Posteriors, conjugacy, and exponential families for completely random measures

Tamara Broderick, Ashia C. Wilson, Michael I. Jordan

MIT Berkeley Berkeley
Models
Models

- Beta process, Bernoulli process (IBP)
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process

Background
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process

Background

- *Parametric* exponential family conjugacy [Diaconis & Ylvisaker 1979]
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process

Background

- *Parametric* exponential family conjugacy [Diaconis & Ylvisaker 1979]

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process

Background

- *Parametric* exponential family conjugacy [Diaconis & Ylvisaker 1979]

\[
p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0
\]

\[
p(\theta) \propto \theta^\alpha (1 + \theta)^{-\alpha-\beta} = \text{BetaPrime}(\theta|\alpha, \beta) \quad \alpha > 0, \beta > 0
\]
Models

• Beta process, Bernoulli process (IBP)
• Gamma process, Poisson likelihood process (DP, CRP)
• Beta process, negative binomial process

Background

• *Parametric* exponential family conjugacy [Diaconis & Ylvisaker 1979]

\[
p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0
\]

\[
p(\theta) \propto \theta^\alpha (1 + \theta)^{-\alpha-\beta} = \text{BetaPrime}(\theta|\alpha, \beta) \quad \alpha > 0, \beta > 0
\]

\[
p(\theta|x) \propto \theta^{\alpha+x} (1 + \theta)^{-(\alpha+x)-(\beta-x+1)}
\]
Models

• Beta process, Bernoulli process (IBP)
• Gamma process, Poisson likelihood process (DP, CRP)
• Beta process, negative binomial process

Background

• *Parametric* exponential family conjugacy [Diaconis & Ylvisaker 1979]
  • Likelihood $\rightarrow$ conjugate prior, straightforward inference
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process

Background

- *Parametric* exponential family conjugacy [Diaconis & Ylvisaker 1979]
  - Likelihood $\rightarrow$ conjugate prior, straightforward inference
  - Integration $\rightarrow$ addition
Models

• Beta process, Bernoulli process (IBP)
• Gamma process, Poisson likelihood process (DP, CRP)
• Beta process, negative binomial process

Want: One framework

• For Bayesian nonparametric models:
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process

Want: One framework

- For Bayesian *nonparametric* models:
  - Likelihood $\rightarrow$ conjugate prior, straightforward inference
Models

• Beta process, Bernoulli process (IBP)
• Gamma process, Poisson likelihood process (DP, CRP)
• Beta process, negative binomial process

Want: One framework

• For Bayesian nonparametric models:
  • Likelihood $\rightarrow$ conjugate prior, straightforward inference
Models

- Beta process, Bernoulli process (IBP)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process

Want: One framework

- For Bayesian *nonparametric* models:
  - Likelihood $\rightarrow$ conjugate prior, straightforward inference
## Clustering

<table>
<thead>
<tr>
<th>Document</th>
<th>Arts</th>
<th>Econ</th>
<th>Sports</th>
<th>Health</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Feature allocation

<table>
<thead>
<tr>
<th></th>
<th>Arts</th>
<th>Econ</th>
<th>Sports</th>
<th>Health</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Indian buffet process (IBP)

For \( n = 1, 2, \ldots, N \)

1. Data point \( n \) has an existing feature \( k \) that has occurred \( ! \) times with probability \( \frac{1}{k} \).

2. Number of new features for data point \( n \): \( S_{n, k} \).
For \( n = 1, 2, ..., N \)

1. Data point \( n \) has an existing feature \( k \) that has occurred \( S_{n-1,k} \) times with probability \( \frac{S_{n-1,k}}{\beta + n - 1} \)

2.
Indian buffet process (IBP)

For $n = 1, 2, \ldots, N$

1. Data point $n$ has an existing feature $k$ that has occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\beta + n - 1}$

2. Number of new features for data point $n$:

$$K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)$$

[Griffiths & Ghahramani 2006]
Indian buffet process (IBP)

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has occurred $S_{n-1,k}$ times with probability

   $$\frac{S_{n-1,k}}{\beta + n - 1}$$

2. Number of new features for data point $n$:

   $$K_n^+ = \text{Poisson}\left(\gamma\frac{\beta}{\beta + n - 1}\right)$$
For $n = 1, 2, \ldots, N$

1. Data point $n$ has an existing feature $k$ that has occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\beta + n - 1}$.

2. Number of new features for data point $n$:

$$K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)$$

[Griffiths & Ghahramani 2006]
Indian buffet process (IBP)

For \( n = 1, 2, ..., N \)

1. Data point \( n \) has an existing feature \( k \) that has occurred \( S_{n-1,k} \) times with probability

\[
\frac{S_{n-1,k}}{\beta + n - 1}
\]

2. Number of new features for data point \( n \):

\[
K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)
\]

[Griffiths & Ghahramani 2006]
Indian buffet process (IBP)

For $n = 1, 2, ..., N$

1. Data point $n$ has an existing feature $k$ that has occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\beta + n - 1}$

2. Number of new features for data point $n$:

$$K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)$$

[Griffiths & Ghahramani 2006]
Indian buffet process (IBP)

For $n = 1, 2, \ldots, N$

1. Data point $n$ has an existing feature $k$ that has occurred $S_{n-1,k}$ times with probability $rac{S_{n-1,k}}{\beta + n - 1}$

2. Number of new features for data point $n$:

$$K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)$$

[Griffiths & Ghahramani 2006]
Indian buffet process (IBP)

For $n = 1, 2, \ldots, N$

1. Data point $n$ has an existing feature $k$ that has occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\beta + n - 1}$

2. Number of new features for data point $n$:

$$K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)$$

[Griffiths & Ghahramani 2006]
Indian buffet process (IBP)

For $n = 1, 2, \ldots, N$

1. Data point $n$ has an existing feature $k$ that has occurred $S_{n-1,k}$ times with probability

$$\frac{S_{n-1,k}}{\beta + n - 1}$$

2. Number of new features for data point $n$:

$$K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)$$

[Griffiths & Ghahramani 2006]
Indian buffet process (IBP)

For \( n = 1, 2, \ldots, N \)

1. Data point \( n \) has an existing feature \( k \) that has occurred \( S_{n-1,k} \) times with probability

\[
\frac{S_{n-1,k}}{\beta + n - 1}
\]

2. Number of new features for data point \( n \):

\[
K_n^+ = \text{Poisson} \left( \gamma \frac{\beta}{\beta + n - 1} \right)
\]

[Griffiths & Ghahramani 2006]
Beta process & Bernoulli process
Beta process & Bernoulli process

For $m = 1, 2, \ldots$
Beta process & Bernoulli process

For $m = 1, 2, ...$
1. Draw

$$K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right)$$

[Refs: Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Beta process & Bernoulli process

For $m = 1, 2, \ldots$

1. Draw

$$K_m^{+} \sim \text{Poisson} \left( \frac{\beta}{\beta + m - 1} \right)$$

2. For $k = 1, \ldots, K_m^{+}$

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Beta process & Bernoulli process

For $m = 1, 2, \ldots$

1. Draw

$$K_m^+ \sim \text{Poisson} \left( \frac{\beta}{\beta + m - 1} \right)$$

2. For $k = 1, \ldots, K_m^+$
   Draw a frequency of size
   $$\theta_k \sim \text{Beta}(1, \beta + m - 1)$$

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Beta process & Bernoulli process

For \( m = 1, 2, \ldots \)

1. Draw

\[
K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right)
\]

2. For \( k = 1, \ldots, K_m^+ \)

Draw a frequency of size

\[
\theta_k \sim \text{Beta}(1, \beta + m - 1)
\]

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Beta process & Bernoulli process

For $m = 1, 2, ...$

1. Draw

$$K_m^+ \sim \text{Poisson} \left( \frac{\gamma \beta}{\beta + m - 1} \right)$$

2. For $k = 1, \ldots, K_m^+$

   Draw a frequency of size
   $$\theta_k \sim \text{Beta}(1, \beta + m - 1)$$

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
For $m = 1, 2, \ldots$

1. Draw

$$K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right)$$

2. For $k = 1, \ldots, K_m^+$

Draw a frequency of size

$$\theta_k \sim \text{Beta}(1, \beta + m - 1)$$
Beta process & Bernoulli process

For \( m = 1, 2, \ldots \)

1. Draw

\[ K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right) \]

2. For \( k = 1, \ldots, K_m^+ \)

Draw a frequency of size

\[ \theta_k \sim \text{Beta}(1, \beta + m - 1) \]

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Beta process & Bernoulli process

For $m = 1, 2, \ldots$

1. Draw

\[ K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right) \]

2. For $k = 1, \ldots, K_m^+$

Draw a frequency of size

\[ \theta_k \sim \text{Beta}(1, \beta + m - 1) \]

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Beta process & Bernoulli process

For $m = 1, 2, \ldots$

1. Draw

$$K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right)$$

2. For $k = 1, \ldots, K_m^+$
   
   Draw a frequency of size
   
   $$\theta_k \sim \text{Beta}(1, \beta + m - 1)$$

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
For $m = 1, 2, ...$

1. Draw

$$K_m^+ \sim \text{Poisson} \left( \frac{\gamma \beta}{\beta + m - 1} \right)$$

2. For $k = 1, \ldots, K_m^+$

   Draw a frequency of size
   $$\theta_k \sim \text{Beta}(1, \beta + m - 1)$$
Beta process & Bernoulli process

For \( m = 1, 2, ... \)

1. Draw

\[ K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right) \]

2. For \( k = 1, \ldots, K_m^+ \)

Draw a frequency of size

\[ \theta_k \sim \text{Beta}(1, \beta + m - 1) \]

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Beta process & Bernoulli process

For $m = 1, 2, \ldots$

1. Draw

$$K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right)$$

2. For $k = 1, \ldots, K_m^+$

   Draw a frequency of size

   $$\theta_k \sim \text{Beta}(1, \beta + m - 1)$$

[Hjort 1990; Kim 1999; Thibaux & Jordan 2007]
Why are these useful?

How do we come up with these models?

\[ k = 1 \quad 2 \quad \ldots \]

\[ n = 1 \quad 2 \quad \ldots \]

\[ N \]
Why are these useful?

- Exchangeable (e.g. Gibbs sampling)

\[
\begin{array}{c c c}
  n & 1 & 2 \\
  2 & \vdots & \ddots \\
  N & \vdots & \\
\end{array}
\]
Why are these useful?

- Exchangeable (e.g., Gibbs sampling)
- Finite but unbounded
Why are these useful?

- Exchangeable (e.g. Gibbs sampling)
- Finite but unbounded
- Hierarchical models
Why are these useful?

- Exchangeable (e.g. Gibbs sampling)
- Finite but unbounded
- Hierarchical models
- (Countable) sequence of finite-dimensional distributions
Why are these useful?

- Exchangeable (e.g. Gibbs sampling)
- Finite but unbounded
- Hierarchical models
- (Countable) sequence of finite-dimensional distributions

How do we come up with these models?
One Framework

Likelihood

[Broderick, Wilson, Jordan 2014]
One Framework

- Conjugate prior

Likelihood

[Broderick, Wilson, Jordan 2014]
One Framework

- Conjugate prior
- Marginal

Likelihood

[Broderick, Wilson, Jordan 2014]
One Framework

Likelihood

- Conjugate prior
- Marginal
- Size-biased atom sequence

[Broderick, Wilson, Jordan 2014]
One Framework

- Conjugate prior
- Marginal
- Size-biased atom sequence

Likelihood (e.g. Bernoulli)  

[Broderick, Wilson, Jordan 2014]
One Framework

- Likelihood (e.g. Bernoulli)
- Conjugate prior (e.g. BP)
- Marginal
- Size-biased atom sequence

[Broderick, Wilson, Jordan 2014]
One Framework

- Likelihood (e.g. Bernoulli)
  [Broderick, Wilson, Jordan 2014]

  - Conjugate prior (e.g. BP)
  - Marginal (e.g. IBP)
  - Size-biased atom sequence
One Framework

- Conjugate prior (e.g. BP)
- Marginal (e.g. IBP)
- Size-biased atom sequence (e.g. BP stick-breaking)

Likelihood (e.g. Bernoulli) [Broderick, Wilson, Jordan 2014]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \]

\[ x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
- Marked Poisson process rate measure
  \[ \nu(d\theta) p(x|\theta) \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1}\]

\[ x \in \{0, 1\} \quad \theta > 0\]

- Poisson process rate measure \( \nu(d\theta) \)
- Marked Poisson process rate measure
  \[ \nu(d\theta)p(x|\theta) \]
- Conjugate prior:
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \]
\[ x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
- Marked Poisson process rate measure
  \[ \nu(d\theta)p(x|\theta) \]
- Conjugate prior:
  - Rate measure
  \[ \nu(d\theta) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha-\beta} d\theta \]
  \[ \alpha \in (-1, 0], \beta > 0, \gamma > 0 \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

- Poisson process rate measure \( \nu(d\theta) \)
- Marked Poisson process rate measure
  \[ \nu(d\theta) p(x|\theta) \]
- Conjugate prior:
  - Rate measure
  \[ \nu(d\theta) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha-\beta} d\theta \]
  \( \alpha \in (-1, 0], \beta > 0, \gamma > 0 \)
- Beta prime fixed atoms
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha-\beta} d\theta \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha-\beta} d\theta \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha - \beta} \, d\theta \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha-\beta} d\theta \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta) = \gamma \theta^{\alpha - 1} (1 - \theta)^{-\alpha - \beta} d\theta \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta|x_1 = 0) = \gamma \theta^{\alpha - 1} (1 - \theta)^{\alpha - (\beta + 1)} \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta|x_1 = 0) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha-1}(\beta+1) \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta|x_{1:2} = 0) = \gamma \theta^{\alpha-1} (1 - \theta)^{-\alpha-(\beta+2)} \]
Example: odds Bernoulli

\[ p(x|\theta) = \theta^x (1 + \theta)^{-1} \quad x \in \{0, 1\} \quad \theta > 0 \]

\[ \nu(d\theta | x_{1:2} = 0) = \gamma \theta^{\alpha - 1} (1 - \theta)^{-\alpha - (\beta + 2)} \]
Size-biased atoms, beta prime process

\( \alpha = 0 \)

For \( m = 1, 2, \ldots \)

1. Draw

\[ K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right) \]

2. For \( k = 1, \ldots, K_m^+ \)

Draw a rate of size

\[ \theta_k \sim \text{BetaPrime}(1, \beta + m - 1) \]
Size-biased atoms, beta prime process

\( \alpha = 0 \)

For \( m = 1, 2, \ldots \)

1. Draw

\[
K_m^+ \sim \text{Poisson}\left(\gamma \frac{\beta}{\beta + m - 1}\right)
\]

2. For \( k = 1, \ldots, K_m^+ \)

Draw a rate of size

\[
\theta_k \sim \text{BetaPrime}(1, \beta + m - 1)
\]
**Size-biased** atoms, beta prime process

\[ \alpha = 0 \]

For \( m = 1, 2, \ldots \)

1. Draw

\[ K_m^+ \sim \text{Poisson} \left( \gamma \frac{\beta}{\beta + m - 1} \right) \]

2. For \( k = 1, \ldots, K_m^+ \)

Draw a rate of size

\[ \theta_k \sim \text{BetaPrime}(1, \beta + m - 1) \]

Marginal process derivation is similar
One Framework

Exponential family likelihood
One Framework

Exponential family likelihood

\[ p(dx|\theta) = \kappa(x) \exp\{\langle \eta(\theta), \phi(x) \rangle - A(\theta) \} \, dx \]
One Framework

Exponential family likelihood

\[ p(dx|\theta) = \kappa(x) \exp\{\langle \eta(\theta), \phi(x) \rangle - A(\theta) \} \, dx \]

[Broderick, Wilson, Jordan 2014]
One Framework

Exponential family
likelihood

\[ p(dx|\theta) = \kappa(x) \exp\{\langle \eta(\theta), \phi(x) \rangle - A(\theta)\} \, dx \]

[Broderick, Wilson, Jordan 2014]

- Conjugate prior

PPP rate measure \( \nu(d\theta) = \gamma \exp\{\langle \xi, \eta(\theta) \rangle + \lambda[-A(\theta)]\} d\theta \)

+ fixed atoms \( f(d\theta) = \exp\{\langle \xi_k, \eta(\theta) \rangle + \lambda_k[-A(\theta)] - B(\xi_k, \lambda_k)\} d\theta \)
One Framework

Exponential family likelihood

\[ p(dx|\theta) = \kappa(x) \exp\{\langle \eta(\theta), \phi(x)\rangle - A(\theta)\} \, dx \]

- Conjugate prior

PPP rate measure

\[ \nu(d\theta) = \gamma \exp\{\langle \xi, \eta(\theta)\rangle + \lambda[-A(\theta)]\} d\theta \]

+ fixed atoms

\[ f(d\theta) = \exp\{\langle \xi_k, \eta(\theta)\rangle + \lambda_k[-A(\theta)] - B(\xi_k, \lambda_k)\} d\theta \]

- Size-biased atom sequence
One Framework

Exponential family likelihood

\[ p(dx|\theta) = \kappa(x) \exp\{\langle \eta(\theta), \phi(x) \rangle - A(\theta) \} \, dx \]

- Conjugate prior
  - PPP rate measure
    \[ \nu(d\theta) = \gamma \exp\{\langle \xi, \eta(\theta) \rangle + \lambda[-A(\theta)]\} \, d\theta \]
    + fixed atoms
    \[ f(d\theta) = \exp\{\langle \xi_k, \eta(\theta) \rangle + \lambda_k[-A(\theta)] - B(\xi_k, \lambda_k)\} \, d\theta \]

- Size-biased atom sequence
  \[ K_m^+ \sim \text{Poisson} \left( \int_{x>0} \gamma \cdot \kappa(0)^{m-1} \cdot \kappa(x) \cdot \exp \{ B(\xi + (m-1)\phi(0) + \phi(x), \lambda + m) \} \, dx \right) \]
  \[ f(d\theta) \propto \int_{x>0} \exp \{ \langle \xi + (m-1)\phi(0) + \phi(x), \eta(\theta) \rangle + (\lambda + m)[-A(\theta)] \} \, dx \]

[Broderick, Wilson, Jordan 2014]
One Framework

Exponential family likelihood

\[ p(dx|\theta) = \kappa(x) \exp\{\langle \eta(\theta), \phi(x) \rangle - A(\theta) \} \, dx \]

- Conjugate prior

PPP rate measure \( \nu(d\theta) = \gamma \exp\{\langle \xi, \eta(\theta) \rangle + \lambda[-A(\theta)]\} \, d\theta \)

+ fixed atoms \( f(d\theta) = \exp\{\langle \xi_k, \eta(\theta) \rangle + \lambda_k[-A(\theta)] - B(\xi_k, \lambda_k)\} \, d\theta \)

- Size-biased atom sequence

\[ K^+_m \sim \text{Poisson} \left( \int_{x>0} \gamma \cdot \kappa(0)^{m-1} \cdot \kappa(x) \cdot \exp \{ B(\xi + (m-1)\phi(0) + \phi(x), \lambda + m) \} \, dx \right) \]

\[ f(d\theta) \propto \int_{x>0} \exp \{\langle \xi + (m-1)\phi(0) + \phi(x), \eta(\theta) \rangle + (\lambda + m)[-A(\theta)]\} \, dx \]

- Marginal process
One Framework

Exponential family
likelihood

\[ p(dx|\theta) = \kappa(x) \exp\{\langle \eta(\theta), \phi(x) \rangle - A(\theta)\} \, dx \]

[Broderick, Wilson, Jordan 2014]

• Conjugate prior

PPP rate measure

\[ \nu(d\theta) = \gamma \exp\{\langle \xi, \eta(\theta) \rangle + \lambda[-A(\theta)]\} d\theta \]

+ fixed atoms

\[ f(d\theta) = \exp\{\langle \xi_k, \eta(\theta) \rangle + \lambda_k[-A(\theta)] - B(\xi_k, \lambda_k)\} d\theta \]

• Size-biased atom sequence

\[
K_m^+ \sim \text{Poisson} \left( \int_{x>0} \gamma \cdot \kappa(0)^{m-1} \cdot \kappa(x) \cdot \exp \left\{ B(\xi + (m-1)\phi(0) + \phi(x), \lambda + m) \right\} dx \right)
\]

\[ f(d\theta) \propto \int_{x>0} \exp \left\{ \langle \xi + (m-1)\phi(0) + \phi(x), \eta(\theta) \rangle + (\lambda + m)[-A(\theta)] \right\} dx \]

• Marginal process

\[ K_n^+ \text{ as above} \]

\[ p(x_n|x_1:(n-1)) = \kappa(x_n) \exp \left\{ -B(\xi + \sum_{m=1}^{n-1} x_m, \lambda + n - 1) + B(\xi + \sum_{m=1}^{n-1} x_m + x_n, \lambda + n) \right\} \]
To satisfy BNP desiderata, likelihood must have a point mass at 0.

Poisson distribution is a direct result of a Poisson process.

Much previous work on conjugacy at a different level of a BNP hierarchy.

Can be used with arbitrary (i.e., discrete, continuous, or other) data likelihood.
Notes

• To satisfy BNP desiderata, likelihood must have a point mass at 0
Notes

- To satisfy BNP desiderata, likelihood must have a point mass at 0
- Poisson distribution direct result of Poisson process
Notes

- To satisfy BNP desiderata, likelihood must have a point mass at 0
- Poisson distribution direct result of Poisson process
- Much previous work on conjugacy at a different level of a BNP hierarchy
To satisfy BNP desiderata, likelihood must have a point mass at 0

Poisson distribution direct result of Poisson process

Much previous work on conjugacy at a different level of a BNP hierarchy
To satisfy BNP desiderata, likelihood must have a point mass at 0.

Poisson distribution direct result of Poisson process.

Much previous work on conjugacy at a different level of a BNP hierarchy.
To satisfy BNP desiderata, likelihood must have a point mass at 0

Poisson distribution direct result of Poisson process

Much previous work on conjugacy at a different level of a BNP hierarchy

Can be used with arbitrary (i.e., discrete, continuous, or other) data likelihood


