On the Bayesian Nonparametric Estimation of Discontinuous Densities

Olivier Binette

October 31, 2018

Université du Québec à Montréal

Introduction

Introduction

The problem:

Estimation of a density f_0 supported on a compact metric space $(\mathrm{M}; d)$ using observations $f X_i g_{i=1}^k$ i.i.d. f_0.

Introduction

The problem:

Estimation of a density f_0 supported on a compact metric space $(\mathrm{M}; d)$ using observations $fX_ig_{i=1}^k$ i.i.d. f_0.

The focus for this talk:

What happens with Bayesian nonparametric procedures when f_0 is discontinuous?

Motivation

Density f_0 supported on a compact metric space $\mathbb{M}$:

Analysis of non-euclidean data; Statistics on manifolds.

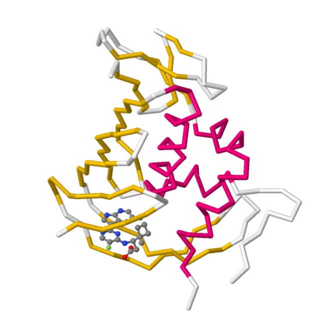
$\mathbb{M}$ could be the circle, the sphere, a quotient space, etc.

We want to account for the particular metric structure of $\mathbb{M}$.

Motivation

Bioinformatics:

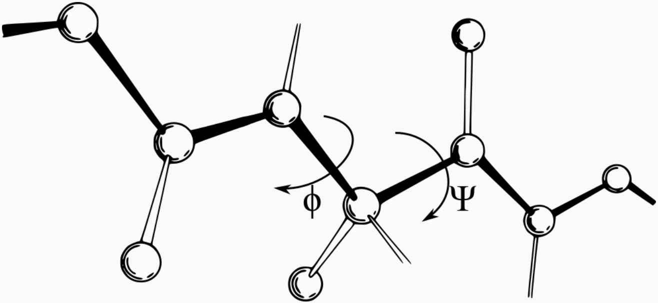
Protein structure prediction.



Motivation

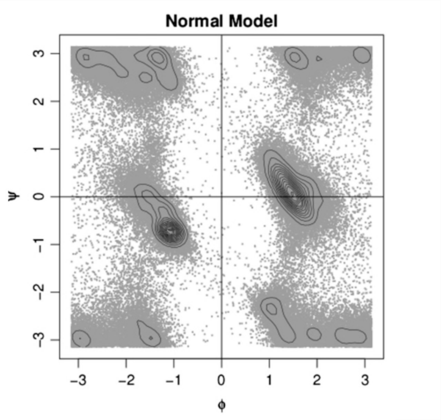
Bioinformatics:

Protein structure prediction.



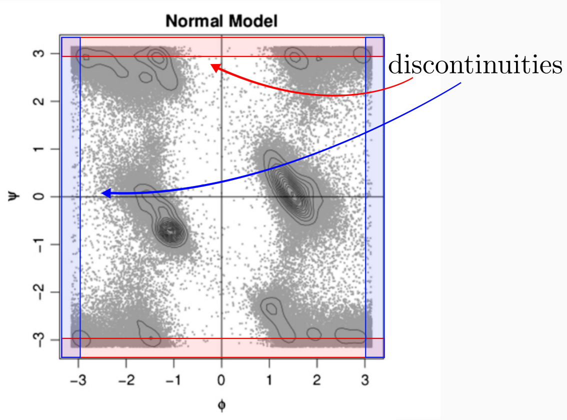
Motivation

Estimation of angle pairs distributions using mixtures of normal distributions (Dhal & al. (2008) *J. Mol. Bio.*).



Motivation

Estimation of angle pairs distributions using mixtures of normal distributions (Dhal & al. (2008) *J. Mol. Bio.*).



Motivation

Failed to take into account the metric structure of the Torus.

Motivation

Failed to take into account the metric structure of the Torus.

The team adjusted their model the next year (Lennox & al. (2009) *JASA*).

Motivation

Previous work:

Bhattacharya & Dunson (*Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to speci c manifolds*, 2012) give conditions on DPMs for consistency at positive continuous densities.

Motivation

Previous work:

Bhattacharya & Dunson (*Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to specific manifolds*, 2012) give conditions on DPMs for consistency at positive continuous densities.

Case $M = [0, 1]$: Petrone and Wasserman (2002) show posterior consistency for random Bernstein polynomial priors at continuous densities.

Motivation

Previous work:

Bhattacharya & Dunson (*Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to specific manifolds*, 2012) give conditions on DPMs for consistency at positive continuous densities.

Case $M = [0;1]$: Petrone and Wasserman (2002) show posterior consistency for random Bernstein polynomial priors at continuous densities.

Wu and Ghosal (2008) study the Kullback-Leibler property of sieve priors, but also require continuity.

Motivation

Previous work:

Bhattacharya & Dunson (*Strong consistency of nonparametric Bayes density estimation on compact metric spaces with applications to speci c manifolds*, 2012) give conditions on DPMs for consistency at positive continuous densities.

Case $M = [0; 1]$: Petrone and Wasserman (2002) show posterior consistency for random Bernstein polynomial priors at continuous densities.

Wu and Ghosal (2008) study the Kullback-Leibler property of sieve priors, but also require continuity.

Constrasts with KDE which are known to be L^1 consistent at bounded discontinuous densities since Devroye & Wagner (1979).

Motivation

What happens when we drop continuity assumptions on f_0?

Contribution

Our contribution

Show that posterior consistency also holds at discontinuous (bounded) densities for a class of sieve priors.

Contribution

Our contribution

Show that posterior consistency also holds at discontinuous (bounded) densities for a class of sieve priors.

In particular, we show that continuity assumptions are not necessary for the random (Bernstein) polynomial priors of Petrone and Wasserman (2002).

Background

Background

F as set of densities; a prior on F.

Kullback-Leibler divergence between f_0 and f:

$$K(f_0; f) = \int_{\{f_0 > 0\}} f_0 \log \frac{f_0}{f}.$$

$\mathcal{F}$ as set of densities; Π a prior on $\mathcal{F}$.

Kullback-Leibler divergence between f_0 and f:

$$K(f_0; f) = \int_{\{f_0 > 0\}} f_0 \log \frac{f_0}{f}.$$

Why is it important? If $\{X_i\}$ is i.i.d. f_0, then

$$K(f_0; f) = L \iff \prod_{i=1}^{k} \frac{f_0(X_i)}{f(X_i)} = e^{Lk + o(k)} \quad \text{a.s. as } k \to \infty.$$

The Kullback-Leibler support of Π is the set of all densities $f_0 \in \mathcal{F}$ such that

$$\Pi(\{f : K(f_0; f) < \epsilon\}) > 0 \quad \forall \epsilon > 0.$$

Background

The Kullback-Leibler support of Π is the set of all densities $f_0 \in \mathcal{F}$ such that

$$\Pi\left(\{f : K(f_0; f) < \epsilon\}\right) > 0 \quad \forall \epsilon > 0.$$

The posterior distribution of Π is strongly consistent at every density of its Kullback-Leibler support provided a complexity constraint on the prior (fundamental work by Andrew Barron and many others; see also Xing & Ranneby (2009)).

Strong posterior consistency means that if $\{X_i\}$ is i.i.d. f_0, then

$$\Pi\left(f : \int |f_0 - f| < \varepsilon \;\middle|\; \{X_i\}_{i=1}^{k} \right) \to 1$$

a.s. as $k \to \infty$ and for every $\varepsilon > 0$.

Implies almost sure L^1 convergence of the Bayes estimator.

The Kullback-Leibler support of sieve priors

General framework

(M, d) a compact metric space together with a finite measure defined on its Borel σ-algebra.

General framework

- $(M; d)$ a compact metric space together with a nite measure de ned on its Borel -algebra.
- F a set of bounded densities.

General framework

- (M, d) a compact metric space together with a finite measure defined on its Borel σ-algebra.
- $\mathcal{F}$ a set of bounded densities.
- Linear operators $T_n : L^1(M) \to L^1(M)$, $n \in \mathbb{N}$ which maps densities to densities and $T_n(\mathcal{F}) \subset \mathcal{F}$.

General framework

- (M, d) a compact metric space together with a finite measure defined on its Borel σ-algebra.
- $\mathcal{F}$ a set of bounded densities.
- Linear operators $T_n : L^1(M) \to L^1(M)$, $n \in \mathbb{N}$ which maps densities to densities and $T_n(\mathcal{F}) \subset \mathcal{F}$.

We thus obtain a model $\mathcal{C} = \bigcup_{n \in \mathbb{N}} \mathcal{C}_n$ with $\mathcal{C}_n := T_n(\mathcal{F})$.

- $(M; d)$ a compact metric space together with a finite measure μ defined on its Borel σ-algebra.
- F a set of bounded densities.
- Linear operators $T_n : L^1(M) \to L^1(M)$, $n \in \mathbb{N}$ which maps densities to densities and $T_n(F) \subset F$.

We thus obtain a model $C = \bigcup_{n \in \mathbb{N}} C_n$ with $C_n := T_n(F)$. Priors Π_n on C_n and distribution ρ on $\mathbb{N}$ yields

$$\Pi = \sum_n \rho(n) \Pi_n.$$

- T_n is linear, maps densities to densities and $T_n(F) \subset F$.
- $= \sum_n (n) \;_n$.

Theorem (Binette & Guilotte, 2018)

Let F, $_n$, and T_n be as before. Suppose also that

- $T_n(F)$ is of finite dimension, and
- $\| T_n f - f \|_\infty \to 0$ for every continuous f.

If $(n) > 0$ and $_n$ has (L^1) support $T_n(F)$, then any (bounded) $f_0 \in F$ is in the Kullback-Leibler support of .

Corollary

Under the hypotheses of the Theorem, suppose further that $(n) < Ce^{cd_n}$ for some increasing sequence $d_n > \dim T_n(F)$ and some $c; C > 0$. Then the posterior distribution of is strongly consistent at any (bounded) f_0 2 F.

Corollary

Under the hypotheses of the Theorem, suppose further that $(n) < Ce^{cd_n}$ for some increasing sequence $d_n > \dim T_n(F)$ and some $c; C > 0$. Then the posterior distribution of is strongly consistent at any (bounded) f_0 2 F.

Idea: If $(n) < Ce^{cd_n}$, then the prior satis es a nite Hausdor -entropy conditions which entails posterior consistency on the Kullback-Leibler support (Xing & Ranneby (2009); same spirit as Walker (2004) and Barron (1986)).

What can this be applied to?

- Sieve priors.
- Random series priors / nite mixtures with a random number of components.
- Dirichlet Process location mixtures with a discrete scale parameter; $T_n f = f \quad {}_{1=n}$.

Random polynomial priors

The Bernstein-Kantorovich operator is defined as

$$T_n f(x) = (n+1) \sum_{i=0}^{n} \int_{\frac{i}{n+1}}^{\frac{i+1}{n+1}} f(t)\,dt\, p_{i;n}(x); \quad x \in [0;1];$$

where $p_{i;n}(u) = \binom{n}{i} u^i (1-u)^{n-i}$ is the ith Bernstein polynomial of degree n.

Random polynomial priors

The Bernstein-Kantorovich operator is defined as

$$T_n f(x) = (n+1) \sum_{i=0}^{n} \int_{\frac{i}{n+1}}^{\frac{i+1}{n+1}} f(t)\, dt\, p_{i;n}(x); \quad x \in [0;1];$$

where $p_{i;n}(u) = \binom{n}{i} u^i (1-u)^{n-i}$ is the ith Bernstein polynomial of degree n. It may be extended to act on a probability measure P by letting

$$T_n P = (n+1) \sum_{i=0}^{n} P\left(\frac{i}{n+1}; \frac{i+1}{n+1}\right) p_{i;n}(x):$$

If P has a density f, then $T_n P = T_n f$.

Random polynomial priors

Bernstein-Dirichlet prior of Petrone and Wasserman (2002): DPM induced by the random density

$$T_N D; \quad \text{where } N \quad ; \; D \quad \text{DP}:$$

Bernstein-Dirichlet prior of Petrone and Wasserman (2002): DPM induced by the random density

$$T_N D; \quad \text{where } N \quad ; \; D \quad \text{DP}.$$

T_n is a linear operator mapping densities to densities; its image is of dimension n and it is well-known that $\|T_n f \quad f\|_\infty \to 0$ $\forall f \in C([0;1])$.

Random polynomial priors

Bernstein-Dirichlet prior of Petrone and Wasserman (2002): DPM induced by the random density

$$T_N D; \quad \text{where } N \quad ; \; D \quad DP:$$

T_n is a linear operator mapping densities to densities; its image is of dimension n and it is well-known that $kT_n f \quad f k_1 \; ! \; 0$ $8f \; 2 \; C([0;1])$.

Provided $T_n(D)$ has full support and $0 < \; (n) < Ce^{\; cn}$, then posterior consistency holds at any bounded f_0.

Random polynomial priors

Bernstein-Dirichlet prior of Petrone and Wasserman (2002): DPM induced by the random density

$$T_N D; \quad \text{where } N \sim \lambda; \; D \sim \text{DP}.$$

T_n is a linear operator mapping densities to densities; its image is of dimension n and it is well-known that $\|T_n f - f\|_\infty \to 0$ $\forall f \in C([0,1])$.

Provided $T_n(D)$ has full support and $0 < \lambda(n) < Ce^{-cn}$, then posterior consistency holds at any bounded f_0.

This shows the continuity assumptions of Petrone and Wasserman (2002) are not necessary; boundedness is sufficient.

Conclusion

Conclusion

For a class of sieve priors, L^1 posterior consistency also holds at bounded discontinuous densities.

- Circular/spherical density estimation
- Copula density estimation and other shape constrained problems.

Notes and slides can be found on my personal page `olivierbinette.ca`.

Full story is in *Bayesian Nonparametrics for Directional Statistics* (Binette & Guillotte, 2018; arXiv:1807.00305).

Thank you!