Some good news about nonparametric priors in density estimation

Luis E. Nieto-Barajas

joint work with I. Prünster

Department of Statistics, ITAM
Mexico

INI-BNR 2007, Cambridge, UK

Outline

- Introduction

Outline

- Introduction
- Models based on mixtures

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis
  - The perturbed mixture model

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis
  - The perturbed mixture model
  - The posterior distribution of the mixing measure

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis
  - The perturbed mixture model
  - The posterior distribution of the mixing measure
  - Posterior simulation

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis
  - The perturbed mixture model
  - The posterior distribution of the mixing measure
  - Posterior simulation
- Sensitivity analysis

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis
  - The perturbed mixture model
  - The posterior distribution of the mixing measure
  - Posterior simulation
- Sensitivity analysis
  - A simulated example

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis
  - The perturbed mixture model
  - The posterior distribution of the mixing measure
  - Posterior simulation
- Sensitivity analysis
  - A simulated example
  - Galaxy data

Outline

- Introduction
- Models based on mixtures
- Tools for the sensitivity Analysis
  - The perturbed mixture model
  - The posterior distribution of the mixing measure
  - Posterior simulation
- Sensitivity analysis
  - A simulated example
  - Galaxy data
  - Acidity data

Introduction

- Density estimation is a NP problem in nature

Introduction

- Density estimation is a NP problem in nature
- Mixture models:

$$f(x) = \int k(x, y)\mathrm{P}(\mathrm{d}y)$$

where $k(x, y)$ is a parametric kernel and P a mixing distribution

Introduction

- Two approaches:

Introduction

- Two approaches:
  - Classical: Take P the e.d.f. ⇒ popular kernel density estimator (Silverman, 1986)

Introduction

- Two approaches:
  - Classical: Take P the e.d.f. $\Rightarrow$ popular kernel density estimator (Silverman, 1986)
  - Bayesian: Take P as a NP prior $\Rightarrow$ $f(x)$ is a random density function and the Bayes estimate becomes

$$\hat{f}_n(x) = \mathrm{E}\left[\int k(x,y)\mathrm{P}(\mathrm{d}y)\Big| X_1,\ldots,X_n\right]$$

Models based on mixtures

- Lo (1984): $\mathrm{P} \sim \mathscr{D}_\alpha \Rightarrow$ MDP
  Lo gave an expression for $\hat{f}_n(x)$ in terms of a sum over partitions

Models based on mixtures

- Lo (1984): $\mathrm{P} \sim \mathscr{D}_\alpha \Rightarrow$ MDP
  Lo gave an expression for $\hat{f}_n(x)$ in terms of a sum over partitions
- Ferguson (1983): Formulated the MDP as a hierarchical model

$$\begin{aligned} X_i \,|\, Y_i &\overset{\text{ind}}{\sim} k(\,\cdot\,, Y_i) \\ Y_i|\mathrm{P} &\overset{\text{iid}}{\sim} \mathrm{P} \\ \mathrm{P} &\sim \mathscr{D}_\alpha \end{aligned}$$

Models based on mixtures

- Lo (1984): $\mathrm{P} \sim \mathscr{D}_\alpha \Rightarrow$ MDP
  Lo gave an expression for $\hat{f}_n(x)$ in terms of a sum over partitions
- Ferguson (1983): Formulated the MDP as a hierarchical model

$$
\begin{aligned}
X_i \,|\, Y_i &\overset{\text{ind}}{\sim} k(\,\cdot\,, Y_i) \\
Y_i | \mathrm{P} &\overset{\text{iid}}{\sim} \mathrm{P} \\
\mathrm{P} &\sim \mathscr{D}_\alpha
\end{aligned}
$$

- Escobar and West (1995): Developed MCMC algorithms $\Rightarrow$ MDP to be exploited in several ways

Potential candidates for P

Potential candidates for P

- Pitman (1996): P $\sim$ Species sampling models

Potential candidates for P

- Pitman (1996): P $\sim$ Species sampling models
- Ishwaran and James (2001): P $\sim$ Stick breaking priors

Potential candidates for P

- Pitman (1996): P $\sim$ Species sampling models
- Ishwaran and James (2001): P $\sim$ Stick breaking priors
- Regazzini, Lijoi and Prünster (2003): P $\sim$ NRMI

Potential candidates for P

- Pitman (1996): P $\sim$ Species sampling models
- Ishwaran and James (2001): P $\sim$ Stick breaking priors
- Regazzini, Lijoi and Prünster (2003): P $\sim$ NRMI
- Pitman (2003): P $\sim$ Poisson-Kingman models

Potential candidates for P

- Pitman (1996): P $\sim$ Species sampling models
- Ishwaran and James (2001): P $\sim$ Stick breaking priors
- Regazzini, Lijoi and Prünster (2003): P $\sim$ NRMI
- Pitman (2003): P $\sim$ Poisson-Kingman models
- James (2006): P $\sim$ Spatial neutral to the right

Mixtures based on specific P's

Mixtures based on specific P's

- De Iorio, Müller, MacEachern and Rosner (2004): Designed an ANOVA model based on DDP's

Mixtures based on specific P's

- De Iorio, Müller, MacEachern and Rosner (2004): Designed an ANOVA model based on DDP's
- Lijoi, Mena and Prünster (2005, 2007): Studied special tractable cases of mixtures of NRMI and highlighted their different cluster behaviour with respect to the MDP's.

Mixtures based on specific P's

- De Iorio, Müller, MacEachern and Rosner (2004): Designed an ANOVA model based on DDP's
- Lijoi, Mena and Prünster (2005, 2007): Studied special tractable cases of mixtures of NRMI and highlighted their different cluster behaviour with respect to the MDP's.
- Dunson and Park (2006): Introduced dependent stick-breaking priors for epidemiologic and bioassay problems.

Another candidate for P

Another candidate for P

- Nieto-Barajas, Prünster and Walker (2004): Introduced a class of normalized IAP driven priors for which under certain conditions the prior can be written as a mixture model with

$$\mathrm{P} \sim \text{normalized weighted IAP},$$

Another candidate for P

- Nieto-Barajas, Prünster and Walker (2004): Introduced a class of normalized IAP driven priors for which under certain conditions the prior can be written as a mixture model with

$$\mathrm{P} \sim \text{normalized weighted IAP},$$

i.e.,

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

where

Another candidate for P

- Nieto-Barajas, Prünster and Walker (2004): Introduced a class of normalized IAP driven priors for which under certain conditions the prior can be written as a mixture model with

$$\mathrm{P} \sim \text{normalized weighted IAP},$$

i.e.,

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

where

  - $\bar{k}$ is a non-negative weighting function

Another candidate for P

- Nieto-Barajas, Prünster and Walker (2004): Introduced a class of normalized IAP driven priors for which under certain conditions the prior can be written as a mixture model with

$$\mathrm{P} \sim \text{normalized weighted IAP},$$

i.e.,

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

where

  - $\bar{k}$ is a non-negative weighting function
  - $A(\cdot)$ in an IAP

$$\Leftrightarrow A(y) \text{ is weighted with } \bar{k} \text{ and normalized}$$

What is a sensitivity analysis?

What is a sensitivity analysis?

- Typically a sensitivity analysis answers the question:

What is a sensitivity analysis?

- Typically a sensitivity analysis answers the question:

  How sensitive is the inference with respect to different prior specifications?

What is a sensitivity analysis?

- Typically a sensitivity analysis answers the question:

  How sensitive is the inference with respect to different prior specifications?

- That is, once I have chosen the mixing distribution P, how much the posterior estimates are affected if we change the hyper-parameters of P?.

What is a sensitivity analysis?

- Typically a sensitivity analysis answers the question:

  How sensitive is the inference with respect to different prior specifications?

- That is, once I have chosen the mixing distribution P, how much the posterior estimates are affected if we change the hyper-parameters of P?.
- Our approach to a sensitivity analysis is to perturb the chosen prior P

What is a sensitivity analysis?

- Typically a sensitivity analysis answers the question:

  How sensitive is the inference with respect to different prior specifications?

- That is, once I have chosen the mixing distribution P, how much the posterior estimates are affected if we change the hyper-parameters of P?.
- Our approach to a sensitivity analysis is to perturb the chosen prior P
- Perturb $\equiv$ external modification that cannot be achieved by a simple modification of the hyper-parameters

$$\Rightarrow \text{ sensitivity } + \text{ robustness}$$

The perturbed mixture model

The perturbed mixture model

- The perturbed model can be expressed in a hierarchical way:

The perturbed mixture model

- The perturbed model can be expressed in a hierarchical way: Let $(X_i)_{i\geq 1}$ be a sequence of observable r.v.'s and let $(Y_i)_{i\geq 1}$ be a sequence of latent r.v.'s, therefore

$$\begin{aligned} X_i \mid Y_i &\overset{\text{ind}}{\sim} k(\,\cdot\,, Y_i) \\ Y_i \mid \mathrm{P} &\overset{\text{iid}}{\sim} \mathrm{P} \\ \mathrm{P} &\sim \mathscr{P} \end{aligned}$$

where $\mathscr{P}$ is the distribution of a normalized weighted IAP

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

The perturbed mixture model

- Why this is a perturbed model?.

The perturbed mixture model

- Why this is a perturbed model?.
- Consider

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

The perturbed mixture model

- Why this is a perturbed model?.
- Consider

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

  - If $\bar{k} = const. \Rightarrow$ P is a NRMI

The perturbed mixture model

- Why this is a perturbed model?.
- Consider

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

  - If $\bar{k} = const. \Rightarrow$ P is a NRMI
  - If $\bar{k} \neq const. \Rightarrow$ P is a perturbed NRMI

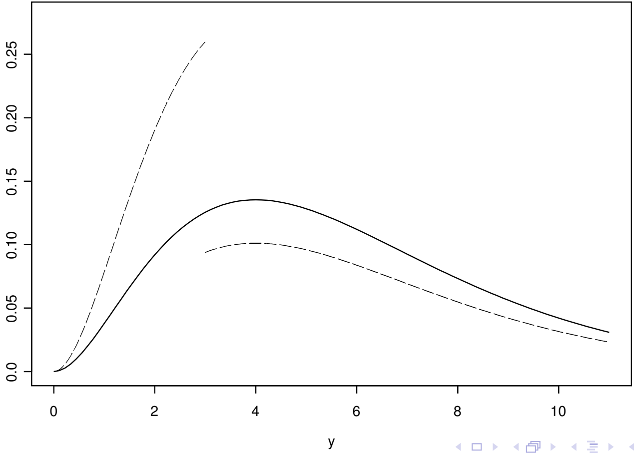
The perturbed mixture model

- Why this is a perturbed model?.
- Consider

$$\mathrm{P}(\mathrm{d}y) = \frac{\bar{k}(y)A(\mathrm{d}y)}{\int \bar{k}(s)A(\mathrm{d}s)}$$

  - If $\bar{k} = const. \Rightarrow$ P is a NRMI
  - If $\bar{k} \neq const. \Rightarrow$ P is a perturbed NRMI
- The role of $\bar{k}$ is to amplify or squeeze the jump heights of the IAP A

The perturbed mixture model



The perturbed mixture model

- The law of $\mathscr{P}$ is uniquely characterized by the weighting function $\bar{k}$ and the Poisson intensity measure ν of A. If we write

$$\nu(\mathrm{d}y, \mathrm{d}v) = \rho(\mathrm{d}v|y)\alpha(\mathrm{d}y)$$

then

The perturbed mixture model

- The law of $\mathscr{P}$ is uniquely characterized by the weighting function $\bar{k}$ and the Poisson intensity measure ν of A. If we write
$$\nu(\mathrm{d}y, \mathrm{d}v) = \rho(\mathrm{d}v|y)\alpha(\mathrm{d}y)$$
then
- If $\rho(\mathrm{d}v|y) = v^{-1}e^{-v}\mathrm{d}v$ and

The perturbed mixture model

- The law of $\mathscr{P}$ is uniquely characterized by the weighting function $\bar{k}$ and the Poisson intensity measure ν of A. If we write

$$\nu(\mathrm{d}y, \mathrm{d}v) = \rho(\mathrm{d}v|y)\alpha(\mathrm{d}y)$$

  then
- If $\rho(\mathrm{d}v|y) = v^{-1}e^{-v}\mathrm{d}v$ and
  - $\bar{k}(y) = c \Rightarrow \mathrm{P} \sim \mathscr{D}_\alpha$

The perturbed mixture model

- The law of $\mathscr{P}$ is uniquely characterized by the weighting function $\bar{k}$ and the Poisson intensity measure ν of A. If we write
$$\nu(\mathrm{d}y, \mathrm{d}v) = \rho(\mathrm{d}v|y)\alpha(\mathrm{d}y)$$
then
- If $\rho(\mathrm{d}v|y) = v^{-1}e^{-v}\mathrm{d}v$ and
  - $\bar{k}(y) = c \Rightarrow \mathrm{P} \sim \mathscr{D}_\alpha$
  - $\bar{k}(y) \neq c \Rightarrow \mathrm{P} \sim$ Perturbed $\mathscr{D}_\alpha$

The perturbed mixture model

- The law of $\mathscr{P}$ is uniquely characterized by the weighting function $\bar{k}$ and the Poisson intensity measure ν of A. If we write
$$\nu(\mathrm{d}y, \mathrm{d}v) = \rho(\mathrm{d}v|y)\alpha(\mathrm{d}y)$$
then
- If $\rho(\mathrm{d}v|y) = v^{-1}e^{-v}\mathrm{d}v$ and
    - $\bar{k}(y) = c \Rightarrow \mathrm{P} \sim \mathscr{D}_\alpha$
    - $\bar{k}(y) \neq c \Rightarrow \mathrm{P} \sim$ Perturbed $\mathscr{D}_\alpha$
- If $\rho(\mathrm{d}v|y) = \delta(\sqrt{2\pi})^{-1}v^{-3/2}e^{-\gamma^2 v/2}\mathrm{d}v$ and

The perturbed mixture model

- The law of $\mathscr{P}$ is uniquely characterized by the weighting function $\bar{k}$ and the Poisson intensity measure ν of A. If we write

$$\nu(\mathrm{d}y, \mathrm{d}v) = \rho(\mathrm{d}v|y)\alpha(\mathrm{d}y)$$

  then

- If $\rho(\mathrm{d}v|y) = v^{-1}e^{-v}\mathrm{d}v$ and
  - $\bar{k}(y) = c \Rightarrow \mathrm{P} \sim \mathscr{D}_\alpha$
  - $\bar{k}(y) \neq c \Rightarrow \mathrm{P} \sim$ Perturbed $\mathscr{D}_\alpha$
- If $\rho(\mathrm{d}v|y) = \delta(\sqrt{2\pi})^{-1}v^{-3/2}e^{-\gamma^2 v/2}\mathrm{d}v$ and
  - $\bar{k}(y) = c \Rightarrow \mathrm{P} \sim$ N-IG$_\alpha$

The perturbed mixture model

- The law of $\mathscr{P}$ is uniquely characterized by the weighting function $\bar{k}$ and the Poisson intensity measure ν of A. If we write

$$\nu(\mathrm{d}y, \mathrm{d}v) = \rho(\mathrm{d}v|y)\alpha(\mathrm{d}y)$$

  then
- If $\rho(\mathrm{d}v|y) = v^{-1}e^{-v}\mathrm{d}v$ and
  - $\bar{k}(y) = c \Rightarrow \mathrm{P} \sim \mathscr{D}_\alpha$
  - $\bar{k}(y) \neq c \Rightarrow \mathrm{P} \sim$ Perturbed $\mathscr{D}_\alpha$
- If $\rho(\mathrm{d}v|y) = \delta(\sqrt{2\pi})^{-1}v^{-3/2}e^{-\gamma^2 v/2}\mathrm{d}v$ and
  - $\bar{k}(y) = c \Rightarrow \mathrm{P} \sim$ N-IG$_\alpha$
  - $\bar{k}(y) \neq c \Rightarrow \mathrm{P} \sim$ Perturbed N-IG$_\alpha$

The posterior distribution

The posterior distribution

- For characterizing the posterior distribution of these kind of models there are usually two ways:

The posterior distribution

- For characterizing the posterior distribution of these kind of models there are usually two ways:
    1. Marginal method: Integrate out P and resort to the predictive distribution

The posterior distribution

- For characterizing the posterior distribution of these kind of models there are usually two ways:
  1. Marginal method: Integrate out P and resort to the predictive distribution
  2. Conditional method: Characterize P conditionally (usually on latent variables)

The posterior distribution

- For characterizing the posterior distribution of these kind of models there are usually two ways:
    1. Marginal method: Integrate out P and resort to the predictive distribution
    2. Conditional method: Characterize P conditionally (usually on latent variables)
- The predictive distributions of the perturbed NRMI are intractable so we will follow the conditional method

The posterior distribution

- Using the hierarchical representation of the perturbed mixture model, we will characterize the un-normalized weighted IAP, say

$$A_k(\mathrm{d}y) := \bar{k}(y)A(\mathrm{d}y)$$

The posterior distribution

- Using the hierarchical representation of the perturbed mixture model, we will characterize the un-normalized weighted IAP, say
$$A_k(\mathrm{d}y) := \bar{k}(y)A(\mathrm{d}y)$$
- Due to the discreteness of $A_k(\mathrm{d}y) \Rightarrow$ ties in the Y_i's. Therefore
$$(Y_1,\ldots,Y_n) \to \left\{ \begin{array}{l} (Y_1^*,\ldots,Y_r^*) \\ (n_1,\ldots,n_r) \end{array} \right. , \text{ con } \sum_{j=1}^{r} n_j = n$$

The posterior distribution

- Using the hierarchical representation of the perturbed mixture model, we will characterize the un-normalized weighted IAP, say

$$A_k(\mathrm{d}y) := \bar{k}(y)A(\mathrm{d}y)$$

- Due to the discreteness of $A_k(\mathrm{d}y) \Rightarrow$ ties in the Y_i's. Therefore

$$(Y_1, \ldots, Y_n) \rightarrow \begin{cases} (Y_1^*, \ldots, Y_r^*) \\ (n_1, \ldots, n_r) \end{cases}, \text{ con } \sum_{j=1}^{r} n_j = n$$

- Apart from the Y_i's we will need another single latent U

The posterior distribution

- Using the hierarchical representation of the perturbed mixture model, we will characterize the un-normalized weighted IAP, say
$$A_k(\mathrm{d}y) := \bar{k}(y)A(\mathrm{d}y)$$
- Due to the discreteness of $A_k(\mathrm{d}y) \Rightarrow$ ties in the Y_i's. Therefore
$$(Y_1,\ldots,Y_n) \rightarrow \begin{cases} (Y_1^*,\ldots,Y_r^*) \\ (n_1,\ldots,n_r) \end{cases}, \text{ con } \sum_{j=1}^{r} n_j = n$$
- Apart from the Y_i's we will need another single latent U
- Characterize the posterior law through conditional distributions
$[A_k(\cdot)|X,U,Y]$, $[U|X,Y,A_k]$ and $[Y|X,U,A_k]$

The posterior distribution

- $[A_k|X, U, Y] = [A_k|U, Y]$ is another weighted IAP, i.e.,

The posterior distribution

- $[A_k|X, U, Y] = [A_k|U, Y]$ is another weighted IAP, i.e.,

$$A_k(\mathrm{d}s)|U, Y = \bar{k}(s)A^{u,y}(\mathrm{d}s)$$

where

$$A^{u,y}(\mathrm{d}s) = A^u_c(\mathrm{d}s) + \sum_{j=1}^{r} J_j^{u,y}\delta_{y_j^*}(\mathrm{d}s),$$

The posterior distribution

- $[A_k|X, U, Y] = [A_k|U, Y]$ is another weighted IAP, i.e.,

$$A_k(\mathrm{d}s)|U, Y = \bar{k}(s)A^{u,y}(\mathrm{d}s)$$

where

$$A^{u,y}(\mathrm{d}s) = A_c^u(\mathrm{d}s) + \sum_{j=1}^{r} J_j^{u,y}\delta_{y_j^*}(\mathrm{d}s),$$

  - $A_c^u(\mathrm{d}s)$ is characterized by $\nu_u(\mathrm{d}s, \mathrm{d}v) = e^{-u\bar{k}(s)v}\nu(\mathrm{d}s, \mathrm{d}v)$

The posterior distribution

- $[A_k|X, U, Y] = [A_k|U, Y]$ is another weighted IAP, i.e.,

$$A_k(\mathrm{d}s)|U, Y = \bar{k}(s)A^{u,y}(\mathrm{d}s)$$

where

$$A^{u,y}(\mathrm{d}s) = A^u_c(\mathrm{d}s) + \sum_{j=1}^{r} J^{u,y}_j \delta_{y^*_j}(\mathrm{d}s),$$

  - $A^u_c(\mathrm{d}s)$ is characterized by $\nu_u(\mathrm{d}s, \mathrm{d}v) = e^{-u\bar{k}(s)v}\nu(\mathrm{d}s, \mathrm{d}v)$
  - y^*_j, $j = 1, \ldots, r$ are fixed points of discontinuity

The posterior distribution

- $[A_k|X, U, Y] = [A_k|U, Y]$ is another weighted IAP, i.e.,

$$A_k(\mathrm{d}s)|U, Y = \bar{k}(s)A^{u,y}(\mathrm{d}s)$$

where

$$A^{u,y}(\mathrm{d}s) = A_c^u(\mathrm{d}s) + \sum_{j=1}^{r} J_j^{u,y}\delta_{y_j^*}(\mathrm{d}s),$$

  - $A_c^u(\mathrm{d}s)$ is characterized by $\nu_u(\mathrm{d}s, \mathrm{d}v) = e^{-u\bar{k}(s)v}\nu(\mathrm{d}s, \mathrm{d}v)$
  - y_j^*, $j = 1, \ldots, r$ are fixed points of discontinuity
  - $J_j^{u,y}$ are abs. continuous with density $f_{y_j^*}(v) \propto v^{n_j}e^{-u\bar{k}(y_j^*)v}\rho(\mathrm{d}v|y_j^*)$

The posterior distribution

- $[A_k|X, U, Y] = [A_k|U, Y]$ is another weighted IAP, i.e.,

$$A_k(\mathrm{d}s)|U, Y = \bar{k}(s)A^{u,y}(\mathrm{d}s)$$

  where

$$A^{u,y}(\mathrm{d}s) = A_c^u(\mathrm{d}s) + \sum_{j=1}^{r} J_j^{u,y}\delta_{y_j^*}(\mathrm{d}s),$$

  - $A_c^u(\mathrm{d}s)$ is characterized by $\nu_u(\mathrm{d}s, \mathrm{d}v) = e^{-u\bar{k}(s)v}\nu(\mathrm{d}s, \mathrm{d}v)$
  - y_j^*, $j = 1, \ldots, r$ are fixed points of discontinuity
  - $J_j^{u,y}$ are abs. continuous with density $f_{y_j^*}(v) \propto v^{n_j}e^{-u\bar{k}(y_j^*)v}\rho(\mathrm{d}v|y_j^*)$
- This result can also be obtained as an extension of James et al. (2005)

The posterior distribution

- $[U|X, Y, A_k] = [U|Y]$ is a univariate continuous density

The posterior distribution

- $[U|X, Y, A_k] = [U|Y]$ is a univariate continuous density

$$f_{U_n|Y}(u) \propto u^{n-1} \exp\{-\psi_{\bar{k}}(u)\} \prod_{j=1}^{r} \tau_{n_j}(u|y_j^*),$$

where $\psi_{\bar{k}}(u) = \int_{\mathbb{R}\times\mathbb{R}^+} (1 - e^{-u\bar{k}(s)v})\rho(\mathrm{d}v|s)\alpha(\mathrm{d}s)$
and $\tau_{n_j}(u|y_j^*) = \int_0^\infty v^{n_j} e^{-u\bar{k}(y_j^*)v} \rho(\mathrm{d}v|y_j^*)$.

The posterior distribution

- $[U|X,Y,A_k] = [U|Y]$ is a univariate continuous density

$$f_{U_n|Y}(u) \propto u^{n-1}\exp\{-\psi_{\bar{k}}(u)\}\prod_{j=1}^{r}\tau_{n_j}(u|y_j^*),$$

where $\psi_{\bar{k}}(u) = \int_{\mathbb{R}\times\mathbb{R}^+}(1-e^{-u\bar{k}(s)v})\rho(\mathrm{d}v|s)\alpha(\mathrm{d}s)$
and $\tau_{n_j}(u|y_j^*) = \int_0^\infty v^{n_j}e^{-u\bar{k}(y_j^*)v}\rho(\mathrm{d}v|y_j^*)$.

- $[Y|X,U,A_k] = [Y|X,A_k]$ are n univariate discrete densities

$$f(y_i|x_i,A) \propto k(x_i,y_i)\bar{k}(y_i)A\{\mathrm{d}y_i\}$$

for $i = 1,\ldots,n$

Posterior simulation

The algorithm for simulating from the previous conditionals is:

Posterior simulation

The algorithm for simulating from the previous conditionals is:

1. Sample the IAP $A_k(y)$ using Ferguson and Klass (1972)' algorithm
$$A_k(y) = \sum_i \bar{k}(\tau_i) J_i I(\tau_i \leq y),$$
where $J_i \in \{J_i^c \cup J_j^{u,y^*}\}$ and $\tau_i \in \{\tau_i^c \cup y_j^*\}$

Posterior simulation

The algorithm for simulating from the previous conditionals is:

1. Sample the IAP $A_k(y)$ using Ferguson and Klass (1972)' algorithm
$$A_k(y) = \sum_i \bar{k}(\tau_i) J_i I(\tau_i \leq y),$$
where $J_i \in \{J_i^c \cup J_j^{u,y^*}\}$ and $\tau_i \in \{\tau_i^c \cup y_j^*\}$
2. Choose the different Y_i's, label them as Y_j^* and record their frequency n_i. For each Y_j^* sample a jump height $J_j^{u,y*}$ according.

Posterior simulation

The algorithm for simulating from the previous conditionals is:

1. Sample the IAP $A_k(y)$ using Ferguson and Klass (1972)' algorithm
$$A_k(y) = \sum_i \bar{k}(\tau_i) J_i I(\tau_i \leq y),$$
where $J_i \in \{J_i^c \cup J_j^{u,y^*}\}$ and $\tau_i \in \{\tau_i^c \cup y_j^*\}$
2. Choose the different Y_i's, label them as Y_j^* and record their frequency n_i. For each Y_j^* sample a jump height $J_j^{u,y*}$ according.
3. Sample a U from its absolutely continuous conditional density using a M-H step.

Posterior simulation

The algorithm for simulating from the previous conditionals is:

1. Sample the IAP $A_k(y)$ using Ferguson and Klass (1972)' algorithm
$$A_k(y) = \sum_i \bar{k}(\tau_i) J_i I(\tau_i \leq y),$$
where $J_i \in \{J_i^c \cup J_j^{u,y^*}\}$ and $\tau_i \in \{\tau_i^c \cup y_j^*\}$
2. Choose the different Y_i's, label them as Y_j^* and record their frequency n_i. For each Y_j^* sample a jump height $J_j^{u,y*}$ according.
3. Sample a U from its absolutely continuous conditional density using a M-H step.
4. Sample a set of n Y_i's, one for each $i = 1, \ldots, n$, from its discrete conditional density.

Posterior simulation

The algorithm for simulating from the previous conditionals is:

1. Sample the IAP $A_k(y)$ using Ferguson and Klass (1972)' algorithm
$$A_k(y) = \sum_i \bar{k}(\tau_i) J_i I(\tau_i \leq y),$$
where $J_i \in \{J_i^c \cup J_j^{u,y^*}\}$ and $\tau_i \in \{\tau_i^c \cup y_j^*\}$
2. Choose the different Y_i's, label them as Y_j^* and record their frequency n_i. For each Y_j^* sample a jump height $J_j^{u,y*}$ according.
3. Sample a U from its absolutely continuous conditional density using a M-H step.
4. Sample a set of n Y_i's, one for each $i = 1, \ldots, n$, from its discrete conditional density.
5. Compute a path of the random density

Setup for the sensitivity analysis

Setup for the sensitivity analysis

- Mixture model: $f(x) = \int k(x, y) \mathrm{P}(\mathrm{d}y)$

Setup for the sensitivity analysis

- Mixture model: $f(x) = \int k(x, y)\mathrm{P}(\mathrm{d}y)$
- The parametric kernel: the typical choice

Setup for the sensitivity analysis

- Mixture model: $f(x) = \int k(x, y) \mathrm{P}(\mathrm{d}y)$
- The parametric kernel: the typical choice

$$k(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x - y)^2\right\}$$

Setup for the sensitivity analysis

- Mixture model: $f(x) = \int k(x, y) \mathrm{P}(\mathrm{d}y)$
- The parametric kernel: the typical choice

$$k(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-y)^2\right\}$$

- The mixing distribution $\mathrm{P}(\mathrm{d}y)$: perturbed NRMI

Setup for the sensitivity analysis

- Mixture model: $f(x) = \int k(x, y)\mathrm{P}(\mathrm{d}y)$
- The parametric kernel: the typical choice

$$k(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x - y)^2\right\}$$

- The mixing distribution $\mathrm{P}(\mathrm{d}y)$: perturbed NRMI
  - A: gamma or inverse Gaussian process with parameter $\alpha(\,\cdot\,)$;

Setup for the sensitivity analysis

- Mixture model: $f(x) = \int k(x,y)\mathrm{P}(\mathrm{d}y)$
- The parametric kernel: the typical choice

$$k(x,y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-y)^2\right\}$$

- The mixing distribution $\mathrm{P}(\mathrm{d}y)$: perturbed NRMI
  - A: gamma or inverse Gaussian process with parameter $\alpha(\,\cdot\,)$;
    - $\alpha(\mathrm{d}y) = a\,P_0(\mathrm{d}y) = a\,\mathrm{Ga}(y|p,q)\,\mathrm{d}y$

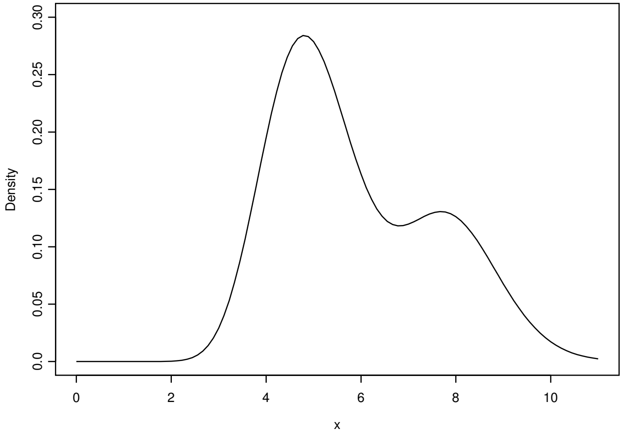
Setup for the sensitivity analysis

- Mixture model: $f(x) = \int k(x,y)\mathrm{P}(\mathrm{d}y)$
- The parametric kernel: the typical choice

$$k(x,y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(x-y)^2\right\}$$

- The mixing distribution $\mathrm{P}(\mathrm{d}y)$: perturbed NRMI
  - A: gamma or inverse Gaussian process with parameter $\alpha(\,\cdot\,)$;
    - $\alpha(\mathrm{d}y) = a\,P_0(\mathrm{d}y) = a\,\mathrm{Ga}(y|p,q)\,\mathrm{d}y$
  - $$\bar{k}(y) = \begin{cases} \kappa & \text{if } y \in B \subset \mathbb{R}^+ \\ 1 & \text{otherwise} \end{cases}$$

A simulated data example



A simulated data example

- Uneven mixture: one dominating group

A simulated data example

- Uneven mixture: one dominating group

$$f(x) = 0.7\,\mathrm{Ga}(x|25,5) + 0.3\,\mathrm{Ga}(x|64,8)$$

A simulated data example

- Uneven mixture: one dominating group

$$f(x) = 0.7\,\text{Ga}(x|25, 5) + 0.3\,\text{Ga}(x|64, 8)$$

- We simulated $n = 30$ obs.

A simulated data example

- Uneven mixture: one dominating group

$$f(x) = 0.7\,\mathrm{Ga}(x|25,5) + 0.3\,\mathrm{Ga}(x|64,8)$$

- We simulated $n = 30$ obs.
- We fixed $\sigma = 1$ (=variance of each component)

A simulated data example

- Uneven mixture: one dominating group

$$f(x) = 0.7\,\mathrm{Ga}(x|25,5) + 0.3\,\mathrm{Ga}(x|64,8)$$

- We simulated $n = 30$ obs.
- We fixed $\sigma = 1$ (=variance of each component)
- The IAP A is a gamma process with $\alpha(\mathrm{d}y) = \mathrm{Ga}(y|3,0.5)\mathrm{d}y$

A simulated data example

- Uneven mixture: one dominating group

$$f(x) = 0.7\,\text{Ga}(x|25,5) + 0.3\,\text{Ga}(x|64,8)$$

- We simulated $n = 30$ obs.
- We fixed $\sigma = 1$ (=variance of each component)
- The IAP A is a gamma process with $\alpha(\text{d}y) = \text{Ga}(y|3,0.5)\text{d}y$
- The perturbing function is

$$\bar{k}(y) = \begin{cases} 10 & \text{if } y \in [0,3] \\ 1 & \text{if } y > 3 \end{cases}$$

A simulated data example

- Implications on the mixing distribution: P

A simulated data example

- Implications on the mixing distribution: P
- The prior mean of the Dirichlet P is

$$\mathrm{P}_0(\mathrm{d}y) = E[\mathrm{P}(\mathrm{d}y)] = \mathrm{Ga}(y|3, 0.5)\mathrm{d}y$$

A simulated data example

- Implications on the mixing distribution: P
- The prior mean of the Dirichlet P is

$$P_0(dy) = E[P(dy)] = \text{Ga}(y|3, 0.5)dy$$

- The prior mean of the perturbed Dirichlet P is

$$P_0(dy) = E[P(dy)] = 2.07\text{Ga}(y|3, 0.5)\mathbb{I}_{[0,3)}(dy) + 0.75\text{Ga}(y|3, 0.5)\mathbb{I}_{[3,\infty)}(dy)$$

A simulated data example

- Implications on the mixing distribution: P
- The prior mean of the Dirichlet P is

$$\mathrm{P}_0(\mathrm{d}y) = E[\mathrm{P}(\mathrm{d}y)] = \mathrm{Ga}(y|3, 0.5)\mathrm{d}y$$

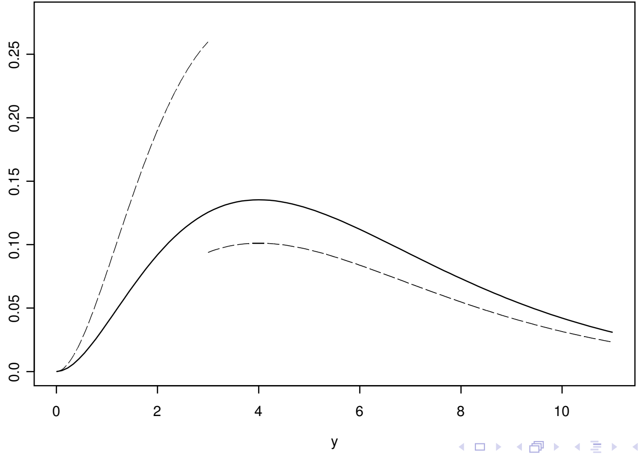
- The prior mean of the perturbed Dirichlet P is

$$\mathrm{P}_0(\mathrm{d}y) = E[\mathrm{P}(\mathrm{d}y)] = 2.07\mathrm{Ga}(y|3, 0.5)\mathbb{I}_{[0,3)}(\mathrm{d}y) + 0.75\mathrm{Ga}(y|3, 0.5)\mathbb{I}_{[3,\infty)}(\mathrm{d}y)$$

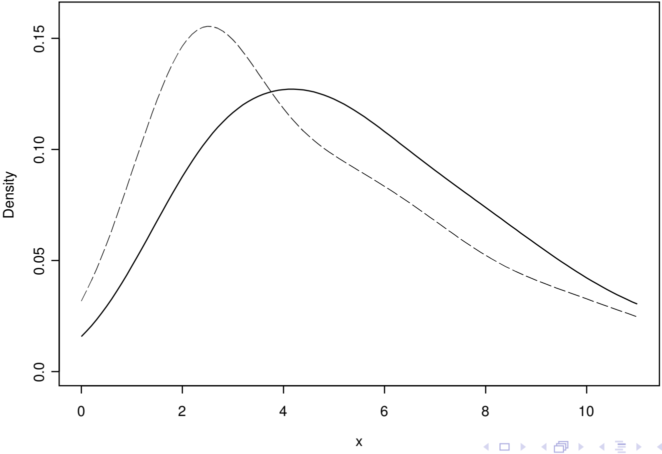
- In any case, the prior density estimates are:

$$\hat{f}(x) = E[f(x)] = \int_{\mathbb{R}^+} (\sqrt{2\pi})^{-1} \mathrm{e}^{-\frac{1}{2}(x-y)^2} \times \mathrm{P}_0(\mathrm{d}y)$$

A simulated data example



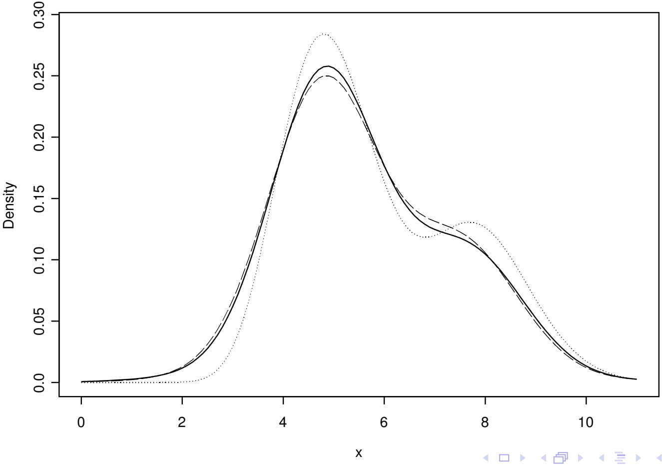
A simulated data example



A simulated data example

- What will be the influence of perturbing the prior and force it to select more means (locations) in the interval $[0, 3]$??

A simulated data example



A simulated data example

- Amazing!!, isn't it?

A simulated data example

- Amazing!!, isn't it?
- Two important facts:

A simulated data example

- Amazing!!, isn't it?
- Two important facts:
  1. Given the small sample size ($n = 30$) the fitting is very good

A simulated data example

- Amazing!!, isn't it?
- Two important facts:
  1. Given the small sample size ($n = 30$) the fitting is very good
  2. The perturbed and un-perturbed estimates are almost identical

A simulated data example

- Amazing!!, isn't it?
- Two important facts:
  1. Given the small sample size ($n = 30$) the fitting is very good
  2. The perturbed and un-perturbed estimates are almost identical
- The big question is:

A simulated data example

- Amazing!!, isn't it?
- Two important facts:
  1. Given the small sample size ($n = 30$) the fitting is very good
  2. The perturbed and un-perturbed estimates are almost identical
- The big question is: Why?

A simulated data example

- Focusing attention to the set $[0, 3]$, a-priori:

A simulated data example

- Focusing attention to the set $[0,3]$, a-priori:
  - For the Dirichlet case: $P_0([0,3]) \cong 0.19$

A simulated data example

- Focusing attention to the set $[0,3]$, a-priori:
  - For the Dirichlet case: $P_0([0,3]) \cong 0.19$
  - For the perturbed-Dirichlet: $P_0([0,3]) \cong 0.4$

A simulated data example

- Focusing attention to the set $[0, 3]$, a-priori:
  - For the Dirichlet case: $P_0([0, 3]) \cong 0.19$
  - For the perturbed-Dirichlet: $P_0([0, 3]) \cong 0.4$
- How is this difference going to be neutralized a-posteriori?

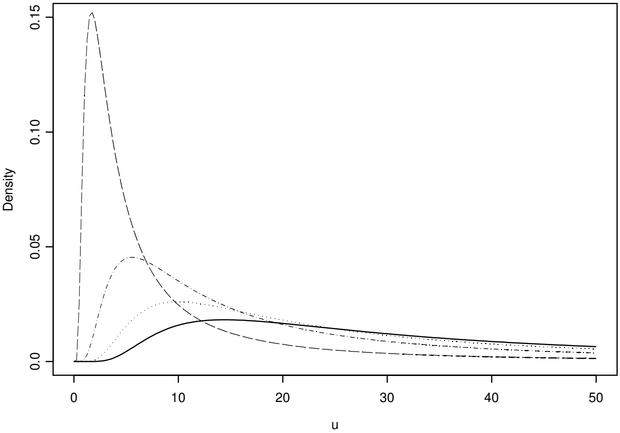
A simulated data example

- Focusing attention to the set $[0,3]$, a-priori:
  - For the Dirichlet case: $P_0([0,3]) \cong 0.19$
  - For the perturbed-Dirichlet: $P_0([0,3]) \cong 0.4$
- How is this difference going to be neutralized a-posteriori?
- The answer depends on the latent U and the configuration of the latents $Y^{(30)}$.

A simulated data example

- Focusing attention to the set $[0,3]$, a-priori:
  - For the Dirichlet case: $P_0([0,3]) \cong 0.19$
  - For the perturbed-Dirichlet: $P_0([0,3]) \cong 0.4$
- How is this difference going to be neutralized a-posteriori?
- The answer depends on the latent U and the configuration of the latents $Y^{(30)}$.
- Consider four cases: 0, 10, 20 and 30 latents Y's fall in $[0,3]$ $\Rightarrow$ a different shape of the density of $U|Y^{(30)}$ in the perturbed-Dirichlet

A simulated data example



A simulated data example

- The distribution of $U|Y^{(30)} \Rightarrow$ infinite mean. So we take the medians as average representatives.

A simulated data example

- The distribution of $U|Y^{(30)} \Rightarrow$ infinite mean. So we take the medians as average representatives.
- We computed the posterior probability of selecting a mean (location) in $[0,3]$ conditional on $U|Y^{(30)}$ and $Y^{(30)}$

$$\mathrm{E}\left[P\left([0,3]\right)\middle| U;\, Y^{(30)}\right] = \mathrm{E}\left[\frac{\int_{[0,3]} \bar{k}(y)A^{u,y^*}(\mathrm{d}y) + \sum_{y_i^* \in [0,3]} k(y_i^*)J_i^{u,y^*}}{\int_{\mathbb{R}^+} \bar{k}(y)A^{u,y^*}(\mathrm{d}y) + \sum_{y_i^* \in \mathbb{R}^+} k(y_i^*)J_i^{u,y^*}}\right]$$

for the Dirichlet and perturbed-Dirichlet

A simulated data example

# latents in $[0, 3]$	Dirichlet (Indep. of U)	Perturbed Dirichlet $u = 1$	$u = 20$	$u = 50$	(Median of U)
30	0.974	0.984	0.970	0.968	(5.10) 0.971
20	0.651	0.767	0.655	0.648	(16.9) 0.657
10	0.328	0.465	0.335	0.325	(29.7) 0.329
0	0.006	0.011	0.005	0.003	(42.6) 0.004

A simulated data example

# latents in [0, 3]	Dirichlet (Indep. of U)	Perturbed Dirichlet $u=1$	$u=20$	$u=50$	(Median of U)
30	0.974	0.984	0.970	0.968	(5.10) 0.971
20	0.651	0.767	0.655	0.648	(16.9) 0.657
10	0.328	0.465	0.335	0.325	(29.7) 0.329
0	0.006	0.011	0.005	0.003	(42.6) 0.004

- What we observed from the Gibbs sampler for the distribution of $U|Y^{(30)}$ was a median of 28.5 $\Rightarrow$ approximately 10 latents fall in $[0,3]$

Galaxy data

Galaxy data

- Roeder (1990): Relative velocities of 82 galaxies from 6 well-separated conic sections of space

Galaxy data

- Roeder (1990): Relative velocities of 82 galaxies from 6 well-separated conic sections of space
- For the kernel: $k(x, y) = N(x|y, \sigma^2)$

Galaxy data

- Roeder (1990): Relative velocities of 82 galaxies from 6 well-separated conic sections of space
- For the kernel: $k(x,y) = N(x|y,\sigma^2)$
- For P: N-IG and perturbed N-IG with $\alpha(\mathrm{d}y) = 0.1\,\mathrm{Ga}(y|1,0.01)\mathrm{d}y$ and

$$\bar{k}(y) = \begin{cases} 20 & \text{if } y \in (11,16] \\ 1 & \text{if } y \in [0,11] \cup (16,\infty) \end{cases}$$

Galaxy data

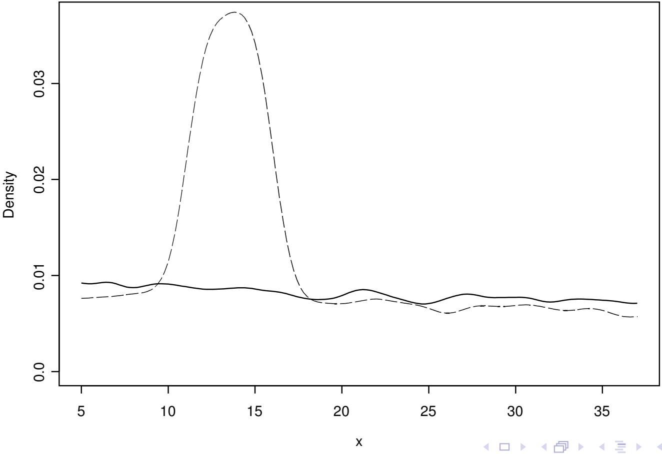
- Roeder (1990): Relative velocities of 82 galaxies from 6 well-separated conic sections of space
- For the kernel: $k(x,y) = N(x|y,\sigma^2)$
- For P: N-IG and perturbed N-IG with $\alpha(\mathrm{d}y) = 0.1\,\mathrm{Ga}(y|1,0.01)\mathrm{d}y$ and

$$\bar{k}(y) = \begin{cases} 20 & \text{if } y \in (11,16] \\ 1 & \text{if } y \in [0,11] \cup (16,\infty) \end{cases}$$

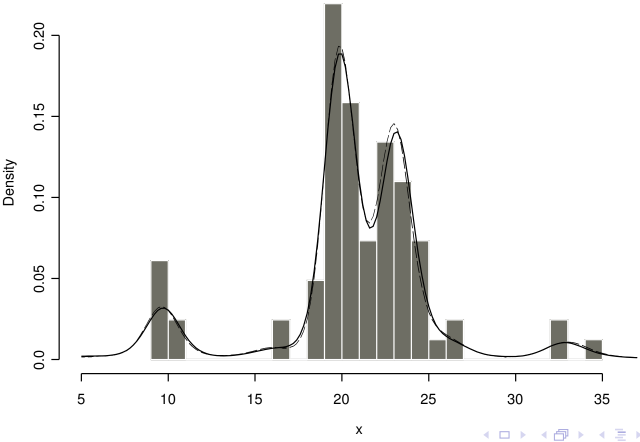
- $+$ a randomization of σ as in Richardson and Green (1997): $\sigma \sim \mathrm{Ga}(1,2) \Rightarrow$ extend the Gibbs sampler to include

$$\pi(\sigma|x,u,y,A_k) \propto \pi(\sigma)\prod_{i=1}^{n} k(x_i|y_i,\sigma)$$

Galaxy data



Galaxy data



Acidity data

Acidity data

- Richardson and Green (1997): Measurements of an acid neutralizing capacity index in a sample of 155 lakes in North-Central Wisconsin (in a log-scale).

Acidity data

- Richardson and Green (1997): Measurements of an acid neutralizing capacity index in a sample of 155 lakes in North-Central Wisconsin (in a log-scale).
- For the kernel: $k(x,y) = N(x|y,\sigma^2)$

Acidity data

- Richardson and Green (1997): Measurements of an acid neutralizing capacity index in a sample of 155 lakes in North-Central Wisconsin (in a log-scale).
- For the kernel: $k(x, y) = N(x|y, \sigma^2)$
- For P: Dirichlet and perturbed Dirichlet with $\alpha(\mathrm{d}y) = \mathrm{Ga}(y|5, 1)\mathrm{d}y$ and

$$\bar{k}(y) = \begin{cases} 0.1 & \text{if } y \in (4, 5] \\ 1 & \text{if } y \in [0, 4] \cup (5, \infty) \end{cases}$$

Note that the perturbing function squeezes by a factor of 10 the jumps of the process in $(4, 5]$

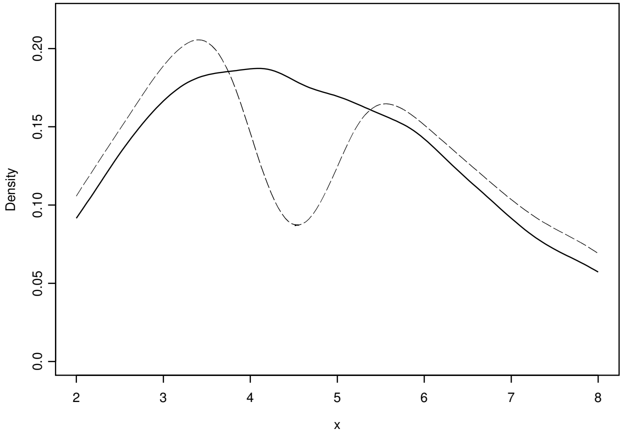
Acidity data

- Richardson and Green (1997): Measurements of an acid neutralizing capacity index in a sample of 155 lakes in North-Central Wisconsin (in a log-scale).
- For the kernel: $k(x, y) = N(x|y, \sigma^2)$
- For P: Dirichlet and perturbed Dirichlet with $\alpha(\mathrm{d}y) = \mathrm{Ga}(y|5, 1)\mathrm{d}y$ and

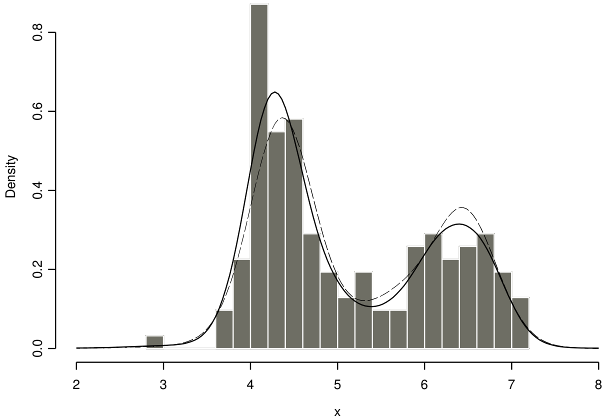
$$\bar{k}(y) = \begin{cases} 0.1 & \text{if } y \in (4, 5] \\ 1 & \text{if } y \in [0, 4] \cup (5, \infty) \end{cases}$$

  Note that the perturbing function squeezes by a factor of 10 the jumps of the process in $(4, 5]$
- $+\ \sigma \sim \mathrm{Ga}(1, 10)$

Acidity data



Acidity data



Summary

Summary

- If you want to do density estimation ...

Summary

- If you want to do density estimation ...
- Mixtures of NRMI are the option ...

Summary

- If you want to do density estimation ...
- Mixtures of NRMI are the option ...
- They are not sensitive and are quite robust