

数値積分のための点過程とサブサンプリング

早川知志 (University of Oxford)

2021/8/19

@実社会に見られる複雑なネットワークと無限粒子系の交差点 II

♣ 空間 \mathcal{X} 上の Borel 確率測度 μ に対して、数値積分

$$\sum_{i=1}^n w_i f(x_i) \approx \int_{\mathcal{X}} f(x) d\mu(x)$$

($w_i \in \mathbb{R}$: 重み、 $x_i \in \mathcal{X}$: サンプル点) を考える

この f がある RKHS (再生核 Hilbert 空間) に入っているとき

- ・ 誤差を小さくするための「良い点配置」は？
- ・ そのときの収束レートは？

という話題について、**サンプリング**の観点から議論する

♣ 集合 \mathcal{X} 上の正定値カーネル $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$

(\Leftrightarrow 任意の $(x_i)_{i=1}^n \subset \mathcal{X}$ について $(k(x_i, x_j))_{i,j=1}^n$ が半正定値対称)

ex) Euclid 内積 $x^\top y$ 、Gaussian $\exp(-|x - y|^2/\sigma^2)$ など

♣ k を再生核にもつ RKHS $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$:

- $\text{span}\{k(\cdot, x)\}_{x \in \mathcal{X}} \subset \mathcal{H}$ は \mathcal{H} 内稠密
- $\forall x \in \mathcal{X}, f \in \mathcal{H}$ について $f(x) = \langle k(\cdot, x) \rangle$ (再生性)

をみたす実 Hilbert 空間

♣ カーネル求積 (kernel quadrature/cubature)

\mathcal{X} を位相空間、 μ を Borel 確率測度として μ の求積公式

$$Q_n(f) = \sum_{i=1}^n w_i f(x_i) \left(\approx \int_{\mathcal{X}} f(x) d\mu(x) =: \mu(f) \right)$$

の最悪誤差 (worst-case error) を次で定める :

$$\text{wce}(Q_n) := \sup_{\|f\| \leq 1} |Q_n(f) - \mu(f)|$$

(ただし $\|\cdot\|$ は RKHS のノルム)

▷ この最悪誤差を小さくする Q_n が欲しい

Why kernel quadrature?

♣ カーネル求積を考えるメリットは？

- ・ 古典的な例 (Sobolev 空間など) を含んでいる
- ・ wce が計算できる (理論保証！)

$$\begin{aligned} \text{wce}(Q_n)^2 &= \sup_{\|f\| \leq 1} \left\langle \mathbf{w}^\top k(\cdot, \mathbf{x}) - \mu(k(\cdot, \cdot)), f \right\rangle^2 \\ &= \mathbf{w}^\top k(\mathbf{x}, \mathbf{x}) \mathbf{w} - 2\mathbf{w}^\top \mu(k(\cdot, \mathbf{x})) + \mu(\mu(k(\cdot, \cdot))) \end{aligned}$$

- ・ カーネルさえ導入してしまえば、空間を一般化できる
 - ・ cubature on Wiener space [Lyons and Victoir, 2004]
 - ・ 時系列データのカーネル [Király and Oberhauser, 2019]

♣ 積分作用素

$$\mathcal{K} : L^2(\mu) \rightarrow L^2(\mu); \quad f \mapsto \int_{\mathcal{X}} k(\cdot, x) f(x) \, d\mu(x)$$

のスペクトル分解を考えると、適当な条件下で

$$k(x, y) = \sum_{m=1}^{\infty} \sigma_m e_m(x) e_m(y)$$

と表示できる、ただし

- (σ_m, e_m) は \mathcal{K} の eigenpair で $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$
- $(e_m)_{m=1}^{\infty} \subset L^2(\mu)$, $(\sqrt{\sigma_m} e_m)_{m=1}^{\infty} \subset \mathcal{H}$ はそれぞれ ONB

▷ 経験則：最適な Q_n は大体 $\text{wce}(Q_n)^2 = \mathcal{O}(\sigma_n)$

カーネル求積：手法の分類

♣ $(x_i)_{i=1}^n$ の配置が肝要：大きく分けて二流派がある

・最適化

- ・ kernel herding [Chen et al., 2010] : wce への貪欲法
- ・ SBQ [Huszár and Duvenaud, 2012] : \uparrow の発展形

✓安定性・✓一般化・△理論保証・△固有値減衰

・サンプリング

- ・ optimized sampling [Bach, 2017] : 重みつき分布から iid
- ・ DPP [Belhadji et al., 2019] : projection DPP
- ・ CVS [Belhadji et al., 2020] : 点数で条件付けた DPP

△安定性・△一般化・✓理論保証・✓固有値減衰

・ $+\alpha$

- ・ Mercer + LP [Hayakawa et al., 2021] : 後に解説 (✓△✓✓)

▷ 本講演はサンプリング $+\alpha$ について

♣ Optimized sampling [Bach, 2017]

初めて kernel quadrature の精度を固有値減衰から示した

Theorem

$\lambda > 0$ に対し、分布 $dq_\lambda \propto \sum_{m=1}^{\infty} \frac{\sigma_m}{\lambda + \sigma_m} e_m^2 d\mu$ を考える
 $d_\lambda := \sum_{m=1}^{\infty} \frac{\sigma_m}{\lambda + \sigma_m}$, $n \geq 5d_\lambda \log(16d_\lambda/\delta)$ の時、 $(x_i)_{i=1}^n \sim_{iid} q_\lambda$ は

$$\mathbb{P}\left(\text{wce}(Q_n^{\mathbf{x}, \text{opt}}) \leq 4\lambda\right) \geq 1 - \delta$$

をみたす、ただし $Q_n^{\mathbf{x}, \text{opt}}$ は $\mathbf{x} = (x_i)_{i=1}^n$ に対して最適な公式

例えば $\sigma_m = \mathcal{O}(m^{-2\alpha})$ のとき、 $\text{wce}(Q_n^{\text{opt}}) = \mathcal{O}((\log m)^\alpha m^{-\alpha})$

△ Mercer 展開が必要 ・ △ サンプルングが難しい？

Kernel quadrature with DPPs

♣ Kernel quadrature with DPPs [Belhadji et al., 2019]

射影カーネル $\kappa_n(x, y) := \sum_{m=1}^n e_m(x)e_m(y)$ を用いる

DPP(κ_n) : $p(\mathbf{x}) \propto \det(\kappa_n(\mathbf{x}, \mathbf{x}))$ で得られる行列式点過程

Theorem

$r_n := \sum_{m=n+1}^{\infty} \sigma_m = o(1/n)$ のとき $(x_i)_{i=1}^n \sim \text{DPP}(\kappa_n)$ は

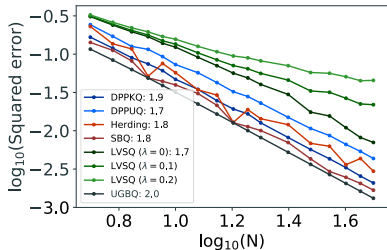
$$\mathbb{E} \left[\text{wce}(Q_n^{\mathbf{X}, \text{opt}})^2 \right] \leq 4nr_n + o(nr_n)$$

をみます

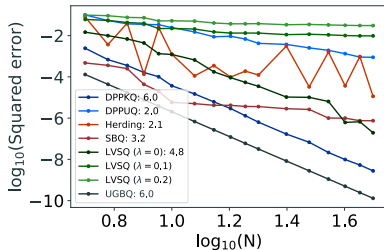
△ Mercer 展開が必要 ・ △ 理論保証が弱い (経験的には優秀)

Kernel quadrature with DPPs

♣ 数值実験 [Belhadji et al., 2019, Figure 1]



(a) Sobolev space, $d = 1, s = 1$



(b) Sobolev space, $d = 1, s = 3$

Continuous volume sampling

♣ Continuous volume sampling [Belhadji et al., 2020]

$\text{CVS}_n(k) : p(\mathbf{x}) \propto \det(k(\mathbf{x}, \mathbf{x}))$ (ただし点数 n) で得られる点過程

Theorem

$B_n := \min_{m \in \{1, \dots, n\}} ((n - m + 1)\sigma_{n+1})^{-1} \sum_{\ell \geq m}^{\infty} \sigma_{\ell}$ について
 $(x_i)_{i=1}^n \sim \text{CVS}_n(k)$ は

$$\mathbb{E} \left[\text{wce}(Q_n^{\mathbf{X}, \text{opt}})^2 \right] \leq (1 + B_n)\sigma_{n+1}$$

をみたす

$\sigma_m \sim m^{-2\alpha}$ ($\alpha > 1/2$) や $\sigma_m \sim \beta^m$ ($\beta \in (0, 1)$) なら $B_n = \mathcal{O}(1)$

✓ Mercer 展開が不要・ Δ サンプリングが非常に難しい

♣ まとめ

手法	$\text{wce}(Q_n^{\mathbf{X}, \text{opt}})^2$	Mercer	サンプリング
Optimized	$\sigma_{\lceil n/\log n \rceil}$	要	?
DPP	$n \sum_{m=n+1}^{\infty} \sigma_m$	要	$\mathcal{O}(n^3) + \text{棄却}^1$
CVS	σ_{n+1}	不要	MCMC 近似? ²

♣ 問題意識

- ・ サンプリングの実用性
- ・ 重み $(w_i)_{i=1}^n$ が非負とは限らない

¹Hough et al. [2006, Algorithm 18] & Bardenet and Hardy [2020, Section 2.4]

²Mixing time $\mathcal{O}(n^5 \log n)$ [Rezaei and Gharan, 2019]

♣ 正重みカーネル求積

重み $(w_i)_{i=1}^n$ を $w_i \geq 0$, $\sum_{i=1}^n w_i = 1$ をみたすようにすると...

- ・ 数値的に安定、 k にオーバーフィットしない
 - ・ 正関数の積分が正になり、定数関数は exact
- ・ Q_n が確率測度になる
 - ・ 解釈しやすくなる
 - ・ カーネル求積の iteration が爆発しない
(cf. cubature on Wiener space)

新しい結果の紹介 [Hayakawa et al., 2021]

Hayakawa, S., Oberhauser, H., & Lyons, T. (2021).
Positively Weighted Kernel Quadrature via Subsampling.
(arXiv:2107.09597).

♣ (古典的な) cubature

$\Omega \subset \mathbb{R}^k$ と Ω 上の確率測度 ν について、 $Q_n = \sum_{i=1}^n w_i \delta_{x_i}$ が次数 t の cubature 公式

$$\Leftrightarrow Q_n(f) = \int_{\Omega} f(x) d\nu(x), \quad \forall f: t \text{ 次以下の多項式}$$

♣ 一般化 cubature

空間 \mathcal{X} とその上の確率測度 μ について、 $Q_n = \sum_{i=1}^n w_i \delta_{x_i}$ がテスト関数 $\varphi = (\varphi_1, \dots, \varphi_d)^\top$ に対する cubature 公式

$$\Leftrightarrow Q_n(\varphi) = \int_{\mathcal{X}} \varphi(x) d\mu(x)$$

▷ Tchakaloff の定理で正重み cubature の存在が保証

♣ どうやって cubature を構成するか

- ・ 古典的な場合：各論的な代数的構成
- ・ 離散測度の場合：recombination [Litterer and Lyons, 2012]
- ・ 一般の場合：ランダムサンプリング +LP [Hayakawa, 2021]

Theorem (Hayakawa 2021)

$\varphi = (\varphi_1, \dots, \varphi_d)^\top \in L^1(\mu)$, $X_1, X_2, \dots \sim \mu$ が独立なとき

$$\mathbb{P}\left(\int_{\mathcal{X}} \varphi(x) d\mu(x) \in \text{conv}\{\varphi(X_1), \dots, \varphi(X_N)\}\right) \rightarrow 1 \quad (N \rightarrow \infty)$$

♣ 提案する正重みカーネル求積の構成手順

- (a) $n - 1$ 個のテスト関数 $\varphi = (\varphi_1, \dots, \varphi_{n-1}) \in L^1(\mu)$ を RKHS をよく近似するようにとる
- (b) 積分値 $\int_{\mathcal{X}} \varphi(x) d\mu(x)$ を計算する
- (c) $X_1, \dots, X_N \sim \mu$ を独立にサンプルする
- (d) LP を解き、 $Q_n = \sum_{i=1}^n w_i \delta_{x_i}$ であって

$$Q_n(\varphi) = \int_{\mathcal{X}} \varphi(x) d\mu(x)$$

と $w_i \geq 0$, $\sum_{i=1}^n w_i = 1$ をみたすものをとる

▷ 各ステップ近似で代用可能

Cubature for eigenfunctions

♣ 再掲 : Mercer 展開

$$k(x, y) = \sum_{m=1}^{\infty} \sigma_m e_m(x) e_m(y)$$

$(\sigma_1 \geq \sigma_2 \geq \dots \geq 0, (e_m)_{m=1}^{\infty}$ は $L^2(\mu)$ の ONB)

Theorem

e_1, \dots, e_{n-1} を exact に積分する n 点正重みカーネル求積公式 Q_n について、 $C := \sup_{m \geq 1} \|e_m\|_{\infty} < \infty$ のとき

$$\text{wce}(Q_n)^2 \leq C^2 \sum_{m=n}^{\infty} \sigma_m$$

♣ 実は sup ノルムによる制約は外せる

Theorem

e_1, \dots, e_{n-1} を exact に積分する n 点正重みカーネル求積公式 Q_n であって

$$\text{wce}(Q_n)^2 \leq 4 \sum_{m=n}^{\infty} \sigma_m$$

をみたすものが存在する

Meta-Algorithm において (d) の LP で $\sum_{i=1}^n w_i k(x_i, x_i)$ を最小化するようにすれば、正の確率 ($\mathbb{P} \rightarrow 1$ as $N \rightarrow \infty$) で構成可能

Nyström 近似

♣ サンプルに基づく手法では Mercer 展開を使う

▷ Nyström 近似で回避可能 : $\mathbf{z} = (z_1, \dots, z_\ell)^\top$ に対し $k(\mathbf{z}, \mathbf{z})$ の $s(\leq \ell)$ ランクの最良近似を $W_s(\mathbf{z})$ とし、その一般逆を $W_s^+(\mathbf{z})$

$$k_s^{\mathbf{z}}(x, y) \approx k(x, \mathbf{z})W_s^+(\mathbf{z})k(\mathbf{z}, y)$$

例えば次が示せる :

Theorem

$k_s^{\mathbf{z}}$ のなす積分作用素 (in $L^2(\mu)$) を $\mathcal{K}_s^{\mathbf{z}}$ とし、 $\mathbf{z} \sim^{iid} \mu$ のとき

$$\|\mathcal{K}_s^{\mathbf{z}} - \mathcal{K}\| \leq \sigma_{s+1} + \frac{2 \sup_x k(x, x)}{\sqrt{\ell}} \left(1 + \sqrt{2 \log \frac{1}{\delta}} \right)$$

が確率 $1 - \delta$ 以上で成り立つ ($\|\cdot\|$: 作用素ノルム)

▷ Nyström 近似で得られる近似固有関数をテスト関数に使用

♣ $[0, 1]$ 上の周期関数にノルム

$$\|f\|^2 = \left(\int_0^1 f(x) dx \right)^2 + (2\pi)^{2r} \int_0^1 f^{(r)}(x)^2 dx$$

を入れる (r は正整数) と、ベルヌーイ多項式を用いて

$$k_r(x, y) = 1 + \frac{(-1)^{r-1}(2\pi)^{2r}}{(2r)!} B_{2r}(|x - y|)$$

と再生核が表示できる [Wahba, 1990]

μ : 一様分布に対して

$$\sigma_m \sim m^{-r}, \quad \text{wce}(Q_n^{\text{opt}})^2 \sim n^{-2r}$$

などが知られている [Novak, 1988]

▷ カーネル求積手法の簡単な評価に使用

数値実験：周期的 Sobolev 空間

♣ 実験結果

- それぞれの n で 50 回の平均をプロット
- ‘Mercer’ では $N = 10n$, ‘Nyström’ では $\ell = N = 10n$
- ‘Uniform Grid’ が最適公式であることが既知

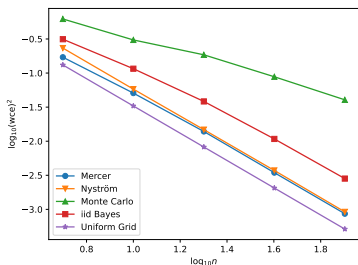


Figure 1: $r = 1$

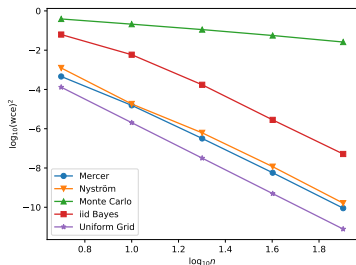
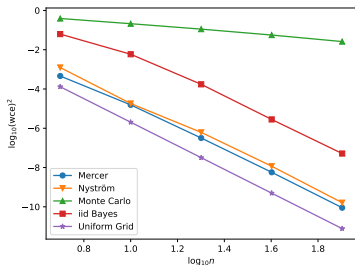
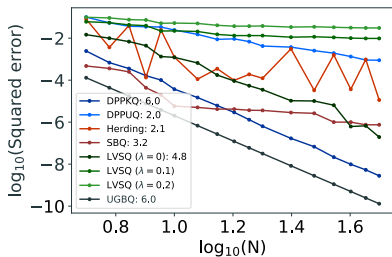


Figure 2: $r = 3$

数値実験：周期的 Sobolev 空間

♣ 実験結果

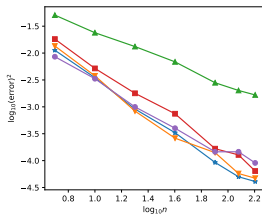
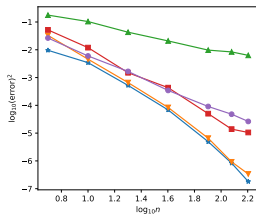
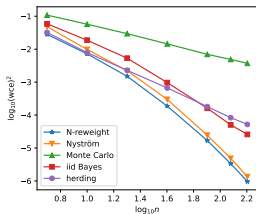
- それぞれの n で 50 回の平均をプロット
- ‘Mercer’ では $N = 10n$, ‘Nyström’ では $\ell = N = 10n$
- ‘Uniform Grid’ が最適公式であることが既知
- $r = 3$ で Belhadji et al. [2019] と比較



数値実験：実データ

♣ UCI Machine Learning Repository から Combined Cycle Power Plant データセット [Kaya et al., 2012; Tüfekci, 2014] を使用

- ・ 9568 点からなる 5 次元データ → 等重み離散測度とみなす
- ・ パラメータを median heuristics で定めた Gauss カーネル
- ・ 左から wce、 $\chi^{(5)}$ の積分誤差、 $\chi^{(5)} \mathbf{1}_{\chi^{(1)}, \chi^{(2)} \geq 0}$ の積分誤差
- ・ $\ell = N = 20n$



まとめ：スパース性から

♣ RKHS に入っている関数を精度よく数値積分したい

理想的には、大きなサンプル $\mathbf{x} = (x_1, \dots, x_N)^\top$ からスパース最適化問題

$$\begin{aligned} & \text{minimize} && \mathbf{w}^\top k(\mathbf{x}, \mathbf{x}) \mathbf{w} - 2\mathbf{w}^\top \int_{\mathcal{X}} k(\mathbf{x}, y) d\mu(y) \\ & \text{subject to} && \mathbf{w} \geq \mathbf{0}, \mathbf{1}^\top \mathbf{w} = 1, |\mathbf{w}|_0 \leq n, \end{aligned}$$

を解いて「良い点配置」を得たい…これは難しいので

- ・ 貪欲法による最適化 (herding, SBQ)
- ・ そもそものサンプリングを工夫する (optimized, DPP, CVS)
- ・ 良いテスト関数の cubature でサブサンプリング (+ α)

まとめ：サンプリング + α

♣ まとめ

手法	$wce(Q_n^{\mathbf{X}, \text{opt}})^2$	Mercer	サンプリング
Optimized	$\sigma_{\lceil n/\log n \rceil}$	要	?
DPP	$n \sum_{m=n+1}^{\infty} \sigma_m$	要	$\mathcal{O}(n^3) + \text{棄却}^3$
CVS	σ_{n+1}	不要	MCMC 近似? ⁴
提案手法	$\sum_{m=n}^{\infty} \sigma_m$	要	iid + LP
Nyström	?	不要	//

- ・ 下二つは**正重み**
- ・ 経験的には DPP・提案手法・Nyström 共に $\mathcal{O}(\sigma_{n+1})$

³Hough et al. [2006, Algorithm 18] & Bardenet and Hardy [2020, Section 2.4]

⁴Mixing time $\mathcal{O}(n^5 \log n)$ [Rezaei and Gharan, 2019]

References

- Bach, F. (2017). On the equivalence between kernel quadrature rules and random feature expansions. *The Journal of Machine Learning Research*, 18(1):714–751.
- Bardenet, R. and Hardy, A. (2020). Monte carlo with determinantal point processes. *The Annals of Applied Probability*, 30(1):368–417.
- Belhadji, A., Bardenet, R., and Chainais, P. (2019). Kernel quadrature with DPPs. In *Advances in Neural Information Processing Systems*, volume 32, pages 12907–12917.
- Belhadji, A., Bardenet, R., and Chainais, P. (2020). Kernel interpolation with continuous volume sampling. In *International Conference on Machine Learning*, pages 725–735. PMLR.
- Chen, Y., Welling, M., and Smola, A. (2010). Super-samples from kernel herding. In *Conference on Uncertainty in Artificial Intelligence*, pages 109–116.

- Hayakawa, S. (2021). Monte Carlo cubature construction. *Japan Journal of Industrial and Applied Mathematics*, 38:561–577.
- Hayakawa, S., Oberhauser, H., and Lyons, T. (2021). Positively weighted kernel quadrature via subsampling. *arXiv preprint arXiv:2107.09597*.
- Hough, J. B., Krishnapur, M., Peres, Y., Virág, B., et al. (2006). Determinantal processes and independence. *Probability surveys*, 3:206–229.
- Huszár, F. and Duvenaud, D. (2012). Optimally-weighted herding is Bayesian quadrature. In *Conference on Uncertainty in Artificial Intelligence*, pages 377–386.
- Kaya, H., Tüfekci, P., and Gürgen, F. S. (2012). Local and global learning methods for predicting power of a combined gas & steam turbine. In *Proceedings of the International Conference on Emerging Trends in Computer and Electronics Engineering*, pages 13–18.
- Kiraly, F. J. and Oberhauser, H. (2019). Kernels for sequentially ordered data. *The Journal of Machine Learning Research*, 20(31):1–45.
- Litterer, C. and Lyons, T. (2012). High order recombination and an application to cubature on Wiener space. *The Annals of Applied Probability*, 22(4):1301–1327.

- Lyons, T. and Victoir, N. (2004). Cubature on Wiener space. *Proceedings of the Royal Society of London Series A*, 460:169–198.
- Novak, E. (1988). *Deterministic and stochastic error bounds in numerical analysis*. Springer.
- Rezaei, A. and Gharan, S. O. (2019). A polynomial time MCMC method for sampling from continuous determinantal point processes. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5438–5447. PMLR.
- Tüfekci, P. (2014). Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods. *International Journal of Electrical Power & Energy Systems*, 60:126–140.
- Wahba, G. (1990). *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.