Efficient learning of smooth probability functions from Bernoulli tests with guarantees

Paul Rolland
paul.rolland@epfl.ch

Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland

July 2019,
Joint work with

Ali Kavis, Alexander Immer, Adish Singla, Volkan Cevher @ LIONS

lions@epfl EPFL FNSNF

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Introduction

- Setup $f : \mathcal{X} \to [0,1]$, $\mathcal{X} \subset \mathbb{R}^d$ compact.
- Observations:
  - ▷ Static setting: $y_i \sim \text{Bernoulli}(f(\mathbf{x}_i))$
  - ▷ Dynamic setting: $y_i \sim \text{Bernoulli}(A_i f(\mathbf{x}_i) + B_i)$, with $0 \leq A_i + B_i \leq 1$
- Goal: Approximate f over $\mathcal{X}$ from observation set $\mathcal{S} = \{(\mathbf{x}_i, y_i)\}_{i=1,\dots,n}$
- Need regularity assumption on f

Logistic Gaussian Process

- Regularity assumption:

$$f(\mathbf{x}) = \sigma(h(\mathbf{x})), \qquad h \sim \mathrm{GP}(\mu, \kappa)$$

where $\sigma(x) = \frac{1}{1+e^{-x}}$.

- Observations: $y_i \sim \mathrm{Bernoulli}(\sigma(h((\mathbf{x}_i)))$

- Issues:

  ▷ No analytically tractable posterior

  ▷ Requires costly Bayesian computations

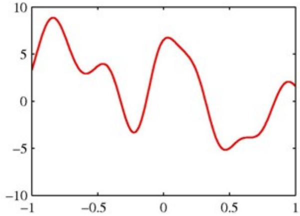


Figure: Sample from GP prior

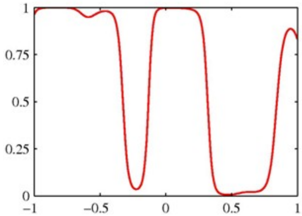


Figure: Sample from LGP prior

Smooth Beta Processes: Static setting

- Regularity assumption: f is L-Lipschitz continuous, i.e.,

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq L\|\mathbf{x} - \mathbf{x}'\|_2 \ \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}$$

- Observations: $y_i \sim \text{Bernoulli}(f(\mathbf{x}_i))$
- Prior: $p(y|\mathbf{x}) = \text{Beta}(\alpha(\mathbf{x}), \beta(\mathbf{x}))$
- Update of $\tilde{f}(\mathbf{x}|X)$ after observing $X = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$ [1]:

$$p(y|X, \mathbf{x}) = \text{Beta}\left(\alpha(\mathbf{x}) + \sum_{i=1}^{n} \delta_{y_i=1}\kappa(\mathbf{x}, \mathbf{x}_i), \beta(\mathbf{x}) + \sum_{i=1}^{n} \delta_{y_i=0}\kappa(\mathbf{x}, \mathbf{x}_i)\right)$$

Theorem (Informal - Convergence of static Beta process)

Using kernel $\kappa(\mathbf{x}, \mathbf{x}') = \delta_{\|\mathbf{x}-\mathbf{x}'\|_2 \leq \Delta_{n,L}}$ *where* $\Delta_{t,L} = L^{-\frac{2}{d+2}} n^{-\frac{1}{d+2}}$*,*

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_X \left(\mathbb{E}\left(\left(\tilde{f}(\mathbf{x}|X) - f(\mathbf{x})\right)^2\right)\right) = \mathcal{O}\left(L^{\frac{2d}{d+2}} n^{-\frac{2}{d+2}}\right).$$

[1] "Continuous Correlated Beta Processes", Goetschalckx et al.

Smooth Beta Processes: Dynamic setting

- Regularity assumption: f is L-Lipschitz continuous, i.e.,

$$|f(\mathbf{x}) - f(\mathbf{x}')| \leq L\|\mathbf{x} - \mathbf{x}'\|_2 \ \forall \mathbf{x}, \mathbf{x}' \in \mathcal{X}$$

- Observations: $y_i \sim \text{Bernoulli}(A_i f(\mathbf{x}_i) + B_i)$, with $0 \leq A_i + B_i \leq 1$.

- Prior: $p(y|\mathbf{x}) = \text{Beta}(\alpha(\mathbf{x}), \beta(\mathbf{x}))$

- Update of $\tilde{f}(\mathbf{x}|X)$ after observing $X = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$:

$$p(y|X, \mathbf{x}) = \sum_{i=1}^{n} C_i^n \text{Beta}\left(\alpha(\mathbf{x}) + i, \beta(\mathbf{x}) + n - i\right)$$

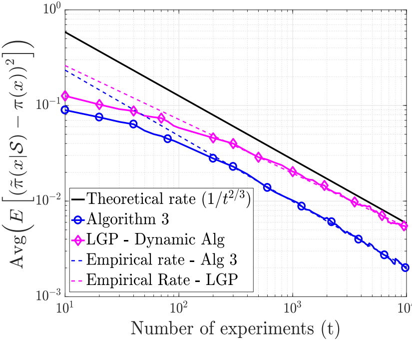
where $\{C_i^n\}_{i=1,\ldots,n}$ depend on $\{A_i, B_i\}_{i=1,\ldots,n}$ and a kernel κ.

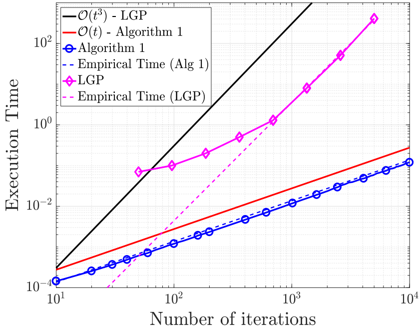
Theorem (Informal - Convergence of dynamic Beta process)

Using kernel $\kappa(\mathbf{x}, \mathbf{x}') = \delta_{\|\mathbf{x}-\mathbf{x}'\|_2 \leq \Delta_{t,L}}$ where $\Delta_{n,L} = L^{-\frac{2}{d+2}} n^{-\frac{1}{d+2}}$, and under the assumption $A_i + B_i = 1$,

$$\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_X \left(\mathbb{E}\left(\left(\tilde{f}(\mathbf{x}|X) - f(\mathbf{x}) \right)^2 \right) \right) = \mathcal{O}\left(L^{\frac{2d}{d+2}} n^{-\frac{2}{d+2}} \right).$$

Numerical results in Dynamic setting





Benefits of SBP

- Fast computation of posterior update
- Can include contextual features directly influencing success probabilities
- Simple to implement

For more details...

Welcome to our poster #233!!