Bayesian nonparametric multiscale mixture models via Hilbert-curve partitioning

Daniele Zago

Department of Statistical Sciences, University of Padova

Joint work with Antonio Canale, University of Padova & Marco Stefanucci, Sapienza Università di Roma

Università degli Studi di Padova

Abstract

We consider the problem of flexible **nonparametric density estimation** using mixtures of densities.

We are motivated by **astrological applications**, where galaxies might be clustered based on their colour spectrum.

We rely on a **multiscale mixture model for the density** in order to **cluster observations at different resolutions**

The multiscale structure is described by using a **sequence of Hilbert curves** in order to **map the multivariate space to a binary tree**

The resulting mixture is **flexible** and can **adapt** very well **to the underlying smoothness** of the data

Motivating application

We focus on relating **DDE exposure**, henceforth x, in pregnant women to the risk of a **premature delivery** (Longnecker et al., 2001)

The data set is obtained from a sub-study of the US Collaborative Perinatal Project (CPP)

The values of x are measured in n pregnant women and y is their gestational age at delivery

Figure 1 shows the histogram of y for interval of x (warning: spoiler ahead!)

In quantitative risk assessment we are interested in quantifying **risk** (Piegorsch and Bailer, 2005)

Risk is defined by the additional **risk function** (Kodell and West, 1993), i.e.

$$R_{\mathsf{A}}(x,a) = \operatorname{pr}(y \le a \mid x) - \operatorname{pr}(y \le a \mid x=0) = F_x(a) - F_0(a)$$

In the above equation a corresponds to a threshold of clinical interest (e.g. $a = 37$ weeks, for premature delivery)

Background

Multiscale models

Multiscale model define a mixture of increasingly concentrated kernels, e.g.

$$f(y) = \sum_{s=0}^{\infty}\sum_{h=1}^{2^s} \pi_{s,h}\mathcal{K}(y;\boldsymbol{\mu}_{s,h},\Omega_{s,h}),$$

where (s,h) **corresponds to a node of a binary tree** and $\mathcal{K}$ is a multivariate scale and location kernel.

$\mathcal{K}$ is **increasingly concentrated** as s increases, and the location parameter $\boldsymbol{\mu}_{s,h}$ should **span the whole space** as h moves between the values $1,\dots,2^s$.

The nonparametric prior distribution for the mixture weights $\pi_{s,h}$ has been developed by Canale and Dunson [2016]

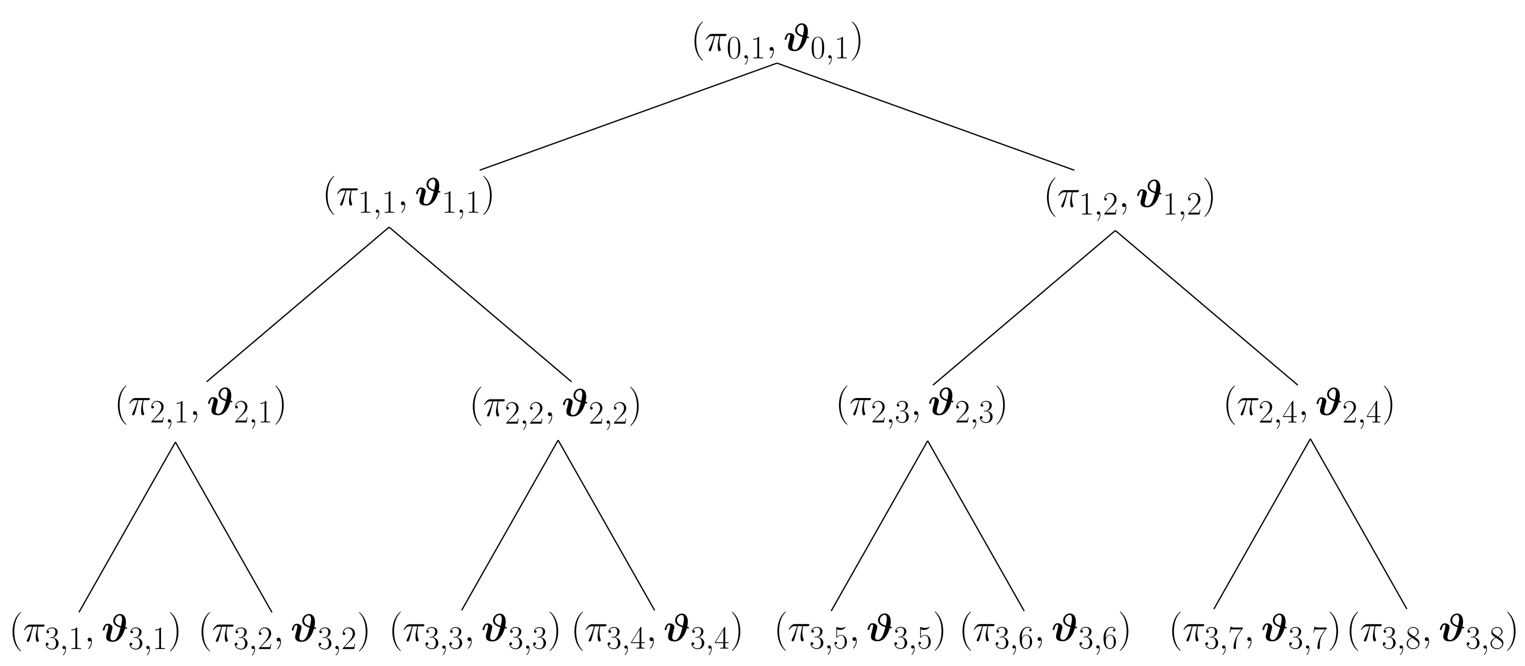


Figure 1

☺ Pro: flexibility and adaptability to data smoothness

☹ Con: Hard to generalize the binary tree to multivariate mixtures

Solution: use Hilbert curve partitioning

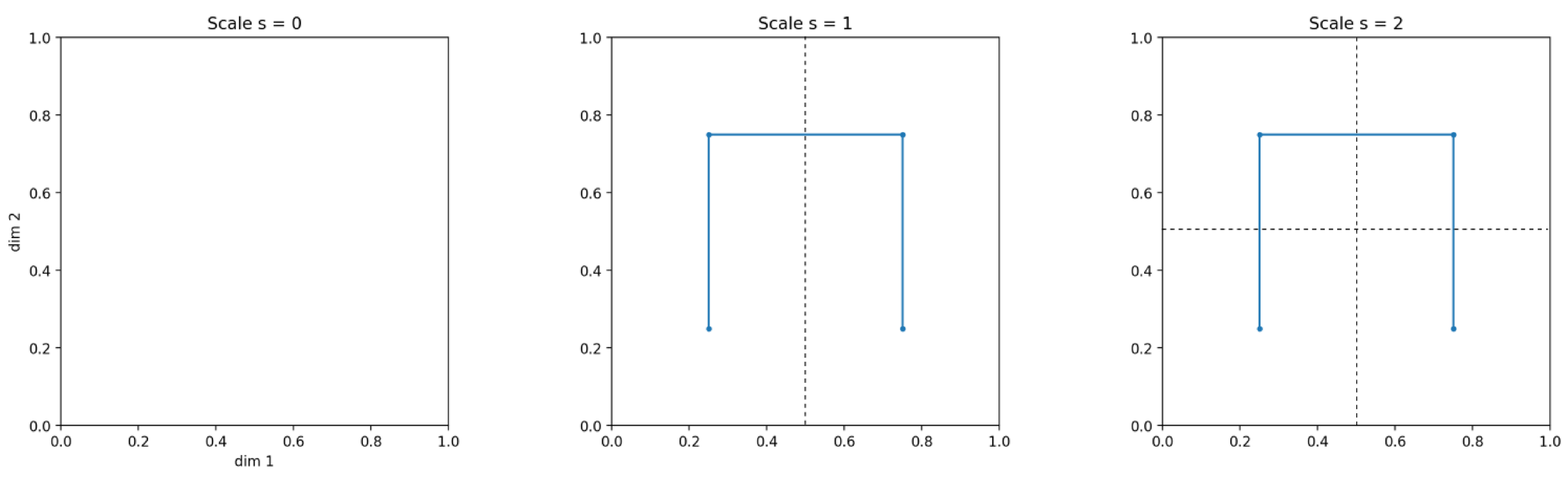
We partition the location space $\Theta_{\boldsymbol{\mu}}$ so that we span the whole space in h and the partitions are nested,

$$\Theta_{\boldsymbol{\mu}} = \bigcup_{h=1}^{2^s}\Theta_{\boldsymbol{\mu};s,h}, \quad \Theta_{\boldsymbol{\mu};s,h} = \Theta_{\boldsymbol{\mu};s+1,2h-1}\cup\Theta_{\boldsymbol{\mu};s+1,2h}.$$

The partition is done using the following two-stage procedure:

– Partition the cube $[0,1]^d$ using the Hilbert curve (Figure 2).

– Apply conditional quantiles of G_0 to the extremes of each subcube to obtain the partition.



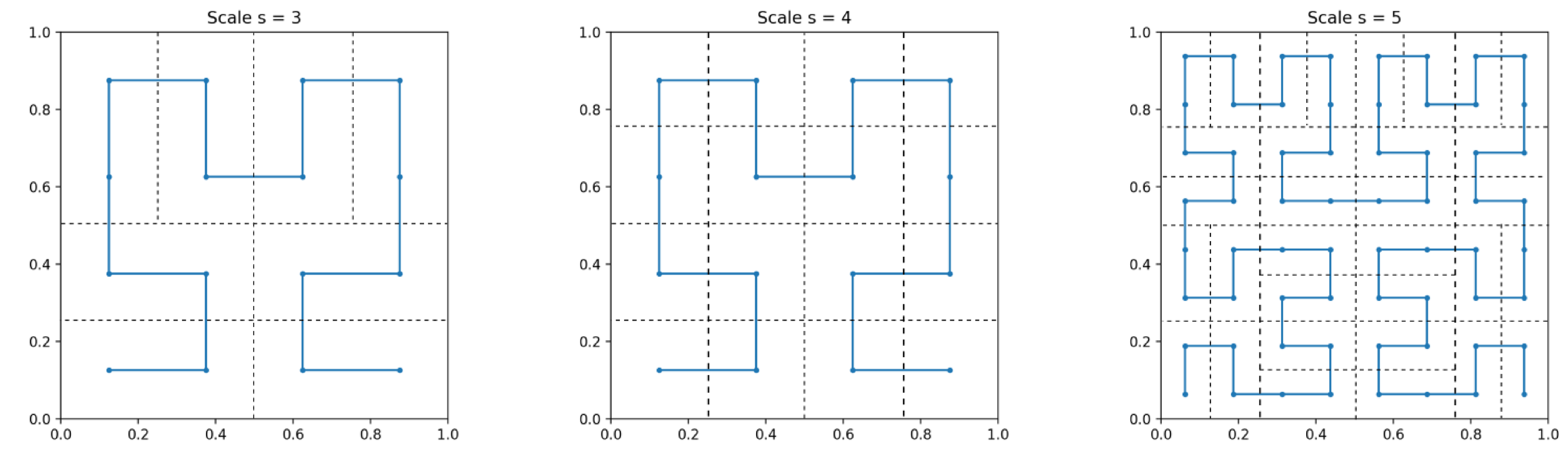


Figure 2: Dyadic partition of $[0,1]^2$ obtained by the application of the Hilbert curve, for $s = 1,\dots,5$.

Sampling at each node from G_0 truncated to $\Theta_{\boldsymbol{\mu};s,h}$ yields the **prior for the location parameter**, whereas $\Omega_{s,h}$ are sampled from a distribution H_0 **scaled by a deterministic monotone decreasing sequence** in s,

$$\Omega_{s,h} = \operatorname{diag}\big(c(s),\dots,c(s)\big)W_{s,h}, \quad W_{s,h} \overset{\text{iid}}{\sim} H_0.$$

Interpretation

· Nodes higher in the tree correspond to **coarser kernels** whereas deeper nodes correspond to **more localized kernels**. The posterior adapts the kernels to the smoothness of the data.

· We show that the random location measure $G = \sum_{s=0}^{\infty}\sum_{h=1}^{2^s}\pi_{s,h}\delta_{\boldsymbol{\mu}_{s,h}}$ is **centered around** G_0 *a priori*,

$$\mathbb{E}[G(A)] = G_0(A) \quad \text{for all } A \subseteq \Theta_{\boldsymbol{\mu}}.$$

Performance in simulated datasets

· Scenarios: 1) correctly specified, 2) misspecified —DDP, 3) misspecified —no x effect

· Competitors: COMIRE, DDP [?], probit stick-breaking [?]

· Inference on true additional risk function $R_{\mathsf{A}}(x,37)$ (shaded areas 95% credible bands) in Figure 2

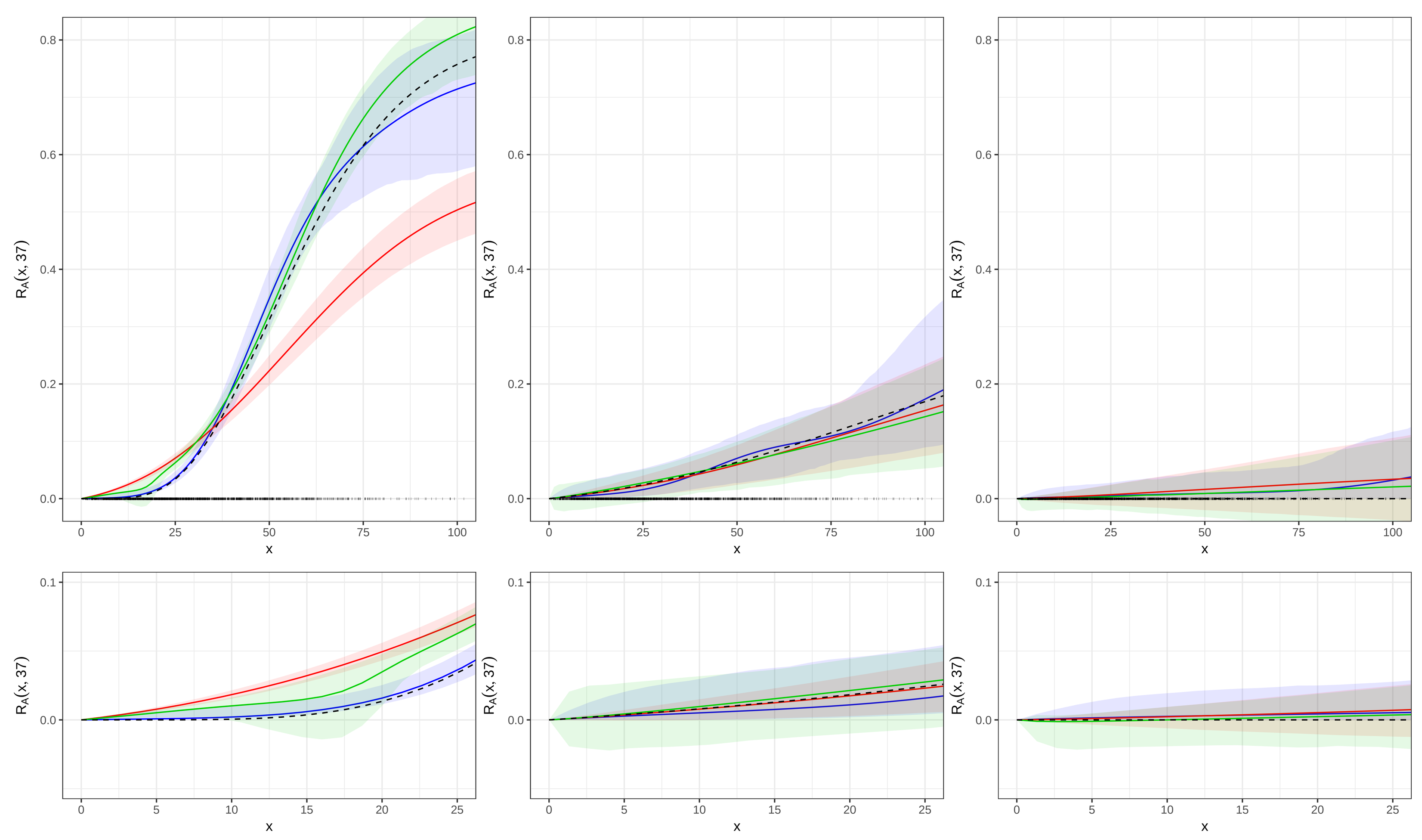


Figure 3: Inference on the additional risk function for the three scenarios.

Analysis of the CPP Data

· According to Figure 1, the conditional density of the gestational age at delivery is far from being Gaussian and displays variability, skewness, and multimodality

· at low–doses, the probability mass is concentrated around normal pregnancies, with the posterior mean (and 95% c.i.) for $\mu(0)$ being 40.20 $(40.01, 40.34)$

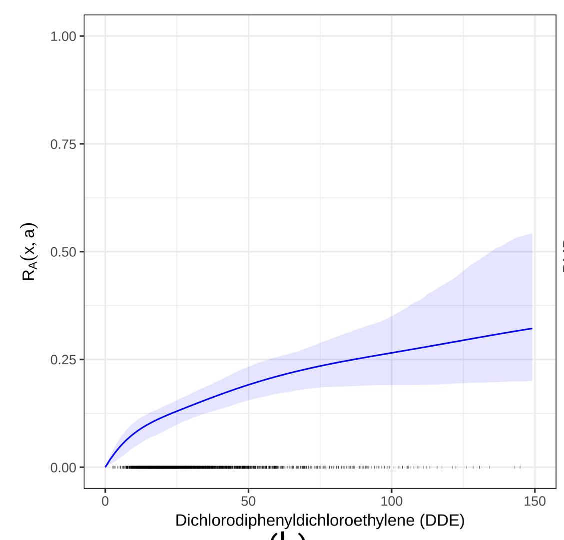
· as DDE grows the negative skewness is still maintained, and preterm deliveries increasingly inflates

· for **benchmark dose analyses** the posterior mean and the 95% c.i. of the BMD$_q$ are reported in Figure 3.

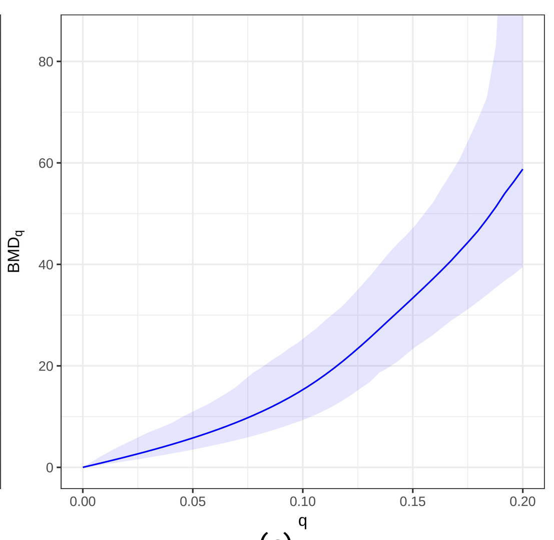
Benchmark doses q

q	**0.01**	**0.05**	**0.10**
BMD$_q$	1.03	5.79	15.29
BMDL$_q$	0.64	3.70	9.85

(a)



(b)



(c)

Figure 4: Benchmark doses for different values of risk q (a); posterior mean (solid lines), and pointwise 95% credible bands (shaded areas) for (b) $R_{\mathsf{A}}(x,37)$ and (c) the related BMD$_q$. In the x axis in (b), the observed exposures.

Scan the QR code below to download the published paper from Biometrics!

References

Antonio Canale and David B. Dunson. Multiscale Bernstein polynomials for densities. *Statistica Sinica*, 26(3):1175–1195, 2016. ISSN 1017-0405.