BAYESIAN NONPARAMETRICS
FOR TIME SERIES MODELING

DOCTORAL THESIS

Francisco Jesús Rodríguez Ruiz

June 30th, 2015
OUTLINE

1 Introduction

2 Bayesian Nonparametrics

3 Contributions
   - Infinite Factorial Unbounded-State HMM
   - Infinite Factorial Finite State Machine

4 Conclusions
Motivation

Sources

Observed signal
Motivation

Sources → Observed signal

Observed signal → Source separation
Motivation

Applications:

- Power disaggregation.
- Multiuser communication systems.
- Speech separation.
- Multi-target tracking.
- Electroencephalography (EEG).
- ...
Motivation

How many hidden sources?

- Classical approaches require either
  - known number of sources.
  - upper bound.
  - model selection.

Bayesian nonparametrics can infer the number of latent sources from the data and avoid model selection.
**Motivation**

How many hidden sources?

- Classical approaches require either
  - known number of sources.
  - upper bound.
  - model selection.

- Bayesian nonparametrics can
  - infer the number of latent sources from the data.
  - avoid model selection.
Motivation

How many hidden sources?

- Classical approaches require either
  - known number of sources.
  - upper bound.
  - model selection.

- Bayesian nonparametrics can
  - infer the number of latent sources from the data.
  - avoid model selection.

Our Approach

Bayesian nonparametric modeling of discrete-time series for source separation problems
Why BNP?
Why BNP?

- Pick a large enough $\#$sources.
Why BNP?

- Pick a large enough #sources.
- Model selection (AIC, BIC).
Why BNP?

- Pick a large enough #sources.
- Model selection (AIC, BIC).
- Bayesian model selection.
Why BNP?

- Pick a large enough #sources.
- Model selection (AIC, BIC).
- Bayesian model selection.
- BNP:
  - Model complexity grows with data size.
  - Unbounded #sources.


State Of The Art

- Many BNP models for discrete-time series.
  - e.g., infinite HMM.
- Not many BNP models for source separation.
State Of The Art

- Many BNP models for discrete-time series.
  - e.g., infinite HMM.
- Not many BNP models for source separation.
  - Infinite ICA.
State Of The Art

• Many BNP models for discrete-time series.
  • e.g., infinite HMM.
• Not many BNP models for source separation.
  • Infinite ICA.
  • Infinite Factorial HMM (IFHMM).
STATE OF THE ART

- Many BNP models for discrete-time series.
  - e.g., infinite HMM.
- Not many BNP models for source separation.
  - Infinite ICA.
  - Infinite Factorial HMM (IFHMM).
- ICA-IFHMM.
State Of The Art

Lack of BNP models for source separation:

- Infinite factorial HMM with non-binary hidden states.
  - e.g., power disaggregation.
State Of The Art

Lack of BNP models for source separation:

- Infinite factorial HMM with non-binary hidden states.
  - e.g., power disaggregation.
- Model that accounts for multipath propagation.
  - e.g., multiuser communications systems.
State Of The Art

Lack of BNP models for source separation:

- Infinite factorial HMM with **non-binary** hidden states.
  - e.g., power disaggregation.

- Model that accounts for **multipath propagation**.
  - e.g., multiuser communications systems.

- Model with continuous-valued states that captures **temporal dependencies**.
  - e.g., speech separation.
Contributions

Infinite Factorial Unbounded-State HMM

1. Non-binary IFHMM.
   - Can infer the number of HMMs in a factorial model.

2. IFUHMM.
   - Can additionally infer the cardinality of the state space.

Applications:
- Power disaggregation.
- Multiuser communication systems.
Contributions

Infinite Factorial Unbounded-State HMM

1. Non-binary IFHMM.
   - Can infer the number of HMMs in a factorial model.

2. IFUHMM.
   - Can additionally infer the cardinality of the state space.

Applications:
- Power disaggregation.
- Multiuser communication systems.

Infinite Factorial Finite State Machine

- Can infer the number of FSMs in a factorial model.
- Naturally account for multipath, echo, . . .

Applications:
- Multiuser communication systems.
OUTLINE

1 INTRODUCTION

2 BAYESIAN NONPARAMETRICS

3 CONTRIBUTIONS
   Infinite Factorial Unbounded-State HMM
   Infinite Factorial Finite State Machine

4 CONCLUSIONS
Bayesian Nonparametrics

- Bayesian framework for **model selection**.
- Prior over **infinite-dimensional** parameter space.
- Only a **finite subset** of the parameters is used for any finite dataset.
- The model complexity is allowed to grow with data size.
- Rely on **stochastic processes**:
  - Gaussian process.
  - Dirichlet process.
  - Beta process.
  - ...
Bayesian Nonparametrics

- Bayesian framework for **model selection**.
- Prior over **infinite-dimensional** parameter space.
- Only a **finite subset** of the parameters is used for any finite dataset.
- The model complexity is allowed to grow with data size.
- Rely on **stochastic processes**:
  - Gaussian process.
  - Dirichlet process.
  - Beta process.
  - ...
**Indian Buffet Process**

- Prior over binary matrices with infinite columns.
- Rows $\equiv$ Data points. Columns $\equiv$ Features.
- $\mathbf{S} \sim \text{IBP}(\alpha)$.
- $\alpha$: Concentration parameter.
- Each element $s_{tm} \in \{0, 1\}$ indicates whether the $m$-th feature contributes to the $t$-th data point.
- Only a finite number of columns $M_+$ active for any finite number of rows.

$$
\mathbf{S} = \begin{bmatrix}
    s_{11} & s_{12} & \cdots & s_{1M_+} & 0 & 0 & \cdots \\
    s_{21} & s_{22} & \cdots & s_{2M_+} & 0 & 0 & \cdots \\
    \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\
    s_{T1} & s_{T2} & \cdots & s_{TM_+} & 0 & 0 & \cdots \\
\end{bmatrix}
$$

$M_+$ non-zero columns

$M$ columns (features)
Markov Indian Buffet Process

- Prior over binary matrices with infinite columns.
- Each column follows a Markov process.
- For any $T$, only $M_+$ chains become active.
- The probability $p(S)$ vanishes, but $p([S]) > 0$.
  - $[S]$: set of matrices equivalent to $S$.
- Useful to build a (binary) infinite factorial HMM.

$$ S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M_+} & 0 & 0 & \cdots \\ s_{21} & s_{22} & \cdots & s_{2M_+} & 0 & 0 & \cdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\ s_{T1} & s_{T2} & \cdots & s_{TM_+} & 0 & 0 & \cdots \end{bmatrix}$$

$S$ consists of $M_+$ non-zero columns, $T$ time instants, and $M$ columns (chains)
**Markov Indian Buffet Process**

\[ S \sim \text{MIBP}(\alpha, \beta_0, \beta_1) \]

- Can be obtained by defining the transition probabilities

\[
A^m = \begin{bmatrix} a^m & 1 - a^m \\ b^m & 1 - b^m \end{bmatrix}
\]

\[
a^m = p(s_{tm} = 0 | s_{(t-1)m} = 0)
\]

\[
b^m = p(s_{tm} = 0 | s_{(t-1)m} = 1)
\]
**Markov Indian Buffet Process**

\[ S \sim MIBP(\alpha, \beta_0, \beta_1) \]

- Can be obtained by defining the transition probabilities

\[
A^m = \begin{bmatrix} a^m & 1 - a^m \\ b^m & 1 - b^m \end{bmatrix}
\]

\[ a^m = p(s_{tm} = 0|s_{(t-1)m} = 0) \]

\[ b^m = p(s_{tm} = 0|s_{(t-1)m} = 1) \]

- ... with priors

\[ a^m \sim \text{Beta}(1, \frac{\alpha}{M}) \]

\[ b^m \sim \text{Beta}(\beta_0, \beta_1) \]

- ... and let \( M \to \infty \)
Markov Indian Buffet Process

\[ S \sim \text{MIBP}(\alpha, \beta_0, \beta_1) \]

- Can be obtained by defining the transition probabilities

\[
A^m = \begin{bmatrix}
a^m & 1 - a^m \\
b^m & 1 - b^m
\end{bmatrix}
\]

\[
a^m = p(s_{tm} = 0|s_{(t-1)m} = 0) \quad b^m = p(s_{tm} = 0|s_{(t-1)m} = 1)
\]

- ... with priors

\[
a^m \sim \text{Beta}(1, \frac{\alpha}{M}) \quad b^m \sim \text{Beta}(\beta_0, \beta_1)
\]

- ... and let \( M \to \infty \)
- After integrating out \( a^m \) and \( b^m \):

\[
\lim_{M \to \infty} p([S]) = \frac{\alpha^M \Gamma(\beta_0 + \beta_1)\Gamma(\beta_0 + n_{10}^m)\Gamma(\beta_1 + n_{11}^m)}{\prod_{h=1}^{2T} M_h! n_{00}^m n_{01}^m n_{10}^m n_{11}^m \Gamma(\beta_0 + \beta_1 + n_{10}^m + n_{11}^m)}
\]

- Markov exchangeable in the rows.
OUTLINE

1 INTRODUCTION

2 BAYESIAN NONPARAMETRICS

3 CONTRIBUTIONS
   Infinite Factorial Unbounded-State HMM
   Infinite Factorial Finite State Machine

4 CONCLUSIONS
OUTLINE

1. INTRODUCTION

2. BAYESIAN NONPARAMETRICS

3. CONTRIBUTIONS

   Infinite Factorial Unbounded-State HMM

   Infinite Factorial Finite State Machine

4. CONCLUSIONS
Non-Binary Infinite Factorial HMM

- Generalization of the MIBP for non-binary matrices.
- Each state $s_{tm} \in \{0, 1, \ldots, Q - 1\}$.
- Inactive state ($s_{tm} = 0$).

$$
S = \begin{bmatrix}
  s_{11} & s_{12} & \cdots & s_{1M+} & 0 & 0 & \cdots \\
  s_{21} & s_{22} & \cdots & s_{2M+} & 0 & 0 & \cdots \\
  \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots \\
  s_{T1} & s_{T2} & \cdots & s_{TM+} & 0 & 0 & \cdots \\
\end{bmatrix}
$$
Non-Binary Infinite Factorial HMM

- Can be obtained by defining the transition probabilities

\[
A^m = \begin{bmatrix}
a_{00}^m & a_{01}^m & \cdots & a_{0(Q-1)}^m \\
a_{10}^m & a_{11}^m & \cdots & a_{1(Q-1)}^m \\
\vdots & \vdots & \ddots & \vdots \\
a_{(Q-1)0}^m & a_{(Q-1)1}^m & \cdots & a_{(Q-1)(Q-1)}^m
\end{bmatrix}
\]
NON-BINARY INFINITE FACTORIAL HMM

- Can be obtained by defining the transition probabilities

\[ A^m = \begin{bmatrix}
  a_{00}^m & a_{01}^m & \cdots & a_{0(Q-1)}^m \\
  a_{10}^m & a_{11}^m & \cdots & a_{1(Q-1)}^m \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{(Q-1)0}^m & a_{(Q-1)1}^m & \cdots & a_{(Q-1)(Q-1)}^m
\end{bmatrix} \]

- Prior distribution:

\[ a^m \sim \text{Beta} \left( 1, \frac{\alpha}{M} \right) \]
\[ p^m \sim \text{Dirichlet}(\gamma) \]
\[ a_0^m = [a^m \ (1 - a^m)p^m] \]

\[ a_q^m \sim \text{Dirichlet}(\beta_0, \beta, \ldots, \beta), \quad q = 1, \ldots, Q - 1 \]
**Limit of** \( p([S]) \)

\[
\lim_{{M \to \infty}} p([S]) = \frac{(Q - 1)!}{(Q - N_Q)!N_f} \frac{\alpha^{M_+}}{q^{T-1}} e^{-\alpha H_T} \prod_{h=1}^{M_h} M_h!
\]

\[
\times \prod_{m=1}^{M_+} \frac{\Gamma(n_{00}^m + 1) \Gamma \left( \sum_{i=1}^{Q-1} n_{0i}^m \right)}{\Gamma(n_{0\bullet}^m + 1)} \frac{\Gamma ((Q - 1)\gamma) \prod_{i=1}^{Q-1} \Gamma(n_{0i}^m + \gamma)}{\Gamma \left( \sum_{i=1}^{Q-1} (n_{0i}^m + \gamma) \right) (\Gamma(\gamma))^{Q-1}}
\]

\[
\times \prod_{q=1}^{Q-1} \frac{\Gamma(\beta_0 + (Q - 1)\beta)}{\Gamma(\beta_0) (\Gamma(\beta))^{Q-1}} \frac{\Gamma(n_{q0}^m + \beta_0) \prod_{i=1}^{Q-1} \Gamma(n_{qi}^m + \beta)}{\Gamma \left( n_{q\bullet}^m + \beta_0 + (Q - 1)\beta \right)}
\]
Culinary Metaphor

\[ Q = 3 \text{ states (1 inactive + 2 active)} \]
**Culinary Metaphor**

\[ Q = 3 \text{ states (1 inactive + 2 active)} \]

\[ t = 1 \]
Culinary Metaphor

\[ Q = 3 \text{ states (1 inactive + 2 active)} \]
Culinary Metaphor

\[ Q = 3 \text{ states (1 inactive + 2 active)} \]
Culinary Metaphor

$Q = 3$ states (1 inactive + 2 active)

$t = 1$

$t = 2$
Culinary Metaphor

$Q = 3$ states (1 inactive $+$ 2 active)
**Culinary Metaphor**

\[ Q = 3 \text{ states (1 inactive + 2 active)} \]

<table>
<thead>
<tr>
<th></th>
<th>( t = 1 )</th>
<th>( t = 2 )</th>
<th>( t = T - 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>2</td>
<td>✔️</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td>3</td>
<td>✗</td>
<td>✔️</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

...
**Culinary Metaphor**

$Q = 3$ states (1 inactive + 2 active)

$t = 1$

$t = 2$

$t = T - 1$

$t = T$

$p(s_{T1} = 0) \propto \beta_0 + n^{1}_{20}$

$p(s_{T1} = 1) \propto \beta + n^{1}_{21}$

$p(s_{T1} = 2) \propto \beta + n^{1}_{22}$
CULINARY METAPHOR

\[ Q = 3 \text{ states (1 inactive + 2 active)} \]

\[
p(s_{T2} = 0) \propto 1 + n_{00}^2
\]
\[
p(s_{T2} \neq 0) \propto n_{01}^2 + n_{02}^2
\]
**Culinary Metaphor**

$Q = 3$ states (1 inactive + 2 active)

\[
p(s_{T2} = 1) \propto \gamma + n_{01}^2 \\
p(s_{T2} = 2) \propto \gamma + n_{02}^2
\]
### Culinary Metaphor

$Q = 3$ states (1 inactive + 2 active)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$t = 1$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$t = 2$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$t = T - 1$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$t = T$

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Culinary Metaphor

$Q = 3$ states (1 inactive + 2 active)

<table>
<thead>
<tr>
<th></th>
<th>t = 1</th>
<th>t = 2</th>
<th>\ldots</th>
<th>t = T - 1</th>
<th>t = T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t = 1$</td>
<td><img src="food_item1.png" alt="Food Item" /></td>
<td><img src="food_item2.png" alt="Food Item" /></td>
<td><img src="food_item3.png" alt="Food Item" /></td>
<td><img src="food_item4.png" alt="Food Item" /></td>
<td><img src="food_item5.png" alt="Food Item" /></td>
</tr>
<tr>
<td>$t = 2$</td>
<td><img src="food_item6.png" alt="Food Item" /></td>
<td><img src="food_item7.png" alt="Food Item" /></td>
<td><img src="food_item8.png" alt="Food Item" /></td>
<td><img src="food_item9.png" alt="Food Item" /></td>
<td><img src="food_item10.png" alt="Food Item" /></td>
</tr>
<tr>
<td>$t = T - 1$</td>
<td><img src="food_item11.png" alt="Food Item" /></td>
<td><img src="food_item12.png" alt="Food Item" /></td>
<td><img src="food_item13.png" alt="Food Item" /></td>
<td><img src="food_item14.png" alt="Food Item" /></td>
<td><img src="food_item15.png" alt="Food Item" /></td>
</tr>
<tr>
<td>$t = T$</td>
<td><img src="food_item16.png" alt="Food Item" /></td>
<td><img src="food_item17.png" alt="Food Item" /></td>
<td><img src="food_item18.png" alt="Food Item" /></td>
<td><img src="food_item19.png" alt="Food Item" /></td>
<td><img src="food_item20.png" alt="Food Item" /></td>
</tr>
</tbody>
</table>
## Culinary Metaphor

$Q = 3$ states (1 inactive + 2 active)

<table>
<thead>
<tr>
<th>$t = 1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="food1.png" alt="Food1" /></td>
<td><img src="food2.png" alt="Food2" /></td>
<td><img src="food3.png" alt="Food3" /></td>
<td><img src="food4.png" alt="Food4" /></td>
<td><img src="food5.png" alt="Food5" /></td>
<td><img src="food6.png" alt="Food6" /></td>
<td><img src="food7.png" alt="Food7" /></td>
<td><img src="food8.png" alt="Food8" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t = 2$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="food1.png" alt="Food1" /></td>
<td><img src="food2.png" alt="Food2" /></td>
<td><img src="food3.png" alt="Food3" /></td>
<td><img src="food4.png" alt="Food4" /></td>
<td><img src="food5.png" alt="Food5" /></td>
<td><img src="food6.png" alt="Food6" /></td>
<td><img src="food7.png" alt="Food7" /></td>
<td><img src="food8.png" alt="Food8" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t = T - 1$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="food1.png" alt="Food1" /></td>
<td><img src="food2.png" alt="Food2" /></td>
<td><img src="food3.png" alt="Food3" /></td>
<td><img src="food4.png" alt="Food4" /></td>
<td><img src="food5.png" alt="Food5" /></td>
<td><img src="food6.png" alt="Food6" /></td>
<td><img src="food7.png" alt="Food7" /></td>
<td><img src="food8.png" alt="Food8" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$t = T$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="food1.png" alt="Food1" /></td>
<td><img src="food2.png" alt="Food2" /></td>
<td><img src="food3.png" alt="Food3" /></td>
<td><img src="food4.png" alt="Food4" /></td>
<td><img src="food5.png" alt="Food5" /></td>
<td><img src="food6.png" alt="Food6" /></td>
<td><img src="food7.png" alt="Food7" /></td>
<td><img src="food8.png" alt="Food8" /></td>
</tr>
</tbody>
</table>
**Culinary Metaphor**

\[ Q = 3 \text{ states } (1 \text{ inactive } + 2 \text{ active}) \]

<table>
<thead>
<tr>
<th></th>
<th>( t = 1 )</th>
<th>( t = 2 )</th>
<th>( t = T - 1 )</th>
<th>( t = T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>...</td>
<td>2</td>
</tr>
</tbody>
</table>
Culinary Metaphor

$Q = 3$ states (1 inactive + 2 active)

<table>
<thead>
<tr>
<th>$t = 1$</th>
<th>0</th>
<th>2</th>
<th>0</th>
<th>...</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t = 2$</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<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>
</tr>
<tr>
<td>$t = T - 1$</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>...</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$t = T$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>...</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
**Culinary Metaphor**

\[ Q = 3 \text{ states (1 inactive + 2 active)} \]

\[
\begin{array}{cccccc}
& 0 & 1 & 0 & \cdots & 2 & 0 \\
\hline
\ \ t=1 \ 
& 2 & 1 & 0 & \cdots & 0 & 0 \\
\ \ t=2 \ 
& \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\ \ t=T-1 \ 
& 0 & 1 & 0 & \cdots & 0 & 2 \\
\ \ t=T \ 
& 0 & 2 & 1 & \cdots & 2 & 2 \\
\end{array}
\]
**Inference**

**Fixed Q**

1. **MCMC:**
   - Sample from the posterior.
   - Blocked sampling approach.
   - Slice sampling → Stick-breaking construction.
   - FFBS for each Markov chain.

2. **Variational:**
   - Approximate the posterior.
   - Structured approach.
   - Involves a forward-backward algorithm.
**Inference**

### Fixed $Q$

1. **MCMC:**
   - Sample from the posterior.
   - Blocked sampling approach.
   - Slice sampling → Stick-breaking construction.
   - FFBS for each Markov chain.

2. **Variational:**
   - Approximate the posterior.
   - Structured approach.
   - Involves a forward-backward algorithm.

### Infinite Factorial Unbounded-State HMM

- Prior over the number of states:
  \[
  Q = 2 + Q', \quad Q' \sim \text{Poisson}(\lambda)
  \]
Inference

**Fixed Q**

1. **MCMC:**
   - Sample from the posterior.
   - Blocked sampling approach.
   - Slice sampling $\rightarrow$ Stick-breaking construction.
   - FFBS for each Markov chain.

2. **Variational:**
   - Approximate the posterior.
   - Structured approach.
   - Involves a forward-backward algorithm.

**Unknown Q**

1. **MCMC:**
   - Based on reversible jump MCMC.
   - Integrate out dimension-changing variables.
   - Updating variables:
     - $Q$: Split/merge, birth/death.
     - $M_+$: Slice sampling.

**Infinite Factorial Unbounded-State HMM**

- Prior over the number of states:
  
  \[ Q = 2 + Q', \quad Q' \sim \text{Poisson}(\lambda) \]
Power Disaggregation

- Estimate the power consumption of each device.
- Non-invasive measurements.
  - Improve efficiency of consumers.
  - Detect faulty equipment.
- Two datasets.
  - REDD (1 day, 5 houses, 6 devices).
  - AMP (2 days, 1 house, 8 devices).
**Power Disaggregation**

Gaussian observation model

\[
p_m \sim \text{Gaussian}(\phi_m, \sigma^2_m)
\]

where \( \phi_m \) and \( \sigma^2_m \) are parameters of the Gaussian distribution. The diagram above illustrates the relationship between the parameters and the observations, with arrows indicating the flow of information from the prior distributions to the posteriors.
Results for the AMP database (2 days, 8 devices):

- **Day 1**
- **Day 2**

---

**Accuracy**

\[
\text{accuracy} = 1 - \frac{1}{T} \sum_{t=1}^{T} \sum_{m=1}^{M} | y(m)_t - \hat{y}(m)_t |^2
\]
Power Disaggregation

Results for the AMP database (2 days, 8 devices):

![Graph showing consumption for different devices across two days]
## Power Disaggregation

Results for the AMP database (2 days, 8 devices):

\[
\text{accuracy} = 1 - \frac{\sum_{t=1}^{T} \sum_{m=1}^{M} |y_{t}^{(m)} - \hat{y}_{t}^{(m)}|}{2 \sum_{t=1}^{T} \sum_{m=1}^{M} y_{t}^{(m)}}
\]

<table>
<thead>
<tr>
<th>Model</th>
<th>FHMM</th>
<th>Var-Q4</th>
<th>IFHMM-Q4</th>
<th>IFUHMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>0.36 ± 0.05</td>
<td>0.48 ± 0.06</td>
<td>0.58 ± 0.11</td>
<td>0.69 ± 0.10</td>
</tr>
</tbody>
</table>
**Multiuser Communication System**

![Multiuser Communication System Diagram](image)
Multiuser Communication System

Multipath propagation
OUTLINE

1 INTRODUCTION

2 BAYESIAN NONPARAMETRICS

3 CONTRIBUTIONS
   Infinite Factorial Unbounded-State HMM
   Infinite Factorial Finite State Machine

4 CONCLUSIONS
Finite-Memory Finite State Machine

Finite-Memory FSM: The state depends on the last $L$ inputs $x_t$.

HMM with $Q = 4$ states. Dense transition probability matrix.

FSM with memory length $L = 2$. Sparse transition probability matrix.
**Infinite Factorial Finite State Machine**

- HMM representation of an FSM:

  \[
  \begin{bmatrix}
  x_0 \\
  x_1
  \end{bmatrix} \xrightarrow{L} \begin{bmatrix}
  x_1 \\
  x_2
  \end{bmatrix} \xrightarrow{L} \begin{bmatrix}
  x_2 \\
  x_3
  \end{bmatrix} \rightarrow \cdots \xrightarrow{L} \begin{bmatrix}
  x_{T-1} \\
  x_T
  \end{bmatrix}
  \]

  \[y_1 \quad y_2 \quad y_3 \quad \cdots \quad y_T\]

  State space cardinality: \(|\mathcal{X}|^L\).

- Alternative representation (likelihood accounts for the memory):

  \[x_0 \xrightarrow{L} x_1 \xrightarrow{L} x_2 \xrightarrow{L} x_3 \rightarrow \cdots \xrightarrow{L} x_{T-1} \xrightarrow{L} x_T\]

  \[y_1 \quad y_2 \quad y_3 \quad \cdots \quad y_T\]
**Infinite Factorial Finite State Machine**

- The likelihood accounts for the memory.

- Infinite Factorial FSM:
  - $M \rightarrow \infty$ parallel FSMs.
  - $S \sim \text{MIBP}(\alpha, \beta_0, \beta_1)$.
  - Auxiliary variables $s_{tm}$ indicate activity/inactivity.
  - $x_{tm} = 0$ if $s_{tm} = 0$ and $x_{tm} \in A$ otherwise.
MCMC inference algorithm:

1. Propose new parallel FSMs.
   - Slice sampling.
   - Stick-breaking construction.

2. Update hidden states $x_{tm}$, $s_{tm}$.
   - Particle Gibbs with ancestor sampling.

3. Remove inactive FSMs.

4. Sample global variables.
**Inference**

**MCMC inference algorithm:**

1. Propose new parallel FSMs.
   - Slice sampling.
   - Stick-breaking construction.
2. Update hidden states $x_{tm}, s_{tm}$.
   - Particle Gibbs with ancestor sampling.
3. Remove inactive FSMs.
4. Sample global variables.

**Particle Gibbs with ancestor sampling**

- Combines MCMC and SMC.
- Better mixing properties than FFBS.
- Outperforms FFBS:
  - Quadratic complexity with memory $L$.
  - Can handle more general models.
**Generalization Of The Model**

- Extensions that we can easily handle:
  - States $x_{tm}$ do not necessarily belong to finite set.
  - The state $x_{tm}$ depends on $x_{(t-1)m}$.
**Generalization Of The Model**

- Extensions that we can easily handle:
  - States $x_{tm}$ do not necessarily belong to finite set.
  - The state $x_{tm}$ depends on $x_{(t-1)m}$.

- Applications:
  - Multi-target tracking.
  - Speech separation.
  - ...
Multiuser Communication System

- Estimate the number of users and the transmitted symbols.
- Machine-to-machine communications:
  - Transmitters switching on and off asynchronously.
  - Short bursts of symbols.
  - Reduce message overhead.
  - 5G systems.
Multiuser Communication System

Multipath propagation

Gaussian observation model

\[ y_t = \sum_{m=1}^{M_+} \sum_{\ell=1}^{L} h_{m}^{\ell} x(t - \ell + 1)_m + n_t \]
Multiuser Communication System

Synthetic experiment with 5 transmitters and 20 receivers.

\[ L = 1 \]
Multiuser Communication System

Synthetic experiment with 5 transmitters and 20 receivers.

$L = 1$
Multiuser Communication System

Synthetic experiment with 5 transmitters and 20 receivers.

$L = 1$

Varying $L$ ($-9$ dB)
MULTIUSER COMMUNICATION SYSTEM

Wi-Fi experiment:

- Ray-tracing software (WISE).
- 6 transmitters, 12 receivers.
- Office at Bell Labs Crawford Hill.
Multiuser Communication System

Wi-Fi experiment:

• Ray-tracing software (WISE).
• 6 transmitters, 12 receivers.
• Office at Bell Labs Crawford Hill.

• Recovered transmitters / Inferred $M_+$:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PGAS</td>
<td>6/6</td>
</tr>
<tr>
<td>FFBS</td>
<td>3/11</td>
</tr>
</tbody>
</table>
Multiuser Communication System

Wi-Fi experiment:

- Ray-tracing software (WISE).
- 6 transmitters, 12 receivers.
- Office at Bell Labs Crawford Hill.

- Recovered transmitters / Inferred $M_+$:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PGAS</td>
<td>6/6</td>
</tr>
<tr>
<td>FFBS</td>
<td>3/11</td>
</tr>
</tbody>
</table>

- MSE ($\times10^{-6}$) of the first channel tap ($\ell = 1$):

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PGAS</td>
<td>2.58</td>
</tr>
<tr>
<td>FFBS</td>
<td>2.79</td>
</tr>
</tbody>
</table>

(noise variance is $\sim 10^{-8}$)
OUTLINE

1. INTRODUCTION

2. BAYESIAN NONPARAMETRICS

3. CONTRIBUTIONS
   - Infinite Factorial Unbounded-State HMM
   - Infinite Factorial Finite State Machine

4. CONCLUSIONS
CONCLUSIONS

CONTRIBUTIONS

1. Non-Binary Infinite Factorial HMM.
   - MCMC inference.
   - Variational inference.

2. Infinite Factorial Unbounded-State HMM.
   - MCMC inference.

3. Infinite Factorial Finite State Machine.
   - Particle MCMC inference.

Future Work
- Doubly nonparametric IFHMM.
- Semi-Markov approaches.
- Inference:
  - Scalability.
  - Mixing of MCMC.
  - Online.
- Other applications.
- Time-varying channels.
CONCLUSIONS

CONTRIBUTIONS

1. Non-Binary Infinite Factorial HMM.
   - MCMC inference.
   - Variational inference.
2. Infinite Factorial Unbounded-State HMM.
   - MCMC inference.
3. Infinite Factorial Finite State Machine.
   - Particle MCMC inference.

FUTURE WORK

- Doubly nonparametric IFHMM.
- Semi-Markov approaches.
- Inference:
  - Scalability.
  - Mixing of MCMC.
  - Online.
- Other applications.
- Time-varying channels.
Thanks for your attention!
**Binary IFHMM For Power Disaggregation**

- REDD dataset (5 houses, 1 day, 6 devices).
- Binary IFHMM ($Q = 2$).
- Histogram of inferred $M_+$: