On a Class of Nonparametric Bayesian Autoregressive Models

Maria Anna Di Lucca\textsuperscript{1}, Alessandra Guglielmi\textsuperscript{2}, Peter Müller\textsuperscript{3}, Fernando A. Quintana\textsuperscript{4}

Karolinska Institutet\textsuperscript{1}
Politecnico di Milano\textsuperscript{2}
University of Texas, Austin\textsuperscript{3}
Pontificia Universidad Católica de Chile\textsuperscript{4}

ICERM Workshop, Providence, RI, USA, September 17–21, 2012
1. Motivation

2. DDP Models

3. The Model
   - Some Previous Work
   - The Model: Continuous Case
   - The Model: Binary Case

4. Data Illustrations
   - Old Faithful Geyser
   - Data from Multiple Binary Sequences

5. Final Comments
Outline

1 Motivation

2 DDP Models

3 The Model
   - Some Previous Work
   - The Model: Continuous Case
   - The Model: Binary Case

4 Data Illustrations
   - Old Faithful Geyser
   - Data from Multiple Binary Sequences

5 Final Comments
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.
Autoregressive models are very popular.

We want to generalize usual assumptions ⇒ 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 → 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.
Autoregressive models are very popular. We want to generalize usual assumptions ⇒ 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 → 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.
Autoregressive models are very popular.
We want to generalize usual assumptions ⇒ 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

DDP Models

The Model
- Some Previous Work
- The Model: Continuous Case
- The Model: Binary Case

Data Illustrations
- Old Faithful Geyser
- Data from Multiple Binary Sequences

Final Comments
Dependent Dirichlet Processes (DDP)

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

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

Barrientos et al. (2012) studied the case

\[
w_j(x) = V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)), \text{ where } \{V_j(x)\}_{x \in \mathcal{X}} \text{ are i.i.d. stochastic processes (s.p.) such that } V_j(x) \sim \text{Beta}(1, M_x) \text{ for every } x \in \mathcal{X} \text{ using copulas!}
\]

\[
\{\theta_j(x)\}_{x \in \mathcal{X}} \text{ are i.i.d. s.p. with } \theta_j(x) \sim G_0 \text{ using copulas too!}
\]

\[
\{V_j(x)\} \text{ and } \{\theta_j(x)\} \text{ vary smoothly with } x.
\]
Dependent Dirichlet Processes (DDP)

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

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

Barrientos et al. (2012) studied the case

- \( w_j(x) = V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \), where \( \{ V_j(x) \}_{x \in X} \) are i.i.d. stochastic processes (s.p.) such that \( V_j(x) \sim \text{Beta}(1, M_x) \) for every \( x \in X \) using copulas!
- the \( \{ \theta_j(x) \}_{x \in X} \) 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 \( \{x : x \in \mathcal{X}\} \), MacEachern (1999, 2000) proposed to consider

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

Barrientos et al. (2012) studied the case

- \( w_j(x) = V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \), where \( \{V_j(x)\}_{x \in \mathcal{X}} \) are i.i.d. stochastic processes (s.p.) such that \( V_j(x) \sim \text{Beta}(1, M_x) \) for every \( x \in \mathcal{X} \) using copulas!
- the \( \{\theta_j(x)\}_{x \in \mathcal{X}} \) 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 \( x \).
Given a set of indices \( \{x : x \in \mathcal{X}\} \), MacEachern (1999, 2000) proposed to consider

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

Barrientos et al. (2012) studied the case

- \( w_j(x) = V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \), where \( \{V_j(x)\}_{x \in \mathcal{X}} \) are i.i.d. stochastic processes (s.p.) such that \( V_j(x) \sim \text{Beta}(1, M_x) \) for every \( x \in \mathcal{X} \) using copulas!
- the \( \{\theta_j(x)\}_{x \in \mathcal{X}} \) 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 \( \{x : x \in \mathcal{X}\} \), MacEachern (1999, 2000) proposed to consider

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

Barrientos et al. (2012) studied the case

- \( w_j(x) = V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \), where \( \{V_j(x)\}_{x \in \mathcal{X}} \) are i.i.d. stochastic processes (s.p.) such that \( V_j(x) \sim \text{Beta}(1, M_x) \) for every \( x \in \mathcal{X} \) using copulas!
- The \( \{\theta_j(x)\}_{x \in \mathcal{X}} \) 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 \( x \).
Given a set of indices \( \{x : x \in \mathcal{X}\} \), MacEachern (1999, 2000) proposed to consider
\[
G_x(\cdot) = \sum_{j=1}^{\infty} w_j(x) \delta_{\theta_j(x)}(\cdot), \quad x \in \mathcal{X}.
\]

Barrientos et al. (2012) studied the case
- \( w_j(x) = V_j(x) \prod_{i=1}^{j-1} (1 - V_i(x)) \), where \( \{V_j(x)\}_{x \in \mathcal{X}} \) are i.i.d. stochastic processes (s.p.) such that \( V_j(x) \sim \text{Beta}(1, M_x) \) for every \( x \in \mathcal{X} \) using copulas!
- The \( \{\theta_j(x)\}_{x \in \mathcal{X}} \) 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 \( \mathcal{N}(0, 1) \) marginals, say. For instance, a continuous AR(1) when \( \mathcal{X} = \mathbb{R} \).
- define \( V_j(x) = B_x^{-1}(\Phi(Z_j(x))) \) where \( B_x \): CDF for the Beta(1, \( M_x \)) distribution and \( \Phi: \mathcal{N}(0, 1) \) 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 \( x \in \mathcal{X} \).
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 \( \mathcal{X} = \mathbb{R} \).

- define \( V_j(x) = B_x^{-1}(\Phi(Z_j(x))) \) where \( B_x \): CDF for the Beta(1, \( M_x \)) distribution and \( \Phi \): \( N(0,1) \) CDF.

- define \( \theta_j(x) = G_0^{-1}(\Phi(U_j(x))) \).

- define

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

\( G_x \sim DP(\mathcal{M}_x, G_0) \) for every \( x \in \mathcal{X} \).
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 \( \mathcal{X} = \mathbb{R} \).

- define \( V_j(x) = B_x^{-1}(\Phi(Z_j(x))) \) where \( B_x \): CDF for the Beta\((1, M_x)\) distribution and \( \Phi: N(0, 1) \) 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 \( x \in \mathcal{X} \).
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 \( \mathcal{X} = \mathbb{R} \).
- Define \( V_j(x) = B_x^{-1}(\Phi(Z_j(x))) \) where \( B_x \): CDF for the Beta\((1, M_x)\) distribution and \( \Phi: N(0, 1) \) 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 \( x \in \mathcal{X} \).
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 $\mathcal{X} = \mathbb{R}$.
- define $V_j(x) = B^{-1}_x(\Phi(Z_j(x)))$ where $B_x$: CDF for the Beta$(1, M_x)$ distribution and $\Phi$: $N(0,1)$ 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 $x \in \mathcal{X}$. 
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 \( \mathcal{X} = \mathbb{R} \).

- define \( V_j(x) = B_x^{-1}(\Phi(Z_j(x))) \) where \( B_x \): CDF for the \( \text{Beta}(1, M_x) \) distribution and \( \Phi: N(0,1) \) 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 \( x \in \mathcal{X} \).
Particular cases:

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

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

Result

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

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

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

Result

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

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

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

Result

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

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

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

Result

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

- “single weights”: \( V_j(x) \equiv V_j \) for all \( x \in \mathcal{X} \);
- “single atoms”: \( \theta_j(x) \equiv \theta_j \) for all \( x \in \mathcal{X} \);
- “single everything”: \( V_j(x) \equiv V_j \) and \( \theta_j(x) \equiv \theta_j \) for all \( x \in \mathcal{X} \) ⇒ the usual DP.

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

**Result**

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

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

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

Result

Adequate construction of DDPs implies good properties (Barrientos et al., 2012), in particular, full weak support in \( \mathcal{P}(\Theta)^{\mathcal{X}} \). True also for the single-weights or the single-atoms models.
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 \( \{f_x : x \in \mathcal{X}\} \) is \( \prod_{x \in \mathcal{X}} \{ \int_{\Theta} k(\cdot \mid \theta) dP_x(\theta) : P_x \in \mathcal{P}(\Theta) \} \) 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).
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 \( \{ f_x : x \in \mathcal{X} \} \) is \( \prod_{x \in \mathcal{X}} \left\{ \int_{\Theta} k(\cdot \mid \theta) \, dP_x(\theta) : P_x \in \mathcal{P}(\Theta) \right\} \) 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).
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 \( \{ f_x : x \in X \} \) is \( \prod_{x \in X} \{ \int_{\Theta} k(\cdot \mid \theta) \, dP_x(\theta) : P_x \in \mathcal{P}(\Theta) \} \) 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).
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 \( \{ f_x : x \in \mathcal{X} \} \) is

\[ \prod_{x \in \mathcal{X}} \left\{ \int_{\Theta} k(\cdot \mid \theta) \, dP_x(\theta) : P_x \in \mathcal{P}(\Theta) \right\} \]

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

1. Motivation

2. DDP Models

3. The Model
   - Some Previous Work
   - The Model: Continuous Case
   - The Model: Binary Case

4. Data Illustrations
   - Old Faithful Geyser
   - Data from Multiple Binary Sequences

5. Final Comments
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.
Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$.

We propose, in general,

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

where

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

Equivalent representation:

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

Similar to Müller, West and MacEachern (1997).

Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.
Given \( p \geq 1 \), we want a flexible model for \( Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y \).

We propose, in general,

\[
Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y, m_t \sim N(Y_t \mid m_t, \sigma^2), \quad m_t \sim G_y,
\]

where

\[
G_y(\cdot) = \sum_{h=1}^{\infty} w_h(y) \delta_{\theta_h(y)}(\cdot).
\]

Equivalent representation:

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

Similar to Müller, West and MacEachern (1997).

Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.
Given $p \geq 1$, we want a flexible model for $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$.

We propose, in general,

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

where

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

Equivalent representation:

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

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 $Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$.
- We propose, in general,
  \[ Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y, m_t \sim N(Y_t \mid m_t, \sigma^2), \quad m_t \sim G_Y, \]
  where
  \[ G_Y(\cdot) = \sum_{h=1}^{\infty} w_h(y) \delta_{\theta_h(y)}(\cdot). \]
- Equivalent representation:
  \[ Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y \sim \sum_{h \geq 1} w_h(y) N(Y_t \mid \theta_h(y), \sigma^2). \]
- Similar to Müller, West and MacEachern (1997).
- Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.
Given \( p \geq 1 \), we want a flexible model for \( Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y \).

We propose, in general,

\[
Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y, m_t \sim N(Y_t \mid m_t, \sigma^2), \quad m_t \sim G_y,
\]

where

\[
G_y(\cdot) = \sum_{h=1}^{\infty} w_h(y) \delta_{\theta_h(y)}(\cdot).
\]

Equivalent representation:

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

Similar to Müller, West and MacEachern (1997).

Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.
Given \( p \geq 1 \), we want a flexible model for \( Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y \).

We propose, in general,

\[
Y_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y, m_t \sim N(Y_t \mid m_t, \sigma^2), \quad m_t \sim G_y,
\]

where

\[
G_y(\cdot) = \sum_{h=1}^{\infty} w_h(y) \delta_{\theta_h(y)}(\cdot).
\]

Equivalent representation:

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

Similar to Müller, West and MacEachern (1997).

Different from Mena and Walker (2004), where they focus on stationary models with a given stationary distribution.
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

$$p(Y_t | Y_{t-1} = y, (\beta_t, \alpha_t), \sigma^2) = N(Y_t | \beta_t + \alpha_t y, \sigma^2)$$

$$(\beta_t, \alpha_t) \mid G \overset{\text{i.i.d.}}{\sim} G \quad G \sim DP(M, G_0)$$

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

\[
p(Y_t \mid Y_{t-1} = y, (\beta_t, \alpha_t), \sigma^2) = N(Y_t \mid \beta_t + \alpha_t y, \sigma^2)
\]

\[
(\beta_t, \alpha_t) \mid G \sim G \quad G \sim DP(M, G_0)
\]

(DP mixture model where atoms are given by linear trajectories, similar to Lau and So, 2008).
It may be computationally convenient to consider truncated version of model:

- Redefine the weights as $w_h(y) = \prod_{i<h} (1 - V_i(y)) V_h(y)$, for $h = 1, \ldots, H$, considering $V_h(y)$ as before, and $V_H(y) \equiv 1$, which guarantees $P(\sum_{h=1}^{H} w_h(y) = 1) = 1$ for all $y \in \mathcal{Y}$ (Ishwaran and James, 2001).

- Hierarchical version of the former (linear atoms case):

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

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

General thought

Despite the great generality of the proposed construction, it is in practice useful to resort to simple and manageable specifications.
It may be computationally convenient to consider truncated version of model:

- Redefine the weights as \( w_h(y) = \prod_{i<h}(1 - V_i(y))V_h(y) \), for \( h = 1, \ldots, H \), con \( V_h(y) \) as before, and \( V_H(y) \equiv 1 \), which guarantees \( P(\sum_{h=1}^H w_h(y) = 1) = 1 \) for all \( y \in \mathcal{Y} \) (Ishwaran and James, 2001).

- Hierarchical version of the former (linear atoms case):

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

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

**General thought**

Despite the great generality of the proposed construction, it is in practice useful to resort to simple and manageable specifications.
It may be computationally convenient to consider truncated version of model:

- Redefine the weights as $w_h(y) = \prod_{i<h}(1 - V_i(y))V_h(y)$, for $h = 1, \ldots, H$, con $V_h(y)$ as before, and $V_H(y) \equiv 1$, which guarantees $P(\sum_{h=1}^{H} w_h(y) = 1) = 1$ for all $y \in \mathcal{Y}$ (Ishwaran and James, 2001).

- Hierarchical version of the former (linear atoms case):

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

$$P(r_t = h) = w_h(y), \quad (\beta_h, \alpha_h) \overset{\text{i.i.d.}}{\sim} G_0, \quad h = 1, \ldots, H.$$
It may be computationally convenient to consider truncated version of model:

- Redefine the weights as $w_h(y) = \prod_{i<h}(1 - V_i(y))V_h(y)$, for $h = 1, \ldots, H$, consider $V_h(y)$ as before, and $V_H(y) \equiv 1$, which guarantees $P(\sum_{h=1}^{H} w_h(y) = 1) = 1$ for all $y \in \mathcal{Y}$ (Ishwaran and James, 2001).

- Hierarchical version of the former (linear atoms case):

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

$$P(r_t = h) = w_h(y), \quad (\beta_h, \alpha_h) \overset{i.i.d.}{\sim} G_0, \quad h = 1, \ldots, H.$$
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 $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$ (Markovian of order $p$!); or
2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z$ (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:**
    - Consider $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$ (Markovian of order $p$!); or
    - Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z$ (can be easily extended to ordinal 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 $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$ (Markovian of order $p$!); or
2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z$ (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 $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$ (Markovian of order $p$!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z$ (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 $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$ (Markovian of order $p$!); or
  2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z$ (can be easily extended to ordinal 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 $Z_t \mid (Y_{t-1}, \ldots, Y_{t-p}) = y$ (Markovian of order $p$!); or
2. Consider $Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z$ (can be easily extended to ordinal outcomes).
“Completely latent” definition: \( Y_t = I\{Z_t > 0\} \) with

\[
Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z, m_t \sim N(Z_t \mid m_t, \sigma^2), \quad m_t \sim G_z,
\]

where

\[
G_z(\cdot) = \sum_{h=1}^{\infty} w_h(z) \delta_{\theta_h}(z).
\]

The other case is similar.

We can adopt the same previous simplifications, i.e. truncation, single weights or atoms, etc.
“Completely latent” definition: \( Y_t = I\{Z_t > 0\} \) with

\[
Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z, m_t \sim N(Z_t \mid m_t, \sigma^2), \quad m_t \sim G_z,
\]

where

\[
G_z(\cdot) = \sum_{h=1}^{\infty} w_h(z) \delta_{\theta_h}(z).
\]

The other case is similar.

We can adopt the same previous simplifications, i.e. truncation, single weights or atoms, etc.
“Completely latent” definition: \( Y_t = I\{Z_t > 0\} \) with

\[
Z_t \mid (Z_{t-1}, \ldots, Z_{t-p}) = z, m_t \sim N(Z_t \mid m_t, \sigma^2), \quad m_t \sim G_z,
\]

where

\[
G_z(\cdot) = \sum_{h=1}^{\infty} w_h(z) \delta_{\theta_h}(z).
\]

The other case is similar.

We can adopt the same previous simplifications, i.e. truncation, single weights or atoms, etc.
1 Motivation

2 DDP Models

3 The Model
   - Some Previous Work
   - The Model: Continuous Case
   - The Model: Binary Case

4 Data Illustrations
   - Old Faithful Geyser
   - Data from Multiple Binary Sequences

5 Final Comments
Old Faithful Geyser

- Available on-line in R.
- Consider \( \{y_t, t = 1, \ldots, 272\} \), where \( y_t \): waiting time until \( t \)th eruption of the geyser.
Old Faithful Geyser (cont.): $y_t$ vs. $y_{t-1}$
Old Faithful Geyser (cont.): $\tilde{F}_y = E(F_y \mid \text{data})$, 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.): $\bar{F}_y = E(F_y \mid \text{data}), \text{AR}(1)\text{ model, single weights, linear atoms}$

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).

Data Illustrations: Old Faithful Geyser
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).
Old Faithful Geyser (cont.)

\[ y_{t-1} = 50 \]

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

\[ w_h(y) = V_h(y) \prod_{i < j} (1 - V_h(y)) \]

with \( V_h(y) = \text{logit}(\eta_{h1} + \eta_{h2}y) \).
One draw of all the atoms $\theta_h$, $h = 1, \ldots, H$ 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 $\leq 12$ 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 \( \leq 12 \) 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 \( i \in \text{group } P \), 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 \( \leq 12 \) 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 $\leq 12$ 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 $\leq 12$ 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 $\leq 12$ 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.
Recurrent tumors are removed at each visit, then treatment continues.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
<th>Time 5</th>
<th>Time 6</th>
<th>Time 7</th>
<th>Time 8</th>
<th>Time 9</th>
<th>Time 10</th>
<th>Time 11</th>
<th>Time 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (P)</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 (P)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 (P)</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 (P)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45 (P)</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>46 (T)</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47 (T)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>81 (T)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Model: Multiple Binary Sequences with covariates

- \( Y_i = (Y_{i1}, \ldots, Y_{in_i}), \ Z_i = (Z_{i1}, \ldots, Z_{in_i}) \): sequences of responses and latent variables for patient \( i = 1, \ldots, m \), with \( Y_{it} = 1 \Leftrightarrow Z_{it} > 0 \).

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

- \( \sigma^2 \) is fixed due to identifiability reasons.
Model: Multiple Binary Sequences with covariates

- \( Y_i = (Y_{i1}, \ldots, Y_{in_i}) \), \( Z_i = (Z_{i1}, \ldots, Z_{in_i}) \): sequences of responses and latent variables for patient \( i = 1, \ldots, m \), with \( Y_{it} = 1 \Leftrightarrow Z_{it} > 0 \).

- Latent AR(1) model: \( \{Z_i\} \) are conditionally independent:

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

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

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

- \( \sigma^2 \) is fixed due to identifiability reasons.
Model: Multiple Binary Sequences with covariates

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

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

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

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

- \( \sigma^2 \) is fixed due to identifiability reasons.
Model: Multiple Binary Sequences with covariates

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

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

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

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

- \( \sigma^2 \) is fixed due to identifiability reasons.
Models are completed by defining

- $G_0(\alpha) \equiv N_2(\alpha; \alpha_0, V_\alpha)$ and $\alpha_0 \sim N_2(\alpha_{00}, V)$.
- $(\beta_0, \beta_1) \sim N(\beta_0, V_\beta)$;
- 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).
Models are completed by defining
- \( G_0(\alpha) \equiv N_2(\alpha; \alpha_0, V_\alpha) \) and \( \alpha_0 \sim N_2(\alpha_00, V) \).
- \((\beta_0, \beta_1) \sim N(\beta_0, V_\beta)\);
- 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

\[
M = 1
\]

\[
M \sim U(0.5, 10)
\]

\[
M \sim \text{trunc-IG}(2, 2)
\]

<table>
<thead>
<tr>
<th></th>
<th>3P</th>
<th>4P</th>
<th>3P</th>
<th>4P</th>
<th>4P</th>
<th>4P</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.2171</td>
<td>0.0410</td>
<td>-0.2221</td>
<td>0.0439</td>
<td>-0.2206</td>
<td>0.0433</td>
</tr>
<tr>
<td>sd</td>
<td>0.0410</td>
<td>0.0439</td>
<td>0.0433</td>
<td>0.0429</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(\beta_0)</th>
<th>(\beta_1)</th>
<th>(\alpha_{01})</th>
<th>(\alpha_{02})</th>
<th>(\mu_1)</th>
<th>(D)</th>
<th>(K)</th>
<th>(M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.1348</td>
<td>0.0749</td>
<td>0.3576</td>
<td>0.9326</td>
<td>0.4703</td>
<td>0.9552</td>
<td>0.4128</td>
</tr>
<tr>
<td>sd</td>
<td>0.0749</td>
<td>0.1299</td>
<td>0.9326</td>
<td>0.9552</td>
<td>0.4128</td>
<td>0.9386</td>
<td></td>
</tr>
</tbody>
</table>

| mean        | -0.2642     | 0.9937         | -0.2642      | 0.9937     | -0.1596 | 0.9635 | -0.1969 |
| sd          | 0.9937      | 0.9635         | 0.9937       | 0.9635     | 0.9562 |

| mean        | -0.4275     | 0.0890         | -0.4275      | 0.0890     | -0.4252 | 0.0883 | -0.4249 |
| sd          | 0.0890      | 0.0883         | 0.0883       | 0.0883     | 0.0882 |

| mean        | 0.1475      | 0.0811         | 0.1483       | 0.0816     | 0.1482 | 0.0815 | 0.1465 |
| sd          | 0.0811      | 0.0816         | 0.0816       | 0.0815     | 0.0809 |

| mean        | 4.0524      | 1.5484         | 4.2164       | 1.6007     | 3.7666 | 1.6754 | 4.2758 |
| sd          | 1.5484      | 1.6007         | 1.6007       | 1.6754     | 1.6719 |

| mean        | -           | -              | -            | -          | 0.8411 | 0.3331 | 1.1115 |
| sd          | -           | -              | -            | -          | 0.2748 |

3P and 4P Models; \(\sigma^2=0.25, H = 30\).
$H = 30$ and $M = 1$, for models 4P (continuous) and 3P (segments).
## Results - Latent AR(1) Model

The results for the Latent AR(1) Model are presented in a tabular format. The table compares three scenarios:

- **$M = 1$**
- **$M \sim U(0.5, 10)$**
- **$M \sim \text{trunc-IG}(2, 2)$**

### Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Case 1 ($M = 1$)</th>
<th>Case 2 ($M \sim U(0.5, 10)$)</th>
<th>Case 3 ($M \sim \text{trunc-IG}(2, 2)$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>-1.0797</td>
<td>-1.0818</td>
<td>-1.0816</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.4039</td>
<td>-0.4009</td>
<td>-0.4007</td>
</tr>
<tr>
<td>$\alpha_{01}$</td>
<td>0.8921</td>
<td>0.8870</td>
<td>0.8851</td>
</tr>
<tr>
<td>$\alpha_{02}$</td>
<td>0.2114</td>
<td>0.2234</td>
<td>0.2136</td>
</tr>
<tr>
<td>$\mu_1$</td>
<td>-0.7454</td>
<td>-0.7479</td>
<td>-0.7465</td>
</tr>
<tr>
<td>$D$</td>
<td>0.2143</td>
<td>0.2173</td>
<td>0.2157</td>
</tr>
<tr>
<td>$K$</td>
<td>4.3454</td>
<td>3.9334</td>
<td>4.8270</td>
</tr>
<tr>
<td>$M$</td>
<td>-</td>
<td>0.8615</td>
<td>1.1450</td>
</tr>
</tbody>
</table>

### Additional Information

- **Case $H = 30$ and $\sigma^2 = 1$.**
Results - Latent AR(1) Model

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

Data Illustration: Data from Multiple Binary Sequences
Comparison of predictions for both models (4P case)

Prediction for a new P and T patient.

Data Illustrations: Data from Multiple Binary Sequences
Comparison of predictions (cont.)
Outline

1 Motivation

2 DDP Models

3 The Model
   - Some Previous Work
   - The Model: Continuous Case
   - The Model: Binary Case

4 Data Illustrations
   - Old Faithful Geyser
   - Data from Multiple Binary Sequences

5 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.
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.
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!