Nonparametric Bayesian Methods: Models, Algorithms, and Applications (Part III)

Tamara Broderick
ITT Career Development Assistant Professor
Electrical Engineering & Computer Science
MIT
Marginal cluster assignments
Marginal cluster assignments

- Pólya urn
Marginal cluster assignments

- Pólya urn
  - Choose any ball with prob proportional to its mass
  - Replace and add ball of same color
Marginal cluster assignments

- Pólya urn
  - Choose any ball with prob proportional to its mass
  - Replace and add ball of same color

\[ \text{PolyaUrn}(a_{\text{orange}}, a_{\text{green}}) \]
Marginal cluster assignments

- Pólya urn
  - Choose any ball with prob proportional to its mass
  - Replace and add ball of same color

\[
\lim_{n \to \infty} \frac{\# \text{ orange}}{\# \text{ total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})
\]
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

[Blackwell, MacQueen 1973; Hoppe 1984]
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass

[Blackwell, MacQueen 1973; Hoppe 1984]
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color

[Blackwell, MacQueen 1973; Hoppe 1984]
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn
  • Choose ball with prob proportional to its mass
  • If black, replace and add ball of new color
  • Else, replace and add ball of same color

[Blackwell, MacQueen 1973; Hoppe 1984]
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

Step 0

[Blackwell, MacQueen 1973; Hoppe 1984]
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
    - If black, replace and add ball of new color
    - Else, replace and add ball of same color
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

Step 0 | Step 1 | Step 2
---|---|---
[Blackwell, MacQueen 1973; Hoppe 1984]
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn

• Choose ball with prob proportional to its mass
  • If black, replace and add ball of new color
  • Else, replace and add ball of same color

Step 0 | Step 1 | Step 2 | Step 3
---|---|---|---
• | • | • | •
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

- Choose ball with prob proportional to its mass
- If black, replace and add ball of new color
- Else, replace and add ball of same color

Step 0

Step 1

Step 2

Step 3

Step 4

[Blackwell, MacQueen 1973; Hoppe 1984]
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

Step 0

Step 1

Step 2

Step 3

Step 4

(\#orange, \#other) = PolyaUrn(1, \alpha)
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

\[
\begin{align*}
\text{Step 0} & \quad \text{Step 1} & \quad \text{Step 2} & \quad \text{Step 3} & \quad \text{Step 4} \\
\text{\#black} & \quad \text{\#black} & \quad \text{\#black} & \quad \text{\#black} & \quad \text{\#black} \\
\text{\#orange, \#other} & \quad \text{\#orange, \#other} & \quad \text{\#orange, \#other} & \quad \text{\#orange, \#other} & \quad \text{\#orange, \#other} \\
\end{align*}
\]

\[(#\text{orange}, #\text{other}) = \text{PolyaUrn}(1, \alpha)\]
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn

  • Choose ball with prob proportional to its mass
  • If black, replace and add ball of new color
  • Else, replace and add ball of same color

\[
\begin{align*}
(\#\text{orange}, \#\text{other}) &= \text{PolyaUrn}(1, \alpha) \\
\text{not orange: } (\#\text{green}, \#\text{other}) &= \text{PolyaUrn}(1, \alpha)
\end{align*}
\]
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

\[
\begin{align*}
\text{Step 0:} & \quad \text{(not orange, \#other)} = \text{PolyaUrn}(1, \alpha) \\
\text{Step 1:} & \quad \text{(not orange, \#other)} = \text{PolyaUrn}(1, \alpha) \\
\text{Step 2:} & \quad \text{(not orange, \#other)} = \text{PolyaUrn}(1, \alpha) \\
\text{Step 3:} & \quad \text{(not orange, \#other)} = \text{PolyaUrn}(1, \alpha) \\
\text{Step 4:} & \quad \text{(not orange, \#other)} = \text{PolyaUrn}(1, \alpha)
\end{align*}
\]
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

\[
\begin{array}{c|c|c|c|c}
\text{Step 0} & \text{Step 1} & \text{Step 2} & \text{Step 3} & \text{Step 4} \\
\hline
\black & \black & \black & \black & \black \\
\hline
\end{array}
\]

\[
(#\text{orange}, #\text{other}) = \text{PolyaUrn}(1, \alpha)
\]

• not orange:
  - not orange, green: \((#\text{green}, #\text{other}) = \text{PolyaUrn}(1, \alpha)\)
  - not orange, green: \((#\text{red}, #\text{other}) = \text{PolyaUrn}(1, \alpha)\)
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

```
Step 0

(#{orange}, #{other}) = PolyaUrn(1, \alpha)
```

```
Step 1

- not orange: (#{green}, #{other}) = PolyaUrn(1, \alpha)
```

```
Step 2

- not orange, green: (#{red}, #{other}) = PolyaUrn(1, \alpha)
```

```
Step 3

Step 4
```
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

\[
V_k \overset{\text{iid}}{\sim} \text{Beta}(1, \alpha)
\]

\[
(#\text{orange}, #\text{other}) = \text{PolyaUrn}(1, \alpha)
\]
- not orange: \((#\text{green}, #\text{other}) = \text{PolyaUrn}(1, \alpha)\)
- not orange, green: \((#\text{red}, #\text{other}) = \text{PolyaUrn}(1, \alpha)\)
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

<table>
<thead>
<tr>
<th>Step 0</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Initial Step" /></td>
<td><img src="image2" alt="Step 1" /></td>
<td><img src="image3" alt="Step 2" /></td>
<td><img src="image4" alt="Step 3" /></td>
<td><img src="image5" alt="Step 4" /></td>
</tr>
</tbody>
</table>

\[
V_k \overset{iid}{\sim} \text{Beta}(1, \alpha) \quad \rho_1 = V_1
\]

\((\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)\)
- not orange: \((\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)\)
- not orange, green: \((\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)\)
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn
  - Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

\[
V_k \overset{iid}{\sim} \text{Beta}(1, \alpha)
\]

\[
\begin{align*}
\rho_1 &= V_1 \\
\rho_2 &= (1 - V_1)V_2
\end{align*}
\]

- not orange: \((\text{\#green, \#other}) = \text{PolyaUrn}(1, \alpha)\)
- not orange, green: \((\text{\#red, \#other}) = \text{PolyaUrn}(1, \alpha)\)
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn

• Choose ball with prob proportional to its mass
  • If black, replace and add ball of new color
  • Else, replace and add ball of same color

\[ (\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha) \]

\[ V_k \overset{iid}{\sim} \text{Beta}(1, \alpha) \]

\[ \rho_1 = V_1 \]
\[ \rho_2 = (1 - V_1)V_2 \]
\[ \rho_3 = \prod_{k=1}^{2} (1 - V_k)]V_3 \]

• not orange: \((\#\text{green, \#other}) = \text{PolyaUrn}(1, \alpha)\)
• not orange, green: \((\#\text{red, \#other}) = \text{PolyaUrn}(1, \alpha)\)
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
- Sits at existing table with prob proportional to # people there
- Forms new table with prob proportional to $\alpha$
- Marginal for the Categorical likelihood with GEM prior

**Partition of $[8]$**: set of mutually exclusive & exhaustive sets

$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$

$\uparrow 8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}$

$[8] = \{1, \ldots, 8\}$
Chinese restaurant process

- Same thing we just did
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$

\[ \mathbb{P}(Z) = \frac{1}{Z} \prod_{j=1}^{K} \left( \frac{\alpha_j}{\sum_{k=1}^{K} \alpha_k} \right) \]

\[ \mathbb{P}(Z|\mathcal{D}) = \frac{1}{Z} \prod_{j=1}^{K} \left( \frac{\alpha_j}{\sum_{k=1}^{K} \alpha_k} \right) \]
Chinese restaurant process

1. Same thing we just did
2. Each customer walks into the restaurant
   - Sits at existing table with prob proportional to # people there
   - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

1. Same thing we just did
2. Each customer walks into the restaurant
   - Sits at existing table with prob proportional to # people there
   - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

1

\( \phi_1 \)

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to \# people there
  - Forms new table with prob proportional to \( \alpha \)
Chinese restaurant process

• Same thing we just did
• Each customer walks into the restaurant
  • Sits at existing table with prob proportional to # people there
  • Forms new table with prob proportional to $\alpha$
Chinese restaurant process

• Same thing we just did
• Each customer walks into the restaurant
  • Sits at existing table with prob proportional to # people there
  • Forms new table with prob proportional to $\alpha$
Chinese restaurant process

• Same thing we just did
• Each customer walks into the restaurant
  • Sits at existing table with prob proportional to # people there
  • Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

• Same thing we just did
• Each customer walks into the restaurant
  • Sits at existing table with prob proportional to # people there
  • Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
Chinese restaurant process

• Same thing we just did
• Each customer walks into the restaurant
  • Sits at existing table with prob proportional to # people there
  • Forms new table with prob proportional to $\alpha$
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to \# people there
  - Forms new table with prob proportional to \( \alpha \)
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$

[Aldous 1983]
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
- Marginal for the Categorical likelihood with GEM prior
Chinese restaurant process

- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
- Marginal for the Categorical likelihood with GEM prior

So far: Dirichlet process, Chinese restaurant process

- Infinity of parameters, growing number of parameters
Roadmap

• Example problem: clustering
• Example NPBayes model: Dirichlet process
• Chinese restaurant process
• Inference
• Venture further into the wild world of Nonparametric Bayes

• Big questions
  • Why NPBayes?
  • What does an infinite/growing number of parameters really mean (in NPBayes)?
  • Why is NPBayes challenging but practical?
Roadmap

• Example problem: clustering
• Example NPBayes model: Dirichlet process
• Chinese restaurant process
• Inference
• Venture further into the wild world of Nonparametric Bayes

• Big questions
  • Why NPBayes?
  • What does an infinite/growing number of parameters really mean (in NPBayes)?
  • Why is NPBayes challenging but practical?
Roadmap

• Example problem: clustering
• Example NPBayes model: Dirichlet process
• Chinese restaurant process
• Inference
• Venture further into the wild world of Nonparametric Bayes

• Big questions
  • Why NPBayes?
  • What does an infinite/growing number of parameters really mean (in NPBayes)?
  • Why is NPBayes challenging but practical?
Roadmap

• Example problem: clustering
• Example NPBayes model: Dirichlet process
• Chinese restaurant process
• Inference
• Venture further into the wild world of Nonparametric Bayes

• Big questions
  • Why NPBayes?
  • What does an infinite/growing number of parameters really mean (in NPBayes)?
  • Why is NPBayes challenging but practical?
Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes

Big questions
- Why NPBayes? Learn more as acquire more data
- What does an infinite/growing number of parameters really mean (in NPBayes)?
- Why is NPBayes challenging but practical?
Roadmap

• Example problem: clustering
• Example NPBayes model: Dirichlet process
• Chinese restaurant process
• Inference
• Venture further into the wild world of Nonparametric Bayes

• Big questions
  • Why NPBayes? Learn more as acquire more data
  • What does an infinite/growing number of parameters really mean (in NPBayes)? Components vs. clusters; latent vs. realized
  • Why is NPBayes challenging but practical?
Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes

Big questions
- Why NPBayes? Learn more as acquire more data
- What does an infinite/growing number of parameters really mean (in NPBayes)? Components vs. clusters; latent vs. realized
- Why is NPBayes challenging but practical? Infinite dimensional parameter, but finitely many parameters realized
Chinese restaurant process

- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
- Marginal for the Categorical likelihood with GEM prior
Chinese restaurant process

- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to $\alpha$
- Marginal for the Categorical likelihood with GEM prior
  $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$
Chinese restaurant process

- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to \( \alpha \)
- Marginal for the Categorical likelihood with GEM prior

\[
\begin{align*}
\phi_1 & = \{1, 2, 7, 8\}, \\
\phi_2 & = \{3, 5, 6\}, \\
\phi_3 & = \{4\}
\end{align*}
\]
Each customer walks into the restaurant
• Sits at existing table with prob proportional to # people there
• Forms new table with prob proportional to $\alpha$
• Marginal for the Categorical likelihood with GEM prior
  \[ z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3 \]
  \[ \implies \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\} \]
• *Partition of [8]*: set of mutually exclusive & exhaustive sets of [8] := {1, \ldots, 8}
Chinese restaurant process

- Probability of this seating:
Chinese restaurant process

- Probability of this seating:

\[
\frac{\alpha}{\bar{\alpha}}
\]
Chinese restaurant process

• Probability of this seating:

\[ \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \]
Chinese restaurant process

- Probability of this seating:
\[
\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2}
\]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3}
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4}
  \]
Chinese restaurant process

- Probability of this seating:

\[
\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5}
\]
Chinese restaurant process

- Probability of this seating:
\[
\frac{\alpha}{\alpha + 6} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6}
\]
Chinese restaurant process

- Probability of this seating:

\[
\alpha \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
\]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)): 
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)):
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)):
  \[
  \frac{\alpha \cdots (\alpha + N - 1)}{\alpha \cdots (\alpha + N - 1)}
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)):
  \[
  \frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)):
  \[
  \frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]
- Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)):
  \[
  \frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{3}{\alpha+6} \cdot \frac{2}{\alpha+7}
  \]

- Probability of \(N\) customers (\(K_N\) tables, \(n_k\) at table \(k\)):
  \[
  \alpha^{K_N} \over \alpha \cdots (\alpha + N - 1)
  \]
• Probability of this seating:

\[
\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
\]

• Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)):

\[
\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}
\]
Chinese restaurant process

- Probability of this seating:
  \[ \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7} \]

- Probability of \( N \) customers (\( K_N \) tables, \( n_k \) at table \( k \)):
  \[ \frac{\alpha^{K_N} \prod_{k=1}^{K_N} (n_k - 1)!}{\alpha \cdots (\alpha + N - 1)} \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \( \#C \) at table \( C \)):
  \[
  \alpha^{K_N} \frac{\prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)}
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \#\( C \) at table \( C \)):
  \[
  \alpha^{K_N} \prod_{C \in \Pi_N} (#C - 1)! \cdot \frac{\alpha \cdots (\alpha + N - 1)}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]
- Probability of \( N \) customers (\( K_N \) tables, \#\( C \) at table \( C \)):
  \[
  \alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)! \quad \frac{1}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]
- Prob doesn’t depend on customer order: exchangeable
Chinese restaurant process

1. Probability of this seating:
\[
\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
\]

2. Probability of \( N \) customers (\( K_N \) tables, \( \#C \) at table \( C \)):
\[
\alpha^{K_N} \frac{\prod_{C \in \Pi_N}(\#C - 1)!}{\alpha \cdot \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
\]

3. Prob doesn’t depend on customer order: exchangeable
\[
\mathbb{P}(\Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}) = \mathbb{P}(\Pi_8 = \{\{2, 3, 8, 1\}, \{4, 6, 7\}, \{5\}\})
\]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}
  \]

- Probability of \( N \) customers (\( K_N \) tables, \( \#C \) at table \( C \)):
  \[
  \alpha^{K_N} \frac{\prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]

- Prob doesn’t depend on customer order: exchangeable
  \[
  \mathbb{P}(\Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}) = \mathbb{P}(\Pi_8 = \{\{2, 3, 8, 1\}, \{4, 6, 7\}, \{5\}\})
  \]

- Can always pretend \( n \) is the last customer and calculate
  \[
  p(\Pi_N | \Pi_{N,-n})
  \]
Chinese restaurant process

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]
- Probability of \( N \) customers (\( K_N \) tables, \( \#C \) at table \( C \)):
  \[
  \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdot \cdots (\alpha + N - 1)} = P(\Pi_N = \pi_N)
  \]
- Prob doesn’t depend on customer order: exchangeable
  \[
  P(\Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}) = P(\Pi_8 = \{\{2, 3, 8, 1\}, \{4, 6, 7\}, \{5\}\})
  \]
- Can always pretend \( n \) is the last customer and calculate
  \[
  p(\Pi_N | \Pi_{N,n})
  \]
  - e.g. \( \Pi_{8,-5} = \{\{1, 2, 7, 8\}, \{3, 6\}, \{4\}\} \)
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):
  \[
  \alpha^{K_N} \frac{\prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]
- So:
  \[
p(\Pi_N | \Pi_{N,-n}) =
  \]
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, #C at table $C$):
  \[
  \alpha^{K_N} \frac{\prod_{C \in \Pi_N} (#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]

- So:
  \[
  p(\Pi_N|\Pi_{N,-n}) = \left\{
  \right.
  \]
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):
  \[
  \frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)} \prod_{C \in \Pi_N} (\#C - 1)! = \mathbb{P}(\Pi_N = \pi_N)
  \]

- So:
  \[
  p(\Pi_N | \Pi_{N,-n}) = \begin{cases} 
    \text{if } n \text{ joins cluster } C \\
    \text{if } n \text{ starts a new cluster}
  \end{cases}
  \]
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):
  \[
  \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]

- So:
  \[
  p(\Pi_N | \Pi_{N,-n}) = \begin{cases} 
  \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\
  \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster}
  \end{cases}
  \]
• Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

• So:

$$p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, $#C$ at table $C$):
  \[
  \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} \cdot \mathbb{P}(\Pi_N = \pi_N)
  \]
- So:
  \[
  p(\Pi_N | \Pi_{N,-n}) = \begin{cases} 
  \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\
  \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster}
  \end{cases}
  \]
- Gibbs sampling review:
Chinese restaurant process

1. Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$): 
   \[
   \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
   \]

2. So: 
   \[
   p(\Pi_N | \Pi_{N,-n}) = \begin{cases} 
   \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\
   \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster}
   \end{cases}
   \]

3. Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):
  \[ \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N) \]

- So:
  \[ p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases} \]

- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):
  $$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$
- So:
  $$p(\Pi_N | \Pi_N, -n) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$
- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - $t^{th}$ step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$
Chinese restaurant process

- Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- So:

$$p(\Pi_N|\Pi_N,-n) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases}$$

- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$

- Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$

- $t^{th}$ step: $v_1^{(t)} \sim p(v_1|v_2^{(t-1)}, v_3^{(t-1)})$
Chinese restaurant process

• Probability of $N$ customers ($K_N$ tables, $#C$ at table $C$):

\[
\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)! \frac{\alpha \cdots (\alpha + N - 1)}{\alpha + N - 1} = \mathbb{P}(\Pi_N = \pi_N)
\]

• So:

\[
p(\Pi_N | \Pi_N, -n) = \begin{cases} 
\frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\
\frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster}
\end{cases}
\]

• Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$

  Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$

  $t^{th}$ step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

• $v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$

• $v_3^{(t)} \sim p(v_3 | v_1^{(t)}, v_2^{(t)})$
CRP mixture model: inference
CRP mixture model: inference

- Data $x_{1:N}$
CRP mixture model: inference

- Data $x_{1:N}$
CRP mixture model: inference

- Data $x_{1:N}$
- Generative model
CRP mixture model: inference

• Data $x_1:N$

• Generative model

$\Pi_N \sim \text{CRP}(N, \alpha)$
CRP mixture model: inference

- Data $x_{1:N}$
- Generative model $\Pi_N \sim \text{CRP}(N, \alpha)$
CRP mixture model: inference

- **Data** \( x_{1:N} \)
- **Generative model**
  \[
  \Pi_N \sim \text{CRP}(N, \alpha) \\
  \forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)
  \]
CRP mixture model: inference

• Data $x_{1:N}$

• Generative model
  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$
CRP mixture model: inference

- Data $x_1:N$
- Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0) \]
CRP mixture model: inference

- **Data** $x_1:N$

- **Generative model**

  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]

  \[ \forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0) \]

  \[ \forall C \in \Pi_N, \forall n \in C, x_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu_C, \Sigma) \]
CRP mixture model: inference

- Data $x_{1:N}$

- Generative model

  $\Pi_N \sim \text{CRP}(N, \alpha)$

  $\forall C \in \Pi_N, \mu_C \sim N(\mu_0, \Sigma_0)$

  $\forall C \in \Pi_N, \forall n \in C, x_n \sim N(\mu_C, \Sigma)$
CRP mixture model: inference

- Data $x_{1:N}$

- Generative model
  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0)$
  $\forall C \in \Pi_N, \forall n \in C, x_n \overset{iid}{\sim} \mathcal{N}(\mu_C, \Sigma)$

- Want: posterior
CRP mixture model: inference

- Data $x_{1:N}$
- Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0) \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \overset{\text{indep}}{\sim} \mathcal{N}(\mu_C, \Sigma) \]

- Want: posterior $p(\Pi_N|x_{1:N})$
CRP mixture model: inference

- Data \( x_{1:N} \)

- Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0) \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma) \]

- Want: posterior \( p(\Pi_N | x_{1:N}) \)

- Gibbs sampler:
CRP mixture model: inference

- Data $x_{1:N}$
- Generative model
  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0)$
  $\forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma)$

- Want: posterior $p(\Pi_N|x_{1:N})$
- Gibbs sampler:

$$p(\Pi_N|\Pi_{N,-n}, x)$$
CRP mixture model: inference

- **Data** $x_{1:N}$
- **Generative model**
  \[ 
  \begin{align*}
  \Pi_N & \sim \text{CRP}(N, \alpha) \\
  \forall C \in \Pi_N, \mu_C & \sim \mathcal{N}(\mu_0, \Sigma_0) \\
  \forall C \in \Pi_N, \forall n \in C, x_n & \overset{\text{iid}}{\sim} \mathcal{N}(\mu_C, \Sigma)
  \end{align*}
  \]
- **Want:** posterior $p(\Pi_N|x_{1:N})$
- **Gibbs sampler:**
  \[
  p(\Pi_N|\Pi_{N,-n}, x) = \left\{ \right. \]

\[
  \begin{align*}
  \mu_1 & \\
  \mu_2 & \\
  \mu_3 & \\
  1 & \\
  2 & \\
  3 & \\
  4 & \\
  5 & \\
  6 & \\
  7 & \\
  8 & \\
  \end{align*}
  \]
CRP mixture model: inference

- Data $x_{1:N}$
- Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0) \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma) \]

- Want: posterior $p(\Pi_N|x_{1:N})$

- Gibbs sampler:
  \[ p(\Pi_N|\Pi_{N,-n}, x) = \left\{ \begin{array}{ll}
  \frac{1}{\text{if } n \text{ joins cluster } C} & \\
  & \\
\end{array} \right. \]
CRP mixture model: inference

- Data \(x_1 : N\)

- Generative model
  \[\Pi_N \sim \text{CRP}(N, \alpha)\]
  \[\forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0)\]
  \[\forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma)\]

- Want: posterior \(p(\Pi_N | x_1 : N)\)

- Gibbs sampler:
  \[p(\Pi_N | \Pi_{N,-n}, x) = \{\]
  \[\text{if } n \text{ joins cluster } C\]
  \[\text{if } n \text{ starts a new cluster}\]
CRP mixture model: inference

- **Data** \( x_{1:N} \)
- **Generative model**
  \[
  \Pi_N \sim \text{CRP}(N, \alpha) \\
  \forall C \in \Pi_N, \mu_C \sim \mathcal{N} (\mu_0, \Sigma_0) \\
  \forall C \in \Pi_N, \forall n \in C, x_n \sim \text{indep} \mathcal{N} (\mu_C, \Sigma)
  \]

- Want: posterior \( p(\Pi_N | x_{1:N}) \)
- Gibbs sampler:
  \[
  p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} 
  \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\
  & \text{if } n \text{ starts a new cluster}
  \end{cases}
  \]
CRP mixture model: inference

- Data \( x_{1:N} \)

- Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0) \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma) \]

- Want: posterior \( p(\Pi_N | x_{1:N}) \)

- Gibbs sampler:

\[
p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} 
\frac{\#C}{\alpha + N-1} p(x_{\mathcal{C} \cup \{n\}} | x_{\mathcal{C}}) & \text{if } n \text{ joins cluster } C \\
\frac{\alpha}{\alpha + N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster}
\end{cases}
\]
CRP mixture model: inference

- **Data** \( x_{1:N} \)

- **Generative model**
  - \( \Pi_N \sim \text{CRP}(N, \alpha) \)
  - \( \forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0) \)
  - \( \forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma) \)

- **Want**: posterior \( p(\Pi_N| x_{1:N}) \)

- **Gibbs sampler**:

\[
p(\Pi_N|\Pi_{N,-n}, x) = \begin{cases} 
\frac{\#C}{\alpha + N - 1} p(x_{C\cup\{n\}}|x_C) & \text{if } n \text{ joins cluster } C \\
\frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster}
\end{cases}
\]

- **For completeness**: \( p(x_{C\cup\{n\}}|x_C) = \)
CRP mixture model: inference

• Data $x_{1:N}$

• Generative model

  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0)$
  $\forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma)$

• Want: posterior $p(\Pi_N|x_{1:N})$

• Gibbs sampler:

  \[
  p(\Pi_N|\Pi_{N,-n}, x) = \begin{cases} 
  \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}}|x_C) & \text{if } n \text{ joins cluster } C \\
  \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster}
  \end{cases}
  \]

• For completeness: $p(x_{C \cup \{n\}}|x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$
CRP mixture model: inference

- Data $x_{1:N}$
- Generative model
  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$
  $\forall C \in \Pi_N, \forall n \in C, x_n \overset{\text{indep}}{\sim} \mathcal{N}(\mu_C, \Sigma)$

- Want: posterior $p(\Pi_N|x_{1:N})$

- Gibbs sampler:
  
  $$p(\Pi_N|\Pi_{N,-n}, x) = \begin{cases} 
  \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}}|x_C) & \text{if } n \text{ joins cluster } C \\
  \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster}
  \end{cases}$$

- For completeness: $p(x_{C\cup\{n\}}|x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$
  
  $$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
  $$\tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)$$
CRP mixture model: inference

• Data $x_{1:N}$

• Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0) \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma) \]

• Want: posterior $p(\Pi_N|x_{1:N})$

• Gibbs sampler:
  \[
p(\Pi_N|\Pi_{N,-n}, x) = \begin{cases} 
\frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}}|x_C) & \text{if } n \text{ joins cluster } C \\
\frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster}
\end{cases}
\]

• For completeness: $p(x_{C\cup\{n\}}|x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$
  \[
  \tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (#C)\Sigma^{-1} \\
  \tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)
\]

[MacEachern 1994; Neal 1992; Neal 2000]
CRP mixture model: inference

- Data \( x_{1:N} \)

- Generative model
  \[
  \Pi_N \sim \text{CRP}(N, \alpha) \quad \forall C \in \Pi_N, \phi_C \sim G_0 \\
  \forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma) \\
  \]

- Want: posterior \( p(\Pi_N|x_{1:N}) \)

- Gibbs sampler:
  \[
  p(\Pi_N|\Pi_{N,-n}, x) = \left\{ \begin{array}{ll}
  \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}}|x_C) & \text{if } n \text{ joins cluster } C \\
  \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \\
  \end{array} \right.
  \]

- For completeness:
  \[
  p(x_{C \cup \{n\}}|x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma) \\
  \tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (#C)\Sigma^{-1} \\
  \tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)
  \]

[MacEachern 1994; Neal 1992; Neal 2000]
CRP mixture model: inference

- **Data** $x_{1:N}$

- **Generative model**
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \phi_C \sim \text{iid } G_0 \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \overset{\text{indep}}{\sim} F(\phi_C) \]

- **Want:** posterior $p(\Pi_N | x_{1:N})$

- **Gibbs sampler:**
  \[
p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} 
    \frac{\#C}{\alpha+N-1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\
    \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster}
  \end{cases}
\]

- **For completeness:**
  \[
p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)
\]
  \[
  \tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (#C)\Sigma^{-1}
  \]
  \[
  \tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)
\]
CRP mixture model: inference

• Data $x_{1:N}$

• Generative model

  $\Pi_N \sim \text{CRP}(N, \alpha)$

  $\forall C \in \Pi_N, \phi_C \sim \text{iid } G_0$

  $\forall C \in \Pi_N, \forall n \in C, x_n \sim \text{indep } F(\phi_C)$

• Want: posterior $p(\Pi_N | x_{1:N})$

• Gibbs sampler:

  
  $$p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} 
  \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\
  \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster}
  \end{cases}$$

[MacEachern 1994; Neal 1992; Neal 2000]
CRP mixture model: inference

- Data $x_{1:N}$
- Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0) \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma) \]

- Want: posterior $p(\Pi_N|x_{1:N})$

- Gibbs sampler:
  \[ p(\Pi_N|\Pi_{N,-n}, x) = \begin{cases} 
  \frac{\#C}{\alpha+N-1} p(x_{C \cup \{n\}}|x_C) & \text{if } n \text{ joins cluster } C \\
  \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} 
\end{cases} \]

- For completeness:
  \[ p(x_{C \cup \{n\}}|x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma) \]
  \[ \tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (#C)\Sigma^{-1} \]
  \[ \tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right) \]

[MacEachern 1994; Neal 1992; Neal 2000]
CRP mixture model: inference

• Data $x_{1:N}$

• Generative model

$\Pi_N \sim \text{CRP}(N, \alpha)$

$\forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0)$

$\forall C \in \Pi_N, \forall n \in C, x_n \sim \text{indep} \; \mathcal{N}(\mu_C, \Sigma)$

• Want: posterior $p(\Pi_N|x_{1:N})$

• Gibbs sampler:

$p(\Pi_N|\Pi_{N,-n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}}|x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$

• For completeness: $p(x_{C\cup\{n\}}|x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (#C)\Sigma^{-1}$

$\tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)$

[MacEachern 1994; Neal 1992; Neal 2000]
More Markov Chain Monte Carlo
More Markov Chain Monte Carlo

• Slice sampling
More Markov Chain Monte Carlo

• Slice sampling
  • auxiliary variable $\rightarrow$ finite conditionals
More Markov Chain Monte Carlo

• Slice sampling
• auxiliary variable $\rightarrow$ finite conditionals
More Markov Chain Monte Carlo

- Slice sampling
- auxiliary variable $\rightarrow$ finite conditionals
More Markov Chain Monte Carlo

• Slice sampling
  • auxiliary variable $\rightarrow$ finite conditionals

• Approximate with truncated distribution

[Ishwaran, James 2001; Campbell*, Huggins*, Broderick 2016]
More Markov Chain Monte Carlo

• Slice sampling
  • auxiliary variable $\rightarrow$ finite conditionals

• Approximate with truncated distribution
  • E.g., Hamiltonian Monte Carlo

[Ishwaran, James 2001; Campbell*, Huggins*, Broderick 2016]
Variational Bayes
Variational Bayes

- Variational Bayes (VB)

Variational Bayes (VB) is an approximation method for posterior distribution. It minimizes Kullback-Leibler (KL) divergence by factorizing the posterior distribution into a product of simpler distributions. This is particularly useful when dealing with exponential family distributions and truncation.
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
- “Close”: Minimize Kullback-Liebler (KL) divergence:
  $$KL(q||p(\cdot|x))$$
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
- “Close”: Minimize Kullback-Liebler (KL) divergence:
  $$KL(q||p(\cdot|x))$$
Variational Bayes

- Variational Bayes (VB)
- Approximation \( q^*(\theta) \) for posterior \( p(\theta|x) \)
- “Close”: Minimize Kullback-Liebler (KL) divergence:
  \[
  KL(q\|p(\cdot|x))
  \]
- “Nice”: factorizes, exponential family, truncation
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
- “Close”: Minimize Kullback-Liebler (KL) divergence:
  $$KL(q||p(\cdot|x))$$
- “Nice”: factorizes, exponential family, truncation
- VB practical success
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
- “Close”: Minimize Kullback-Liebler (KL) divergence:
  $$KL(q\|p(\cdot|x))$$
- “Nice”: factorizes, exponential family, truncation

- VB practical success
  - point estimates and prediction
Variational Bayes

- Variational Bayes (VB)
- Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
- “Close”: Minimize Kullback-Liebler (KL) divergence:
  $$KL(q||p(\cdot|x))$$
- “Nice”: factorizes, exponential family, truncation

- VB practical success
  - point estimates and prediction
  - fast, streaming, distributed

[Broderick, Boyd, Wibisono, Wilson, Jordan 2013]
Exercises

• Data $x_{1:N}$
Exercises

- Code a CRP mixture model simulator

Data $x_{1:N}$
Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
Exercises

• Code a CRP mixture model simulator
• Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
• Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
• Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
Exercises

• Code a CRP mixture model simulator
• Derive the CRP mixture model Gibbs sampler in the slides; derive \( p(x_{C \cup \{n\}} | x_C) \) explicitly for a Gaussian mixture
• Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
• Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
• Read [Neal 2000] and code a DPMM Gibbs sampler
Exercises

• Code a CRP mixture model simulator
• Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
• Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
• Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
• Read [Neal 2000] and code a DPMM Gibbs sampler
• Read [Walker 2007; Kalli, Griffin, Walker 2011] and code a DPMM slice sampler
Exercises

• Code a CRP mixture model simulator
• Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
• Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
• Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
• Read [Neal 2000] and code a DPMM Gibbs sampler
• Read [Walker 2007; Kalli, Griffin, Walker 2011] and code a DPMM slice sampler
• Read [Blei, Jordan 2006] and code variational inference for the DPMM
# 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><strong>Document 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Document 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Document 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Document 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Document 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Document 6</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Document 7</strong></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
Feature allocation

- Indian buffet process

Griffiths, Ghahramani 2005
Feature allocation

- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005]
Feature allocation

- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005, Hjort 1990]
Feature allocation

- Indian buffet process
- Beta process

Power laws
Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
Power laws

- $K_N := \#$ clusters occupied by $N$ data points
- CRP: $K_N \sim \alpha \log N$ w.p. 1
Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N$ w.p. 1
  - vs. Heaps’ law, Herdan’s law, etc

$K_N \sim \alpha \log N$ w.p. 1, $\sim \frac{N}{j}$ w.p. 1, $\sim C_{\times j}$, $j \rightarrow 1$, w.p. 1

Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N$ w.p. 1
- vs. Heaps’ law, Herdan’s law, etc

[Gnedin, et al 2007]
Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N \text{ w.p. 1}$
  - vs. Heaps’ law, Herdan’s law, etc
- Pitman-Yor process:

[Gnedin, et al 2007]
Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N \text{ w.p. 1}$
  - vs. Heaps’ law, Herdan’s law, etc
- Pitman-Yor process:

Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N$ w.p. 1
- vs. Heaps’ law, Herdan’s law, etc
- Pitman-Yor process:
  
  $K_N \sim S_\alpha N^\sigma$ w.p. 1

Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N$ w.p. 1
  - vs. Heaps’ law, Herdan’s law, etc
- Pitman-Yor process:
  
  
  $K_N \sim S_\alpha N^\sigma$ w.p. 1

Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N \text{ w.p. 1}$
  - vs. Heaps’ law, Herdan’s law, etc
- Pitman-Yor process:
  - $K_N \sim S_{\alpha} N^\sigma \text{ w.p. 1}$
  - related to Zipf’s law (ranked frequencies)

Power laws

- $K_N := \# \text{ clusters occupied by } N \text{ data points}$
- CRP: $K_N \sim \alpha \log N$ w.p. 1
  - vs. Heaps’ law, Herdan’s law, etc
- Pitman-Yor process:
  - $K_N \sim S_\alpha N^\sigma$ w.p. 1
  - related to Zipf’s law (ranked frequencies)
- Not just clusters

Hierarchies
Hierarchies

- Hierarchical Dirichlet process

Hierarchies

- Hierarchical Dirichlet process

[Teh et al 2006, Rodríguez et al 2008]
Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008]
Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008]
Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise
- Hierarchical beta process

[Teh et al 2006]
Hierarchies

• Hierarchical Dirichlet process
• Chinese restaurant franchise
• Hierarchical beta process

Genealogy, trees, beyond trees

[Wakeley 2008]
Genealogy, trees, beyond trees

- Kingman coalescent

[Diagram of a genealogical tree with labels 1 to 9 and nodes T2, T3, T4.

[Wakeley 2008]]
Genealogy, trees, beyond trees

- Kingman coalescent
Genealogy, trees, beyond trees

- Kingman coalescent
- Fragmentation
- Coagulation

1. Genealogy, trees, beyond trees
2. Kingman coalescent
3. Fragmentation
4. Coagulation

[Wakeley 2008]
[Kingman 1982]
Genealogy, trees, beyond trees

- Kingman coalescent
- Fragmentation
- Coagulation

[Figure credit: Wakeley 2008, Kingman 1982, Bertoin 2006, Teh et al 2011]
Genealogy, trees, beyond trees

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

Genealogy, trees, beyond trees

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

Conjugacy & Poisson point processes

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

Posteriors, conjugacy, and exponential families for completely random measures


Orbanz 2009, Orbanz 2010
Conjugacy & Poisson point processes

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

Conjugacy & Poisson point processes

• Beta process, Bernoulli process (Indian buffet)

• Gamma process, Poisson likelihood process (DP, CRP)
Conjugacy & Poisson point processes

- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process
Conjugacy & Poisson point processes

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

Conjugacy & Poisson point processes

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

Conjugacy & Poisson point processes

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

• Posteriors, conjugacy, and exponential families for completely random measures

Conjugacy & Poisson point processes

• Beta process, Bernoulli process (Indian buffet)

• Gamma process, Poisson likelihood process (DP, CRP)

• Beta process, negative binomial process

• Posteriors, conjugacy, and exponential families for completely random measures

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian
  \[ P(\text{parameters}|\text{data}) \propto P(\text{data}|\text{parameters})P(\text{parameters}) \]
- Not parametric (i.e. not finite parameter, unbounded/growing/infinite number of parameters)

[Wikipedia](http://en.wikipedia.org)
[Ed Bowlby, NOAA](http://example.com)
[Fox, et al 2014](http://example.com)
[Arjas, Gasbarra 1994](http://example.com)
[Escobar, West 1995; Ghosal, et al 1999](http://example.com)
[Saria et al 2010](http://example.com)
[Ewens, 1972; Hartl, Clark 2003](http://example.com)
[Sudderth, Jordan 2009](http://example.com)
References (page 1 of 4)


References (page 2 of 4)


