# BNP density estimation on compact metric spaces with a view towards circular statistics

NSERC CRSNG UQÀM

Olivier Binette & Simon Guillotte

## Key points

- We propose a general framework of density estimation on compact metric spaces using sieve priors.
- Posterior consistency and convergence rates are deduced from the approximation properties of the sieves, as they are accounted for by shape-preserving linear operators.
- This is applied in the context of circular statistics: we study a circular analogue to the Bernstein polynomial densities and resulting BNP estimators. The density basis is of independent interest for circular statistics applications.

## Some definitions

Let $(\mathbb{M}, d)$ be a (finitely measured) compact metric space and let $\mathbb{F}$ be the set of all bounded density functions on $\mathbb{M}$. A sieve prior on $\mathbb{F}$ with sieves $\mathcal{S}_n \subset \mathbb{F}$ is of the form

$$\Pi(A) = \sum_{n \in \mathbb{N}} \rho(n) \Pi_n(A \cap \mathcal{S}_n)$$

where each $\Pi_n$ is a prior on $\mathcal{S}_n$ and $\rho$ is a prior on $\mathbb{N}$.

## Motivation

*Many papers on the properties of particular sieve priors for density estimation on compact metric spaces (random Bernstein polynomials, their tensor product in higher dimension, thinned basis functions, adaptations to compositional data, Feller priors)*

1. Need for a unified treatment that is easily applicable to other models (splines with fixed knots, models for circular and directional data, shape space data).
2. Does posterior consistency also hold for **data-generating distributions** with discontinuous densities?
3. Can these models be helpful for the analysis of non-linear data, such as angular or circular data?

## Our approach

We consider sieves of the form

$$\mathcal{S}_n = T_n(\mathbb{F}) \subset \mathbb{F}$$

where $T_n : L^1(\mathbb{M}) \to L^1(\mathbb{M})$ is a positive linear operator of finite increasing rank $d_n$ with $T_n(1) \to 1$. Typically,

$$T_n(\mathbb{F}) \subset \left\{ \sum_{j=0}^{d_n} c_j \varphi_{j,n} \mid \sum_j c_j = 1 \right\}$$

is a finite mixture model with fixed density basis functions $\varphi_{j,n}$. Now assume there is a $d > 0$ such that

$$d_n \asymp n^d, \; \log \rho(n) \asymp -d_n \log(d_n),$$

and, for simplicity on this poster, let $\Pi_n$ be "uniformly distributed" on the sieves $\mathcal{S}_n = T_n(\mathbb{F})$.

### Theorem

*Under the above hypotheses, if $X_i \sim^{i.i.d.} f_0$ for some $f_0 \in \mathbb{F}$ satisfying $\| \log f_0 \|_\infty < \infty$ and*

$$\|T_n f_0 - f_0\|_\infty = \mathcal{O}\left(n^{-\beta}\right)$$

*for some $\beta > 0$, then the posterior distribution of $\Pi$ given $\{X_i\}_{i=1}^n$ contracts around $f_0$ with respect to the Hellinger distance at the rate*

$$\varepsilon_n = (n/\log n)^{-\frac{\beta}{2\beta+d}}.$$

**Remarks:**

- If the metric space has Minkowski dimension $d$, then operators corresponding to random histogram models can be constructed so that (roughly) $d_n \asymp n^d$ and $\|T_n f - f\|_\infty = \mathcal{O}(\omega_f(n^{-1}))$ for any $f$, where $\omega_f$ is a modulus of continuity of $f$. This implies the convergence rate $(n/\log n)^{-\beta/(2\beta+d)}$ over the class of $\beta$-smooth density functions.
- In the case of the random Bernstein polynomials, the model can be summarized using the *Bernstein-Kantorovich* operator $K_n$. Well-known approximation results such as
$$\|K_n f - f\|_\infty = \mathcal{O}(\omega_f(n^{-1/2}))$$
immediately yield known rates for this model.
- The use of a linear operator to summarize the approximation properties of the sieves is suboptimal when $f_0$ is an element of the model, and our rate is typically not sharp in that case.

**Consistency theorem**

In addition to rates at continuous densities, we have posterior consistency at any (bounded) $f_0 \in \mathbb{F}$.

### Theorem

*Under the same hypotheses as before, the posterior distribution of $\Pi$ is Hellinger consistent at any $f_0 \in \mathbb{F}$.*

See Binette & Guillotte (2019) for more general statements.

## Circular statistics

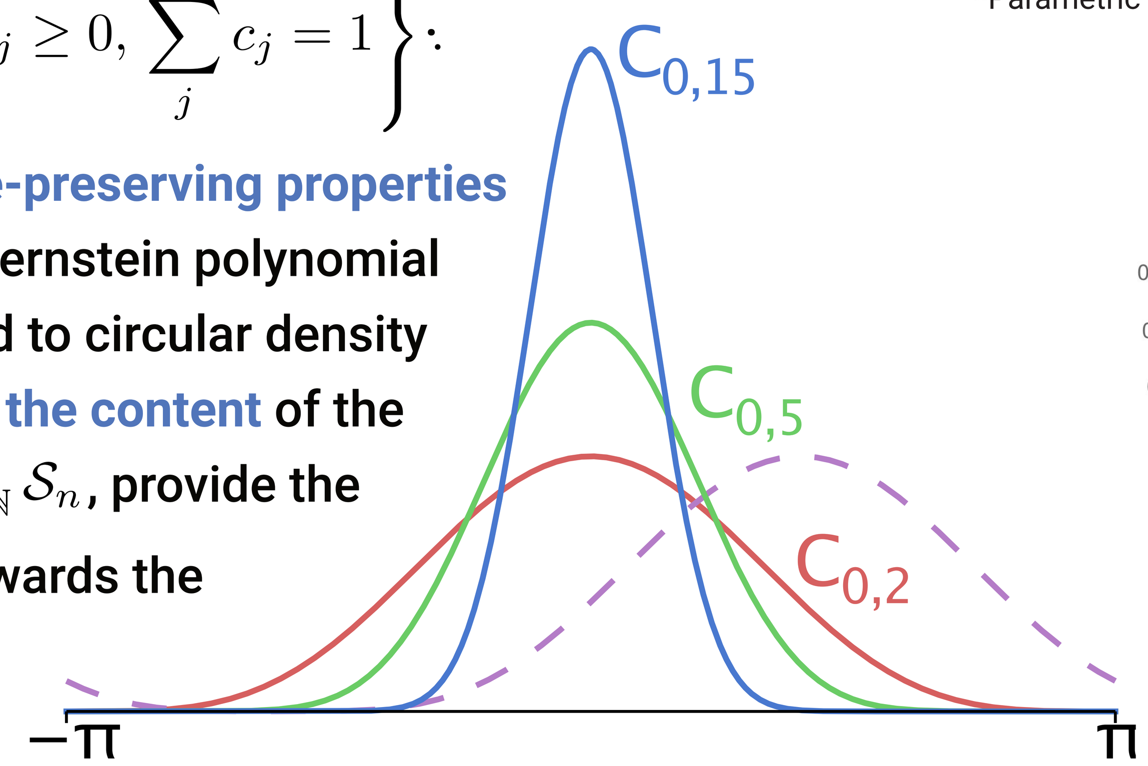
We study the *De la Vallée Poussin* basis on the circle

$$C_{j,n}(u) \propto \left(1 + \cos\left(u - 2\pi j/(2n+1)\right)\right)^n$$

and resulting trigonometric density mixture models

$$\mathcal{S}_n = \left\{ \sum_{j=0}^{2n} c_{j,n} C_{j,n} \mid c_j \geq 0, \; \sum_j c_j = 1 \right\}.$$

These densities have shape-preserving properties analogous to those of the Bernstein polynomial densities, while being suited to circular density modelling. We characterize the content of the nonparametric model $\bigcup_{n \in \mathbb{N}} \mathcal{S}_n$, provide the change of basis formula towards the truncated Fourier series representation and show how to generate random variates following the density basis. This also provides a negative mixture sampling algorithm to simulate trigonometric polynomial densities. See Binette & Guillotte (2019) for full story.

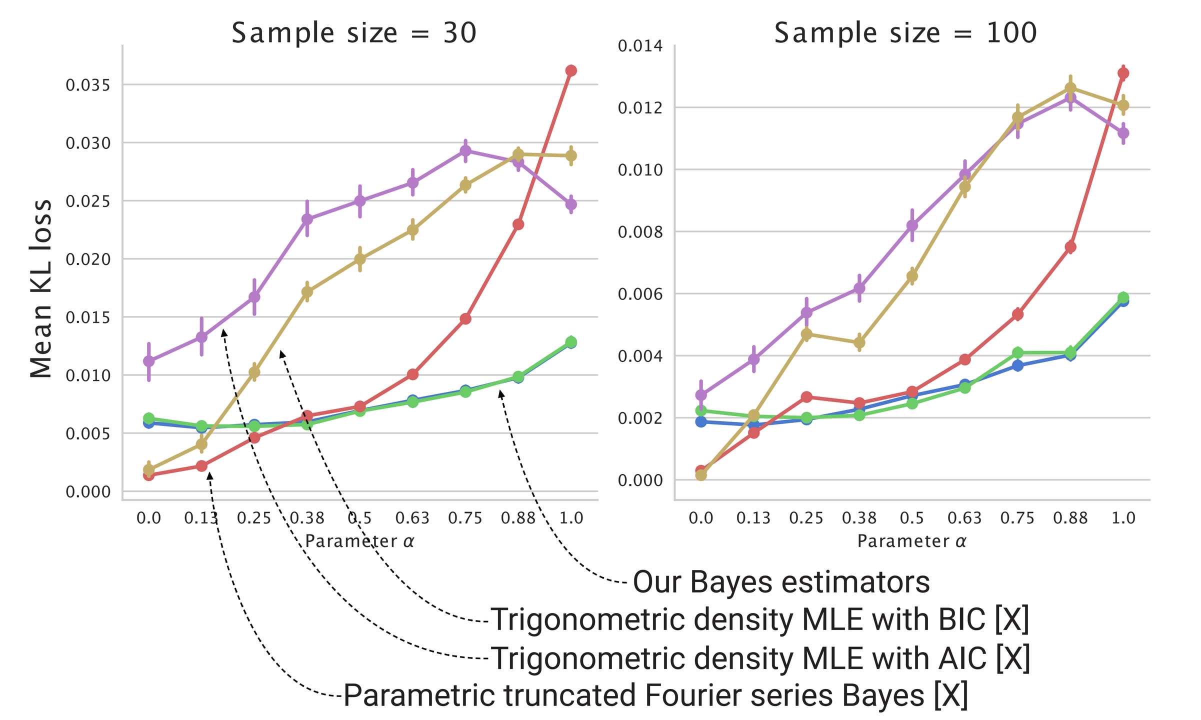


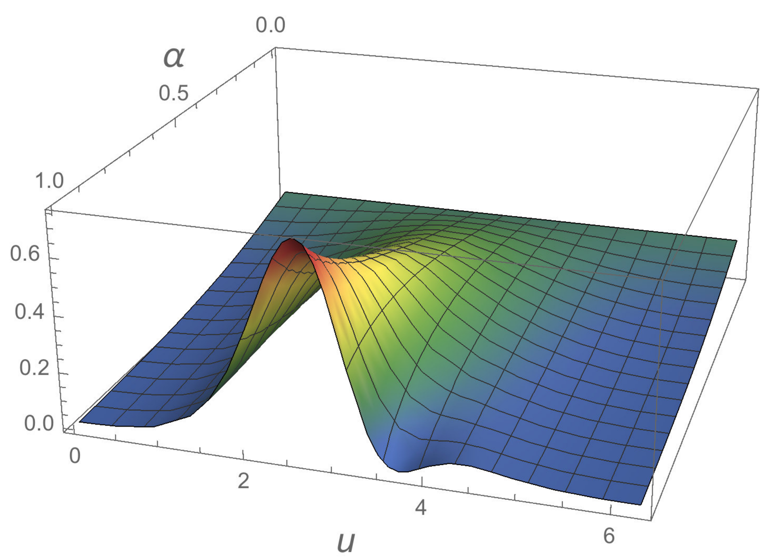

## Example

Through Dirichlet priors on the coefficients and a prior $\rho$ on $n$, we obtain priors on the set of all bounded circular densities and posterior mean estimates.

Convergence rates follow immediately in our framework from known approximation properties of the "De la Vallée Poussin means" operator. They're the same as for the random Bernstein polynomials.

Finite sample performance compares favourably to other circular density estimators (complete experiments and details in Binette & Guillotte (2019)).




Family of target densities

References are appended.