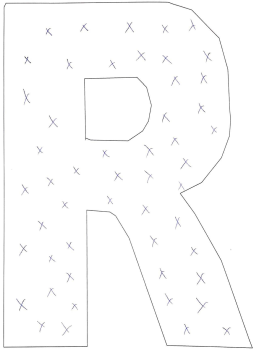


Are you sure you you're not in favor of repulsion ?

- Here are 100 points drawn randomly in . . .



- Then, I asked my daughter to draw a few points



Outline (no interest)

1. Introduction
2. Spatial point processes and DPPs
3. Main result
4. Simulation study
5. Conclusion

Outline (no interest)

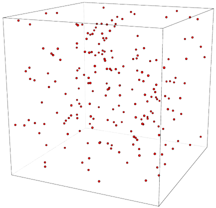
Space-filling design

Computer experiments

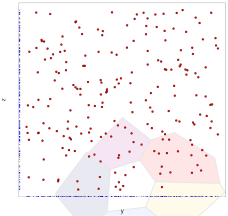
(1) The design must cover "nicely" $B = [0,1]^d$;

- Additional objective : study influence of a subset of inputs on output **without constructing a new design** ;

(2) Cover "nicely" the projection $B_I = [0,1]^\iota$ where $I \subseteq \{1,\dots,d\}$ with $|I| = \iota < d$).



↓



Spatial point processes for computer experiments

(1) Use of ***repulsive*** point processes $\mathbf{X}$ to produce the design ;

(2) Find a repulsive point process $\mathbf{X}$ such that its projection $\mathbf{X}_I$ onto B_I remains a repulsive point process.

Integral estimation problem "reformulation"

How to estimate $\mu(f_I) = \int_{[0,1]^\iota} f_I(u)\mathrm{d}u$ for any $f_I : [0,1]^\iota \to \mathbb{R}$, $I \subseteq \{1,\ldots,d\}$ and $\iota = |I|$, **using quadrature points defined in dimension** d ?

That is let $U_1,\ldots,U_N \in [0,1]^d$ (to be chosen). We want to estimate $\mu(f_I)$ by

$$\hat{\mu}(f_I) = \frac{1}{N}\sum_{j=1}^{N} f_I\left((U_j)_I \right)^{a}$$

where for any $u \in \mathbb{R}^d$, $u_I = (u_i)_{i\in I}$.

a. eventually weigthed

- The d-dimensional design is defined once for all, and used to estimate any ι-dimensional integral.
- In this problem, the design is therefore independent of the integrand (hence a "homogeneous" design makes sense).

The simplest solution

Standard Monte-Carlo

- Let $U_1, \ldots, U_N \sim \mathcal{U}([0,1]^d)$ be iid random variables
- Estimate $\mu(f_I)$ by

$$\widehat{\mu}(f_I) = \frac{1}{N} \sum_{j=1}^{N} f_I \left((U_j)_I \right)$$

- Since $(U_j)_I$ are still iid uniformly distributed on $[0,1]^\iota \ldots$ for any $I \subseteq \{1, \ldots, d\}$

$$\mathrm{E}\,\widehat{\mu}(f_I) = \mu(f_I) \qquad \text{and} \qquad \mathrm{Var}\,\widehat{\mu}(f_I) = N^{-1} \left\{ \mu(f_I^2) - \mu(f_I)^2 \right\}$$

and a central limit theorem holds (and actually much more ...)

More than 70 years of research in one slide

Improvements of standard MC : huge literature (already in the situation $\iota = d$)

- Stratified Monte-Carlo methods
- MCMC, importance sampling
- Bayesian quadrature
- Quasi Monte-Carlo, Randomized Quasi Monte-Carlo methods
- + control variates, antitihetic methods, variance-reduction methods
- …

General remarks : Either methods are

- quite stable when $d >> 1$, computationally efficient, but CLT with variance decrasing as N^{-1} ;
- or have MSE much faster ($N^{-1-2/d}$, $N^{-3-2/d+\varepsilon}$) but more unstable when $d >> 1$, not straightforward to implement, CLT does not always hold, require strong assumptions on f ($f \in C^1$ or C^d).

More than 70 years of research in one slide

Improvements of standard MC : huge literature (already in the situation $\iota = d$)

- Stratified Monte-Carlo methods
- MCMC, importance sampling
- Bayesian quadrature
- Quasi Monte-Carlo, Randomized Quasi Monte-Carlo methods
- + control variates, antitihetic methods, variance-reduction methods
- ...

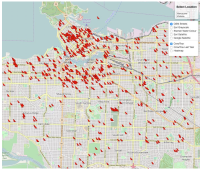
Objective

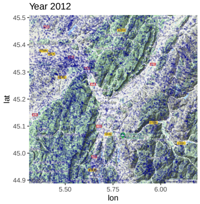
- provide a faster estimator than the standard MC one, for any f_I, under minimal assumptions on f_I (no differentiability)
- using a stochastic model ...in particular a class of repulsive spatial point processes ...and in particular a specific Determinantal Point Process.

Outline (no interest)

(Continuous) SPP on $\mathbb{R}^d$

- Let $\mathbf{X}$ be a spatial point process defined on $[0,1]^d$, viewed as a locally finite random measure : $\mathbf{x} = \{x_1, \ldots, x_m\}, \qquad x_j \in [0,1]^d$





Intensity functions (the first two ones ...and informally

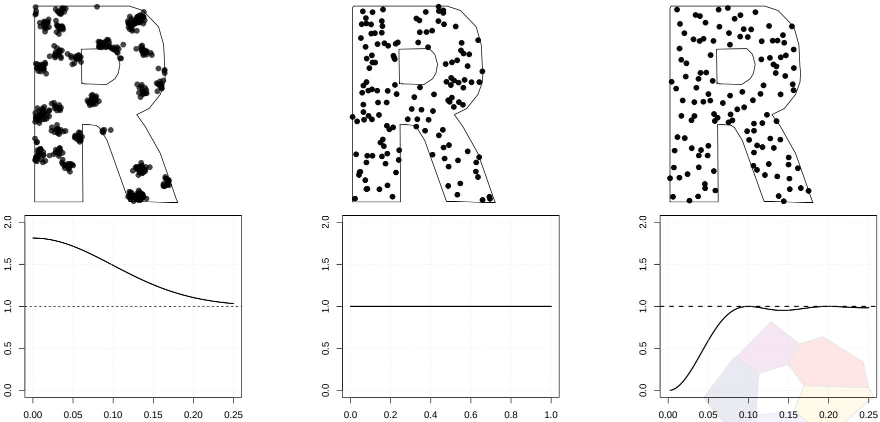
$$\rho(u) = \lim_{|\mathrm{d}u|\to 0} \frac{\mathrm{E}\{N(\mathrm{d}u)\}}{|\mathrm{d}u||\mathrm{d}v|} \qquad \text{and} \qquad \rho^{(2)}(u,v) = \lim_{|\mathrm{d}u|,|\mathrm{d}v|\to 0} \frac{\mathrm{E}\{N(\mathrm{d}u)N(\mathrm{d}v)\}}{|\mathrm{d}u||\mathrm{d}v|}$$

- $\rho(u)\mathrm{d}u \approx$ Prob. to observe a point in $B(u, \mathrm{d}u)$.
- $\rho^{(2)}(u,v)\mathrm{d}u\mathrm{d}v \approx$ Prob. to observe two distinct points in $B(u,\mathrm{d}u)$ and $B(v,\mathrm{d}v)$.
- if $\rho(\cdot) = \rho$, $\mathbf{X}$ is said to be homogeneous ; ρ=mean number of point per unit volume.

Repulsiveness and its statistical interest

- Pair correlation function (assume $\rho(\cdot) > 0$) :

$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} = \frac{\rho^{(2)}(u,v)}{\rho^2} \quad \text{(homog.)} \quad = g_0(\|v-u\|) \quad \text{(isotr.)}$$



Repulsiveness and its statistical interest

- Pair correlation function (assume $\rho(\cdot) > 0$) :

$$g(u,v) = \frac{\rho^{(2)}(u,v)}{\rho(u)\rho(v)} = \frac{\rho^{(2)}(u,v)}{\rho^2} \quad \text{(homog.)} \quad = g_0(\|v-u\|) \;\; \text{(isotr.)}$$

SPP for MC integration ($\iota = d$ for now)

Let $\mathbf{X}$ be a SPP on $[0,1]^d$ with intensity parameter N and pcf g

$$\widehat{\mu}(f) = N^{-1} \sum_{u \in \mathbf{X}} f(u) \quad \text{is such that } \mathrm{E}\, \widehat{\mu}(f) = \mu(f)$$

and

$$\mathrm{Var}\, \widehat{\mu}(f) = N^{-1} \int_{[0,1]^d} f(u)^2 \mathrm{d}u + \int_{[0,1]^d} \int_{[0,1]^d} \{g(u,v) - 1\} f(u) f(v) \mathrm{d}u \mathrm{d}v.$$

(Continuous) Determinantal point processes

$\mathbf{X} \sim \mathrm{DPP}_{B^d}(K)$ for some kernel K …

…if its kth order intensity ($k \geq 1$) writes

$$\rho^{(k)}(u_1, \dots, u_k) = \det\left[\Big(K(u_i, u_j)\Big)_{i,j=1,\dots,k}\right].$$

where K admits the Mercer decomposition $K(u, v) = \sum_{j \in \mathcal{N}_d} \lambda_j \phi_j(u) \bar{\phi}_j(v)$, where $\{\phi_j\}_j$ forms an orthonormal basis of $L^2(B^d)$; $\lambda_j \in [0, 1]$= eigenvalues.

- Introduced by O. Macchi to model fermions ;
- Appear in the study of eigenvalues of certain random matrices, zeroes of Gaussian Analytic Functions (e.g. Permantle, Peres, Hough, Johanson, Soshnikov,…)
- Very tractable class of models of repulsive point processes. Assume $K(u, u) = N$, then $\rho(u) = N$, and

$$g(u, v) = N^{-2} \det\begin{pmatrix} N & K(u, v) \\ K(u, v) & N \end{pmatrix} = 1 - \frac{|K(u, v)|^2}{N^2} < 1 \text{ !!}$$

Why DPPs are interesting ?

Assume $K(u,u) = N$ (also valid for inhomogeneous)

$$\mathrm{Var}\,\widehat{\mu}(f) = N^{-1}\int_{[0,1]^d} f(u)^2 \mathrm{d}u \underbrace{- N^{-2}\sum_{j,k\in\mathcal{N}_d}\lambda_j\lambda_k\left|\int_{[0,1]^d} f(u)\mathrm{d}u\,\phi_j(u)\bar{\phi_k}(u)\right|^2}_{\leq 0\ \forall\ \mathrm{sign}(f)}$$

- Remark : Projection DPP, $\lambda_j \in \{0,1\}$, in which case N **is an integer**.

Why DPPs are interesting ?

Assume $K(u,u) = N$ (also valid for inhomogeneous)

$$\mathrm{Var}\,\widehat{\mu}(f) = N^{-1}\int_{[0,1]^d} f(u)^2 \mathrm{d}u \underbrace{-N^{-2}\sum_{j,k\in\mathcal{N}_d} \lambda_j\lambda_k \left|\int_{[0,1]^d} f(u)\mathrm{d}u\phi_j(u)\bar{\phi_k}(u)\right|^2}_{\leq 0\ \forall\ \mathrm{sign}(f)}$$

- Remark : Projection DPP, $\lambda_j \in \{0,1\}$, in which case N **is an integer**.

Bardenet and Hardy'19

- $\iota = d$, build an ad-hoc OPE (Leg. polyn.)
- **Pros** : $\mathrm{Var}\,\widehat{\mu}(f) \propto N^{-1-1/d}$; CLT for $\mu(f)$;
- **Cons** : $f \in C_b^1([0,1]^d)$ and compactly supported (due to inhomog. kernel) ; $\iota < d$ cannot be considered ; proofs very long.

Outline (no interest)

Our approach : DPP wrt Fourier basis

- $N = n_1 \times \cdots \times n_d$ and $E_N = \{1, \dots, n_1\} \times \dots \{1, \dots, n_d\}$ ($\# E_N = N$)
- $\phi_j(u) = \mathrm{e}^{2\mathrm{i}\pi j^\top u}$ for $u \in [0,1]^d$
- then we consider the (homogeneous) kernel K defined by

$$K(u,v) = \sum_{j \in E_N} \mathrm{e}^{2\mathrm{i}\pi j^\top (u-v)} = \prod_{\ell=1}^{d} K_\ell(u_\ell, v_\ell) \quad \text{where} \quad \underbrace{K_\ell(u_\ell, v_\ell) = \sum_{j=1}^{n_\ell} \mathrm{e}^{2\mathrm{i}\pi j (u_\ell - v_\ell)}}_{\propto \text{Dirichlet kernel in Sign. Proc.}} .$$

Our approach : DPP wrt Fourier basis

- $N = n_1 \times \cdots \times n_d$ and $E_N = \{1, \ldots, n_1\} \times \ldots \{1, \ldots, n_d\}$ ($\#E_N = N$)
- $\phi_j(u) = \mathrm{e}^{2\mathrm{i}\pi j^\top u}$ for $u \in [0,1]^d$
- then we consider the (homogeneous) kernel K defined by

$$K(u,v) = \sum_{j \in E_N} \mathrm{e}^{2\mathrm{i}\pi j^\top (u-v)} = \prod_{\ell=1}^{d} K_\ell(u_\ell, v_\ell) \quad \text{where} \quad \underbrace{K_\ell(u_\ell, v_\ell) = \sum_{j=1}^{n_\ell} \mathrm{e}^{2\mathrm{i}\pi j (u_\ell - v_\ell)}}_{\propto \text{Dirichlet kernel in Sign. Proc.}} .$$

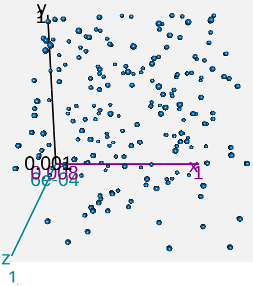
Theorem (MCA'19) : Why taking E_N as a rectangular set ?

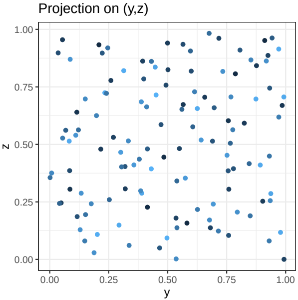
Let $\mathbf{X} \sim \mathrm{DPP}_{[0,1]^d}(K)$ and let $I \subseteq \{1, \ldots, d\}$, then

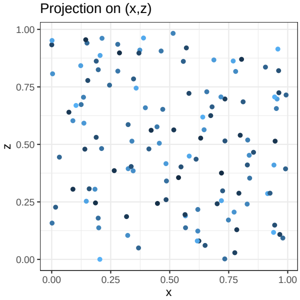
$$\mathbf{X}_I \;\sim\; (-1/N_{I^c}) - \mathrm{DPP}_{B^\iota}(N_{I^c} K_\iota)$$

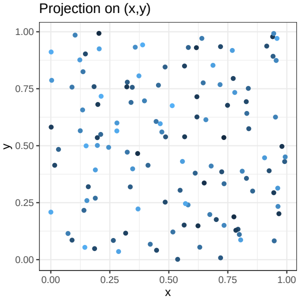
where $N = N_I \times N_{I^c}$ and $K_I = \prod_{\ell \in I} K_\ell$. In particular, $g_{\mathbf{X}_I} < 1$, $\forall I$! !

An example of a 3d-realization









Our main result

- Fourier coefficient : $\hat{f}_I(k) =< f_I, \phi_k >$
- Assume $f_I \in \mathcal{H}^s([0,1]^\iota) = \left\{ f_I \in L^2 : \sum_{k\in\mathbb{Z}^\iota} (1+\|k\|_\infty)^{2s} |\hat{f}(k)|^2 < \infty \right\}$.

Theorem (CMA'19+)

Let $d > 1$ and assume $N = n_1 \times \cdots \times n_d$ s.t. $n_\ell N^{-1/d} \to 1$

(i) If $s > 1/2$ and $\|f_I\|_\infty < \infty$ then

$$\mathrm{Var}\left(\widehat{\mu}(f_I) \right) \sim N^{-1-1/d} \, \sigma^2(f_I) \quad \text{with} \quad \sigma^2(f_I) = \sum_{k\in\mathbb{Z}^\iota} \|k\|_1 \, |\hat{f}_I(k)|^2.$$

$$\sqrt{N^{1+1/d}} \left(\widehat{\mu}(f_I) - \mu(f_I) \right) \xrightarrow{d} N\left(0, \sigma^2(f_I)\right).$$

(ii) If $0 < s < 1/2$, then

$$\mathrm{Var}\left(\widehat{\mu}(f_I) \right) = \mathcal{O}\left(N^{-1-\frac{2s}{d}} \right) = o\left(N^{-1} \right)$$

Outline (no interest)

Setting

Estimate $\mu(f)$ for different d

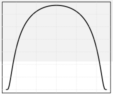
1. For $N = 100, \ldots, 1000$ an $d = 1, \ldots, 6$: 2500 replications of a Dirichlet DPP producing N points in $[0,1]^d$;
2. For each d and some given f :
   1. Estimate $\mu(f)$ for each N using the design constructed in dimension d ;
   2. Linear regression of empirical variance vs. $\log(N)$ and estimate the slope.

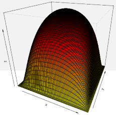
Estimate $\mu(f_I)$ for different ι

1. For $N = 100, \ldots, 1000$: 2500 replications of a Dirichlet DPP producing N points in $[0,1]^6$;
2. For $\iota \in \{6, \ldots, 1\}$:
   1. For a ι-dimensional function f_I, estimate $\mu(f_I)$ for each N using the design constructed in dimension 6 ;
   2. Linear regression of empirical vs. $\log(N)$ and estimate the slope.

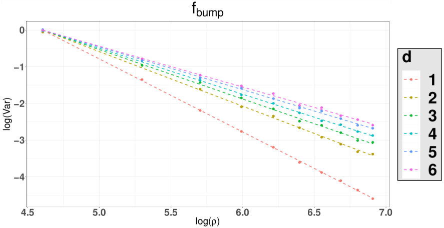
Bump : $f \in \mathcal{H}^s$, for any $s > 0$

$f(x) \propto \prod_{i=1}^{d} \exp\left(-\frac{0.1}{.25-(x_i-.5)^2}\right)$





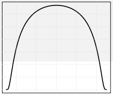
Estimation of $\mu(f)$ via $\mathbf{X}$

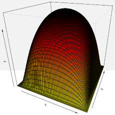


d	Conf. interval	Estimation	−1−1/d
1	[−2.016 ; −1.966]	−1.991	−2.000
2	[−1.536 ; −1.438]	−1.487	−1.500
3	[−1.391 ; −1.302]	−1.347	−1.333
4	[−1.282 ; −1.225]	−1.254	−1.250
5	[−1.2 ; −1.129]	−1.164	−1.200
6	[−1.153 ; −1.073]	−1.113	−1.167

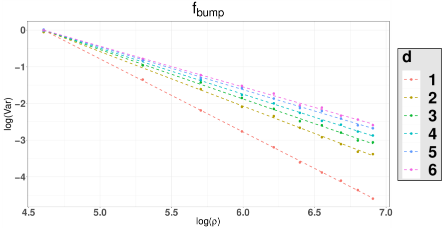
Bump : $f \in \mathcal{H}^s$, for any $s > 0$

$$f(x) \propto \prod_{i=1}^{d} \exp\left(-\frac{0.1}{.25-(x_i-.5)^2}\right)$$



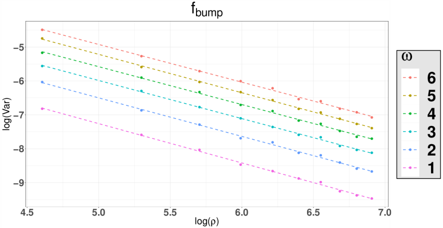


Estimation of $\mu(f)$ via $\mathbf{X}$



d	Conf. interval	Estimation	−1−1/d
1	[−2.016 ; −1.966]	−1.991	−2.000
2	[−1.536 ; −1.438]	−1.487	−1.500
3	[−1.391 ; −1.302]	−1.347	−1.333
4	[−1.282 ; −1.225]	−1.254	−1.250
5	[−1.2 ; −1.129]	−1.164	−1.200
6	[−1.153 ; −1.073]	−1.113	−1.167

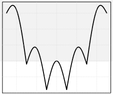
Estimation of $\mu(f_I)$ via $\mathbf{X}_I$ $(d = 6)$

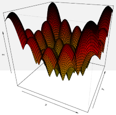


ω	Conf. interval	Estimation	−1−1/d
6	[−1.153 ; −1.073]	−1.113	−1.167
5	[−1.162 ; −1.104]	−1.133	−1.167
4	[−1.157 ; −1.077]	−1.117	−1.167
3	[−1.16 ; −1.09]	−1.125	−1.167
2	[−1.186 ; −1.087]	−1.136	−1.167
1	[−1.204 ; −1.116]	−1.160	−1.167

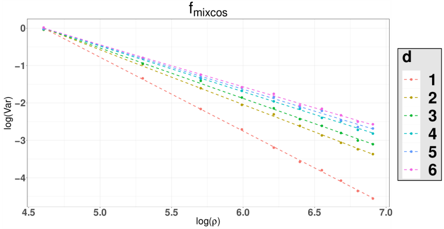
Mix-cos : $f \in \mathcal{H}^{3/2}$

$$f(x) \propto \sum_{i=1}^{d} \{0.1|\cos(5\pi(x_i - 1/2))| + (x_i - 1/2)^2\}$$



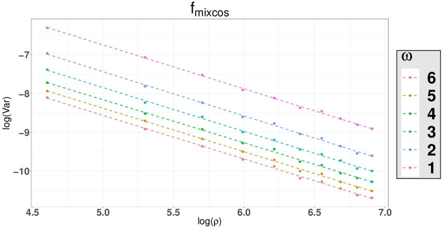


Estimation of $\int f_d$ via $\mathbf{X}$



d	Conf. interval	Estimation	−1−1/d
1	[−2.003 ; −1.942]	−1.972	−2.000
2	[−1.5 ; −1.429]	−1.464	−1.500
3	[−1.384 ; −1.31]	−1.347	−1.333
4	[−1.253 ; −1.183]	−1.218	−1.250
5	[−1.219 ; −1.123]	−1.171	−1.200
6	[−1.159 ; −1.089]	−1.124	−1.167

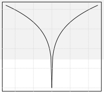
Estimation of $\int f_I$ via $\mathbf{X}_I$ $(d = 6)$

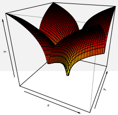


ω	Conf. interval	Estimation	−1−1/d
6	[−1.159 ; −1.089]	−1.124	−1.167
5	[−1.154 ; −1.091]	−1.123	−1.167
4	[−1.151 ; −1.072]	−1.111	−1.167
3	[−1.168 ; −1.092]	−1.130	−1.167
2	[−1.18 ; −1.102]	−1.141	−1.167
1	[−1.163 ; −1.104]	−1.134	−1.167

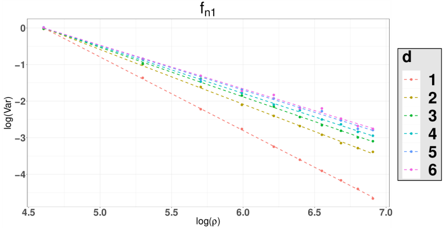
"L^γ-norm" : $f(x) \propto \sum_{i=1}^d |u_i - 1/2|^{1/4}$,

$f \in H^s$ for any $s < 3/4$



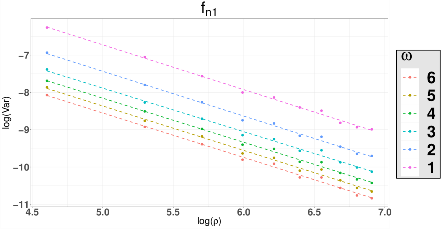


Estimation of $\mu(f)$ via $\mathbf{X}$



d	Conf. interval	Estimation	−1−1/d
1	[−2.033 ; −1.986]	−2.009	−2.000
2	[−1.529 ; −1.458]	−1.493	−1.500
3	[−1.393 ; −1.306]	−1.349	−1.333
4	[−1.318 ; −1.246]	−1.282	−1.250
5	[−1.272 ; −1.161]	−1.217	−1.200
6	[−1.262 ; −1.123]	−1.192	−1.167

Estimation $\mu(f_I)$ via $\mathbf{X}_I$ $(d = 6)$



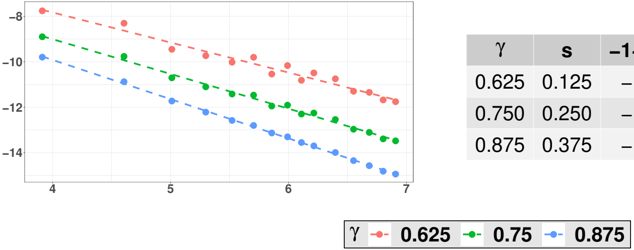
ω	Conf. interval	Estimation	−1−1/d
6	[−1.262 ; −1.123]	−1.192	−1.167
5	[−1.269 ; −1.114]	−1.191	−1.167
4	[−1.248 ; −1.118]	−1.183	−1.167
3	[−1.243 ; −1.094]	−1.168	−1.167
2	[−1.272 ; −1.129]	−1.200	−1.167
1	[−1.263 ; −1.14]	−1.201	−1.167

A last example in the situation $\mathcal{H}^s$ with $s < 1/2$

- Particular case $d = 1$, $B = [0, 1]$,

$$h_\gamma(x) = \sum_{j \geqslant 1} \frac{\cos(2\pi j x)}{2\pi j^\gamma}, \qquad \gamma > 1/2.$$

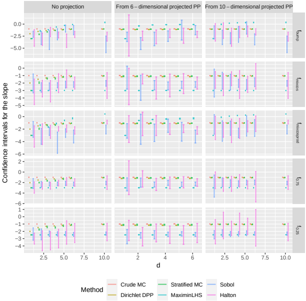
Then $h_\gamma \in \mathcal{H}^s$ for any $0 < s < \gamma - 1/2$



γ	s	−1−2s/d	Estimation	Conf. interval
0.625	0.125	−1.25	−1.32	[−1.48 ; −1.17]
0.750	0.250	−1.50	−1.53	[−1.61 ; −1.45]
0.875	0.375	−1.75	−1.73	[−1.76 ; −1.69]

More simulations . . .

- Simulations were made for $d = 6$ or 10 and any $\iota = 1, \ldots, d$ for additional functions, and for our `DirDPP` design and for : `BH-DPP` [BH'19], `crudeMC`, `stratified MC`, `Sobol`, `MaximinLHS`, `Halton` designs.



Just impossible to sum up >20 figures (please refer to EJS'21) . . . **but** . . .

- for a specific function, sample size, dimension : there were **as expected** better methods (and actually much faster)
- `DirDPP` was the only one to remain **stable** in terms of behaviour for all considered functions, sample size, dimension d and for any $\iota = 1, \ldots, d$.

Outline (no interest)

Summary :

- A simple MC estimator based on a homogeneous projection DPP.
- Universal result : unbiased est. and CLT with rate $\sqrt{N^{1+1/d}}$
- Results available
  - for $f \in \mathcal{H}^{1/2}$, i.e. for some non-differentiable functions
  - on lower dimensional spaces

Summary :

- A simple MC estimator based on a homogeneous projection DPP.
- Universal result : unbiased est. and CLT with rate $\sqrt{N^{1+1/d}}$
- Results available
  - for $f \in \mathcal{H}^{1/2}$, i.e. for some non-differentiable functions
  - on lower dimensional spaces

What I didn't say :

- Asymptotic normality "checked" (qqplots and Shapiro-Wilk test).
- Asympt. CI for $\mu(f_I)$? Interesting fact when $f_I \in C_b^1([0,1]^\iota)$

$$\sigma^2(f_I) = \sum_{j \in Z^\iota} \|j\|_1 |\hat{f}(j)|^2 \le 4\pi^2 \int_{[0,1]^\iota} \|\nabla f_I(u)\|^2 \mathrm{d}u =: \mathcal{J}_I$$

$\Rightarrow \mathcal{J}_I$ is easily estimated (using again $\mathbf{X}_I$!!), leading to a conservative asympt. CI.

Where to go ? (just a few perspectives)

- for the same problem : other projection kernels, infinite integrals
- for a given f : inhomogeneous kernels to exploit f
- lots of possible statistical applications (density estimation, regression,...) when a random design is involved

- d-dimensional sphere, torus, ...?
- simulations=$\mathcal{O}(dN^3)$ (for now), not sequential but code and replications are available : `https://github.com/AdriMaz`

Tricky Fourier basis

Just to give an idea, let's consider $d = 1$

$$\begin{aligned}
\mathrm{Var}\widehat{I_1}(f_1) &= \frac{1}{N}\sum_{j\in\mathbb{Z}} |<f,\phi_j>|^2 - \frac{1}{N^2}\sum_{j,k=1}^{N} |<f,\phi_j\bar{\phi}_k>|^2 \\
&= \frac{1}{N}\sum_{j\in\mathbb{Z}} |<f,\phi_j>|^2 - \frac{1}{N^2}\sum_{j,k=1}^{N} |<f,\phi_{j-k}>|^2 \\
&= \frac{1}{N}\sum_{j\in\mathbb{Z}} |\hat{f}(j)|^2 - \frac{1}{N^2}\sum_{|j|\le N} (N-|j|)|\hat{f}(j)|^2 \\
&= \frac{1}{N}\sum_{|j|\ge N} |\hat{f}(j)|^2 + \frac{1}{N^2}\sum_{|j|\le N} |j|\,|\hat{f}(j)|^2 \\
&\sim \frac{1}{N^2}\sum_{j\in\mathbb{Z}} |j|\,|\hat{f}(j)|^2
\end{aligned}$$