On a Class of Nonparametric Bayesian Autoregressive Models

Maria Anna Di Lucca[1], Alessandra Guglielmi[2], Peter Müller[3], Fernando A. Quintana[4]

Karolinska Institutet[1]
Politecnico di Milano[2]
University of Texas, Austin[3]
Pontificia Universidad Católica de Chile[4]

ICERM Workshop, Providence, RI, USA, September 17–21, 2012

Outline

Outline

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Motivation

- Autoregressive models are very popular.
- We want to generalize usual assumptions $\Rightarrow$ parametric case limits the scope and extent of inference.
- Instead, we want to define a notion of "flexible autoregressive model".
- For instance, for order 1 dependence, we would like to replace $Y_t = \beta + \alpha Y_{t-1} + \epsilon_t$ by $Y_t \mid Y_{t-1} = y \sim F_y$.
- Proposal is based on dependent Dirichlet processes (DDP) but method can be extended to other types of random probability measures.

Outline

Dependent Dirichlet Processes (DDP)

Given a set of indices $\{x : x \in \mathscr{X}\}$, MacEachern (1999, 2000) proposed to consider

$$G_x(\cdot) = \sum_{j=1}^{\infty} w_j(x)\delta_{\theta_j(x)}(\cdot), \quad x \in \mathscr{X}.$$

Barrientos et al. (2012) studied the case

- $w_j(x) = V_j(x)\prod_{i=1}^{j-1}(1 - V_i(x))$, where $\{V_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. stochastic processes (s.p.) such that $V_j(x) \sim \text{Beta}(1, M_x)$ for every $x \in \mathscr{X}$ using copulas!
- the $\{\theta_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. s.p. with $\theta_j(x) \sim G_0$ using copulas too!
- $\{V_j(x)\}$ and $\{\theta_j(x)\}$ vary *smoothly* with x.

Dependent Dirichlet Processes (DDP)

Given a set of indices $\{x : x \in \mathscr{X}\}$, MacEachern (1999, 2000) proposed to consider

$$G_x(\cdot) = \sum_{j=1}^{\infty} w_j(x)\delta_{\theta_j(x)}(\cdot), \quad x \in \mathscr{X}.$$

Barrientos et al. (2012) studied the case

- $w_j(x) = V_j(x)\prod_{i=1}^{j-1}(1 - V_i(x))$, where $\{V_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. stochastic processes (s.p.) such that $V_j(x) \sim \text{Beta}(1, M_x)$ for every $x \in \mathscr{X}$ using copulas!
- the $\{\theta_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. s.p. with $\theta_j(x) \sim G_0$ using copulas too!
- $\{V_j(x)\}$ and $\{\theta_j(x)\}$ vary *smoothly* with x.

Dependent Dirichlet Processes (DDP)

Given a set of indices $\{x : x \in \mathscr{X}\}$, MacEachern (1999, 2000) proposed to consider

$$G_x(\cdot) = \sum_{j=1}^{\infty} w_j(x)\delta_{\theta_j(x)}(\cdot), \quad x \in \mathscr{X}.$$

Barrientos et al. (2012) studied the case

- $w_j(x) = V_j(x)\prod_{i=1}^{j-1}(1 - V_i(x))$, where $\{V_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. stochastic processes (s.p.) such that $V_j(x) \sim \text{Beta}(1, M_x)$ for every $x \in \mathscr{X}$ using copulas!
- the $\{\theta_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. s.p. with $\theta_j(x) \sim G_0$ using copulas too!
- $\{V_j(x)\}$ and $\{\theta_j(x)\}$ vary *smoothly* with x.

Dependent Dirichlet Processes (DDP)

Given a set of indices $\{x : x \in \mathscr{X}\}$, MacEachern (1999, 2000) proposed to consider

$$G_x(\cdot) = \sum_{j=1}^{\infty} w_j(x)\delta_{\theta_j(x)}(\cdot), \quad x \in \mathscr{X}.$$

Barrientos et al. (2012) studied the case

- $w_j(x) = V_j(x)\prod_{i=1}^{j-1}(1 - V_i(x))$, where $\{V_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. stochastic processes (s.p.) such that $V_j(x) \sim \mathrm{Beta}(1, M_x)$ for every $x \in \mathscr{X}$ using copulas!
- the $\{\theta_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. s.p. with $\theta_j(x) \sim G_0$ using copulas too!
- $\{V_j(x)\}$ and $\{\theta_j(x)\}$ vary *smoothly* with x.

Dependent Dirichlet Processes (DDP)

Given a set of indices $\{x : x \in \mathscr{X}\}$, MacEachern (1999, 2000) proposed to consider

$$G_x(\cdot) = \sum_{j=1}^{\infty} w_j(x) \delta_{\theta_j(x)}(\cdot), \quad x \in \mathscr{X}.$$

Barrientos et al. (2012) studied the case

- $w_j(x) = V_j(x) \prod_{i=1}^{j-1}(1 - V_i(x))$, where $\{V_j(x)\}_{x \in \mathscr{X}}$ are i.i.d. stochastic processes (s.p.) such that $V_j(x) \sim \mathrm{Beta}(1, M_x)$ for every $x \in \mathscr{X}$ using copulas!
- the $\{\theta_j(x)\}_{x \in \mathscr{X}}$ are i.i.d. s.p. with $\theta_j(x) \sim G_0$ using copulas too!
- $\{V_j(x)\}$ and $\{\theta_j(x)\}$ vary *smoothly* with x.

Dependent Dirichlet Processes (DDP)

Given a set of indices $\{x : x \in \mathscr{X}\}$, MacEachern (1999, 2000) proposed to consider

$$G_x(\cdot) = \sum_{j=1}^{\infty} w_j(x)\delta_{\theta_j(x)}(\cdot), \quad x \in \mathscr{X}.$$

Barrientos et al. (2012) studied the case

- $w_j(x) = V_j(x)\prod_{i=1}^{j-1}(1 - V_i(x))$, where $\{V_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. stochastic processes (s.p.) such that $V_j(x) \sim \mathrm{Beta}(1, M_x)$ for every $x \in \mathscr{X}$ using copulas!
- the $\{\theta_j(x)\}_{x\in\mathscr{X}}$ are i.i.d. s.p. with $\theta_j(x) \sim G_0$ using copulas too!
- $\{V_j(x)\}$ and $\{\theta_j(x)\}$ vary *smoothly* with x.

DDPs (Cont.)

Generic form to construct DDPs:

- use real-valued i.i.d. Gaussian processes $\{Z_j(x)\}$ and $\{U_j(x)\}$, $j \geq 1$, with $N(0,1)$ marginals, say. For instance, a continuous AR(1) when $\mathscr{X} = \mathbb{R}$.
- define $V_j(x) = B_x^{-1}(\Phi(Z_j(x)))$ where B_x: CDF for the Beta$(1, M_x)$ distribution and Φ: $N(0,1)$ CDF.
- define $\theta_j(x) = G_0^{-1}(\Phi(U_j(x)))$.
- define

$$G_x(\cdot) = \sum_{j=1}^{\infty} \left\{ V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \right\} \delta_{\theta_j(x)}(\cdot).$$

$G_x \sim DP(M_x, G_0)$ for every $x \in \mathscr{X}$.

DDPs (Cont.)

Generic form to construct DDPs:

- use real-valued i.i.d. Gaussian processes $\{Z_j(x)\}$ and $\{U_j(x)\}$, $j \geq 1$, with $N(0,1)$ marginals, say. For instance, a continuous AR(1) when $\mathscr{X} = \mathbb{R}$.
- define $V_j(x) = B_x^{-1}(\Phi(Z_j(x)))$ where B_x: CDF for the $\text{Beta}(1, M_x)$ distribution and Φ: $N(0,1)$ CDF.
- define $\theta_j(x) = G_0^{-1}(\Phi(U_j(x)))$.
- define

$$G_x(\cdot) = \sum_{j=1}^{\infty} \left\{ V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \right\} \delta_{\theta_j(x)}(\cdot).$$

$G_x \sim DP(M_x, G_0)$ for every $x \in \mathscr{X}$.

DDPs (Cont.)

Generic form to construct DDPs:

- use real-valued i.i.d. Gaussian processes $\{Z_j(x)\}$ and $\{U_j(x)\}$, $j \geq 1$, with $N(0,1)$ marginals, say. For instance, a continuous AR(1) when $\mathscr{X} = \mathbb{R}$.
- define $V_j(x) = B_x^{-1}(\Phi(Z_j(x)))$ where B_x: CDF for the $\mathrm{Beta}(1, M_x)$ distribution and Φ: $N(0,1)$ CDF.
- define $\theta_j(x) = G_0^{-1}(\Phi(U_j(x)))$.
- define

$$G_x(\cdot) = \sum_{j=1}^{\infty} \left\{ V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \right\} \delta_{\theta_j(x)}(\cdot).$$

$G_x \sim DP(M_x, G_0)$ for every $x \in \mathscr{X}$.

DDPs (Cont.)

Generic form to construct DDPs:

- use real-valued i.i.d. Gaussian processes $\{Z_j(x)\}$ and $\{U_j(x)\}$, $j \geq 1$, with $N(0,1)$ marginals, say. For instance, a continuous AR(1) when $\mathscr{X} = \mathbb{R}$.
- define $V_j(x) = B_x^{-1}(\Phi(Z_j(x)))$ where B_x: CDF for the $\mathrm{Beta}(1, M_x)$ distribution and Φ: $N(0,1)$ CDF.
- define $\theta_j(x) = G_0^{-1}(\Phi(U_j(x)))$.
- define

$$G_x(\cdot) = \sum_{j=1}^{\infty} \left\{ V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \right\} \delta_{\theta_j(x)}(\cdot).$$

$G_x \sim DP(M_x, G_0)$ for every $x \in \mathscr{X}$.

DDPs (Cont.)

Generic form to construct DDPs:

- use real-valued i.i.d. Gaussian processes $\{Z_j(x)\}$ and $\{U_j(x)\}$, $j \geq 1$, with $N(0,1)$ marginals, say. For instance, a continuous AR(1) when $\mathscr{X} = \mathbb{R}$.
- define $V_j(x) = B_x^{-1}(\Phi(Z_j(x)))$ where B_x: CDF for the $\mathrm{Beta}(1, M_x)$ distribution and Φ: $N(0,1)$ CDF.
- define $\theta_j(x) = G_0^{-1}(\Phi(U_j(x)))$.
- define

$$G_x(\cdot) = \sum_{j=1}^{\infty} \left\{ V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \right\} \delta_{\theta_j(x)}(\cdot).$$

$G_x \sim DP(M_x, G_0)$ for every $x \in \mathscr{X}$.

DDPs (Cont.)

Generic form to construct DDPs:

- use real-valued i.i.d. Gaussian processes $\{Z_j(x)\}$ and $\{U_j(x)\}$, $j \geq 1$, with $N(0,1)$ marginals, say. For instance, a continuous AR(1) when $\mathscr{X} = \mathbb{R}$.
- define $V_j(x) = B_x^{-1}(\Phi(Z_j(x)))$ where B_x: CDF for the $\mathrm{Beta}(1, M_x)$ distribution and Φ: $N(0,1)$ CDF.
- define $\theta_j(x) = G_0^{-1}(\Phi(U_j(x)))$.
- define

$$G_x(\cdot) = \sum_{j=1}^{\infty} \left\{ V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \right\} \delta_{\theta_j(x)}(\cdot).$$

$G_x \sim DP(M_x, G_0)$ for every $x \in \mathscr{X}$.

DDPs (Cont.)

Particular cases:

- "single weights": $V_j(x) \equiv V_j$ for all $x \in \mathscr{X}$;
- "single atoms": $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X}$;
- "single everything": $V_j(x) \equiv V_j$ and $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X} \Rightarrow$ the usual DP.

Let Θ: support of baseline measure; $\mathscr{P}(\Theta)$: set of all probability measures supported on Θ; $\mathscr{P}(\Theta)^{\mathscr{X}}$: all $\mathscr{P}(\Theta)$-valued functions defined on $\mathscr{X}$.

Result

Adequate construction of DDPs implies good properties (Barrientos et al., 2012), in particular, full weak support in $\mathscr{P}(\Theta)^{\mathscr{X}}$. True also for the single-weights or the single-atoms models.

DDPs (Cont.)

Particular cases:

- "single weights": $V_j(x) \equiv V_j$ for all $x \in \mathscr{X}$;
- "single atoms": $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X}$;
- "single everything": $V_j(x) \equiv V_j$ and $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X} \Rightarrow$ the usual DP.

Let Θ: support of baseline measure; $\mathscr{P}(\Theta)$: set of all probability measures supported on Θ; $\mathscr{P}(\Theta)^{\mathscr{X}}$: all $\mathscr{P}(\Theta)$-valued functions defined on $\mathscr{X}$.

Result

Adequate construction of DDPs implies good properties (Barrientos et al., 2012), in particular, full weak support in $\mathscr{P}(\Theta)^{\mathscr{X}}$. True also for the single-weights or the single-atoms models.

DDPs (Cont.)

Particular cases:

- "single weights": $V_j(x) \equiv V_j$ for all $x \in \mathscr{X}$;
- "single atoms": $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X}$;
- "single everything": $V_j(x) \equiv V_j$ and $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X} \Rightarrow$ the usual DP.

Let Θ: support of baseline measure; $\mathscr{P}(\Theta)$: set of all probability measures supported on Θ; $\mathscr{P}(\Theta)^{\mathscr{X}}$: all $\mathscr{P}(\Theta)$-valued functions defined on $\mathscr{X}$.

Result

Adequate construction of DDPs implies good properties (Barrientos et al., 2012), in particular, full weak support in $\mathscr{P}(\Theta)^{\mathscr{X}}$. True also for the single-weights or the single-atoms models.

DDPs (Cont.)

Particular cases:

- "single weights": $V_j(x) \equiv V_j$ for all $x \in \mathscr{X}$;
- "single atoms": $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X}$;
- "single everything": $V_j(x) \equiv V_j$ and $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X} \Rightarrow$ the usual DP.

Let Θ: support of baseline measure; $\mathscr{P}(\Theta)$: set of all probability measures supported on Θ; $\mathscr{P}(\Theta)^{\mathscr{X}}$: all $\mathscr{P}(\Theta)$-valued functions defined on $\mathscr{X}$.

Result

Adequate construction of DDPs implies good properties (Barrientos et al., 2012), in particular, full weak support in $\mathscr{P}(\Theta)^{\mathscr{X}}$. True also for the single-weights or the single-atoms models.

DDPs (Cont.)

Particular cases:

- "single weights": $V_j(x) \equiv V_j$ for all $x \in \mathscr{X}$;
- "single atoms": $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X}$;
- "single everything": $V_j(x) \equiv V_j$ and $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X} \Rightarrow$ the usual DP.

Let Θ: support of baseline measure; $\mathscr{P}(\Theta)$: set of all probability measures supported on Θ; $\mathscr{P}(\Theta)^{\mathscr{X}}$: all $\mathscr{P}(\Theta)$-valued functions defined on $\mathscr{X}$.

Result

Adequate construction of DDPs implies good properties (Barrientos et al., 2012), in particular, full weak support in $\mathscr{P}(\Theta)^{\mathscr{X}}$. True also for the single-weights or the single-atoms models.

DDPs (Cont.)

Particular cases:

- "single weights": $V_j(x) \equiv V_j$ for all $x \in \mathscr{X}$;
- "single atoms": $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X}$;
- "single everything": $V_j(x) \equiv V_j$ and $\theta_j(x) \equiv \theta_j$ for all $x \in \mathscr{X} \Rightarrow$ the usual DP.

Let Θ: support of baseline measure; $\mathscr{P}(\Theta)$: set of all probability measures supported on Θ; $\mathscr{P}(\Theta)^{\mathscr{X}}$: all $\mathscr{P}(\Theta)$-valued functions defined on $\mathscr{X}$.

Result

Adequate construction of DDPs implies good properties (Barrientos et al., 2012), in particular, full weak support in $\mathscr{P}(\Theta)^{\mathscr{X}}$. True also for the single-weights or the single-atoms models.

DDPs (Cont.)

We typically want to use mixture model

$$f_x(\cdot \mid G_x) = \int k(\cdot \mid \theta)\, dG_x(\theta)$$

for some convenient kernel density function $k(\cdot \mid \theta)$ (e.g. location-scale family).

Result

Under adequate assumptions on $k(\cdot \mid \theta)$, Hellinger support of $\{f_x : x \in \mathscr{X}\}$ is $\prod_{x \in \mathscr{X}} \left\{ \int_\Theta k(\cdot \mid \theta) dP_x(\theta) : P_x \in \mathscr{P}(\Theta) \right\}$ valid for DDPs, single-atoms or single-weights models.

It is even possible to obtain large Kullback-Leibler support under further conditions on $k(\cdot \mid \theta)$ (similar to Wu and Ghosal, 2008).

DDPs (Cont.)

We typically want to use mixture model

$$f_x(\cdot \mid G_x) = \int k(\cdot \mid \theta)\, dG_x(\theta)$$

for some convenient kernel density function $k(\cdot \mid \theta)$ (e.g. location-scale family).

Result

Under adequate assumptions on $k(\cdot \mid \theta)$, Hellinger support of $\{f_x : x \in \mathscr{X}\}$ is $\prod_{x \in \mathscr{X}} \left\{\int_\Theta k(\cdot \mid \theta) dP_x(\theta) : P_x \in \mathscr{P}(\Theta)\right\}$ valid for DDPs, single-atoms or single-weights models.

It is even possible to obtain large Kullback-Leibler support under further conditions on $k(\cdot \mid \theta)$ (similar to Wu and Ghosal, 2008).

DDPs (Cont.)

We typically want to use mixture model

$$f_x(\cdot \mid G_x) = \int k(\cdot \mid \theta)\, dG_x(\theta)$$

for some convenient kernel density function $k(\cdot \mid \theta)$ (e.g. location-scale family).

Result

Under adequate assumptions on $k(\cdot \mid \theta)$, Hellinger support of $\{f_x : x \in \mathscr{X}\}$ is $\prod_{x \in \mathscr{X}} \left\{ \int_\Theta k(\cdot \mid \theta) dP_x(\theta) : P_x \in \mathscr{P}(\Theta) \right\}$ valid for DDPs, single-atoms or single-weights models.

It is even possible to obtain large Kullback-Leibler support under further conditions on $k(\cdot \mid \theta)$ (similar to Wu and Ghosal, 2008).

DDPs (Cont.)

We typically want to use mixture model

$$f_x(\cdot \mid G_x) = \int k(\cdot \mid \theta)\, dG_x(\theta)$$

for some convenient kernel density function $k(\cdot \mid \theta)$ (e.g. location-scale family).

Result

Under adequate assumptions on $k(\cdot \mid \theta)$, Hellinger support of $\{f_x : x \in \mathscr{X}\}$ is $\prod_{x \in \mathscr{X}} \left\{ \int_\Theta k(\cdot \mid \theta) dP_x(\theta) : P_x \in \mathscr{P}(\Theta) \right\}$ valid for DDPs, single-atoms or single-weights models.

It is even possible to obtain large Kullback-Leibler support under further conditions on $k(\cdot \mid \theta)$ (similar to Wu and Ghosal, 2008).

Outline

Some recent references

Caron et al. (2008a): linear dynamic models with Dirichlet process mixtures for hidden states and observations.

Caron et al. (2008b): propose a stationary sequence of urn models, each marginally following a DPM.

Rodríguez and ter Horst (2008): propose time-dependent stick-breaking weights (but focus on the single-weights case) and Markovian dependence in the atoms using a dynamic linear model.

Lau and So (2008): propose an infinite mixture of autoregressive models.

Fox et al. (2011): propose a modified version of the HDP-HMM of Teh et al. (2006) applied to speaker diarization data, to allow persistence of states in time (i.e., sticky states).

Rodríguez and Dunson (2011): propose a probit stick-breaking approach, with atoms defined in terms of a latent Markov random field.

Nieto-Barajas et al. (2012): a time dependence is introduced in the weights of stick-breaking representation.

The Model: Continuous Case

- Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$.
- We propose, in general,

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}, m_t \sim N(Y_t \mid m_t, \sigma^2), \qquad m_t \sim G_{\boldsymbol{y}},$$

where

$$G_{\boldsymbol{y}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{y}) \delta_{\theta_h(\boldsymbol{y})}(\cdot).$$

- Equivalent representation:

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y} \sim \sum_{h \geq 1} w_h(\boldsymbol{y}) N(Y_t \mid \theta_h(\boldsymbol{y}), \sigma^2).$$

- Similar to Müller, West and MacEachern (1997).
- Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.

The Model: Continuous Case

- Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$.
- We propose, in general,

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}, m_t \sim N(Y_t \mid m_t, \sigma^2), \qquad m_t \sim G_{\boldsymbol{y}},$$

where

$$G_{\boldsymbol{y}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{y}) \delta_{\theta_h(\boldsymbol{y})}(\cdot).$$

- Equivalent representation:

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y} \sim \sum_{h \geq 1} w_h(\boldsymbol{y}) N(Y_t \mid \theta_h(\boldsymbol{y}), \sigma^2).$$

- Similar to Müller, West and MacEachern (1997).
- Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.

The Model: Continuous Case

- Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$.
- We propose, in general,

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}, m_t \sim N(Y_t \mid m_t, \sigma^2), \qquad m_t \sim G_{\boldsymbol{y}},$$

where

$$G_{\boldsymbol{y}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{y}) \delta_{\theta_h(\boldsymbol{y})}(\cdot).$$

- Equivalent representation:

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y} \sim \sum_{h \geq 1} w_h(\boldsymbol{y}) N(Y_t \mid \theta_h(\boldsymbol{y}), \sigma^2).$$

- Similar to Müller, West and MacEachern (1997).
- Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.

The Model: Continuous Case

- Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$.
- We propose, in general,

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}, m_t \sim N(Y_t \mid m_t, \sigma^2), \qquad m_t \sim G_{\boldsymbol{y}},$$

where

$$G_{\boldsymbol{y}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{y}) \delta_{\theta_h(\boldsymbol{y})}(\cdot).$$

- Equivalent representation:

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y} \sim \sum_{h \geq 1} w_h(\boldsymbol{y}) N(Y_t \mid \theta_h(\boldsymbol{y}), \sigma^2).$$

- Similar to Müller, West and MacEachern (1997).
- Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.

The Model: Continuous Case

- Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$.
- We propose, in general,

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}, m_t \sim N(Y_t \mid m_t, \sigma^2), \qquad m_t \sim G_{\boldsymbol{y}},$$

  where

$$G_{\boldsymbol{y}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{y}) \delta_{\theta_h(\boldsymbol{y})}(\cdot).$$

- Equivalent representation:

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y} \sim \sum_{h \geq 1} w_h(\boldsymbol{y}) N(Y_t \mid \theta_h(\boldsymbol{y}), \sigma^2).$$

- Similar to Müller, West and MacEachern (1997).
- Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.

The Model: Continuous Case

- Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$.
- We propose, in general,

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}, m_t \sim N(Y_t \mid m_t, \sigma^2), \qquad m_t \sim G_{\boldsymbol{y}},$$

where

$$G_{\boldsymbol{y}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{y}) \delta_{\theta_h(\boldsymbol{y})}(\cdot).$$

- Equivalent representation:

$$Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y} \sim \sum_{h \geq 1} w_h(\boldsymbol{y}) N(Y_t \mid \theta_h(\boldsymbol{y}), \sigma^2).$$

- Similar to Müller, West and MacEachern (1997).
- Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.

The Model: Continuous Case (cont.)

- Example: if $p = 1$, $w_h(y) = w_h$ and if $\theta_h(y) = \beta_h + \alpha_h y$ the model can be represented as

$$
\begin{aligned}
p(Y_t \mid Y_{t-1} = y, (\beta_t, \alpha_t), \sigma^2) &= N(Y_t \mid \beta_t + \alpha_t y, \sigma^2) \\
(\beta_t, \alpha_t) \mid G \overset{\text{i.i.d.}}{\sim} G & \qquad G \sim DP(M, G_0)
\end{aligned}
$$

(DP mixture model where atoms are given by linear trajectories, similar to Lau and So, 2008).

The Model: Continuous Case (cont.)

- Example: if $p = 1$, $w_h(y) = w_h$ and if $\theta_h(y) = \beta_h + \alpha_h y$ the model can be represented as

$$
\begin{aligned}
p(Y_t \mid Y_{t-1} = y, (\beta_t, \alpha_t), \sigma^2) &= N(Y_t \mid \beta_t + \alpha_t y, \sigma^2) \\
(\beta_t, \alpha_t) \mid G \overset{\text{i.i.d.}}{\sim} G & \qquad G \sim DP(M, G_0)
\end{aligned}
$$

(DP mixture model where atoms are given by linear trajectories, similar to Lau and So, 2008).

The Model: Continuous Case (cont.)

It may be computationally convenient to consider truncated version of model:

- Redefine the weights as $w_h(\boldsymbol{y}) = \prod_{i<h}(1 - V_i(\boldsymbol{y}))V_h(\boldsymbol{y})$, for $h = 1, \ldots, H$, con $V_h(\boldsymbol{y})$ as before, and $V_H(\boldsymbol{y}) \equiv 1$, which guarantees $P(\sum_{h=1}^{H} w_h(\boldsymbol{y}) = 1) = 1$ for all $\boldsymbol{y} \in \mathscr{Y}$ (Ishwaran and James, 2001).
- Hierarchical version of the former (linear atoms case):

$$Y_t \mid Y_{t-1} = \boldsymbol{y}, r_t = h, \{(\beta_j, \alpha_j)\}, \sigma^2 \sim N(\beta_h + \alpha_h y, \sigma^2),$$

$$P(r_t = h) = w_h(\boldsymbol{y}), \quad (\beta_h, \alpha_h) \overset{\text{i.i.d.}}{\sim} G_0, \quad h = 1, \ldots, H.$$

General thought

Despite the great generality of the proposed construction, it is in practice useful to resort to simple and manageable specifications.

The Model: Continuous Case (cont.)

It may be computationally convenient to consider truncated version of model:

- Redefine the weights as $w_h(\boldsymbol{y}) = \prod_{i<h}(1 - V_i(\boldsymbol{y}))V_h(\boldsymbol{y})$, for $h = 1, \ldots, H$, con $V_h(\boldsymbol{y})$ as before, and $V_H(\boldsymbol{y}) \equiv 1$, which guarantees $P(\sum_{h=1}^{H} w_h(\boldsymbol{y}) = 1) = 1$ for all $\boldsymbol{y} \in \mathscr{Y}$ (Ishwaran and James, 2001).
- Hierarchical version of the former (linear atoms case):

$$Y_t \mid Y_{t-1} = \boldsymbol{y}, r_t = h, \{(\beta_j, \alpha_j)\}, \sigma^2 \sim N(\beta_h + \alpha_h y, \sigma^2),$$

$$P(r_t = h) = w_h(\boldsymbol{y}), \quad (\beta_h, \alpha_h) \overset{\text{i.i.d.}}{\sim} G_0, \quad h = 1, \ldots, H.$$

General thought

Despite the great generality of the proposed construction, it is in practice useful to resort to simple and manageable specifications.

The Model: Continuous Case (cont.)

It may be computationally convenient to consider truncated version of model:

- Redefine the weights as $w_h(\boldsymbol{y}) = \prod_{i<h}(1 - V_i(\boldsymbol{y}))V_h(\boldsymbol{y})$, for $h = 1, \ldots, H$, con $V_h(\boldsymbol{y})$ as before, and $V_H(\boldsymbol{y}) \equiv 1$, which guarantees $P(\sum_{h=1}^{H} w_h(\boldsymbol{y}) = 1) = 1$ for all $\boldsymbol{y} \in \mathscr{Y}$ (Ishwaran and James, 2001).
- Hierarchical version of the former (linear atoms case):

$$Y_t \mid Y_{t-1} = \boldsymbol{y}, r_t = h, \{(\beta_j, \alpha_j)\}, \sigma^2 \sim N(\beta_h + \alpha_h y, \sigma^2),$$
$$P(r_t = h) = w_h(\boldsymbol{y}), \quad (\beta_h, \alpha_h) \overset{\text{i.i.d.}}{\sim} G_0, \quad h = 1, \ldots, H.$$

General thought

Despite the great generality of the proposed construction, it is in practice useful to resort to simple and manageable specifications.

The Model: Continuous Case (cont.)

It may be computationally convenient to consider truncated version of model:

- Redefine the weights as $w_h(\boldsymbol{y}) = \prod_{i<h}(1 - V_i(\boldsymbol{y}))V_h(\boldsymbol{y})$, for $h = 1, \ldots, H$, con $V_h(\boldsymbol{y})$ as before, and $V_H(\boldsymbol{y}) \equiv 1$, which guarantees $P(\sum_{h=1}^{H} w_h(\boldsymbol{y}) = 1) = 1$ for all $\boldsymbol{y} \in \mathscr{Y}$ (Ishwaran and James, 2001).
- Hierarchical version of the former (linear atoms case):

$$Y_t \mid Y_{t-1} = \boldsymbol{y}, r_t = h, \{(\beta_j, \alpha_j)\}, \sigma^2 \sim N(\beta_h + \alpha_h y, \sigma^2),$$
$$P(r_t = h) = w_h(\boldsymbol{y}), \quad (\beta_h, \alpha_h) \overset{\text{i.i.d.}}{\sim} G_0, \quad h = 1, \ldots, H.$$

General thought

Despite the great generality of the proposed construction, it is in practice useful to resort to simple and manageable specifications.

Model for Binary Outcomes

- Purpose: to extend the previous constructions to time series of binary outcomes.
- Idea: use the previous model in a latent scale.
- Albert and Chib (1993): introduce Z_t (continuous) such that

$$Y_t = 1 \iff Z_t > 0,$$

(so that $P(Y_t = 1) = P(Z_t > 0)$).

- Latent sequence $\{Z_t\}$ defines binary sequence $\{Y_t\}$.
- Two options:
  1. Consider $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$ (Markovian of order p!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}$ (can be easily extended to ordinal outcomes).

Model for Binary Outcomes

- Purpose: to extend the previous constructions to time series of binary outcomes.
- Idea: use the previous model in a latent scale.
- Albert and Chib (1993): introduce Z_t (continuous) such that

$$Y_t = 1 \iff Z_t > 0,$$

(so that $P(Y_t = 1) = P(Z_t > 0)$).

- Latent sequence $\{Z_t\}$ defines binary sequence $\{Y_t\}$.
- Two options:
  1. Consider $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$ (Markovian of order p!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}$ (can be easily extended to ordinal outcomes).

Model for Binary Outcomes

- Purpose: to extend the previous constructions to time series of binary outcomes.
- Idea: use the previous model in a latent scale.
- Albert and Chib (1993): introduce Z_t (continuous) such that

$$Y_t = 1 \Longleftrightarrow Z_t > 0,$$

(so that $P(Y_t = 1) = P(Z_t > 0)$).

- Latent sequence $\{Z_t\}$ defines binary sequence $\{Y_t\}$.
- Two options:
  1. Consider $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$ (Markovian of order p!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}$ (can be easily extended to ordinal outcomes).

Model for Binary Outcomes

- Purpose: to extend the previous constructions to time series of binary outcomes.
- Idea: use the previous model in a latent scale.
- Albert and Chib (1993): introduce Z_t (continuous) such that

$$Y_t = 1 \Longleftrightarrow Z_t > 0,$$

(so that $P(Y_t = 1) = P(Z_t > 0)$).

- Latent sequence $\{Z_t\}$ defines binary sequence $\{Y_t\}$.
- Two options:
  1. Consider $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$ (Markovian of order p!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}$ (can be easily extended to ordinal outcomes).

Model for Binary Outcomes

- Purpose: to extend the previous constructions to time series of binary outcomes.
- Idea: use the previous model in a latent scale.
- Albert and Chib (1993): introduce Z_t (continuous) such that

$$Y_t = 1 \iff Z_t > 0,$$

(so that $P(Y_t = 1) = P(Z_t > 0)$).

- Latent sequence $\{Z_t\}$ defines binary sequence $\{Y_t\}$.
- Two options:
  1. Consider $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$ (Markovian of order p!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}$ (can be easily extended to ordinal outcomes).

Model for Binary Outcomes

- Purpose: to extend the previous constructions to time series of binary outcomes.
- Idea: use the previous model in a latent scale.
- Albert and Chib (1993): introduce Z_t (continuous) such that

$$Y_t = 1 \iff Z_t > 0,$$

(so that $P(Y_t = 1) = P(Z_t > 0)$).

- Latent sequence $\{Z_t\}$ defines binary sequence $\{Y_t\}$.
- Two options:
  1. Consider $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = \boldsymbol{y}$ (Markovian of order p!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}$ (can be easily extended to ordinal outcomes).

Model for Binary Outcomes (cont.)

- "Completely latent" definition: $Y_t = I\{Z_t > 0\}$ with

$$Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}, m_t \sim N(Z_t \mid m_t, \sigma^2), \quad m_t \sim G_{\boldsymbol{z}},$$

  where

$$G_{\boldsymbol{z}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{z}) \delta_{\theta_h(\boldsymbol{z})}.$$

- The other case is similar.
- We can adopt the same previous simplifications, i.e. truncation, single weights or atoms, etc.

Model for Binary Outcomes (cont.)

- "Completely latent" definition: $Y_t = I\{Z_t > 0\}$ with

$$Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}, m_t \sim N(Z_t \mid m_t, \sigma^2), \quad m_t \sim G_{\boldsymbol{z}},$$

  where

$$G_{\boldsymbol{z}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{z}) \delta_{\theta_h(\boldsymbol{z})}.$$

- The other case is similar.
- We can adopt the same previous simplifications, i.e. truncation, single weights or atoms, etc.

Model for Binary Outcomes (cont.)

- "Completely latent" definition: $Y_t = I\{Z_t > 0\}$ with

$$Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = \boldsymbol{z}, m_t \sim N(Z_t \mid m_t, \sigma^2), \quad m_t \sim G_{\boldsymbol{z}},$$

  where

$$G_{\boldsymbol{z}}(\cdot) = \sum_{h=1}^{\infty} w_h(\boldsymbol{z}) \delta_{\theta_h(\boldsymbol{z})}.$$

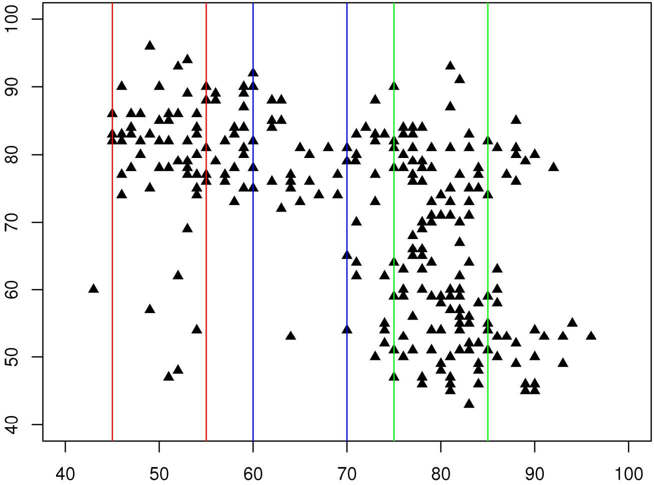
- The other case is similar.
- We can adopt the same previous simplifications, i.e. truncation, single weights or atoms, etc.

Outline

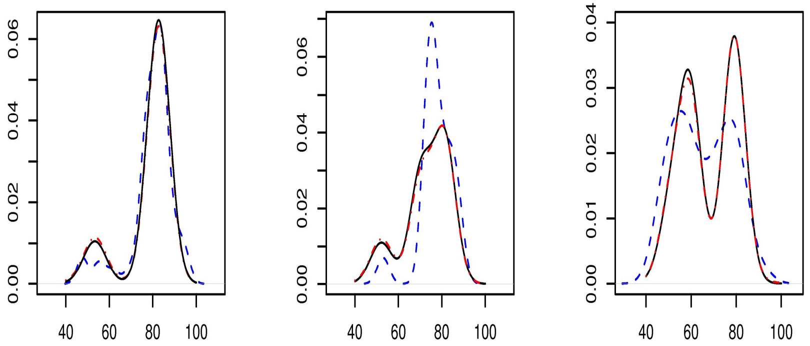
Old Faithful Geyser

- Data discussed in Härdle (1991).
- Available on-line in R.
- Consider $\{y_t, t = 1, \ldots, 272\}$, where y_t: waiting time until tth eruption of the geyser.

Old Faithful Geyser (cont.): y_t vs. y_{t-1}

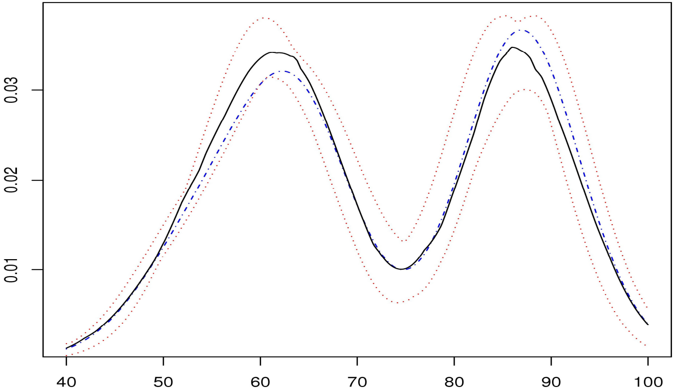


Old Faithful Geyser (cont.): $\bar{F}_y = E(F_y \mid \text{data})$, AR(1) model, single weights, linear atoms



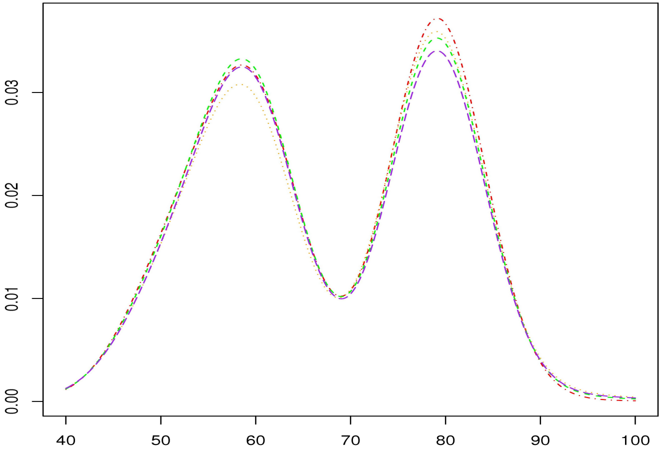
Density of the posterior mean $\bar{f}_{y_{t-1}}(y_t)$ for $y_{t-1} = 50$ (left), 65 (center) and 80 (right). Black line: prior $\sigma^{-2} \sim Ga(2,2)$; red line: $\sigma^2 = 25$; blue line: kernel estimator.

Old Faithful Geyser (cont.): $\bar{F}_y = E(F_y \mid \text{data})$, AR(1) model, single weights, linear atoms

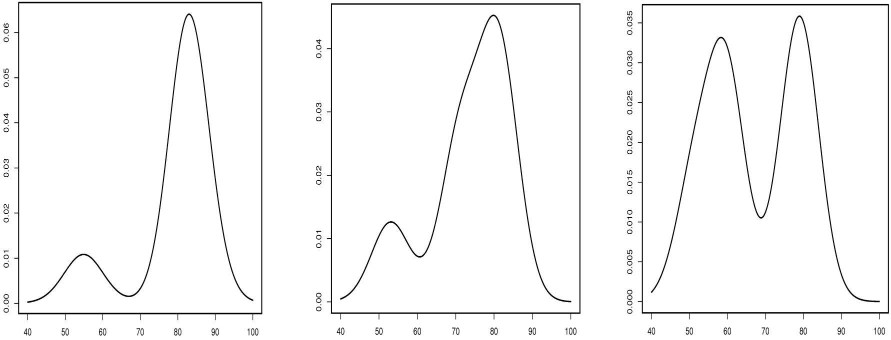


Density of the posterior mean $\bar{f}_{y_{t-1}}(\cdot)$ for $y_{t-1} = 85$ (blue), with pointwise 95% credibility bands (red) and median (black).

Old Faithful Geyser (cont.)



Density of the posterior mean $\bar{f}_{y_{t-1}}(\cdot)$ for $y_{t-1} = 85$ with $M = 1$, $H = 20$ (red), for $M = 10$, $H = 20$ (orange), for $M = 1$, $H = 50$ (green) and for $M = 10$, $H = 50$ (blue).

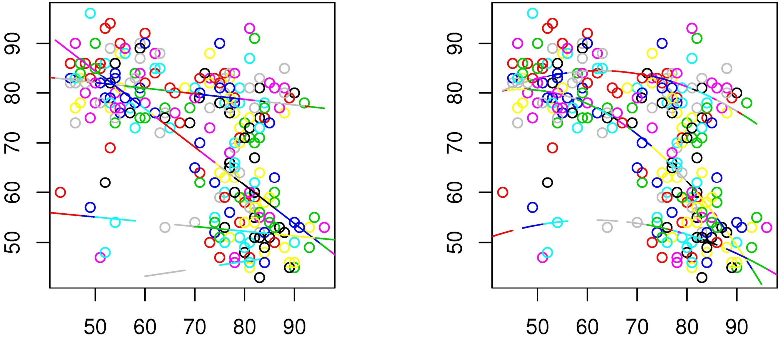


$y_{t-1} = 50$ $y_{t-1} = 65$ $y_{t-1} = 80$

Posterior means $\bar{f}_{y_{t-1}}(\cdot)$ under AR(1)-DDP model with $H = \infty$, and with varying weights

$$w_h(\boldsymbol{y}) = V_h(y) \prod_{i<j} (1 - V_h(y)) \text{ with } V_h(\boldsymbol{y}) = \text{logit}(\eta_{h1} + \eta_{h2} y).$$

Old Faithful Geyser (cont.)



One draw of all the atoms θ_h, $h = 1, \ldots, H$ in the linear case $\theta_h(y) = \beta_h + \alpha_h y$ (left) and the quadratic case $\theta_h(y) = \beta_h + \alpha_h y + \gamma_h y^2$ (right). Colors identify points in the same cluster.

Bladder Cancer Data

- Data from a bladder cancer study carried out by the Veteran's Administration Cooperative Urological Research Group, VACURG (Byar et al., 1977, Davis and Wei, 1988, Giardina et al. 2011).
- Target: compare efficacy of 2 treatments (placebo and thiotepa) in prevention of bladder cancer recurrence.
- $m = 81$ patients with ≤ 12 observations (3-months periodicity).
- Two groups (thiotepa treatment; placebo): T (36 patients), P (45 patients).
- We record indicator of cancerous tumor recurrence.
  - $y_{it} = 1$ if # detected tumors at time t increased for patient i, $y_{it} = 0$ otherwise, $t = 1, \ldots, n_i$, $i = 1, 2, \ldots, m$.
- $x_i = 0$ if patient $i \in$ group P, and $x_i = 1$ otherwise.

Bladder Cancer Data

- Data from a bladder cancer study carried out by the Veteran's Administration Cooperative Urological Research Group, VACURG (Byar et al., 1977, Davis and Wei, 1988, Giardina et al. 2011).
- Target: compare efficacy of 2 treatments (placebo and thiotepa) in prevention of bladder cancer recurrence.
- $m = 81$ patients with ≤ 12 observations (3-months periodicity).
- Two groups (thiotepa treatment; placebo): T (36 patients), P (45 patients).
- We record indicator of cancerous tumor recurrence.
  - $y_{it} = 1$ if # detected tumors at time t increased for patient i, $y_{it} = 0$ otherwise, $t = 1, \ldots, n_i$, $i = 1, 2, \ldots, m$.
- $x_i = 0$ if patient $i \in$ group P, and $x_i = 1$ otherwise.

Bladder Cancer Data

- Data from a bladder cancer study carried out by the Veteran's Administration Cooperative Urological Research Group, VACURG (Byar et al., 1977, Davis and Wei, 1988, Giardina et al. 2011).
- Target: compare efficacy of 2 treatments (placebo and thiotepa) in prevention of bladder cancer recurrence.
- $m = 81$ patients with ≤ 12 observations (3-months periodicity).
- Two groups (thiotepa treatment; placebo): T (36 patients), P (45 patients).
- We record indicator of cancerous tumor recurrence.
  - $y_{it} = 1$ if # detected tumors at time t increased for patient i, $y_{it} = 0$ otherwise, $t = 1, \ldots, n_i$, $i = 1, 2, \ldots, m$.
- $x_i = 0$ if patient $i \in$ group P, and $x_i = 1$ otherwise.

Bladder Cancer Data

- Data from a bladder cancer study carried out by the Veteran's Administration Cooperative Urological Research Group, VACURG (Byar et al., 1977, Davis and Wei, 1988, Giardina et al. 2011).
- Target: compare efficacy of 2 treatments (placebo and thiotepa) in prevention of bladder cancer recurrence.
- $m = 81$ patients with ≤ 12 observations (3-months periodicity).
- Two groups (thiotepa treatment; placebo): T $(36$ patients), P $(45$ patients).
- We record indicator of cancerous tumor recurrence.
  - $y_{it} = 1$ if $\#$ detected tumors at time t increased for patient i, $y_{it} = 0$ otherwise, $t = 1, \ldots, n_i$, $i = 1, 2, \ldots, m$.
- $x_i = 0$ if patient $i \in$ group P, and $x_i = 1$ otherwise.

Bladder Cancer Data

- Data from a bladder cancer study carried out by the Veteran's Administration Cooperative Urological Research Group, VACURG (Byar et al., 1977, Davis and Wei, 1988, Giardina et al. 2011).
- Target: compare efficacy of 2 treatments (placebo and thiotepa) in prevention of bladder cancer recurrence.
- $m = 81$ patients with ≤ 12 observations (3-months periodicity).
- Two groups (thiotepa treatment; placebo): T $(36$ patients), P $(45$ patients).
- We record indicator of cancerous tumor recurrence.
  - $y_{it} = 1$ if # detected tumors at time t increased for patient i, $y_{it} = 0$ otherwise, $t = 1, \dots, n_i$, $i = 1, 2, \dots, m$.
- $x_i = 0$ if patient $i \in$ group P, and $x_i = 1$ otherwise.

Bladder Cancer Data

- Data from a bladder cancer study carried out by the Veteran's Administration Cooperative Urological Research Group, VACURG (Byar et al., 1977, Davis and Wei, 1988, Giardina et al. 2011).
- Target: compare efficacy of 2 treatments (placebo and thiotepa) in prevention of bladder cancer recurrence.
- $m = 81$ patients with ≤ 12 observations (3-months periodicity).
- Two groups (thiotepa treatment; placebo): T (36 patients), P (45 patients).
- We record indicator of cancerous tumor recurrence.
  - $y_{it} = 1$ if # detected tumors at time t increased for patient i, $y_{it} = 0$ otherwise, $t = 1, \ldots, n_i$, $i = 1, 2, \ldots, m$.
- $x_i = 0$ if patient $i \in$ group P, and $x_i = 1$ otherwise.

Data

Recurrent tumors are removed at each visit, then treatment continues.

	TIME											
Patient	1	2	3	4	5	6	7	8	9	10	11	12
1 (P)	0	0										
2 (P)	0	0	0									
3 (P)	0	1	0									
4 (P)	0	0	0	0	0							
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
45 (P)	1	0	1	1	1	1	1	0	0	0	1	0
46 (T)	1	0										
47 (T)	0	0	0									
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
81 (T)	0	0	0	0	0	0	0	0	0	0	0	0

Model: Multiple Binary Sequences with covariates

- $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$, $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{in_i})$: sequences of responses and latent variables for patient $i = 1, \ldots, m$, with $Y_{it} = 1 \Leftrightarrow Z_{it} > 0$.
- Latent AR(1) model: $\{\boldsymbol{Z}_i\}$ are conditionally independent:

$$Z_{it} | Z_{i\,t-1} = z_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 z_{i\,t-1} + \alpha_2 x_i z_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- Latent-Y AR(1) model (Markovian):

$$Z_{it} | Y_{i\,t-1} = y_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 y_{i\,t-1} + \alpha_2 x_i y_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- σ^2 is fixed due to identifiability reasons.

Model: Multiple Binary Sequences with covariates

- $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$, $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{in_i})$: sequences of responses and latent variables for patient $i = 1, \ldots, m$, with $Y_{it} = 1 \Leftrightarrow Z_{it} > 0$.
- Latent AR(1) model: $\{\boldsymbol{Z}_i\}$ are conditionally independent:

$$Z_{it} | Z_{i\,t-1} = z_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 z_{i\,t-1} + \alpha_2 x_i z_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- Latent-Y AR(1) model (Markovian):

$$Z_{it} | Y_{i\,t-1} = y_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 y_{i\,t-1} + \alpha_2 x_i y_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- σ^2 is fixed due to identifiability reasons.

Model: Multiple Binary Sequences with covariates

- $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})$, $\boldsymbol{Z}_i = (Z_{i1}, \dots, Z_{in_i})$: sequences of responses and latent variables for patient $i = 1, \dots, m$, with $Y_{it} = 1 \Leftrightarrow Z_{it} > 0$.
- Latent AR(1) model: $\{\boldsymbol{Z}_i\}$ are conditionally independent:

$$Z_{it} | Z_{i\,t-1} = z_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 z_{i\,t-1} + \alpha_2 x_i z_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- Latent-Y AR(1) model (Markovian):

$$Z_{it} | Y_{i\,t-1} = y_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 y_{i\,t-1} + \alpha_2 x_i y_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- σ^2 is fixed due to identifiability reasons.

Model: Multiple Binary Sequences with covariates

- $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})$, $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{in_i})$: sequences of responses and latent variables for patient $i = 1, \ldots, m$, with $Y_{it} = 1 \Leftrightarrow Z_{it} > 0$.
- Latent AR(1) model: $\{\boldsymbol{Z}_i\}$ are conditionally independent:

$$Z_{it} | Z_{i\,t-1} = z_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 z_{i\,t-1} + \alpha_2 x_i z_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- Latent-Y AR(1) model (Markovian):

$$Z_{it} | Y_{i\,t-1} = y_{i\,t-1}, x_i, \beta_0, \beta_1 \sim$$
$$\int_{\mathbb{R}^2} N(\beta_0 + \beta_1 x_i + \alpha_1 y_{i\,t-1} + \alpha_2 x_i y_{i\,t-1}, \sigma^2) dG(\alpha_1, \alpha_2), \quad G \sim DP(M, G_0)$$

- σ^2 is fixed due to identifiability reasons.

Model (cont.)

- Models are completed by defining
  - $G_0(\boldsymbol{\alpha}) \equiv N_2(\boldsymbol{\alpha}; \boldsymbol{\alpha_0}, V_\alpha)$ and $\boldsymbol{\alpha_0} \sim N_2(\boldsymbol{\alpha}_{00}, V)$.
  - $(\beta_0, \beta_1) \sim N(\boldsymbol{\beta}_0, V_\beta)$;
  - Initial value for each sequence:

$$Z_{i1}|x_i, \mu_{x_i} \sim N(\mu_{x_i}, \sigma_1^2), \quad i = 1, \ldots, m, \quad x_i = 0, 1,$$

with prior such that $\mu_0 = \mu_1 + D$ and $P(D > 0) = 1$.

- We consider also a simplified version with no interaction term (3P model).

Model (cont.)

- Models are completed by defining
  - $G_0(\boldsymbol{\alpha}) \equiv N_2(\boldsymbol{\alpha}; \boldsymbol{\alpha_0}, V_\alpha)$ and $\boldsymbol{\alpha_0} \sim N_2(\boldsymbol{\alpha}_{00}, V)$.
  - $(\beta_0, \beta_1) \sim N(\boldsymbol{\beta}_0, V_\beta)$;
  - Initial value for each sequence:

$$Z_{i1}|x_i, \mu_{x_i} \sim N(\mu_{x_i}, \sigma_1^2), \quad i = 1, \ldots, m, \quad x_i = 0, 1,$$

  with prior such that $\mu_0 = \mu_1 + D$ and $P(D > 0) = 1$.

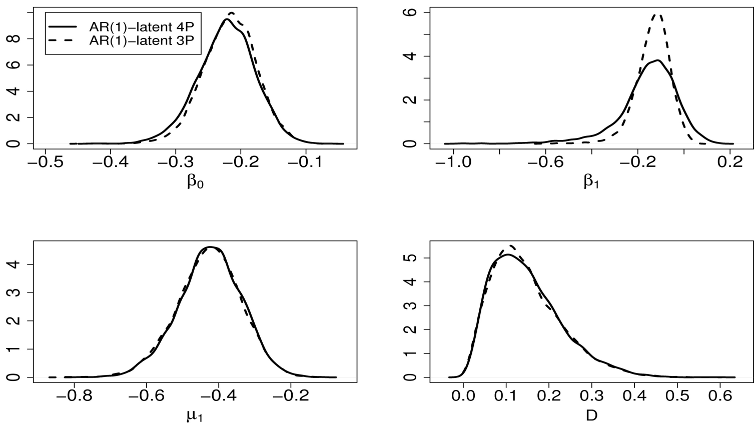
- We consider also a simplified version with no interaction term (3P model).

Results - Latent-Y AR(1) Model

	$M=1$				$M \sim U(0.5, 10)$		$M \sim$ trunc-$\mathcal{IG}(2,2)$	
	3P		4P		4P		4P	
	mean	sd	mean	sd	mean	sd	mean	sd
β_0	-0.2171	0.0410	-0.2221	0.0439	-0.2206	0.0433	-0.2207	0.0429
β_1	-0.1348	0.0749	-0.1547	0.1299	-0.1301	0.1038	-0.1286	0.0995
α_{01}	0.0798	3.1894	0.3576	0.9326	0.4703	0.9552	0.4128	0.9386
α_{02}	-	-	-0.2642	0.9937	-0.1596	0.9635	-0.1969	0.9562
μ_1	-0.4275	0.0890	-0.4240	0.0876	-0.4252	0.0883	-0.4249	0.0882
D	0.1475	0.0811	0.1483	0.0816	0.1482	0.0815	0.1465	0.0809
K	4.0524	1.5484	4.2164	1.6007	3.7666	1.6754	4.2758	1.6719
M	-	-	-	-	0.8411	0.3331	1.1115	0.2748

3P and 4P Models; σ^2=0.25, $H = 30$.

Results - Latent-Y AR(1) Model (cont.)



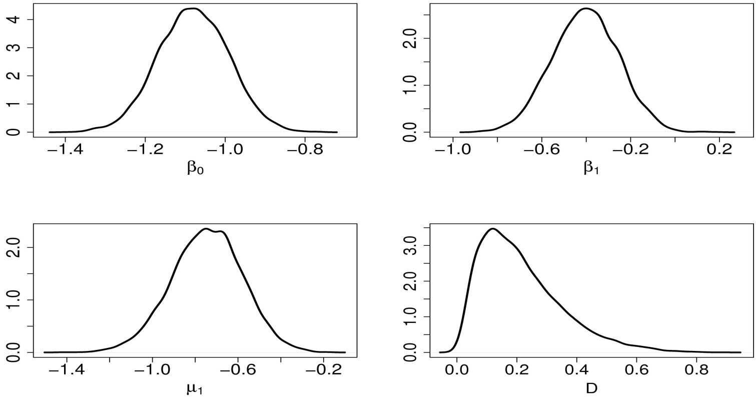
$H = 30$ and $M = 1$, for models 4P (continuous) and 3P (segments).

Results - Latent AR(1) Model

	$M=1$		$M \sim U(0.5,10)$		$M \sim$ trunc-$\mathcal{IG}(2,2)$	
	mean	sd	mean	sd	mean	sd
β_0	-1.0797	0.0881	-1.0818	0.0891	-1.0816	0.0891
β_1	-0.4039	0.1483	-0.4009	0.1532	-0.4007	0.1497
α_{01}	0.8921	0.9371	0.8870	0.9370	0.8851	0.9219
α_{02}	0.2114	0.9766	0.2234	0.9521	0.2136	0.9411
μ_1	-0.7454	0.1656	-0.7479	0.1675	-0.7465	0.1667
D	0.2143	0.1361	0.2173	0.1376	0.2157	0.1373
K	4.3454	1.6996	3.9334	1.8607	4.8270	2.0100
M	-	-	0.8615	0.3582	1.1450	0.3103

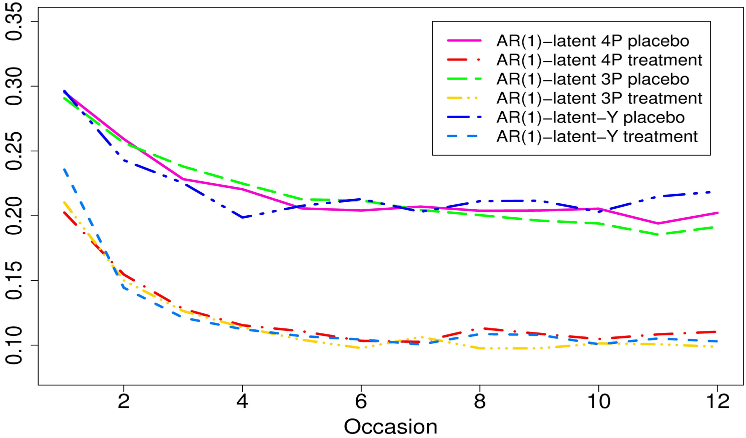
Case $H=30$ and σ^2=1.

Results - Latent AR(1) Model



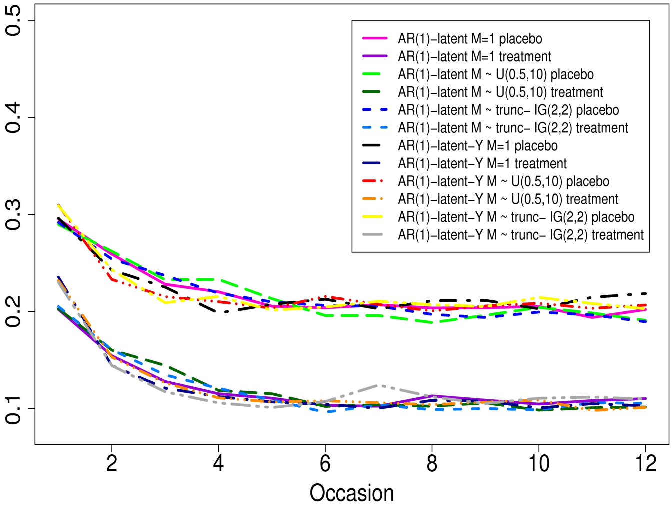
Case $H = 30$ and $M = 1$, for $\sigma^2 = 1$.

Comparison of predictions for both models (4P case)



Prediction for a new P and T patient.

Comparison of predictions (cont.)



Outline

Final Comments

- We presented a flexible autoregressive model for both continuous and binary/ordinal data.
- Model is characterized as an infinite/finite mixture of autoregressive terms, with a fixed number of lags.
- Some possible extensions (future research):
  - multivariate model formulation;
  - estimate the number of lags (so, make them random!);
  - study more properties of autoregressive models.

Final Comments

- We presented a flexible autoregressive model for both continuous and binary/ordinal data.
- Model is characterized as an infinite/finite mixture of autoregressive terms, with a fixed number of lags.
- Some possible extensions (future research):
  - multivariate model formulation;
  - estimate the number of lags (so, make them random!);
  - study more properties of autoregressive models.

Final Comments

- We presented a flexible autoregressive model for both continuous and binary/ordinal data.
- Model is characterized as an infinite/finite mixture of autoregressive terms, with a fixed number of lags.
- Some possible extensions (future research):
  - multivariate model formulation;
  - estimate the number of lags (so, make them random!);
  - study more properties of autoregressive models.

¡MUCHAS GRACIAS!

THANKS!

More at http://www.mat.puc.cl/˜quintana.