On two ways to use determinantal point processes for Monte Carlo integration

Guillaume Gautier[1 2] Rémi Bardenet[1] and Michal Valko[3 2]

Code available in the DPPy Python library

https://github.com/guilgautier/DPPy/tree/master/notebooks

[1]Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, France
[2]INRIA Lille - Nord Europe, France
[3]DeepMind Paris

Outline

1. Definitions

2. Two unbiased DPP-based estimators

3. Experiments

Goal

Numerical integration with Monte Carlo

$$\int_{\mathbb{X}} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \quad \mathbf{x}_n \in \mathbb{X} \subset \mathbb{R}^d$$

Goal

Numerical integration with Monte Carlo

$$\int_{\mathbb{X}} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \quad \mathbf{x}_n \in \mathbb{X} \subset \mathbb{R}^d$$

Crude Monte Carlo

- $\mathbf{x}_n \overset{\text{i.i.d.}}{\sim} \mu/\mu(\mathbb{X})$ and $\mathbf{w}_n = \mu(\mathbb{X})/N$
- Unbiased and Central Limit Theorem (CLT) with $\mathbb{V}\mathrm{ar} = \mathcal{O}(1/N)$

Goal

Numerical integration with Monte Carlo

$$\int_{\mathbb{X}} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \quad \mathbf{x}_n \in \mathbb{X} \subset \mathbb{R}^d$$

Crude Monte Carlo

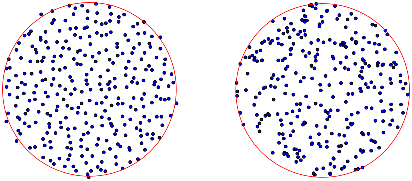
- $\mathbf{x}_n \overset{\text{i.i.d.}}{\sim} \mu/\mu(\mathbb{X})$ and $\mathbf{w}_n = \mu(\mathbb{X})/N$
- Unbiased and Central Limit Theorem (CLT) with $\mathbb{V}\mathrm{ar} = \mathcal{O}(1/N)$

Determinantal Point Processes

- $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}$ and $\mathbf{w}_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$
- Unbiased estimators
  - Bardenet & Hardy (~~2016~~) (2019 Ann. App. Prob. in press) *fast* CLT with $\mathbb{V}\mathrm{ar} = \mathcal{O}(1/N^{1+1/d})$
  - Ermakov & Zolotukhin (**1960**) with striking non asymptotic properties

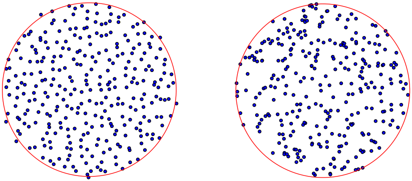
DPPs, heuristically

- **Distribution over configurations of points**



DPPs, heuristically

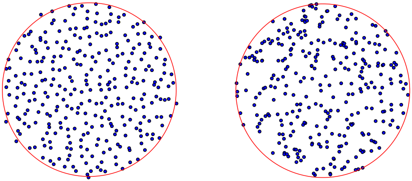
- **Distribution over configurations of points**



- **Defined w.r.t. base measure μ and parametrized by a kernel K**

DPPs, heuristically

- **Distribution over configurations of points**



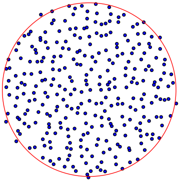
- **Defined w.r.t. base measure μ and parametrized by a kernel K**

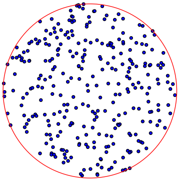
$$``\mathbb{P}\left[\begin{array}{c}\text{there is 1 point in each}\\ B(x_1, \mathrm{d}x_1), \dots, B(x_n, \mathrm{d}x_n)\end{array}\right] = \det(K(x_m, x_p))_{m,p=1}^{n} \mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_n)''$$

DPPs, heuristically

- **Distribution over configurations of points**

DPP (Ginibre)





Uniform (i.i.d.)

- **Defined w.r.t. base measure μ and parametrized by a kernel K**

$$\text{``}\mathbb{P}\left[\begin{array}{c}\text{there is 1 point in each}\\ B(x_1, \mathrm{d}x_1), \ldots, B(x_n, \mathrm{d}x_n)\end{array}\right] = \det(K(x_m, x_p))_{m,p=1}^{n} \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_n)\text{''}$$

- **Negative dependence/repulsion/diversity**

For $n = 2$, "$\ldots = K(x,y)K(y,y) - K(x,y)^2$" $\searrow$ when $K(x,y) \nearrow$
The more similar the points the less likely they co-occur

Projection DPPs

- Orthogonal projection kernel

$$K(x, y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y), \quad \langle \phi_k, \phi_\ell \rangle \triangleq \int_{\mathbb{X}} \phi_k(x)\phi_\ell(x)\mu(\mathrm{d}x) = \delta_{k\ell}$$

Projection DPPs

- Orthogonal projection kernel

$$K(x, y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y), \quad \langle \phi_k, \phi_\ell \rangle \triangleq \int_{\mathbb{X}} \phi_k(x)\phi_\ell(x)\mu(\mathrm{d}x) = \delta_{k\ell}$$

- Building blocks of DPPs (Hough et al., 2006; Kulesza & Taskar, 2012)

Projection DPPs

- Orthogonal projection kernel

$$K(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y), \quad \langle \phi_k, \phi_\ell \rangle \triangleq \int_{\mathbb{X}} \phi_k(x)\phi_\ell(x)\mu(\mathrm{d}x) = \delta_{k\ell}$$

- Building blocks of DPPs (Hough et al., 2006; Kulesza & Taskar, 2012)
- Number of points $=$ rank $K = N$

Projection DPPs

- Orthogonal projection kernel

$$K(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y), \quad \langle \phi_k, \phi_\ell \rangle \triangleq \int_{\mathbb{X}} \phi_k(x)\phi_\ell(x)\mu(\mathrm{d}x) = \delta_{k\ell}$$

- Building blocks of DPPs (Hough et al., 2006; Kulesza & Taskar, 2012)
- Number of points = rank $K = N$
- $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$, when the joint probability distribution of

$$(\mathbf{x}_1, \dots, \mathbf{x}_N) \sim \frac{1}{N!} \det[K(x_p, x_q)]_{p,q=1}^{N} \, \mu(\mathrm{d}x_1) \cdots \mu(\mathrm{d}x_N)$$

Multivariate Jacobi ensemble (Bardenet & Hardy, 2016)

- $\mathbb{X} = [-1, 1]^d$

Multivariate Jacobi ensemble (Bardenet & Hardy, 2016)

- $\mathbb{X} = [-1,1]^d$
- $\mu(\mathrm{d}x) = \prod_{i=1}^d w^i(x^i)\,\mathrm{d}x^i$, with

$$w^i(y) = (1-y)^{a_i}(1+y)^{b_i}$$

Multivariate Jacobi ensemble (Bardenet & Hardy, 2016)

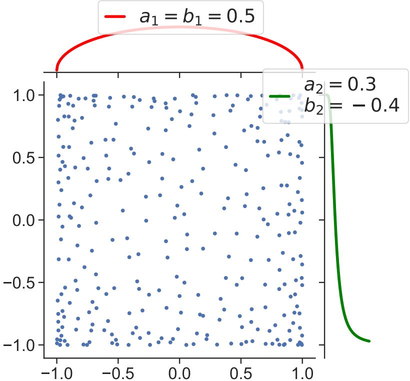
- $\mathbb{X} = [-1,1]^d$
- $\mu(\mathrm{d}x) = \prod_{i=1}^d w^i(x^i)\,\mathrm{d}x^i$, with

$$w^i(y) = (1-y)^{a_i}(1+y)^{b_i}$$

- $\phi_k(x) = \prod_{i=1}^d \phi^i_{k_i}(x^i)$ product of (univariate) orthonormal Jacobi polynomials

Multivariate Jacobi ensemble (Bardenet & Hardy, 2016)

- $\mathbb{X} = [-1, 1]^d$
- $\mu(\mathrm{d}x) = \prod_{i=1}^{d} w^i(x^i)\,\mathrm{d}x^i$, with
$$w^i(y) = (1-y)^{a_i}(1+y)^{b_i}$$
- $\phi_k(x) = \prod_{i=1}^{d} \phi^i_{k_i}(x^i)$ product of (univariate) orthonormal Jacobi polynomials
- A sample $N = 300$



Bardenet & Hardy (BH, 2016) estimator

- "$\mathbb{P}[\exists$ 1 point in $B(x, \mathrm{d}x)] = K(x,x)\mu(\mathrm{d}x)$"

$$\mathbb{E}\left[\sum_{n=1}^{N} g(\mathbf{x}_n)\right] = \int_{\mathbb{X}} g(x)K(x,x)\mu(\mathrm{d}x)$$

Bardenet & Hardy (BH, 2016) estimator

- "$\mathbb{P}[\exists$ 1 point in $B(x, \mathrm{d}x)] = K(x,x)\mu(\mathrm{d}x)$"

$$\mathbb{E}\left[\sum_{n=1}^{N} g(\mathbf{x}_n)\right] = \int_{\mathbb{X}} g(x)K(x,x)\mu(\mathrm{d}x)$$

- $g(x) = f(x)/K(x,x)$

$$\mathbb{E}\left[\sum_{n=1}^{N} \frac{f(\mathbf{x}_n)}{K(\mathbf{x}_n,\mathbf{x}_n)}\right] = \int_{\mathbb{X}} \frac{f(x)}{\cancel{K(x,x)}}\cancel{K(x,x)}\mu(\mathrm{d}x) = \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$$

Bardenet & Hardy (BH, 2016) estimator

- "$\mathbb{P}[\exists$ 1 point in $B(x, \mathrm{d}x)] = K(x,x)\mu(\mathrm{d}x)$"

$$\mathbb{E}\left[\sum_{n=1}^{N} g(\mathbf{x}_n)\right] = \int_{\mathbb{X}} g(x) K(x,x) \mu(\mathrm{d}x)$$

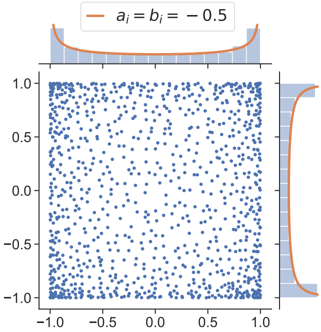
- $g(x) = f(x)/K(x,x)$

$$\mathbb{E}\left[\sum_{n=1}^{N} \frac{f(\mathbf{x}_n)}{K(\mathbf{x}_n, \mathbf{x}_n)}\right] = \int_{\mathbb{X}} \frac{f(x)}{\cancel{K(x,x)}} \cancel{K(x,x)} \mu(\mathrm{d}x) = \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$$

- Natural unbiased estimator of $\int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$

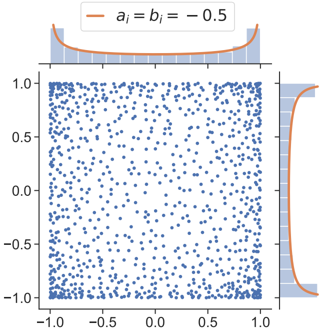
$$\boxed{\widehat{I}_N^{\mathrm{BH}}(f) = \sum_{n=1}^{N} \frac{f(\mathbf{x}_n)}{K(\mathbf{x}_n, \mathbf{x}_n)}}$$

BH estimator and the multivariate Jacobi ensemble

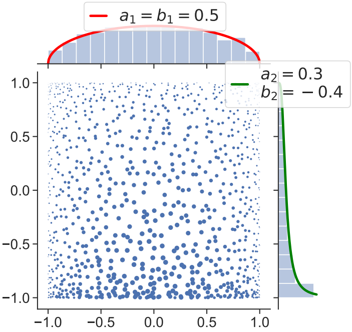


(a) No weight ($N = 1000$)

BH estimator and the multivariate Jacobi ensemble

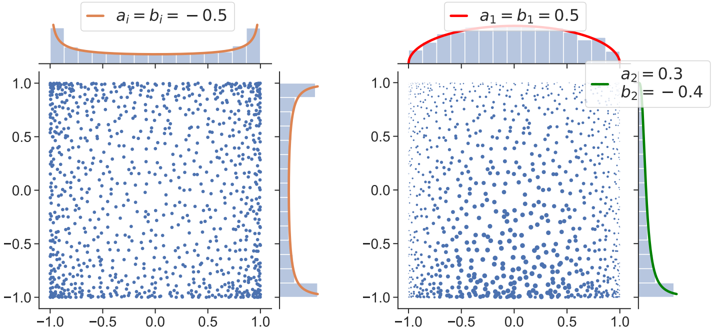


(a) No weight ($N = 1000$)



(b) Weight $1/K(\mathbf{x}_n, \mathbf{x}_n)$ ($N = 1000$)

BH estimator and the multivariate Jacobi ensemble



(a) No weight ($N = 1000$) (b) Weight $1/K(\mathbf{x}_n, \mathbf{x}_n)$ ($N = 1000$)

- Bardenet & Hardy (2016) show fast CLT, for f essentially $\mathcal{C}^1$

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

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

Theorem (Ermakov & Zolotukhin, 1960)

$$f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad M \in \mathbb{N} \cup \{\infty\}$$

Theorem (Ermakov & Zolotukhin, 1960)

$$f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad M \in \mathbb{N} \cup \{\infty\}$$

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

Theorem (Ermakov & Zolotukhin, 1960)

$$f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad M \in \mathbb{N} \cup \{\infty\}$$

1. Sample $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$ with $K(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$
2. Random linear system

$$\begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{N-1}(\mathbf{x}_1) \\ \vdots & & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{N-1}(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}$$

Theorem (Ermakov & Zolotukhin, 1960)

$$f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad M \in \mathbb{N} \cup \{\infty\}$$

1. Sample $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$ with $K(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$
2. Solve

$$\begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{N-1}(\mathbf{x}_1) \\ \vdots & & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{N-1}(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}$$

- $\mathbb{E}[y_k] = \langle f, \phi_k \rangle = \int f(x)\phi_k(x)\mu(\mathrm{d}x)$

Theorem (Ermakov & Zolotukhin, 1960)

$$f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad M \in \mathbb{N} \cup \{\infty\}$$

1. Sample $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$ with $K(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$
2. Solve

$$\begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{N-1}(\mathbf{x}_1) \\ \vdots & & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{N-1}(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}$$

- $\mathbb{E}[y_k] = \langle f, \phi_k \rangle = \int f(x)\phi_k(x)\mu(\mathrm{d}x)$
- $\mathbb{V}\mathrm{ar}[y_k] = \|f\|^2 - \sum_{\ell=0}^{N-1} \langle f, \phi_\ell \rangle^2 = \sum_{\ell=N}^{M-1} \langle f, \phi_\ell \rangle^2$

Theorem (Ermakov & Zolotukhin, 1960)

$$f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad M \in \mathbb{N} \cup \{\infty\}$$

1. Sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$ with $K(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$
2. Solve

$$\begin{pmatrix} \phi_0(\mathbf{x}_1) & \ldots & \phi_{N-1}(\mathbf{x}_1) \\ \vdots & & \vdots \\ \phi_0(\mathbf{x}_N) & \ldots & \phi_{N-1}(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}$$

- $\mathbb{E}[y_k] = \langle f, \phi_k \rangle = \int f(x)\phi_k(x)\mu(\mathrm{d}x)$
- $\mathbb{V}\mathrm{ar}[y_k] = \|f\|^2 - \sum_{\ell=0}^{N-1} \langle f, \phi_\ell \rangle^2 = \sum_{\ell=N}^{M-1} \langle f, \phi_\ell \rangle^2 = 0$ if $M \leq N$

Theorem (Ermakov & Zolotukhin, 1960)

$$f = \sum_{\ell=0}^{M-1} \langle f, \phi_\ell \rangle \phi_\ell, \quad M \in \mathbb{N} \cup \{\infty\}$$

1. Sample $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \operatorname{DPP}(\mu, K)$ with $K(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$
2. Solve

$$\begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{N-1}(\mathbf{x}_1) \\ \vdots & & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{N-1}(\mathbf{x}_N) \end{pmatrix} \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix} = \begin{pmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_N) \end{pmatrix}$$

- $\mathbb{E}[y_k] = \langle f, \phi_k \rangle = \int f(x)\phi_k(x)\mu(\mathrm{d}x)$
- $\mathbb{V}\mathrm{ar}[y_k] = \|f\|^2 - \sum_{\ell=0}^{N-1} \langle f, \phi_\ell \rangle^2 = \sum_{\ell=N}^{M-1} \langle f, \phi_\ell \rangle^2 = 0 \quad \text{if } M \leq N$
- $\approx$ "fitted" control functional (Oates et al., 2017)

Ermakov & Zolotukhin (EZ, 1960) estimator

- If ϕ_0 is constant, e.g., the multivariate Jacobi ensemble,

$$\mathbb{E}[y_0] = \phi_0 \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$$

Ermakov & Zolotukhin (EZ, 1960) estimator

- If ϕ_0 is constant, e.g., the multivariate Jacobi ensemble,

$$\mathbb{E}[y_0] = \phi_0 \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$$

- A less obvious (solve linear system) unbiased estimator of $\int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$

$$\boxed{\widehat{I}_N^{\mathrm{EZ}}(f) = \frac{y_0}{\phi_0}} = \sum_{n=1}^{N} \mathbf{w}_n(\mathbf{x}_1, \dots, \mathbf{x}_N) f(\mathbf{x}_n)$$

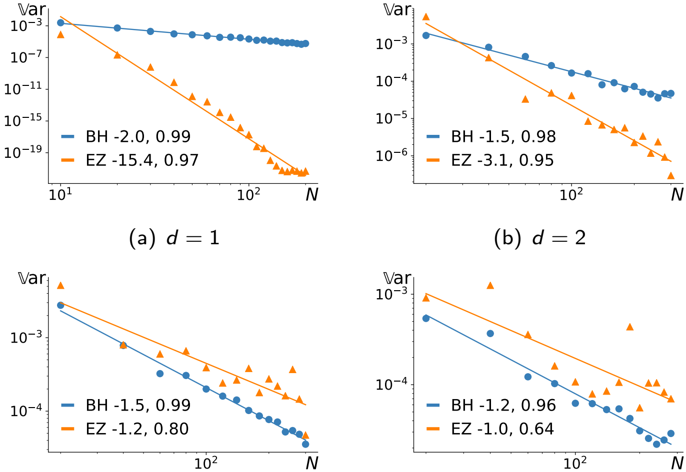
$$f(x) = \prod_{i=1}^{d} \exp\left(-\frac{1}{1-\varepsilon-(x^i)^2}\right)\mathbb{1}_{[-1+\varepsilon,1-\varepsilon]}(x^i)$$

$$f(x) = \prod_{i=1}^{d} \exp\left(-\frac{1}{1-\varepsilon-(x^i)^2}\right)\mathbb{1}_{[-1+\varepsilon,1-\varepsilon]}(x^i)$$

$$\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{BH}}(f)\right] = \mathcal{O}(N^{-\mathbf{(1+1/d)}})$$

$$f(x) = \prod_{i=1}^{d} \exp\left(-\frac{1}{1-\varepsilon-(x^i)^2}\right) \mathbb{1}_{[-1+\varepsilon,1-\varepsilon]}(x^i)$$

$$\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{BH}}(f)\right] = \mathcal{O}(N^{-(\mathbf{1}+\mathbf{1}/\mathbf{d})})$$



(a) $d = 1$ (b) $d = 2$

(c) $d = 3$ (d) $d = 4$

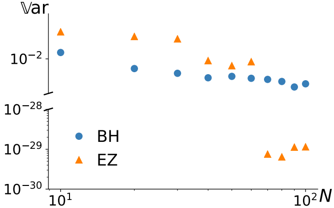
$f(x) = \sum_{\mathfrak{d}(k)=0}^{M-1} \frac{1}{\mathfrak{d}(k)+1} \phi_k(x), \qquad M = 70$

$$f(x) = \sum_{\mathfrak{b}(k)=0}^{M-1} \frac{1}{\mathfrak{b}(k)+1} \phi_k(x), \qquad M = 70$$

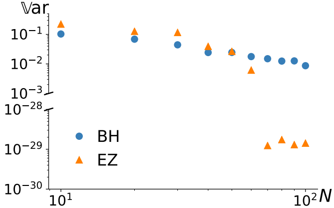
$\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{EZ}}(f)\right] = 0$ once $M \leq N$

$$f(x) = \sum_{\mathfrak{b}(k)=0}^{M-1} \frac{1}{\mathfrak{b}(k)+1} \phi_k(x), \qquad M = 70$$

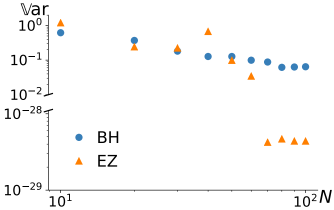
$\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{EZ}}(f)\right] = 0$ once $M \leq N$



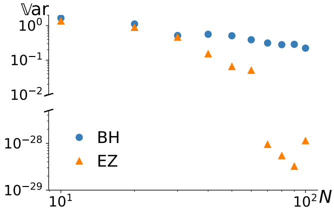
(a) $d = 1$



(b) $d = 2$



(c) $d = 3$



(d) $d = 4$

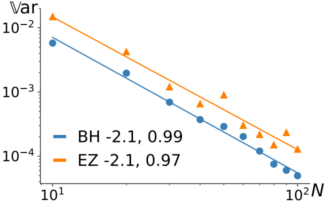
$$f(x) = \sum_{\mathfrak{b}(k)=0}^{M-1} \frac{1}{\mathfrak{b}(k)+1} \phi_k(x), \quad M = N + 1 \implies M \geq N$$

$$f(x) = \sum_{\mathfrak{b}(k)=0}^{M-1} \frac{1}{\mathfrak{b}(k)+1} \phi_k(x), \quad M = N + 1 \implies M \geq N$$

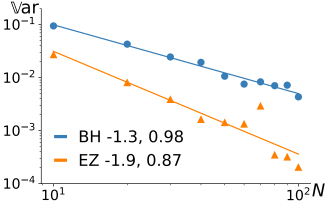
$$\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{EZ}}(f)\right] = \mathcal{O}(N^{-2})$$

$$f(x) = \sum_{\mathfrak{b}(k)=0}^{M-1} \frac{1}{\mathfrak{b}(k)+1} \phi_k(x), \quad M = N + 1 \implies M \geq N$$

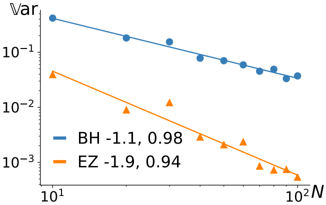
$$\mathbb{V}\mathrm{ar}\left[\widehat{I}_N^{\mathrm{EZ}}(f)\right] = \mathcal{O}(N^{-\mathbf{2}})$$



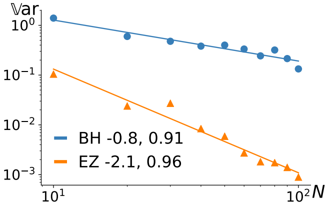
(a) $d = 1$



(b) $d = 2$

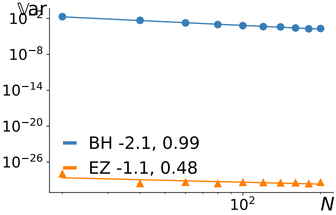


(c) $d = 3$

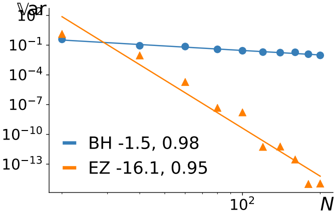


(d) $d = 4$

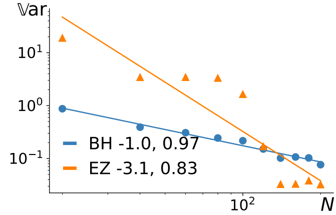
$f(x) = \prod_{i=1}^{d} \cos(\pi x^i)$



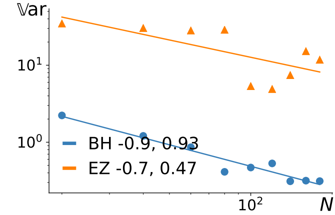
(a) $d = 1$



(b) $d = 2$



(c) $d = 3$



(d) $d = 4$

Monte Carlo with projection DPPs

$$\int_{[-1,1]^d} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \qquad \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$$

Monte Carlo with projection DPPs

$$\int_{[-1,1]^d} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \qquad \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$$

Sampling the multivariate Jacobi ensemble

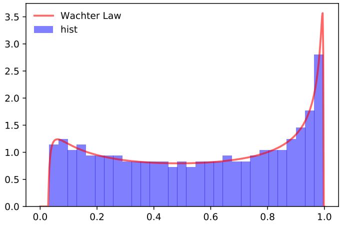
- Remodeled the implementation of the sampler of Bardenet & Hardy (2016)
- Code available in DPPy

Monte Carlo with projection DPPs

$$\int_{[-1,1]^d} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \qquad \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$$

Sampling the multivariate Jacobi ensemble

- Remodeled the implementation of the sampler of Bardenet & Hardy (2016)
- Code available in DPPy
- $d = 1$, eigenvalues of tridiagonal matrix (Killip & Nenciu, 2004), $\mathcal{O}(N^2)$



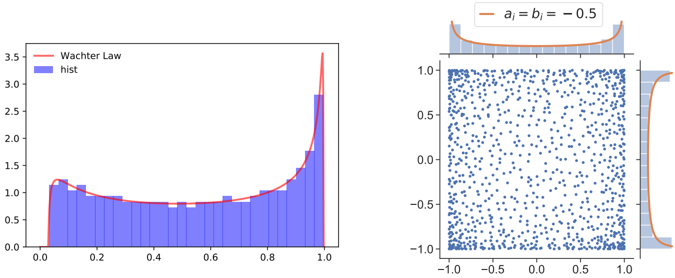
(a) $d = 1$

Monte Carlo with projection DPPs

$$\int_{[-1,1]^d} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \qquad \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$$

Sampling the multivariate Jacobi ensemble

- Remodeled the implementation of the sampler of Bardenet & Hardy (2016)
- Code available in DPPy
- $d = 1$, eigenvalues of tridiagonal matrix (Killip & Nenciu, 2004), $\mathcal{O}(N^2)$
- $d \geq 2$, chain rule (HKPV, 2006) $\mathcal{O}(\mathrm{poly}(N))$ + rejections (CGW, 1994)



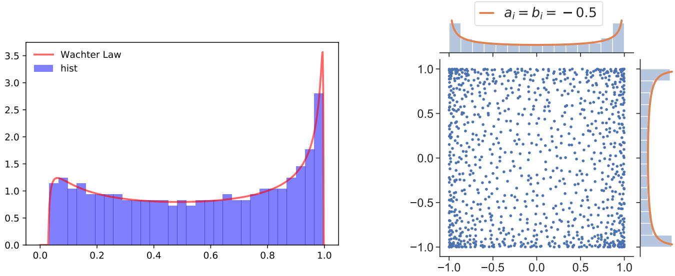
(a) $d = 1$ (b) $d \geq 2$

Monte Carlo with projection DPPs

$$\int_{[-1,1]^d} f(x)\mu(\mathrm{d}x) \approx \sum_{n=1}^{N} \mathbf{w}_n f(\mathbf{x}_n), \qquad \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \sim \mathrm{DPP}(\mu, K)$$

Sampling the multivariate Jacobi ensemble

- Remodeled the implementation of the sampler of Bardenet & Hardy (2016)
- Code available in DPPy
- $d = 1$, eigenvalues of tridiagonal matrix (Killip & Nenciu, 2004), $\mathcal{O}(N^2)$
- $d \geq 2$, chain rule (HKPV, 2006) $\mathcal{O}(\mathrm{poly}(N))$ + rejections (CGW, 1994)
- Sampling more general continuous DPPs, e.g., wavelets?



(a) $d = 1$ (b) $d \geq 2$

Monte Carlo with projection DPPs

Bardenet & Hardy (~~2016~~) (2019 Ann. App. Prob. in press)

- $\mathbf{w}_n = 1/K(\mathbf{x}_n, \mathbf{x}_n) \equiv$ random Gaussian quadrature
- Test $\mathbb{V}\mathrm{ar} = \mathcal{O}(N^{-(1+1/d)})$ in unexplored regimes

Monte Carlo with projection DPPs

Bardenet & Hardy (~~2016~~) (2019 Ann. App. Prob. in press)

- $\mathbf{w}_n = 1/K(\mathbf{x}_n, \mathbf{x}_n) \equiv$ random Gaussian quadrature
- Test $\mathbb{V}\text{ar} = \mathcal{O}(N^{-(1+1/d)})$ in unexplored regimes

Ermakov & Zolotukhin, 1960 $\ll$ 1975, Macchi formalized DPPs

- First connexion to projection DPPs
- Short proof, using modern arguments
- Linear system
- Potential of $\mathbb{V}\text{ar} = \|f\|^2 - \sum_{k=0}^{N-1} \langle f, \phi_k \rangle^2 = 0$ if $f \in \operatorname{span}\{\phi_0, \dots, \phi_{N-1}\}$

Monte Carlo with projection DPPs

Bardenet & Hardy (~~2016~~) (2019 Ann. App. Prob. in press)

- $\mathbf{w}_n = 1/K(\mathbf{x}_n, \mathbf{x}_n) \;\equiv\;$ random Gaussian quadrature
- Test $\mathbb{V}\mathrm{ar} = \mathcal{O}(N^{-(1+1/d)})$ in unexplored regimes

Ermakov & Zolotukhin, 1960 $\ll$ 1975, Macchi formalized DPPs

- First connexion to projection DPPs
- Short proof, using modern arguments
- Linear system
- Potential of $\mathbb{V}\mathrm{ar} = \|f\|^2 - \sum_{k=0}^{N-1} \langle f, \phi_k \rangle^2 = 0$ if $f \in \mathrm{span}\{\phi_0, \dots, \phi_{N-1}\}$

The smoothness of the representation of f on $\phi_0, \phi_1, \dots$
must drive the choice of $K = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$

Monte Carlo with projection DPPs

Bardenet & Hardy (~~2016~~) (2019 Ann. App. Prob. in press)

- $\mathbf{w}_n = 1/K(\mathbf{x}_n, \mathbf{x}_n) \equiv$ random Gaussian quadrature
- Test $\mathbb{V}\text{ar} = \mathcal{O}(N^{-(1+1/d)})$ in unexplored regimes

Ermakov & Zolotukhin, 1960 $\ll$ 1975, Macchi formalized DPPs

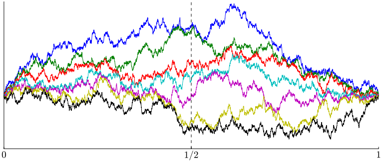
- First connexion to projection DPPs
- Short proof, using modern arguments
- Linear system stability?
- Potential of $\mathbb{V}\text{ar} = \|f\|^2 - \sum_{k=0}^{N-1} \langle f, \phi_k \rangle^2 = 0$ if $f \in \operatorname{span}\{\phi_0, \dots, \phi_{N-1}\}$
- Prove a CLT?

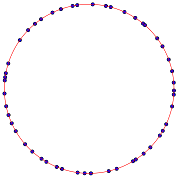
The smoothness of the representation of f on $\phi_0, \phi_1, \dots$
must drive the choice of $K = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y)$

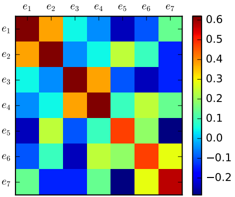
Thanks!

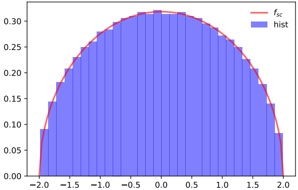
You're welcome to contribute to DPPy ☺

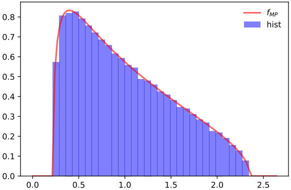
https://github.com/guilgautier/DPPy











References [1]

Bardenet, R. and Hardy, A. Monte Carlo with Determinantal Point Processes. *ArXiv e-prints*, 2016. arXiv:1605.00361.

Chow, Y., Gatteschi, L., and Wong, R. A Bernstein-type inequality for the Jacobi polynomial. *Proceedings of the American Mathematical Society*, 121(3):703–703, 1994.

Ermakov, S. M. and Zolotukhin, V. G. Polynomial Approximations and the Monte-Carlo Method. *Theory of Probability & Its Applications*, 5(4):428–431, jan 1960.

Hough, J. B., Krishnapur, M., Peres, Y., and Virág, B. Determinantal Processes and Independence. In *Probability Surveys*, volume 3, pp. 206–229. The Institute of Mathematical Statistics and the Bernoulli Society, 2006. arXiv:math/0503110.

Johansson, K. Random matrices and determinantal processes. *Les Houches Summer School Proceedings*, 83(C):1–56, 2006.

Killip, R. and Nenciu, I. Matrix models for circular ensembles. *International Mathematics Research Notices*, 2004(50):2665, 2004. arXiv:math/0410034.

Kulesza, A. and Taskar, B. Determinantal Point Processes for Machine Learning. *Foundations and Trends in Machine Learning*, 5(2-3):123–286, 2012. arXiv:1207.6083.

Macchi, O. The coincidence approach to stochastic point processes. *Advances in Applied Probability*, 7(01):83–122, mar 1975.

Oates, C. J., Girolami, M., and Chopin, N. Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):695–718, jun 2017. arXiv:1410.2392.

Setup

- Distribution over configurations of points of $\mathbb{X}$
  - $\mathbb{X}$ compact $\subset \mathbb{R}^d$
- Defined w.r.t. reference measure μ on $(\mathbb{X}, \mathcal{B}(X))$
  - $\mu(\mathbb{X}) < \infty$
  - $\mu(\mathrm{d}x) = w(x)\,\mathrm{d}x$
- Parametrized by a kernel $K : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$
  - Continuous
  - Positive semi-definite:

$$K(y, x) = K(x, y) \quad \text{and} \quad \det(K(x_p, x_q))_{p,q=1}^n \geq 0, \ \forall n \in \mathbb{N}^*$$

Mercer theorem applies

$$K(x, y) = \sum_{k=0}^{\infty} \lambda_k \phi_k(x)\phi_k(y), \quad \langle \phi_k, \phi_\ell \rangle \triangleq \int_{\mathbb{X}} \phi_k(x)\phi_\ell(x)\ \mu(\mathrm{d}x) = \delta_{k\ell}$$

Existence guaranteed when $0 \leq \lambda_k \leq 1$

Linear statistics of projection DPPs

$$\sum_{n=1}^{N} g(x_n), \quad \text{where} \quad \{x_1, \dots, x_N\} \sim \mathrm{DPP}(\mu, K)$$

1. $g(x) = \mathbb{1}_C(x) \implies \#|\{x_1, \dots, x_N\} \cap C|$

- Expectation

$$\mathbb{E}\left[\sum_{n=1}^{N} g(x_n)\right] = \int_{\mathbb{X}} g(x) K(x,x) \mu(\mathrm{d}x)$$

  1. $g(x) = \dfrac{f(x)}{K(x,x)} \implies$ unbiased estimate of $\int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$!

- Variance

$$\mathbb{V}\mathrm{ar}\left[\sum_{n=1}^{N} g(x_n)\right] = \frac{1}{2} \iint (g(x) - g(y))^2 K(x,y)^2 \mu(\mathrm{d}x)\mu(\mathrm{d}x)$$

  Reflects a notion of smoothness of g w.r.t. K

  1. g L-Lipschitz $\implies \mathbb{V}\mathrm{ar} \leq L^2 N$

Ermakov & Zolotukhin (1960) estimator

For constant ϕ_0, e.g., multivariate Jacobi ensemble,

$$\mathbb{E}[y_0] = \phi_0 \int_{\mathbb{X}} f(x)\mu(\mathrm{d}x)$$

A direct application of EZ theorem yields

$$\widehat{I}_N^{\mathrm{EZ}}(f) \triangleq \frac{y_0}{\phi_0} = \sqrt{\mu([-1,1]^d)} \; \frac{\det \boldsymbol{\Phi}_{\phi_0,f}(\mathbf{x}_{1:N})}{\det \boldsymbol{\Phi}(\mathbf{x}_{1:N})}$$

as an unbiased estimator of $\int f(x)\mu(\mathrm{d}x)$

Using $\|\phi_0\| = 1$ and Cramer's rule

$$\boldsymbol{\Phi}_{\phi_0,f} = \begin{pmatrix} f(x_1) & \dots & \psi_{N-1}(x_1) \\ \vdots & & \vdots \\ f(x_N) & \dots & \psi_{N-1}(x_N) \end{pmatrix} \quad \boldsymbol{\Phi} = \begin{pmatrix} \phi_0(x_1) & \dots & \phi_{N-1}(x_1) \\ \vdots & & \vdots \\ \phi_0(x_N) & \dots & \phi_{N-1}(x_N) \end{pmatrix}$$

Sketch proof of Ermakov & Zolotukhin (1960)

Cramer's rule

$$y_0 = \frac{\det \mathbf{\Phi}_{\phi_0,f}(\mathbf{x}_{1:N})}{\det \mathbf{\Phi}(\mathbf{x}_{1:N})}$$

Joint density

$$(\mathbf{x}_1, \dots, \mathbf{x}_N) \sim \frac{1}{N!}(\det \mathbf{\Phi}(\mathbf{x}_{1:N}))^2 \mu^{\otimes N}(x) \tag{1}$$

Moments

$$\mathbb{E}[y_0] \overset{(1)}{=} \frac{1}{N!} \int \det \mathbf{\Phi}_{\phi_0,f}(x_{1:N}) \det \mathbf{\Phi}(x_{1:N})\, \mu^{\otimes N}(\mathrm{d}x)$$

$$\overset{CB}{=} \det \begin{pmatrix} \langle f, \phi_0 \rangle & (\langle f, \phi_\ell \rangle)_{\ell=1}^{N-1} \\ 0_{N-1,1} & I_{N-1} \end{pmatrix} = \langle f, \phi_0 \rangle \tag{2}$$

$$\mathbb{E}\left[y_0^2\right] \overset{(1)}{=} \frac{1}{N!} \int \det \mathbf{\Phi}_{\phi_0,f}(x_{1:N}) \det \mathbf{\Phi}_{\phi_0,f}(x_{1:N})\, \mu^{\otimes N}(\mathrm{d}x)$$

$$\overset{CB}{=} \det \begin{pmatrix} \|f\|^2 & (\langle f, \phi_\ell \rangle)_{\ell=1}^{N-1} \\ (\langle f, \phi_k \rangle)_{k=1}^{N-1} & I_{N-1} \end{pmatrix}$$

$$= \|f\|^2 - \sum_{k=1}^{N-1} \langle f, \phi_k \rangle^2 \tag{3}$$

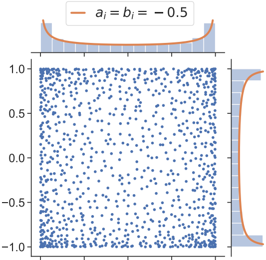
CB, see Johansson (2006, Proposition 2.10)

Sampling the multivariate Jacobi ensemble

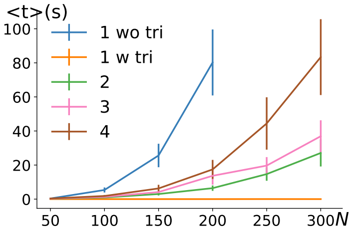
- $d = 1$, for $a_i, b_i > -1$
  - compute eigenvalues of a random tridiagonal matrix (Killip & Nenciu, 2004)
  - $\mathcal{O}(N^2)$
- $d \geq 2$, for $|a_i|, |b_i| \leq \frac{1}{2}$
  - Chain rule of Hough et al. (2006)

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

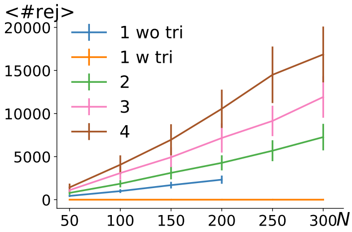
  - Proposal $\omega_{\text{eq}}(x) \, \mathrm{d}x$ (arcsine) + rejection bound (Chow et al., 1994)
  - $\mathcal{O}(\text{poly}(N))$ + rejections



Timings



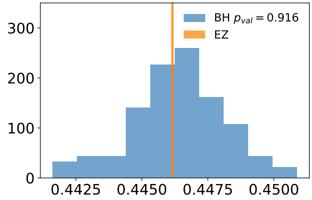
(b) ⟨time⟩ to get one sample



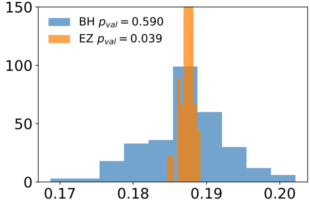
(c) ⟨#rejections⟩

Figure 1: The colors and numbers correspond to the dimension. $a_i, b_i = -1/2$. For $d = 1$, the tridiagonal model (tri) of Killip & Nenciu offers tremendous savings, without it is cheaper to get a sample in larger dimension. The number of rejections grows as $N2^d$.

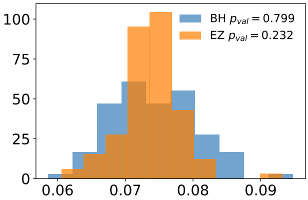
Central Limit Theorem ? KS test $N = 300$



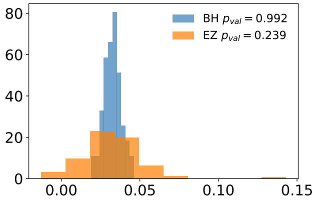
(a) $d = 1$



(b) $d = 2$



(c) $d = 3$



(d) $d = 4$