BAYESIAN NON-PARAMETRICS PRIORS WITH DENSITY ESTIMATION

Andi Wang   Kaspar Märtens   Ho Chung Leon Law

12th February 2016

Introduction

Why non-parametrics, i.e. infinite-dimensional parameters?

Introduction

Why non-parametrics, i.e. infinite-dimensional parameters?

Less assumptions involved.

Introduction

Why non-parametrics, i.e. infinite-dimensional parameters?

Less assumptions involved.

Bayesian non-parametrics: prior on these objects.

Dirichlet process

$$G \sim \text{Prior}$$

is a random measure

$$X_1, \ldots, X_n | G \overset{iid}{\sim} G$$

Dirichlet process

$$G \sim \text{Prior}$$

is a random measure

$$X_1, \ldots, X_n | G \overset{iid}{\sim} G$$

Want a conjugate prior for G.

Dirichlet process

$$G \sim \text{Prior}$$

is a random measure

$$X_1, \ldots, X_n | G \overset{iid}{\sim} G$$

Want a conjugate prior for G.

One solution: Dirichlet process, parameters $\alpha_0 > 0$, G_0, a probability measure.

Dirichlet process

Recall the stick-breaking construction:

Dirichlet process

Recall the stick-breaking construction:

$$\pi'_k \overset{iid}{\sim} \mathrm{Beta}(1, \alpha_0) \text{ and } \phi_k \overset{iid}{\sim} G_0$$

Dirichlet process

Recall the stick-breaking construction:

$$\pi'_k \overset{iid}{\sim} \mathrm{Beta}(1, \alpha_0) \text{ and } \phi_k \overset{iid}{\sim} G_0$$

$$\pi_k = \pi'_k \prod_{l=1}^{k-1} (1 - \pi'_l)$$

Dirichlet process

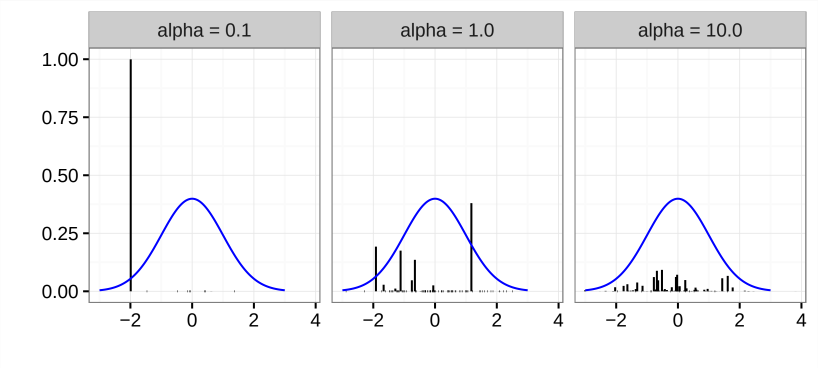
Recall the stick-breaking construction:

$$\pi'_k \overset{iid}{\sim} \mathrm{Beta}(1, \alpha_0) \text{ and } \phi_k \overset{iid}{\sim} G_0$$

$$\pi_k = \pi'_k \prod_{l=1}^{k-1} (1 - \pi'_l)$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

Dirichlet process



Dirichlet process conjugacy

$$G \sim \mathrm{DP}(\alpha_0, G_0)$$

$$X_1, \ldots, X_n | G \overset{iid}{\sim} G$$

Dirichlet process conjugacy

$$G \sim \mathrm{DP}(\alpha_0, G_0)$$

$$X_1, \ldots, X_n | G \overset{iid}{\sim} G$$

$$G | X_1, \ldots, X_n \sim \mathrm{DP}(\hat{\alpha}, \hat{G})$$

Dirichlet process conjugacy

$$G \sim \mathrm{DP}(\alpha_0, G_0)$$

$$X_1, \ldots, X_n | G \stackrel{iid}{\sim} G$$

$$G | X_1, \ldots, X_n \sim \mathrm{DP}(\hat{\alpha}, \hat{G})$$

$$\hat{\alpha} = \alpha + n$$

$$\hat{G} = \hat{\alpha}^{-1}\big(\alpha G_0 + \sum_{i=1}^{n} \delta_{X_i}\big)$$

Dirichlet process mixture

Idea: use the Dirichlet process as a prior on the parameters of a mixture model.

Dirichlet process mixture

Idea: use the Dirichlet process as a prior on the parameters of a mixture model.

$$
\begin{aligned}
G &\sim \mathrm{DP}(\alpha_0, G_0) \\
\phi_i \,|\, G &\sim G \\
x_i \,|\, \phi_i &\sim F(\phi_i)
\end{aligned}
$$

where F is a parametric distribution with density $p(x|\phi)$.

Dirichlet process mixture

Recalling

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

the Dirichlet process mixture specifies

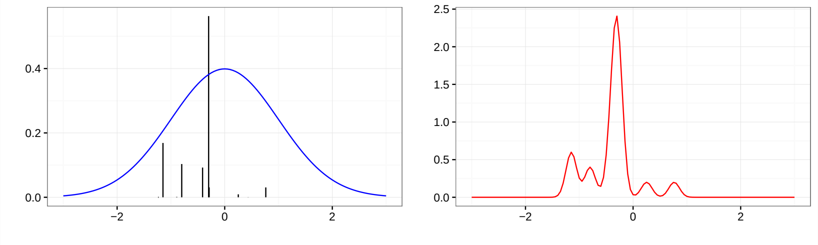
$$p(x) = \sum_{k=1}^{\infty} \pi_k \, p(x|\phi_k)$$

Recalling

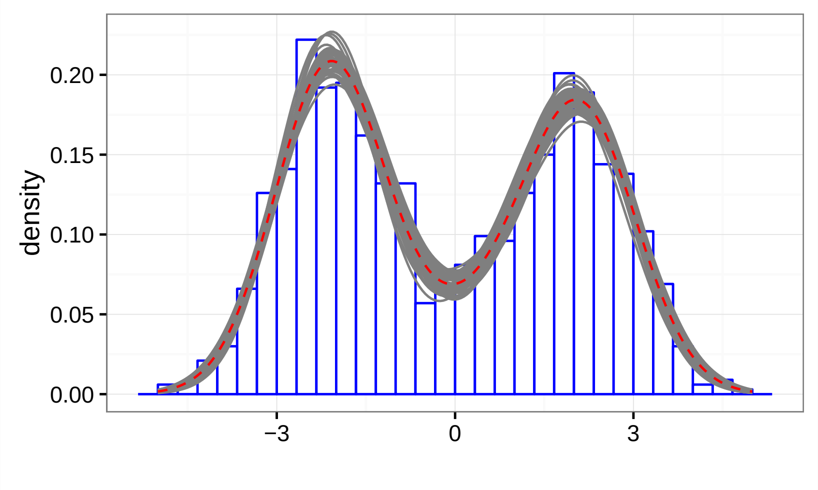
$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

the Dirichlet process mixture specifies

$$p(x) = \sum_{k=1}^{\infty} \pi_k \, p(x|\phi_k)$$



Dirichlet process mixture: fitted to data



Inference for the number of components

When data is generated from a finite mixture,

Inference for the number of components

When data is generated from a finite mixture,

Dirichlet process mixture is a misspecified model

Inference for the number of components

When data is generated from a finite mixture,

Dirichlet process mixture is a misspecified model

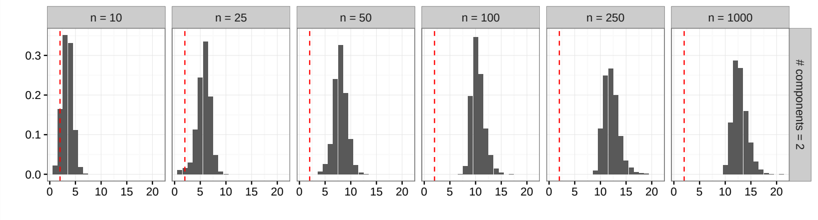
and it is inconsistent for the number of components

Inference for the number of components

When data is generated from a finite mixture,

Dirichlet process mixture is a misspecified model

and it is inconsistent for the number of components



Pòlya tree

Generalisation of the Dirichlet process.

Pòlya tree

Generalisation of the Dirichlet process.

Able to place positive mass on continuous measures.

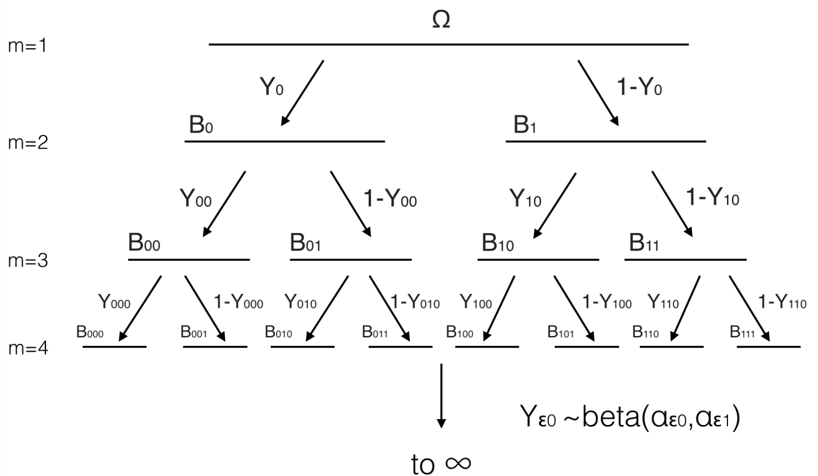
Pòlya tree

Generalisation of the Dirichlet process.

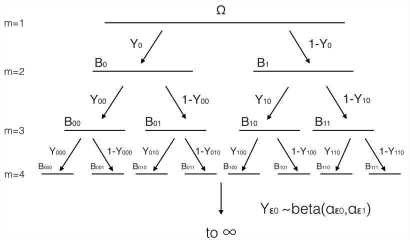
Able to place positive mass on continuous measures.

Popular in survival analysis, where censored data is common.

Pólya tree diagram



Pólya tree diagram



P on Ω have a Pólya tree distribution, with parameters $(\Pi = \{\Pi_m; m = 0, 1, \dots\}, \mathcal{A})$ is written $P \sim \mathrm{PT}(\Pi, \mathcal{A})$

The parameters

We must specify: a set of nested partitions Π and constants $\mathcal{A}$ for the beta random variables.

The parameters

We must specify: a set of nested partitions Π and constants $\mathcal{A}$ for the beta random variables.

Wish to center the tree around a specific G_0, with cdf $F_0 : \mathbb{R} \to [0, 1]$.

The parameters

We must specify: a set of nested partitions Π and constants $\mathcal{A}$ for the beta random variables.

Wish to center the tree around a specific G_0, with cdf $F_0 : \mathbb{R} \to [0, 1]$.

Take the beta parameters $\mathcal{A}$ to be constant on each level, with value $c\,\rho(m)$ on level m.

The parameters

We must specify: a set of nested partitions Π and constants $\mathcal{A}$ for the beta random variables.

Wish to center the tree around a specific G_0, with cdf $F_0 : \mathbb{R} \to [0, 1]$.

Take the beta parameters $\mathcal{A}$ to be constant on each level, with value $c\,\rho(m)$ on level m.

If $\rho(m)$ increases sufficently fast, the distribution is continuous.

The parameters

We must specify: a set of nested partitions Π and constants $\mathcal{A}$ for the beta random variables.

Wish to center the tree around a specific G_0, with cdf $F_0 : \mathbb{R} \to [0, 1]$.

Take the beta parameters $\mathcal{A}$ to be constant on each level, with value $c\,\rho(m)$ on level m.

If $\rho(m)$ increases sufficently fast, the distribution is continuous.

When $\rho(m) = 2^{-m}$ we obtain the Dirichlet process.

Parameters

One choice of Π is *dyadic quantiles*, of the form

$$\left(F_0^{-1}\left(\frac{j-1}{2^m}\right), F_0^{-1}\left(\frac{j}{2^m}\right)\right]$$

Parameters

One choice of Π is *dyadic quantiles*, of the form

$$\left(F_0^{-1}\left(\frac{j-1}{2^m}\right), F_0^{-1}\left(\frac{j}{2^m}\right)\right]$$

We showed this is valid if and only if F_0 is strictly increasing on $\mathbb{R}$.

Parameters

One choice of Π is *dyadic quantiles*, of the form

$$\left(F_0^{-1}\left(\frac{j-1}{2^m}\right), F_0^{-1}\left(\frac{j}{2^m}\right)\right]$$

We showed this is valid if and only if F_0 is strictly increasing on $\mathbb{R}$.

With this choice, we have that for each B, $\mathbb{E}[P(B)] = G_0(B)$, where $P \sim \mathrm{PT}(\Pi, \mathcal{A})$.

Posterior Update

Let us assume the following:

$$x|P \sim P \qquad P \sim \mathrm{PT}(\Pi, \mathcal{A}) \tag{1}$$

Posterior Update

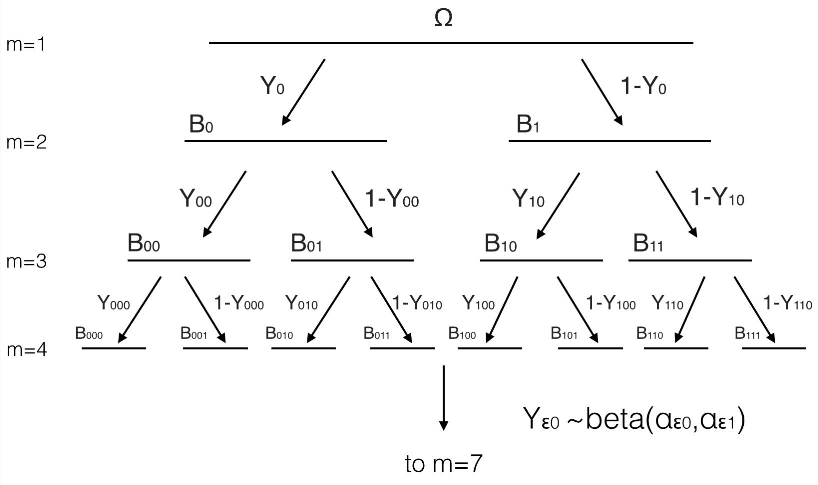
Let us assume the following:

$$x|P \sim P \qquad P \sim \mathrm{PT}(\Pi, \mathcal{A}) \tag{1}$$

Then the posterior is given by a Pólya tree, $P|x \sim \mathrm{PT}(\Pi, \mathcal{A}^*)$ with:

$$\alpha^*_\epsilon = \begin{cases} \alpha_\epsilon + 1, & \text{if } x \in B_\epsilon \\ \alpha_\epsilon, & \text{otherwise} \end{cases} \tag{2}$$

Finite Pólya tree diagram



Density Estimation with Pólya Tree

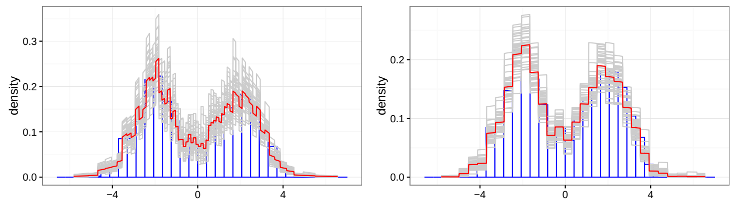


Figure: Pólya trees with two base measures, $G_0 = \mathcal{N}(0, 3^2)$ (left) and $G_0 = \mathcal{N}(0, 10^2)$ (right) with $\alpha_\epsilon = \alpha m^2$, where $\alpha \sim \Gamma(1, 0.01)$.

Summary

We have explored Bayesian non-parametric approaches for density estimation:

1. Dirichlet process mixtures
2. Pólya trees