On two ways to use Determinantal Point Processes for Monte Carlo intergration

Interested in DPPs for a PhD/postdoc in France?
Contact: rbardenet.github.io

Guillaume Gautier[†*] Rémi Bardenet[†] Michal Valko[‡*†]

[†]Univ. Lille, CNRS, Centrale Lille, UMR 9189 – CRIStAL, France, [*]Inria Lille, France, [‡]DeepMind Paris, France

Goal

Numerical integration with Monte Carlo

$$\int_{\mathbb{X}} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \omega_n(\mathbf{x}_1, \dots, \mathbf{x}_N) f(\mathbf{x}_n), \quad (1)$$

where $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$.

Contributions

Comparison between two unbiased estimators:

- Bardenet & Hardy (BH, 2019), see (4)
  - in new experimental regimes (larger N, ...)
- Ermakov & Zolotukhin (EZ, 1960), see (5)
  - analysis from DPP viewpoint
  - slight extension of the origianl result
  - new short and simple proof

Implementation of the sampling scheme in the DPPy toolbox github.com/guilgautier/DPPy

Setup: projection DPPs

Consider $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim$ **projection** $\mathrm{DPP}(\mu, K)$

- reference measure $\mu(\mathrm{d}x) = \omega(x)\,\mathrm{d}x$
- **projection** kernel $K : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$
  - $K(y, x) = K(x, y)$ "$K^{\mathsf{T}} = K$"
  - $K(x, y) = \int_{\mathbb{X}} K(x, z) K(z, y) \mu(\mathrm{d}z)$ "$K^2 = K$"

$$K(x, y) \triangleq \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y) \qquad "K = \sum_{k=0}^{N-1} \phi_k \phi_k^{\mathsf{T}}"$$

$$\langle \phi_k, \phi_\ell \rangle \triangleq \int_{\mathbb{X}} \phi_k(x)\phi_\ell(x)\mu(\mathrm{d}x) = \delta_{k\ell}$$

- number of points $N = \operatorname{rank} K$
- joint distribution of $(\mathbf{x}_1, \dots, \mathbf{x}_N)$

$$\frac{1}{N!} \det(K(x_n, x_p))_{n,p=1}^{N} \mu^{\otimes N}(\mathrm{d}x) \quad (2)$$

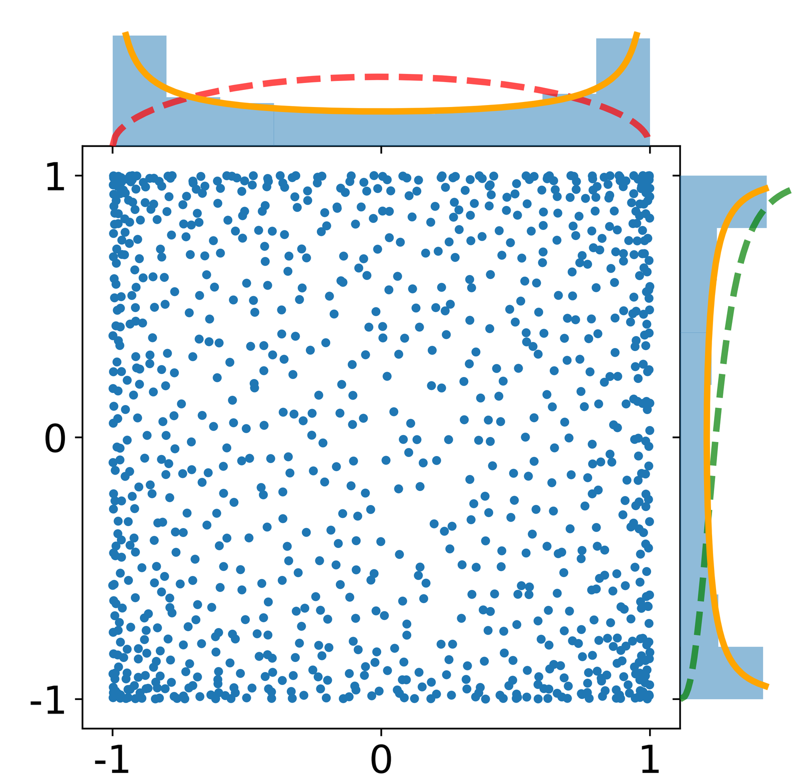
Sampling projection DPPs

To get a valid sample $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim$ projection $\mathrm{DPP}(\mu, K)$, it is enough to apply the chain rule to sample $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ and forget the order the points were selected.

$$(2) = \frac{K(x_1, x_1)}{N} \omega(x_1)\,\mathrm{d}x_1 \prod_{n=2}^{N} \frac{K(x_n, x_n) - \mathbf{K}_{n-1}(x_n)^{\mathsf{T}} \mathbf{K}_{n-1}^{-1} \mathbf{K}_{n-1}(x_n)}{N - (n-1)} \omega(x_n)\,\mathrm{d}x_n \quad (3)$$

where $\mathbf{K}_{n-1}(\cdot) = (K(x_1, \cdot), \dots, K(x_{n-1}, \cdot))^{\mathsf{T}}$, and $\mathbf{K}_{n-1} = (K(x_p, x_q))_{p,q=1}^{n-1}$.

The multivariate Jacobi ensemble



- $\mathbb{X} = [-1, 1]^d$
- $\mu(\mathrm{d}x) = \omega(x)\,\mathrm{d}x = \prod_{i=1}^{d} w^i(x^i)\,\mathrm{d}x^i$, with
  $w^i(z) = (1 - z)^{a^i}(1 + z)^{b^i}, \quad a^i, b^i > -1.$
- $\phi_k(x) = \prod_{i=1}^{d} \phi_{k^i}^i(x^i)$ Jacobi polynomials $\perp w^i$

Sampling

- $d = 1$: eigvals of **random matrix**
- $d \geq 2$: chain rule (3) with rejection sampling

x_1 : prop. $\omega_{\mathrm{eq}}(x)$, rej. const. $\lesssim 2^d$

$x_n \mid x_{1:n-1}$: prop. $\frac{K(x,x)\omega(x)}{N}$, rej. const. $\frac{N}{N-(n-1)}$

Estimators

Bardenet & Hardy (BH, 2019)

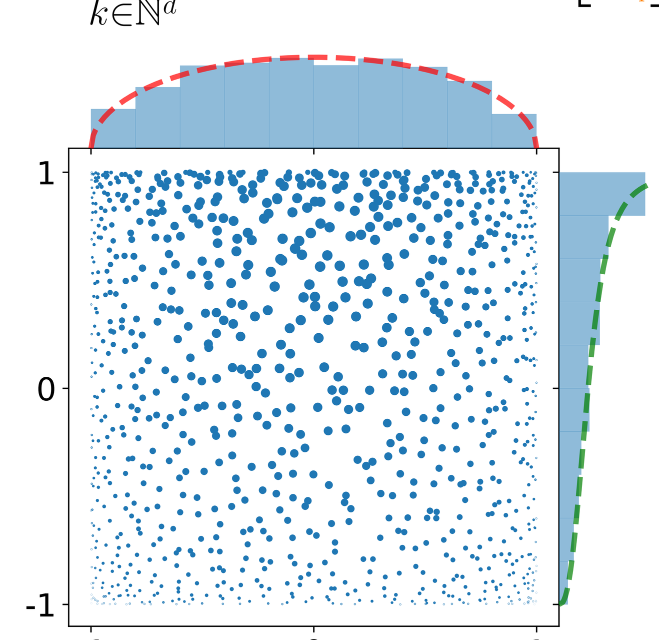
$$\widehat{I}_N^{\mathrm{BH}}(f) \triangleq \sum_{n=1}^{N} \frac{f(\mathbf{x}_n)}{K(\mathbf{x}_n, \mathbf{x}_n)}, \quad (4)$$

- $\mathbb{E}\left[\widehat{I}_N^{\mathrm{BH}}\right] = \int_{\mathbb{X}} \frac{f(x)}{\cancel{K(x,x)}} \cancel{K(x, x)} \mu(\mathrm{d}x) = \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$
- $\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{BH}}\right] = \frac{1}{2} \int_{\mathbb{X}^2} \left(\frac{f(x)}{K(x,x)} - \frac{f(y)}{K(y,y)} \right)^2 K(x, y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}y)$

Fast Central Limit Theorem

$$\sqrt{N^{1+1/d}} \left(\widehat{I}_N^{\mathrm{BH}} - \int_{[-1,1]^d} f(x)\,\mu(\mathrm{d}x) \right) \xrightarrow[N \to \infty]{\text{law}} \mathcal{N}\left(0, \mathbf{\Omega}_{f,\omega}^2\right),$$

with $\mathbf{\Omega}_{f,\omega}^2 \triangleq \frac{1}{2} \sum_{k \in \mathbb{N}^d} (k_1 + \dots + k_d)\, \mathcal{F}\left[\frac{f\,\omega}{\omega_{\mathrm{eq}}}\right](k)^2$



Ermakov & Zolotukhin (EZ, 1960)

Let $f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad 1 \leq M \leq \infty$

Solve the linear system

$$\begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{N-1}(\mathbf{x}_1) \\ \vdots & & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{N-1}(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}$$

The coordinates are unbiaised estimates of the Fourier-like coefficients of f

- $\mathbb{E}[y_k] = \langle f, \phi_k \rangle = \int_{\mathbb{X}} f(x)\phi_k(x)\mu(\mathrm{d}x)$
- $\mathbb{V}\mathrm{ar}[y_k] = \|f\|^2 - \sum_{\ell=0}^{N-1} \langle f, \phi_\ell \rangle^2$
- $\mathbb{C}\mathrm{ov}[y_j, y_k] = 0, j \neq k$

If ϕ_0 is constant

$$\widehat{I}_N^{\mathrm{EZ}}(f) = \frac{y_0}{\phi_0} \quad (5)$$

- $\mathbb{E}\left[\widehat{I}_N^{\mathrm{EZ}}\right] = \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$
- $\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{EZ}}\right] = \mu(\mathbb{X}) \left(\sum_{\ell=N}^{M-1} \langle f, \phi_\ell \rangle^2 \right)$

$\boxed{= 0 \quad \text{if } M \leq N}$

Experiments

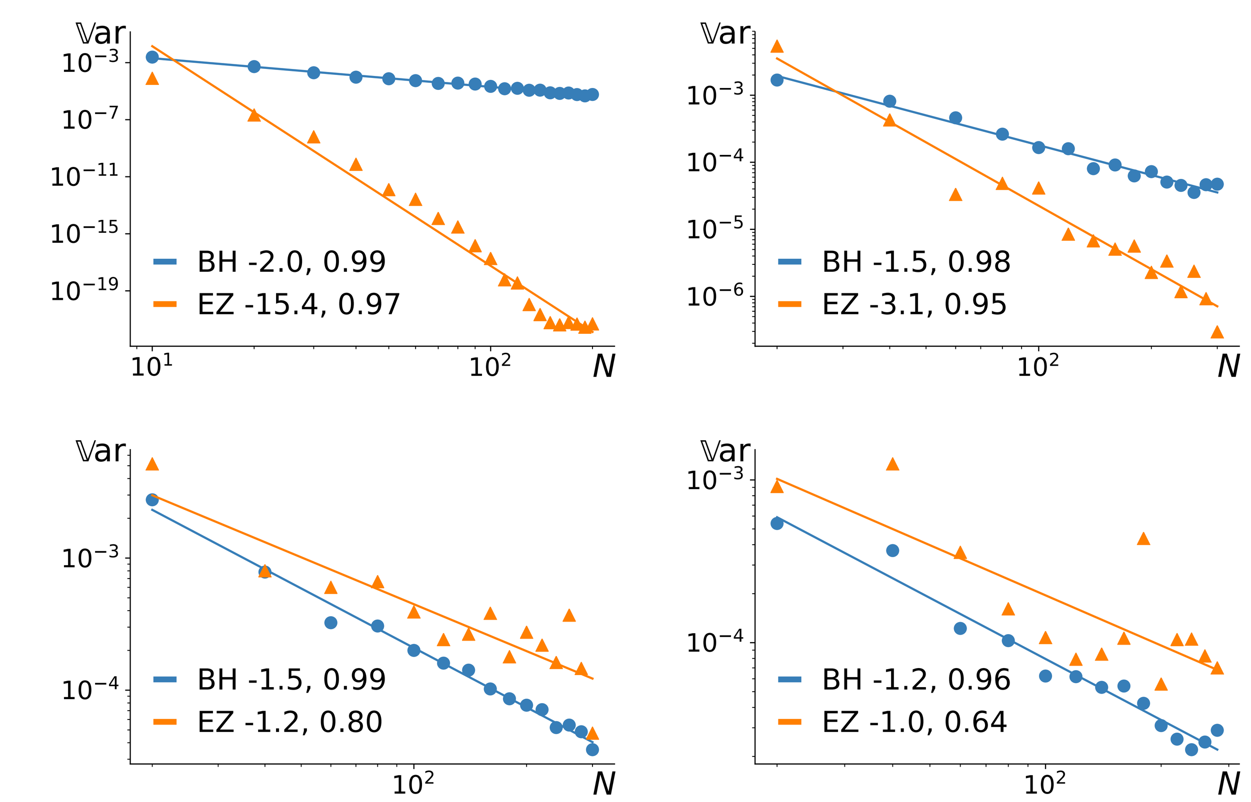


Figure 1: $f =$ smooth bump function $d = 1, 2, 3, 4$

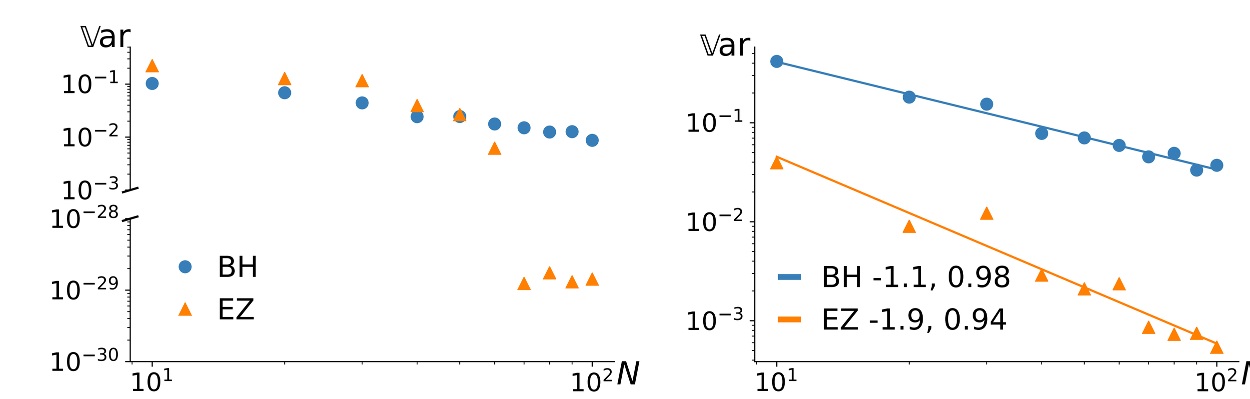


Figure 2: $f = \sum_{k=0}^{70-1} \langle f, \phi \rangle \phi_k$ and $f = \sum_{k=0}^{N+1-1} \frac{1}{k+1} \phi_k$

Punchlines

BH estimator see (4)

- $\omega_n = \frac{1}{K(\mathbf{x}_n, \mathbf{x}_n)}$ **random** Gaussian **quadrature**
- $\mathbb{V}\mathrm{ar} = \mathcal{O}(N^{-(1+1/d)})$

EZ estimator, see (5)

- 1960 $\ll$ Macchi (1975), formalized DPPs
- First connection to **projection** DPPs
- **Interpretable and practical variance**
- **Potential of $\mathbb{V}\mathrm{ar} = 0$ in any dimension**
  $\Longrightarrow$ perfect integration/reconstruction of f
- Linear system stability/regularization?
- Prove asymptotic results, CLT?

Basis $\phi_0, \phi_1, \dots$ where f is smooth must drive the choice of the kernel
$K(x, y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$

Sample from DPP defined from wavelets? ☺