# Bayesian inference for nonparametric mixture models with intractable normalizing constants

Isadora Antoniano-Villalobos [1]    Stephen Walker [2]

[1]Department of Decision Sciences
Bocconi University
Milan, Italy
isadora.antoniano@unibocconi.it
[2]School of Mathematis, Statistics and Actuarial Science
University of Kent
Canterbury, UK

10 May, 2013

# Parametric model with intractable posterior

Consider a Bayesian parametric model

$$Y_i|\theta \overset{iid}{\sim} f(\cdot|\theta)$$
$$\theta \sim \Pi,$$

where $\Pi$ is a prior distribution over the parameter space $\Theta$. The posterior density, given a sample $y_{1:n} = (y_1, \ldots, y_n)$ is

$$\Pi^n(\theta) = \frac{\Pi(\theta)\prod_{i=1}^n f(y_i|\theta)}{\int_\Theta \Pi(\theta)\prod_{i=1}^n f(y_i|\theta)\mathrm{d}\nu(\theta)}.$$

In general, the normalizing constant for this posterior is intractable, the distribution is known but inference is possible via standard MCMC methods (e.g. Gibbs sampler and Metropolis-Hastings)

$$\Pi^n(\theta) \propto \Pi(\theta)\prod_{i=1}^n f(y_i|\theta)$$

# Parametric model with doubly-intractable posterior

Assume the likelihood has an intractable normalizing constant

$$f(y|\theta) = \frac{g(y, \theta)}{Z(\theta)},$$

where

$$Z(\theta) = \int_{\mathbb{Y}} g(y, \theta) \mathrm{d}\nu(y)$$

The resulting posterior is sometimes called a doubly-intractable distribution

$$\Pi^n(\theta) = \frac{\Pi(\theta) f(y|\theta)}{f(y)} = \frac{\Pi(\theta) g(y, \theta)/Z(\theta)}{\int_\Theta \left[\Pi(\theta) g(y, \theta)/Z(\theta)\right] \mathrm{d}\nu(\theta)}.$$

In this case, more elaborate MCMC schemes are needed (e.g. auxiliary variable schemes, importance samplers)

# Bayesian nonparametrics and DP

A Bayesian nonparametric model defines a prior on a space $\mathcal{F}$ of densities which is too large to be indexed by a finite-dimensional parameter.
One way to achieve this is by mixing over some parametric family $K(\cdot|\theta)$, $\theta \in \Theta$

$$Y_i|P \overset{iid}{\sim} f_P(\cdot),$$
$$P \sim \Pi.$$

where

$$f_P(\cdot) = \int_\Theta K(\cdot|\theta) \mathrm{d}P(\theta).$$

The nonparametric prior $\Pi$ is assigned to the mixing distribution $P$.

# Nonparametric mixtures (MSBP and MDP)

Typically, $P$ is assigned a Stick-breking prior, i.e.

$$P = \sum_{j=1}^{\infty} w_j \delta_{\theta_j};$$

$$w_1 = v_1 \quad \text{and} \quad w_j = v_j \prod_{j'<j} (1 - v_{j'}), \forall\, j > 1;$$

$$v_j \overset{iid}{\sim} \text{Be}(\cdot|\alpha_j, \zeta_j);$$

$$\theta_j \overset{iid}{\sim} P_0.$$

In this case, (Lo, 1984; Sethuraman, 1994)

$$f_P(\cdot) = \sum_{j=1}^{\infty} w_j K(\cdot|\theta_j).$$

The Mixture of Dirichlet Process model (MDP) is recovered when $\alpha_j \equiv 1$ and $\zeta_j \equiv \zeta$

# Inference for MDP models

- Marginal methods: The infinite dimensional component (the random mixing distribution) is integrated out. (Escobar, 1988; MacEachern and Müller, 1998; Neal, 2000).

- Conditional methods: The mixing measure is not integrated out. so that posterior simulation is carried out for all the variables involved, i.e. the $w_{1:\infty}$ and the $\theta_{1:\infty}$.
  - Approximate methods: A fixed number $N < \infty$ is considered (Ishwaran and Zarepour, 2000; Ishwaran and James, 2001).
  - Exact methods: No truncation is used.
    - Retrospective sampler: Papaspiliopoulos and Roberts (2008)
    - Slice sampler: Kalli et al. (2011)

# Slice Sampler for MSBP models

The intractability of the likelihood

$$f_P(y) = \sum_{j=1}^{\infty} w_j K(y|\theta_j)$$

is resolved through an augmentation scheme

$$f_P(y, u, d) = w_d K(y|\theta_d) e^{\xi d} \mathbf{1}\left\{u < e^{-\xi d}\right\}.$$

Conditional on all other variables, $d$ can only take values on $\{1, 2, \ldots, J\}$, where

$$J = \lfloor -\xi^{-1} \log u \rfloor$$

.

# Inference for the power likelihood

In some situations, it is convenient to perform inference using a power likelihood

$$\Pi^n(\theta) \propto \Pi(P) \prod_{i=1}^{n} f_P^{1-\alpha}(y_i) = \Pi(P) \prod_{i=1}^{n} \frac{f_P(y_i)}{f_P^{\alpha}(y_i)}$$

This can be interpreted as approximate inference and it can be used to satisfy properties of the estimators, such as consistency (Antoniano-Villalobos and Walker, 2012) or smoothness (Ibrahim and Chen, 2000), and is also common in the context of simulated annealing (Friel and Pettitt, 2008).

# Inference for the power likelihood

When a mixture model is used, the use of a power likelihood leads to a doubly intractable distribution, since

$$\frac{f_P(y)}{f_P^{\alpha}(y)} = \frac{\sum_{j=1}^{\infty} w_j K(y|\theta_j)}{\left[\sum_{j=1}^{\infty} w_j K(y|\theta_j)\right]^{\alpha}}$$

Available algorithms for posterior simulation from mixture models are not useful in this case.

# Nonparametric regression model

The standard linear regression model

$$y = \beta X + \epsilon; \quad \epsilon \sim \mathrm{N}(\epsilon|0, \sigma^2),$$

is too limited to capture the relation between a response variable $y \in \mathbb{Y}$ and some covariate $x \in \mathbb{X}$, in most applications. More flexible models can be constructed through modification of the link function and/ or the error distribution

$$y = m(x, \theta) + \epsilon; \quad \epsilon \sim f_y(\epsilon|y, \theta),$$

A fully nonparametric approach can be adopted in terms of conditional density estimation

$$y|x \sim f_x(y) = f(y|x),$$

# Nonparametric regression model

A popular nonparametric model is the dependent Dirichlet process Mixture (DDP) (MacEachern, 1999, 2000)

$$f_{P_x}(y|x) = \sum_{j=1}^{\infty} w_j(x) K\big(y|x, \theta_j(x)\big);$$

$$P_x = \sum_{j=1}^{\infty} w_j(x) \delta_{\theta_j(x)};$$

$$w_1(x) = v_1(x), \quad \text{and} \quad w_j(x) = v_j(x) \prod_{j'<j} (1 - v_{j'}(x)),\ j > 1,$$

the $\{v_j(x)\}_{x \in \mathbb{X}}$ are independent processes such that, marginally

$$v_j(x) \overset{ind}{\sim} \text{Be}(1, \zeta(x)) \text{ for } j = 1, 2, \dots,$$

and the $\{\theta_j(x)\}_{x \in \mathbb{X}}$ are independent stochastic processes with marginal distribution $P_{0x}$, and independent of the $v_j(x)$.

# Nonparametric regression model

Even when $\theta_j(x) \equiv \theta_j \stackrel{iid}{\sim} P_0$ (the single particle DDP), definition of the $\{v_j(x)\}_{x \in \mathbb{X}}$ is not trivial. Further difficulty arises from multivariate covariates of different types and different ideas have been proposed (e.g. Griffin and Steel, 2006; Dunson and Park, 2008; Dunson and Rodríguez, 2011; Teh et al., 2006). An alternative, is to define, normalized weights, i.e.

$$w_j(x) = \frac{w_j K(x|\phi_j)}{\sum_{j=1}^{\infty} w_{j'} K(x|\phi_{j'})}.$$

Leading again to a nonparametric likelihood with an intractable normalizing constant

$$f_{P_x}(y|x) = \frac{\sum_{j=1}^{\infty} w_j K(x|\phi_j) K\left(y|x, \theta_j\right)}{\sum_{j=1}^{\infty} w_{j'} K(x|\phi_{j'})}.$$

# Stationary autoregressive model

Consider a parametric first order stationary autoregressive model, AR(1), with transition and marginal densities

$$K(y_{n+1}|y_n,\theta);$$
$$K(y_n|\theta)$$

Clearly, there is a bivariate density

$$K(y_{n+1},y_n|\theta)=K(y_{n+1}|y_n,\theta)K(y_n|\theta)$$

such that

$$K(y_{n+1}|\theta)=\int K(y_{n+1},y_n|\theta)\mathrm{d}y_{n+1};\quad K(y_n|\theta)=\int K(y_{n+1},y_n|\theta)\mathrm{d}y_n$$

This model is limited, but can once again be used as the basis for a nonparametric mixture construction.

# Stationary autoregressive model

A non parametric transition can be defined as a mixture of parametric transition kernels (Müller et al., 1997; Tang and Ghosal, 2007)

$$f_{P_{y_n}}(y_{n+1}|y_n) = \sum_{j=1}^{\infty} w_j(y_n) K\big(y_{n+1}|y_n, \theta_j(y_n)\big).$$

However, control over the statistical properties of the process is lost. In particular, stationarity is not satisfied in general.

Alternative aproaches, produce models which are not fully nonparametric (Mena and Walker, 2005), or which are too complex for practical application (Martínez-Ovando and Walker, 2011)

# Stationary autoregressive model

We propose to define a nonparametric mixture over the bivariate parametric kernels

$$f_P(y_{n+1}, y_n) = \sum_{j=1}^{\infty} w_j K\left(y_{n+1}, y_n | \theta_j\right).$$

The stationary density is given by the marginal

$$f_P(y_n) = \sum_{j=1}^{\infty} w_j K\left(y_n | \theta_j\right).$$

and the transition density is simply the conditional

$$f_P(y_{n+1} | y_n) = \frac{\sum_{j=1}^{\infty} w_j K\left(y_{n+1}, y_n | \theta_j\right)}{\sum_{j=1}^{\infty} w_j K\left(y_n | \theta_j\right)}.$$

# Series expansion of the intractable component

For $0 < c < 1$, we can find a positive sequence $\{b_k(\alpha)\}$ such that

$$c^{-\alpha} = \sum_{k=0}^{\infty} b_k(\alpha)(1-c)^k.$$

If $0 < K(y|\theta) < 1$ for all $y$ and $\theta$, we may rewrite

$$\frac{1}{\left[\sum_{j=1}^{\infty} w_j K\left(y|\theta_j\right)\right]^{\alpha}} = \sum_{k=0}^{\infty} b_k(\alpha) \left[1 - \sum_{j=1}^{\infty} w_j K\left(y|\theta_j\right)\right]^k$$

In fact, $b_0(\alpha) = 1$, $b_1(\alpha) = \alpha$ and for $k > 1$,

$$c_k(\alpha) = \frac{\alpha^{(k)}}{k!} = \frac{\alpha(\alpha+1)\dots(\alpha+k-1)}{k!}.$$

And $b_k(1) \equiv 1$.

# Nonparametric stationary normal AR(1) mixture

To make this concrete, we will adopt a particular model based on the normal distribution.

$$K(y_{n+1}, y_n|\theta) = \mathrm{N}_2\Big((y_{n+1}, y_n)|(\mu, \mu), \Sigma\Big)$$

where

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

for some $-1 < \rho < 1$. So $\theta = (\mu, \sigma^2, \rho)$ and

$$K(y_{n+1}|y_n, \theta) = \mathrm{N}\Big(y|\mu + \rho(y_n - \mu), (1 - \rho^2)\sigma^2\Big);$$

$$K(y_n|\theta) = \mathrm{N}(y_n|\mu, \sigma^2).$$

# Nonparametric stationary normal AR(1) mixture

For simplicity, we consider mixing over means only, therefore

$$
\begin{aligned}
f_P(y_{n+1}|y_n) &= \frac{\sum_{j=1}^{\infty} w_j \, \mathrm{N}((y_i, y_{i-1})|(\mu_j, \mu_j), \Sigma)}{\sum_{j=1}^{\infty} w_j \, \mathrm{N}(y_{i-1}|\mu_j, \sigma^2)} \\
&\propto \frac{\sigma \sum_{j=1}^{\infty} w_j \, \mathrm{N}(y_i, y_{i-1}|\mu_j, \sigma^2, \rho)}{\sum_{j=1}^{\infty} w_j \exp\left\{-\frac{1}{2}(y_{i-1} - \mu_j)^2 \sigma^2\right\}}
\end{aligned}
$$

which we can rewrite as

$$
\sigma \sum_{j=1}^{\infty} w_j \, \mathrm{N}(\mathrm{N}(y_i, y_{i-1}|\mu_j, \sigma^2, \rho) \sum_{k=0}^{\infty} \left[ 1 - \sum_{j=1}^{\infty} w_j \exp\left\{ -\frac{1}{2}(y_{i-1} - \mu_j)^2 \sigma^2 \right\} \right]^k
$$

# Introduction of latent variables

We are ready to introduce the first latent variable, $k$,

$$f(y_{n+1}, k|y_n, \theta) \propto \frac{1}{\sigma} \sum_{j=1}^{\infty} w_j \mathrm{N}_2(y_{n+1}, y_n|\mu_j, \sigma^2\rho)$$

$$\sum_{j=1}^{\infty} w_j \left[1 - \exp\left\{-(y_{i-1} - \mu_j)^2/2\sigma^2\right\}\right]^k.$$

We arrive to the full latent model by introducing the indices as auxiliary variables in the usual way:

$$f_P(y_{n+1}, d, k, D_{1:k}) = \sigma w_d \, \mathrm{N}_2\left(y_{n+1}, y_n|\mu_d, \sigma, \rho\right)$$

$$\prod_{l=1}^{k} w_{D_l} \left[1 - \exp\left\{-\frac{1}{2}\left(y_n - \mu_{D_l}\right)^2/\sigma^2\right\}\right],$$

# The prior

We assign a Stick-Breaking process prior to $P$; effectively, the $(w_j, \mu_j)_{j=1}^{\infty}$ and independent priors to $\rho$ and $\sigma$

- For independent and identically distributed $v_j \overset{iid}{\sim} \mathrm{Be}(\alpha, \zeta)$,

$$w_1 = v_1, \quad \text{and for } j > 1, \quad w_j = v_j \prod_{l<j} (1 - v_l).$$

- $\mu_j \overset{iid}{\sim} \mathrm{N}(\cdot | m, t^{-1})$.
- $\tau = \sigma^{-2} \sim \mathrm{Ga}(a, c)$.
- Finally, we define a discrete uniform prior for $\rho$ on some discrete set $R \subset (-1, +1)$

## Updating the Indices, $d_i$ and $D_{i,l}$

Following the slice sampling technique, we introduce additional auxiliary variables through the indicator functions

$$\mathbf{1}\left(\nu_i < e^{-\xi d_i}\right) \quad \text{and} \quad \mathbf{1}\left(\nu_{i,l} < e^{-\xi D_{i,l}}\right),$$

for some $\xi > 0$. Hence, the full conditional distributions for the latent indices are given by

$$\mathbb{P}(d_i = j|\cdots) \propto w_j\, e^{\xi j}\, K_{\theta_j}(y_i, y_{i-1})\, \mathbf{1}\{1 \leq j \leq J_i\}$$

and

$$\mathbb{P}(D_{i,l} = j|\cdots) \propto w_j\, e^{\xi j}\left[1 - \exp\left\{-\frac{1}{2}(y_{i-1} - \mu_j)^2/\sigma^2\right\}\right]\mathbf{1}\{1 \leq j \leq J_{i,l}\}$$

where

$$J_i = \lfloor -\xi^{-1} \log \nu_i \rfloor; \quad J_{i,l} = \lfloor -\xi^{-1} \log \nu_{i,l} \rfloor.$$

# Updating the Mixture Weights, $w_{1:J}$

The weights are updating via the stick breaking representation, by sampling each $v_j$ from the updated distribution

$$f(v_j | \cdots) = \text{Be}(\alpha_j + n_j + N_j, \zeta_j + n_j^+ + N_j^+),$$

where

$$n_j = \sum_{i=1}^{n} \mathbf{1}\,(d_i = j); \quad N_j = \sum_{i=1}^{n} \sum_{l=1}^{k_i} \mathbf{1}\,(D_{i,l} = j);$$
$$n_j^+ = \sum_{i=1}^{n} \mathbf{1}\,(d_i > j); \quad N_j^+ = \sum_{i=1}^{n} \sum_{l=1}^{k_i} \mathbf{1}\,(D_{i,l} > j). \tag{1}$$

# Updating the Correlation Coefficient, $\rho$

A discrete prior for the correlation coefficient $\rho$, results in a discrete full conditional distribution given by

$$\mathbb{P}(\rho = r | \cdots) \propto \pi(r)(1-\rho^2)^{-\frac{n-1}{2}} \exp\left\{ -\frac{1}{2}\tau \sum_{i=1}^{n} \hat{\mu}_i' \Sigma_r^{-1} \hat{\mu}_i \right\}$$

where

$$\hat{\mu}_i = \begin{pmatrix} y_i - \mu_{d_i} \\ y_{i-1} - \mu_{d_i} \end{pmatrix}, \quad \Sigma_r = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}. \tag{2}$$

for every $r \in R$

# Further variable augmentation

We introduce some additional latent variables $u_{1:n,1:k_i}$, which allow us to substitute the terms

$$\left[1 - \exp\left\{-\frac{1}{2}\left(y_{i-1} - \mu_{D_{i,l}}\right)^2/\sigma^2\right\}\right]$$

in the latent likelihood expression, with truncation terms

$$\mathbf{1}\left[u_{i,l} < 1 - \exp\left\{-\frac{1}{2}\left(y_{i-1} - \mu_{D_{i,l}}\right)^2/\sigma^2\right\}\right].$$

This, however, is not necessary; other updating schemes for $\tau$ and the $\mu_j$ are possible.

# Updating the Precision Term, $\tau = \sigma^{-2}$

Now, the full conditional distribution for $\tau$ is a truncated Gamma,

$$f(\tau|\cdots) \propto \mathrm{Ga}(\tau|\hat{a}, \hat{c})\,\mathbf{1}\,(\tau > T),$$

where

$$\hat{a} = a + n/2;$$

$$\hat{c} = c + \frac{1}{2}\sum_{i=1}^{n} \hat{\mu}_i' \Sigma_\rho^{-1} \hat{\mu}_i;$$

$$T = \max\left\{\frac{-2\log(1 - u_{i.l})}{(y_i - \mu_{D_{i,l}})^2} : i = 1, \ldots, n;\ l = 1, \ldots, k_i\right\};$$

# Updating the Kernel Means, $\mu_{1:J}$

For each $j$, the full conditional distribution for $\mu_j$, given the rest of the variables is a truncated Normal

$$f(\mu_j | \cdots) \propto \mathrm{N}(\mu_j | m_j, t_j^{-1}) \mathbf{1}\ \{\mu_j \in \cap_{i=1}^n A_{j,i}\}$$

where

$$m_j = \frac{1}{t_j}\left[mt + \frac{\tau}{1-\rho}\sum_{d_i=j}(y_i + y_{i-1})\right];$$

$$t_j = t + \frac{2\tau n_j}{1+\rho};$$

and the truncation is defined by the sets
$A_{j,i} = (-\infty, y_i - a_{j,i}) \cup (y_i + a_{j,i}, \infty)$,

$$a_{j,i} = \max_l \left\{\sqrt{-2\tau^{-1}\log(1-u_{i,l})} : D_{i,l} = j\right\},$$

with the convention $\max\{\emptyset\} \equiv \infty$, $\min\{\emptyset\} \equiv -\infty$.

# Updating the Latent Model Dimension, $k$

The dimension of the sampling space for the latent index variables ($D_{i,l}$) changes with $k_i$; we are dealing with transdimensional MCMC. We use a M-H scheme due to Godsill (2001) and related to Reversible Jump MCMC (Green, 1995)

- Propose a move from $k_i$ to $k_i + 1$ with probability $1/2$ and accept with probability

$$\min\left\{1, \frac{1-p}{p}\left[1-\exp\left\{-\frac{1}{2}\tau(y_i-\mu_{D_{i,k_i+1}})^2\right\}\right]\right\}.$$

- Otherwise, accept a move from $k_i$ to $k_i - 1$ with probability

$$\min\left\{1, \frac{p}{1-p}\left[1-\exp\left\{-\frac{1}{2}\tau(y_i-\mu_{D_{i,k_i}})^2\right\}\right]^{-1}\right\},$$

# Example 1: Stationary mixture model

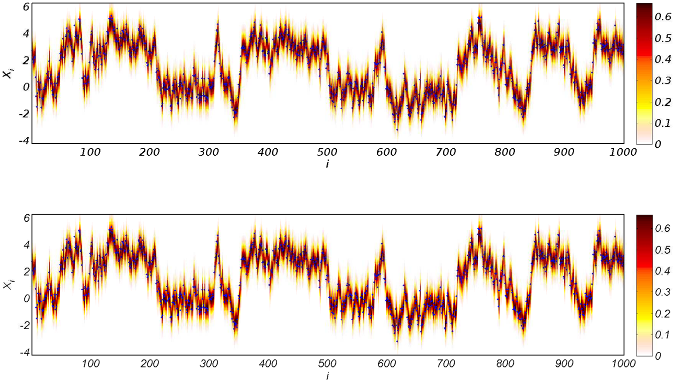
Sample of size $n = 1000$ from the stationary mixture model with true parameters:

$$w = \begin{bmatrix} 0.1 \\ 0.4 \\ 0.5 \end{bmatrix} \quad \mu = \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix}$$

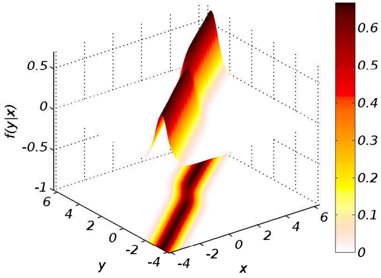
$$\sigma^2 = 1$$

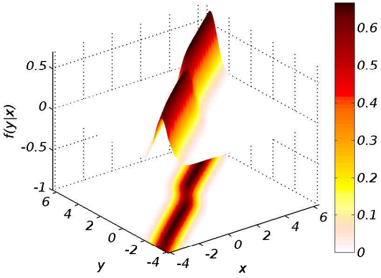
$$\rho = 0.8$$

# Example 1: Data and transition densities

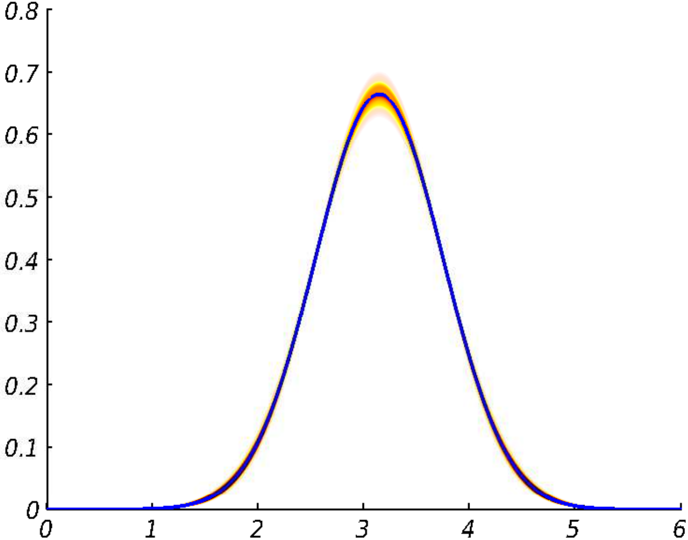


# Example 1: Conditional density surface

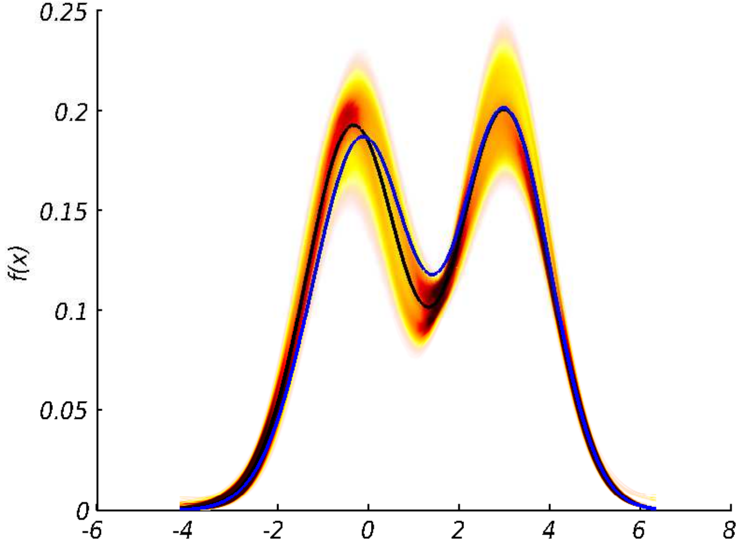


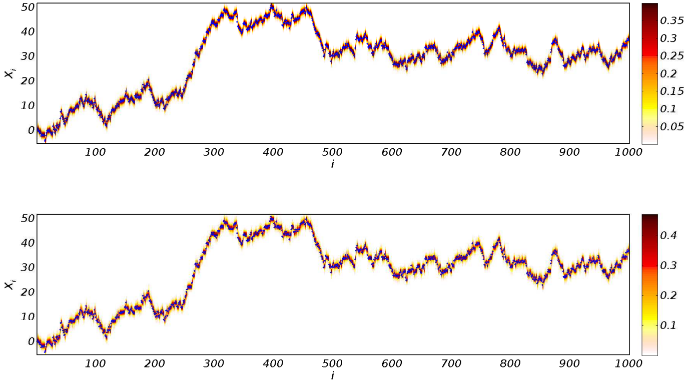
# Example 1: Transition density



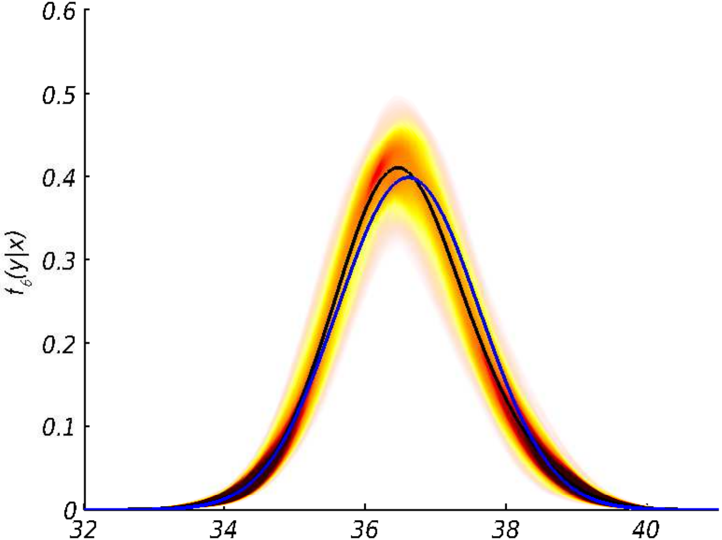
# Example 1: Stationary density



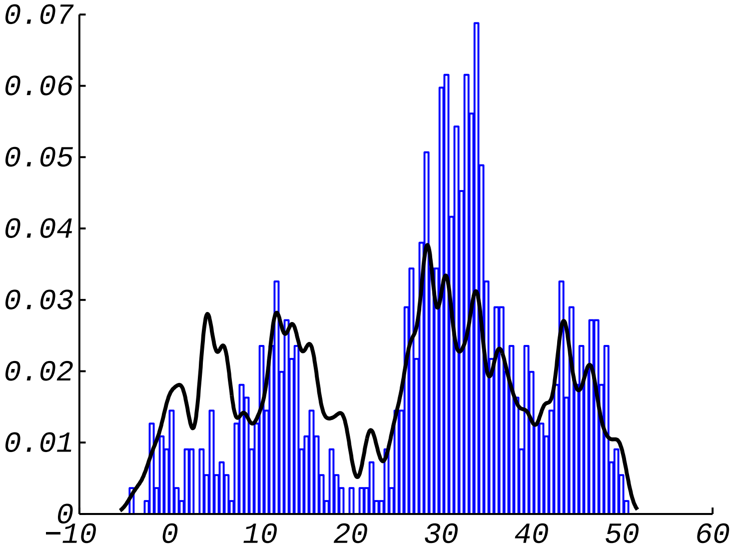
# Example 2: Standard Brownian motion



# Example 2: Transition density



# Example 2: Histogram of the data



# Discussion

- The introduction of latent variables can enable MCMC estimation for intractable models, even when the simulation involves infinite-dimensional variables. In this case, an infinite-dimensional latent variable may be convenient.
- The choice of an appropriate latent variable can be facilitated by series representations of the unknown quantity.
- More efficient simulation algorithms could be devised. In particular, the autocorrelation structure of the MCMC sample should be controlled.
- We focused on the case of normalizing constant for infinite mixture models, but the idea can be applied in more general set-ups.
- Better solutions to the problem of transdimensional MCMCare yet to be found.
- It is possible to define simple, flexible, interpretable models. But they are useless without inference methods.

# Bibliography I

I. Antoniano-Villalobos and S. G. Walker. Bayesian Nonparametric Inference for the Power Likelihood. *Journal of Computational and Graphical Statistics*, 2012. Accepted for publication.

D. B. Dunson and J. H. Park. Kernel Stick-Breaking Processes. *Biometrika*, 95(2):307–323, June 2008.

D. B. Dunson and A. Rodríguez. Nonparametric Bayesian Models Through Probit Stick-Breaking Processes. *Bayesian Analysis*, 6:145–178, 2011.

M. D. Escobar. *Estimating the Means of Several Normal Populations by Nonparametric Estimation of the Distribution of the Means*. PhD thesis, Department of Statistics, Yale University, New Haven, 1988.

N. Friel and A. N. Pettitt. Marginal Likelihood Estimation via Power Posteriors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):589–607, 2008.

S. J. Godsill. On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty. *Journal of Computational and Graphical Statistics*, 10(2):230–248, 2001.

P. J. Green. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, 82(4):711–732, 1995.

J. E. Griffin and M. F. J. Steel. Order-Based Dependent Dirichlet Processes. *Journal of the American Statistical Association*, 101(473):179–194, 2006.

Joseph G. Ibrahim and Ming H. Chen. Power Prior Distributions for Regression Models. *Statistical Science*, 15(1):46–60, 2000.

H. Ishwaran and L. F. James. Gibbs Sampling Methods for Stick Breaking Priors. *Journal of the American Statistical Association*, 96(453):161–173, 2001.

H. Ishwaran and M. Zarepour. Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-Parameter Process Hierarchical Models. *Biometrika*, 87(2):371–390, 2000.

M. Kalli, J. E. Griffin, and S. G. Walker. Slice Sampling Mixture Models. *Statistics and Computing*, 21:93–105, 2011.

# Bibliography II


A. Y. Lo. On a Class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics*, 12(1):351–357, 1984.

S. N. MacEachern. Dependent Nonparametric Processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, pages 50–55. American Statistical Association, 1999.

S. N. MacEachern. Dependent Dirichlet Processes. Technical report, Department of Statistics, Ohio State University, 2000.

S. N. MacEachern and P. Müller. Estimating Mixture of Dirichlet Process Models. *Journal of Computational and Graphical Statistics*, 7(2):223–238, 1998.

J. C. Martínez-Ovando and S. G. Walker. Time-Series Modelling, Stationarity and Bayesian Nonparametric Methods. Technical report, Banco de México, 2011.

R. H. Mena and S. G. Walker. Stationary Autoregressive Models via a Bayesian Nonparametric Approach. *Journal of Time Series Analysis*, 26(6):789–805, 2005.

P. Müller, M. West, and S. MacEachern. Bayesian Models for Non-Linear Auto-Regressions. *Journal of Time Series Analysis*, 18:593–614, 1997.

R. M. Neal. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

O. Papaspiliopoulos and G. O. Roberts. Retrospective Markov Chain Monte Carlo Methods for Dirichlet Process Hierarchical Models. *Biometrika*, 95(1):169–186, 2008.

J. Sethuraman. A Constructive Definition of Dirichlet Priors. *Statistica Sinica*, 4:639–650, 1994.

Y. Tang and S. Ghosal. A Consistent Nonparametric Bayesian Procedure for Eestimating Autoregressive Conditional Densities. *Computational Statistics and Data Analysis*, 51: 4424–4437, 2007.

Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.