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

Guillaume Gautier, Rémi Bardenet, and Michal Valko

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

Abstract

Compare two DPP-based estimators of

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

- Bardenet & Hardy (BH, 2019), see (3)
  - in new experimental regimes (larger N, ...)
- Ermarkov & Zolotukhin (EZ, **1960**), see (4)
  - analysis from DPP viewpoint
  - slight extension of the origianl result
  - new short and simple proof

Provide a sampler for **a specific DPP**

DPPy: DPP sampling with Python
github.com/guilgautier/DPPy
JMLR-MLOSS, in press, 2019.

If f is sparse or has fast-decaying coefficients in a given basis then adapt your DPP kernel and go for EZ, otherwise it is safer to use BH

Setup: projection DPPs

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

- **Repulsive random set of points**
- joint distribution of $(\mathbf{x}_1,\dots,\mathbf{x}_N)$

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

- $K : \mathbb{X}\times\mathbb{X} \to \mathbb{R}$ projection kernel

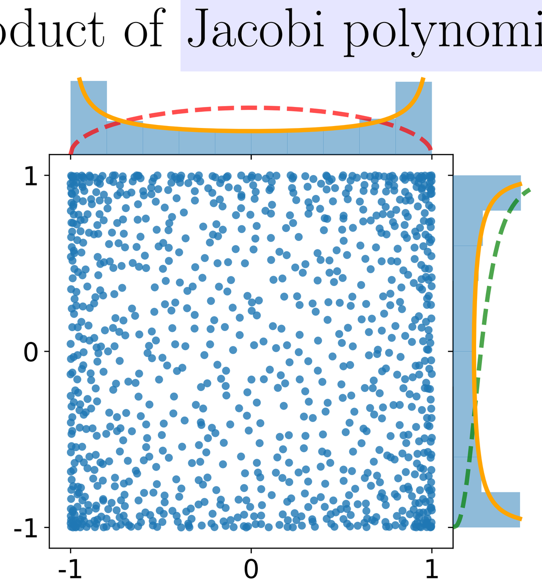
$$K(x,y) \triangleq \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y),$$

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

- reference measure μ on $\mathbb{X}$

The multivariate Jacobi ensemble

- $\mathbb{X} = [-1,1]^d$
- $\mu = \omega(x)\,\mathrm{d}x = \mathrm{Beta}(a_1,b_1)\otimes\cdots\otimes\mathrm{Beta}(a_d,b_d)$
- $\phi_k(x) =$ product of Jacobi polynomials



Sampling projection DPPs

Sampling $\equiv$ sequential Gram-Schmidt orthogonalization of feature vectors $\Phi(x_1),\dots,\Phi(x_N)$

where $K(x,y) = \Phi(x)^{\mathsf{T}}\Phi(y)$, with $\Phi(x) = (\phi_0(x)),\dots,\phi_{N-1}(x))$.

Apply the chain rule to sample $(\mathbf{x}_1,\dots,\mathbf{x}_N)$ and forget the order the points were selected

$$(1) = \frac{\|\Phi(x_1)\|^2}{N}\omega(x_1)\,\mathrm{d}x_1 \prod_{n=2}^{N} \frac{\mathrm{distance}^2\big(\Phi(x_n), \mathrm{span}\{\Phi(x_p)\}_{p=1}^{n-1}\big)}{N-(n-1)}\omega(x_n)\,\mathrm{d}x_n \qquad (2)$$

Sampling the Jacobi ensemble

- $d=1$: eigvals of a **random** $\begin{pmatrix} & 0\\ 0 & \end{pmatrix}$ **matrix**
- $d \geq 2$: chain rule (2) with rejection sampling

$$x_1 : \begin{cases} \text{proposal density } \omega_{\mathrm{eq}}(x)\,\mathrm{d}x \\ \text{rejection constant } \lesssim 2^d \end{cases}$$

$$x_n \mid x_{1:n-1} \sim \begin{cases} \text{proposal } N^{-1}K(x,x)\omega(x)\,\mathrm{d}x \\ \text{rejection constant } \frac{N}{N-(n-1)} \end{cases}$$

Total number of rejections $\approx 2^d N \log(N)$

First estimator (BH, 2019)

$$\widehat{I}_N^{\mathrm{BH}}(f) \triangleq \sum_{n=1}^{N} \frac{1}{K(\mathbf{x}_n,\mathbf{x}_n)} f(\mathbf{x}_n) \qquad (3)$$

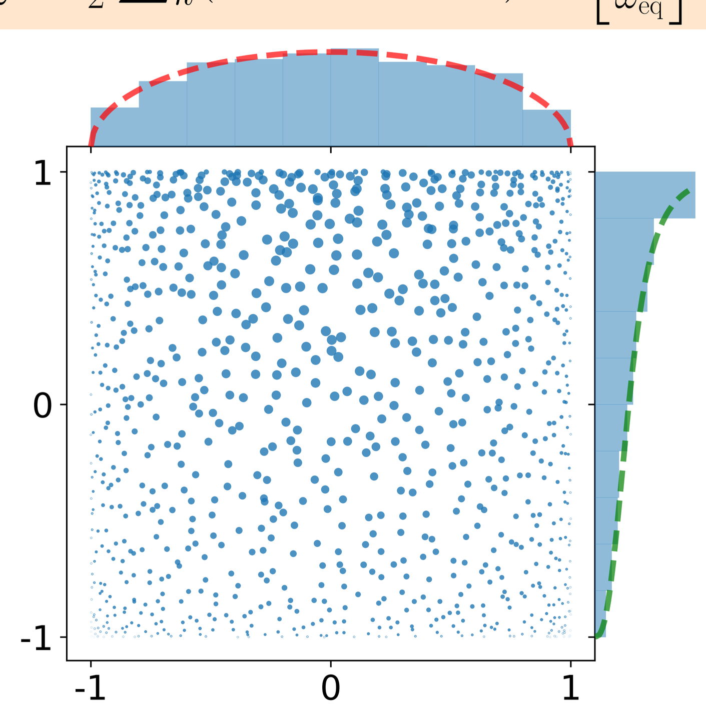
interpretable $\omega_n(\mathbf{x}_n) \equiv$ random Gaussian quadrature

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

Fast Central Limit Theorem

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

with $\mathbf{\Omega}_{f,\omega}^2 \triangleq \frac{1}{2}\sum_k (k_1+\cdots+k_d)\ \mathcal{F}\Big[\frac{f\,\omega}{\omega_{\mathrm{eq}}}\Big](k)^2$



Second estimator (EZ, 1960)

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

Take $\{\mathbf{x}_1,\dots,\mathbf{x}_N\} \sim \mathrm{DPP}\big(\mu, \sum_{k=0}^{N-1}\phi_k(x)\phi_k(y)\big)$

Solve the **linear system**

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

Get **unbiaised estim° of "Fourier coeffs"**

- $\mathbb{E}\big[y_k\big] = \langle f, \phi_{k-1}\rangle = \int_{\mathbb{X}} f(x)\phi_{k-1}(x)\mu(\mathrm{d}x)$

with **interpretable & practical variance**

- $\mathbb{V}\mathrm{ar}\big[y_k\big] = \|f\|^2 - \sum_{\ell=0}^{N-1}\langle f,\phi_\ell\rangle^2 = \mathbf{0}$ if $N \geq M$
- $\mathbb{C}\mathrm{ov}[y_j, y_k] = 0, j \neq k$

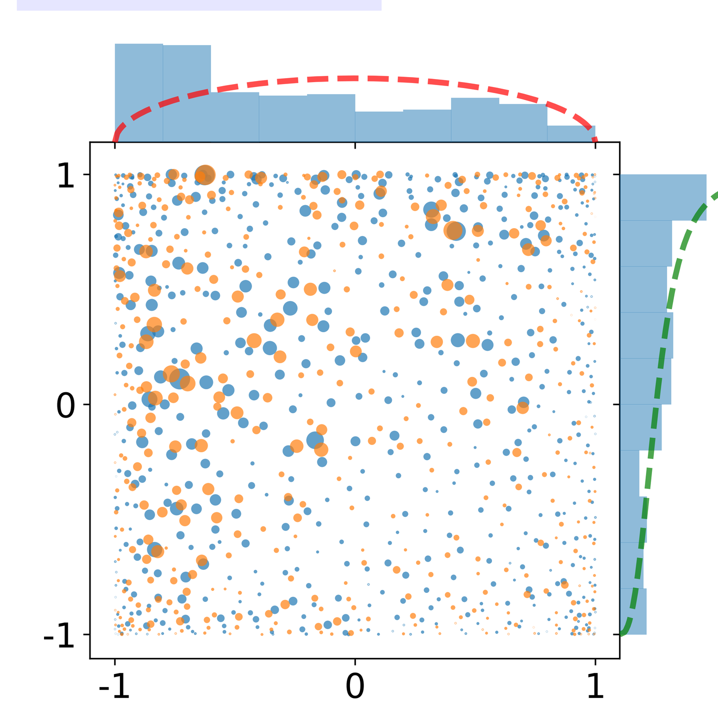
When ϕ_0 is constant, e.g., Jacobi ensemble

$$\widehat{I}_N^{\mathrm{EZ}}(f) \triangleq \frac{y_1}{\phi_0} = \mu(\mathbb{X})^{1/2}\frac{\det \mathbf{\Phi}_{\phi_0\leftarrow f}}{\det \mathbf{\Phi}} \qquad (4)$$

non obvious $\omega_n(\mathbf{x}_1,\dots,\mathbf{x}_n) \lessgtr 0$ $\quad\sum_{n=1}^{N}\omega_n = \mu(\mathbb{X})$

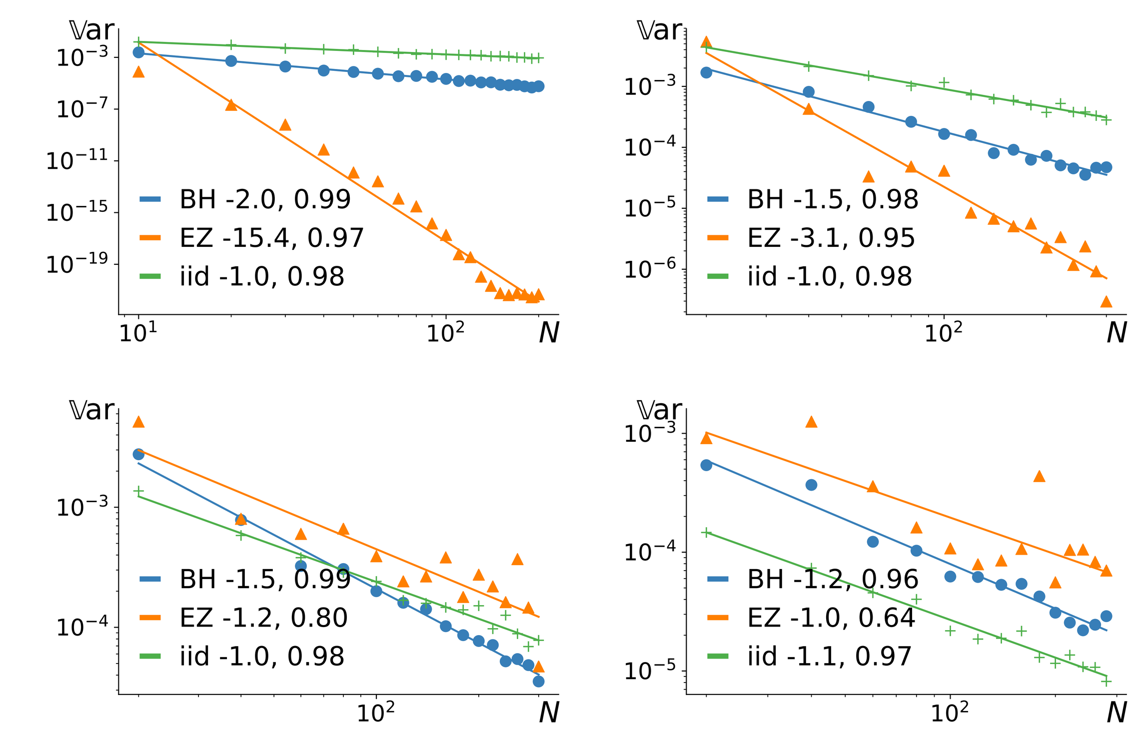
- $\mathbb{E}\big[\widehat{I}_N^{\mathrm{EZ}}\big] = \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$ **unbiased**
- $\mathbb{V}\mathrm{ar}\big[\widehat{I}_N^{\mathrm{EZ}}\big] = \mu(\mathbb{X}) \times \sum_{\ell=N}^{M-1}\langle f,\phi_\ell\rangle^2 = \mathbf{0}$ if $N \geq M$

with the Jacobi ensemble



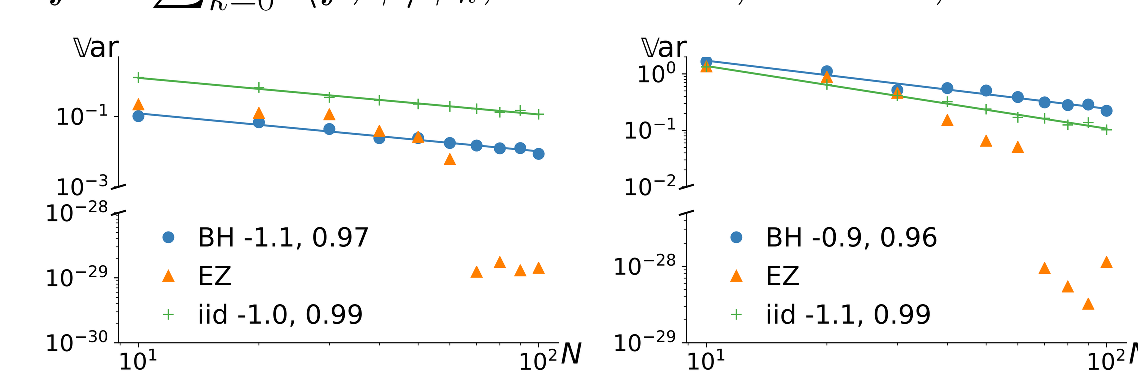
Experiments

$f =$ smooth bump function, $\quad d = 1,2,3,4$



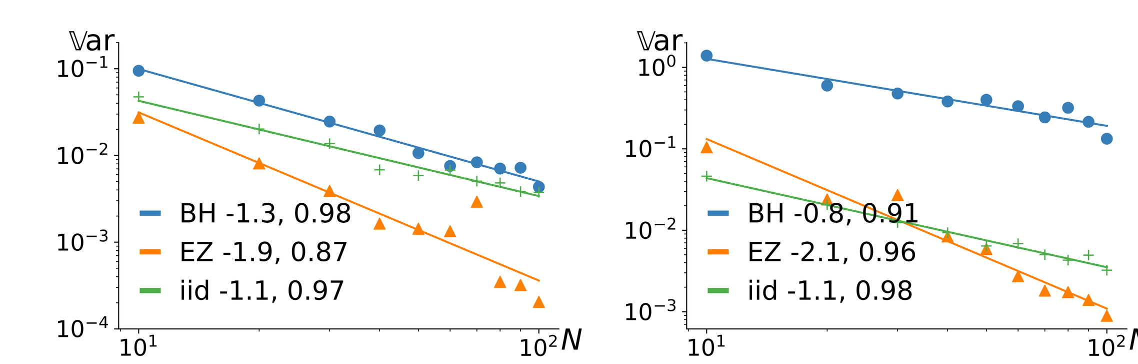
$\mathbb{V}\mathrm{ar}\big[\widehat{I}_N^{\mathrm{BH}}\big] = \mathcal{O}(N^{-(1+1/d)})$ $\ \widehat{I}_N^{\mathrm{EZ}}$: 👍 $d \leq 2$, $d \geq 3$ 👎

$f = \sum_{k=0}^{M-1}\langle f,\phi\rangle\phi_k, \quad M = 70, \quad d = 2,4$



$\mathbb{V}\mathrm{ar}\big[\widehat{I}_N^{\mathrm{EZ}}\big] = \mathbf{0}$ **once** $N \geq M$ **in any dimension**

$f = \sum_{k=0}^{M-1}\frac{1}{k+1}\phi_k, \quad M = N+1, \quad d = 2,4$



$\mathbb{V}\mathrm{ar}\big[\widehat{I}_N^{\mathrm{EZ}}\big] = N^{-2}$ **in any dimension**

Take home message

Follow the titles and the color code ☺

punchlines **contributions**

Jacobi ensemble

Bardenet and Hardy Monte Carlo with DPPs. *Ann. App. Probab., in press*, 2019.

Ermakov and Zolotukhin Polynomial Approximations and the Monte-Carlo Method. *Th. Probab. App. (TVP)*, 1960.

Gautier, Bardenet, and Valko On two ways to use DPPs for Monte Carlo integration. *NeurIPS*, 2019.