

Abstract

Compare two DPP-based estimators of

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

- 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 original result
 - new short and simple proof

Efficient sampler for a specific DPP

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

Setup: projection DPPs

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

▪ Repulsive random set of points

- joint distribution of $(\mathbf{x}_1, \dots, \mathbf{x}_N)$

$$\frac{1}{N!} \det \left(K(x_n, x_p) \right)_{n,p=1}^N \mu^{\otimes N}(dx) \quad (1)$$

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

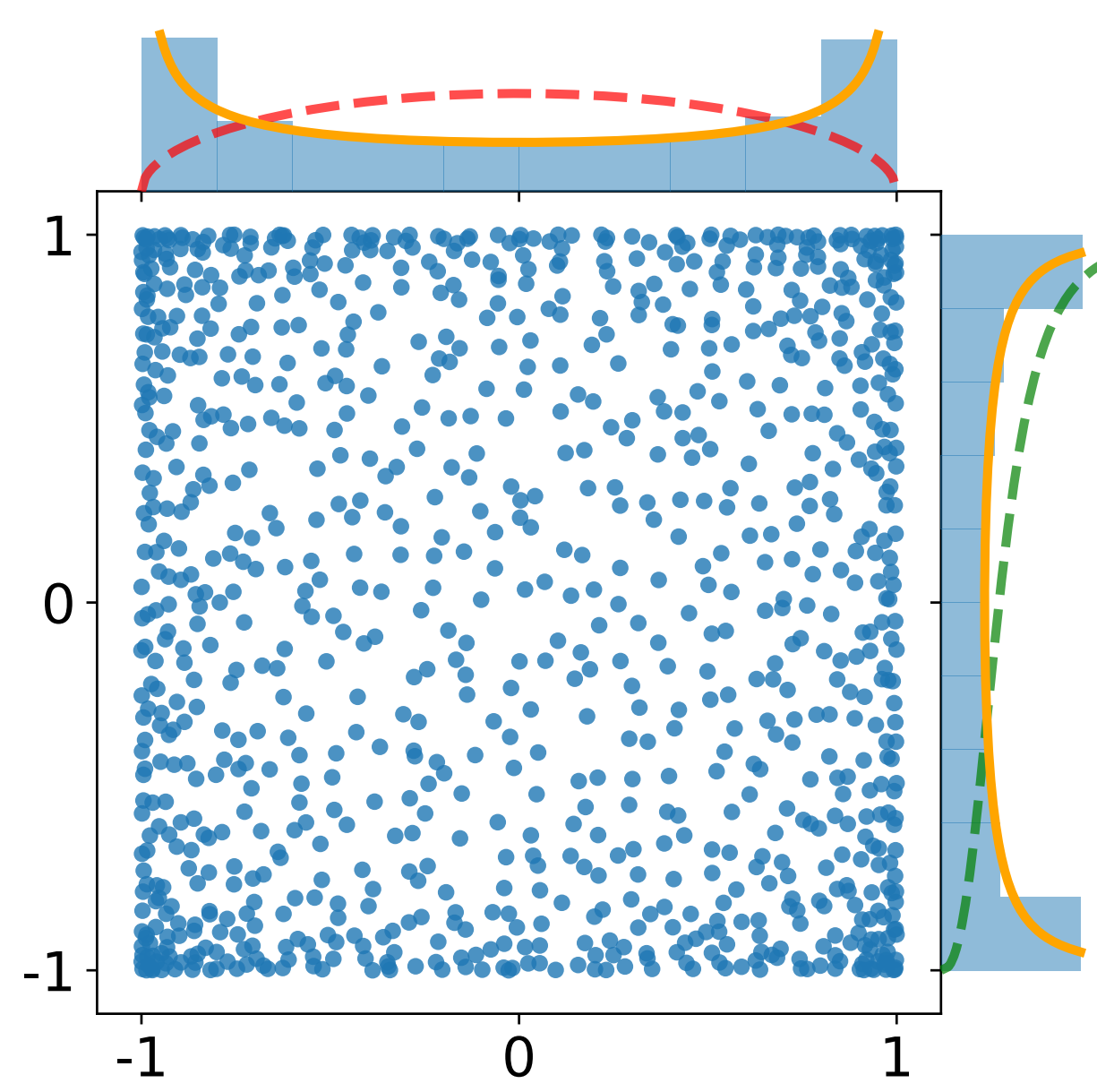
$$K(x, y) \triangleq \sum_{k=0}^{N-1} \phi_k(x) \phi_k(y),$$

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

- reference measure μ on \mathbb{X}

The multivariate Jacobi ensemble

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



Sampling projection DPPs

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

where $K(x, y) = \Phi(x)^T \Phi(y)$, 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

$$(1) = \frac{\|\Phi(x_1)\|^2}{N} \omega(x_1) dx_1 \prod_{n=2}^N \frac{\text{distance}^2(\Phi(x_n), \text{span}\{\Phi(x_p)\}_{p=1}^{n-1})}{N - (n-1)} \omega(x_n) dx_n \quad (2)$$

Sampling the Jacobi ensemble

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

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

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

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

First estimator (BH, 2019)

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

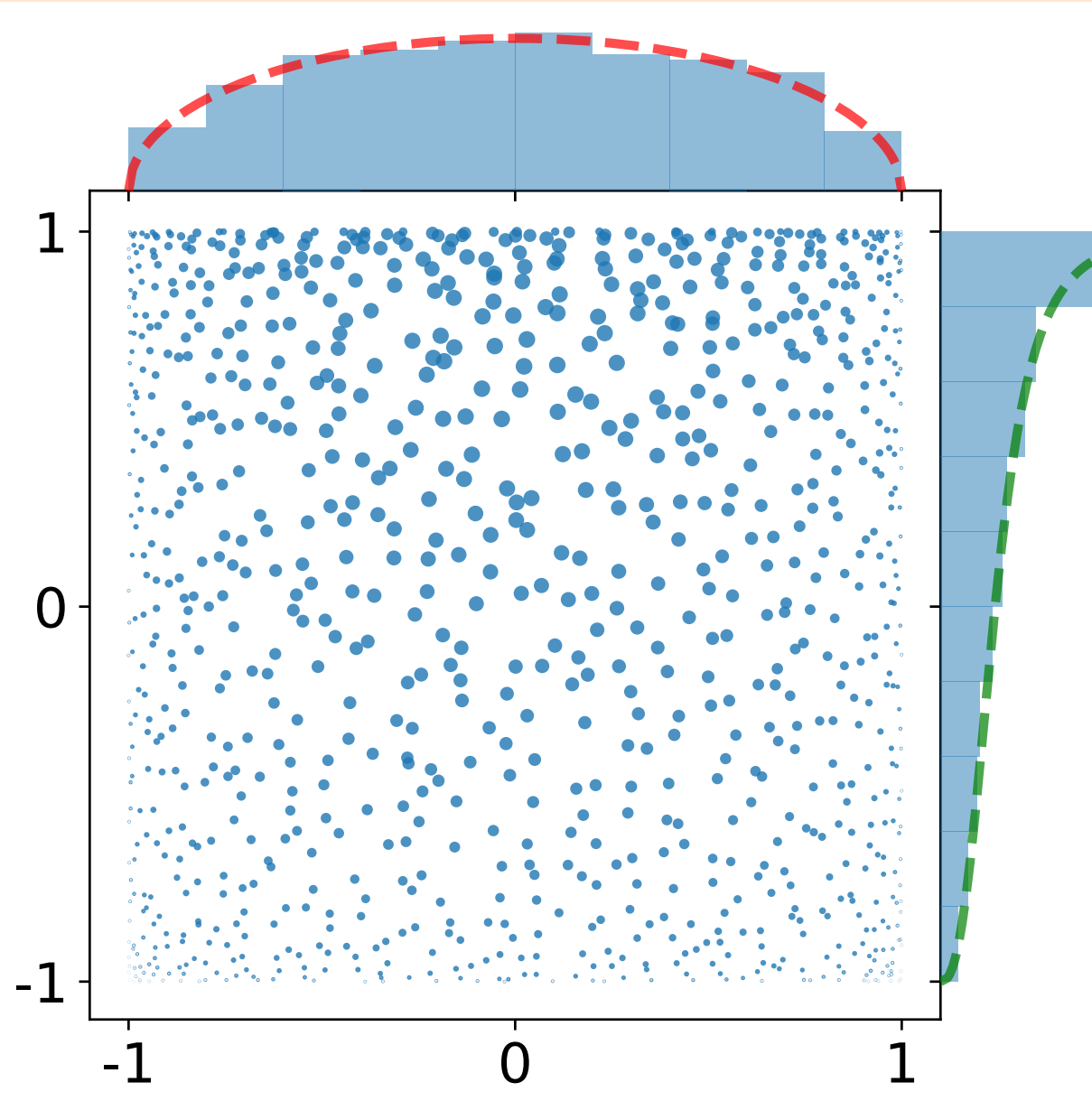
interpretable $\omega_n(\mathbf{x}_n) \equiv$ random Gaussian quadrature

- $\mathbb{E}[\hat{I}_N^{\text{BH}}] = \int_{\mathbb{X}} f(x) \mu(dx)$ **unbiased**
- $\text{Var}[\hat{I}_N^{\text{BH}}] = \frac{1}{2} \int_{\mathbb{X}^2} \left(\frac{f(x)}{K(x, x)} - \frac{f(y)}{K(y, y)} \right)^2 K(x, y)^2 \mu(dx) \mu(dy)$

Fast Central Limit Theorem

$$\sqrt{N^{1+1/d}} \left(\hat{I}_N^{\text{BH}} - \int_{[-1,1]^d} f(x) \omega(x) dx \right) \xrightarrow[N \rightarrow \infty]{\text{law}} \mathcal{N}(0, \Omega_{f, \omega}^2)$$

with $\Omega_{f, \omega}^2 \triangleq \frac{1}{2} \sum_k (k_1 + \dots + k_d) \mathcal{F} \left[\frac{f \omega}{\omega_{\text{eq}}} \right] (k)^2$



Second estimator (EZ, 1960)

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

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

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 **unbiased estim^o of "Fourier coeffs"**

- $\mathbb{E}[y_k] = \langle f, \phi_{k-1} \rangle = \int_{\mathbb{X}} f(x) \phi_{k-1}(x) \mu(dx)$
- with **interpretable & practical variance**
- $\text{Var}[y_k] = \|f\|^2 - \sum_{\ell=0}^{N-1} \langle f, \phi_\ell \rangle^2 = 0$ if $N \geq M$
- $\text{Cov}[y_j, y_k] = 0, j \neq k$

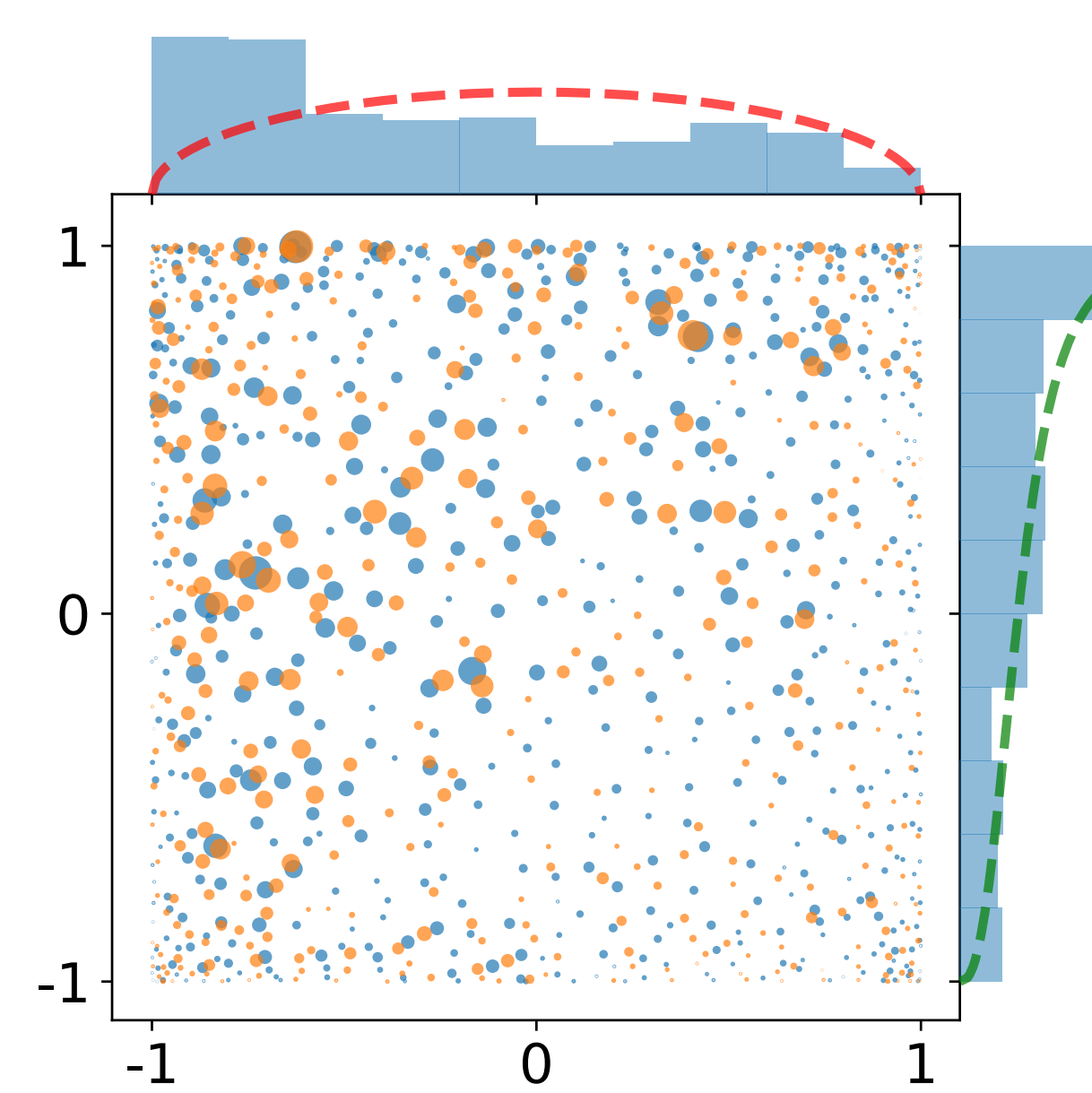
When ϕ_0 is constant, e.g., Jacobi ensemble

$$\hat{I}_N^{\text{EZ}}(f) \triangleq \frac{y_1}{\phi_0} = \mu(\mathbb{X})^{1/2} \frac{\det \Phi_{\phi_0 \leftarrow f}}{\det \Phi} \quad (4)$$

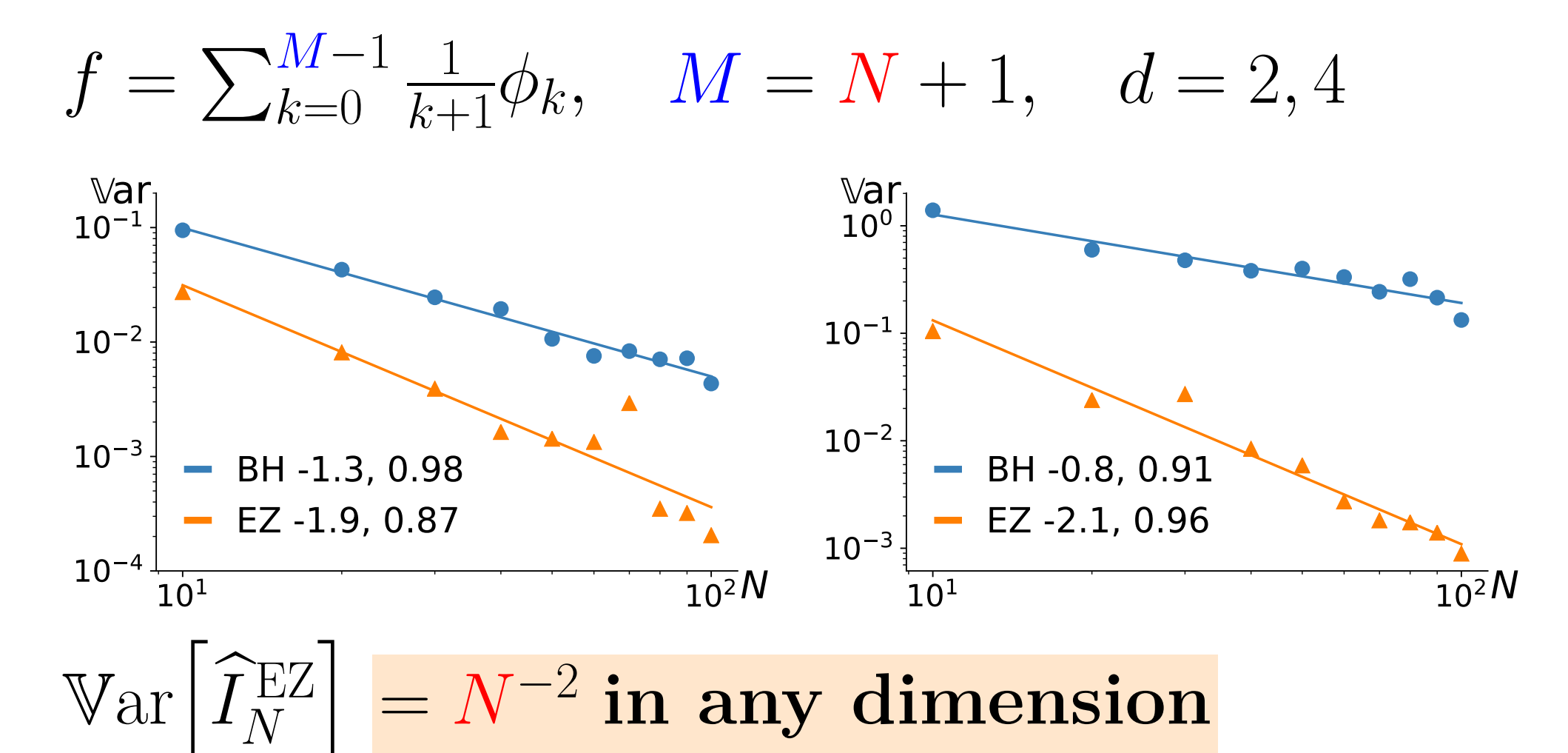
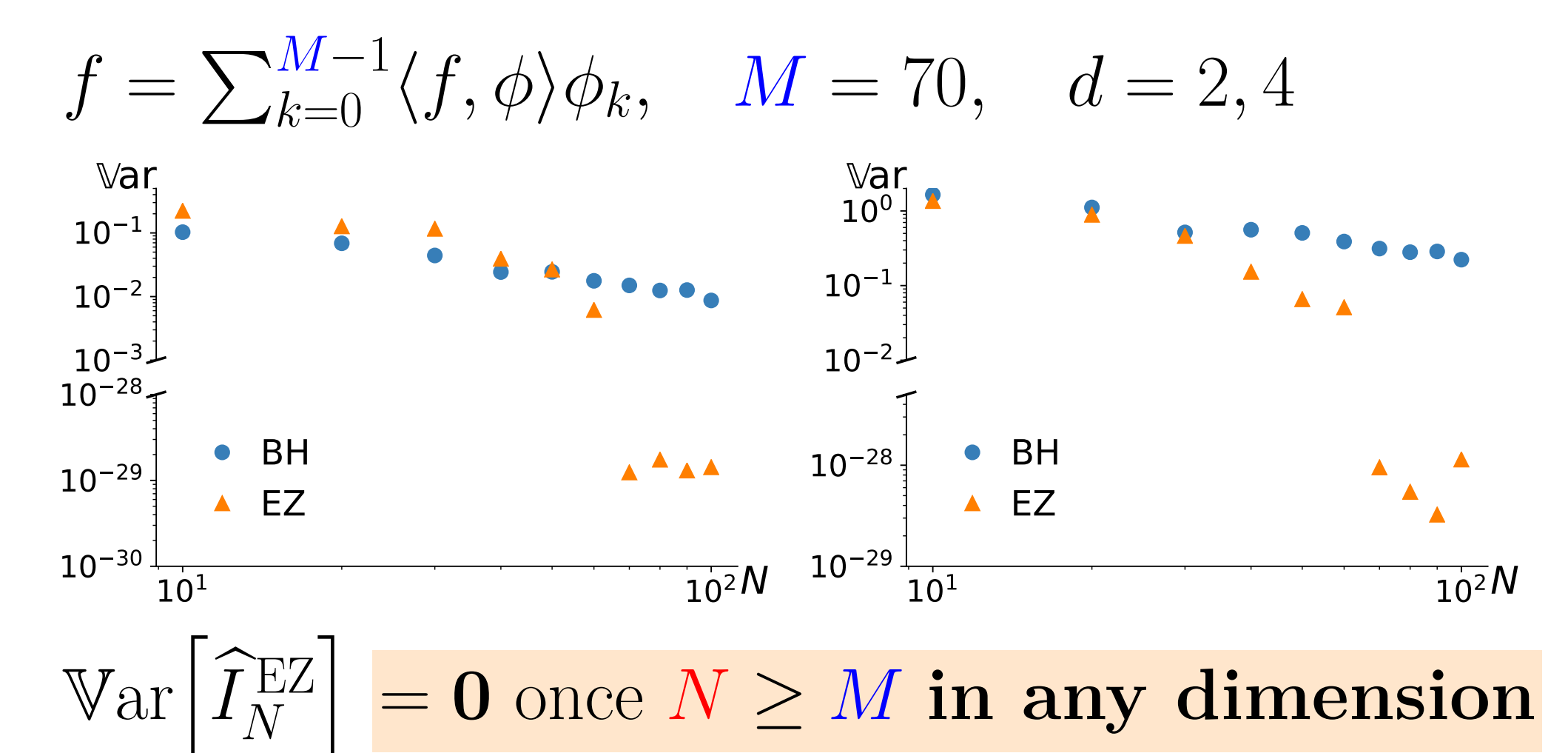
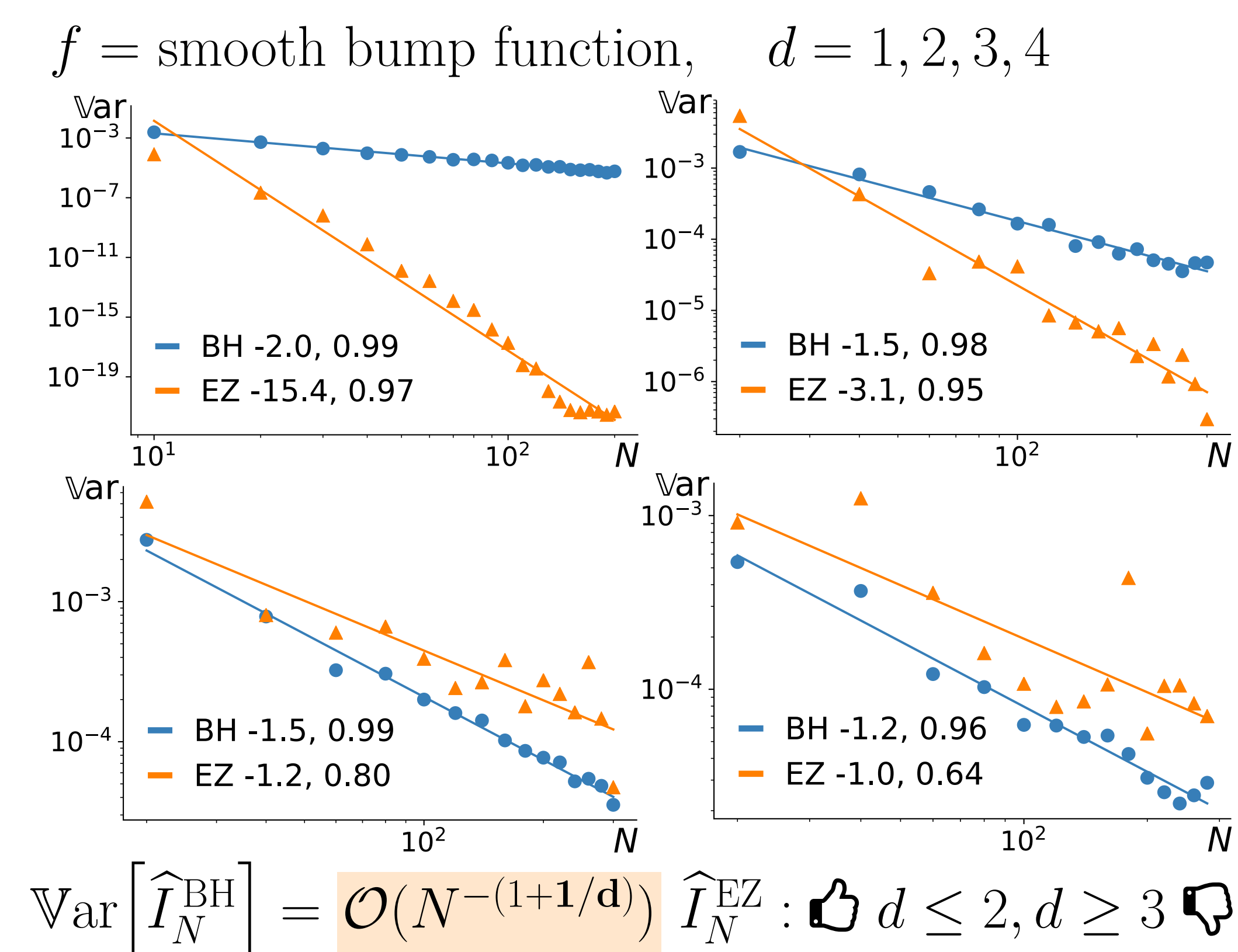
non obvious $\omega_n(\mathbf{x}_1, \dots, \mathbf{x}_n) \leq 0$ $\sum_{n=1}^N \omega_n = \mu(\mathbb{X})$

- $\mathbb{E}[\hat{I}_N^{\text{EZ}}] = \int_{\mathbb{X}} f(x) \mu(dx)$ **unbiased**
- $\text{Var}[\hat{I}_N^{\text{EZ}}] = \mu(\mathbb{X}) \times \sum_{\ell=N}^{M-1} \langle f, \phi_\ell \rangle^2 = 0$ if $N \geq M$

with the Jacobi ensemble



Experiments



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.

Ermarkov 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.