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

[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

[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]
[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

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

\[
\begin{align*}
\text{Step 0} & & \text{Step 1} & & \text{Step 2} & & \text{Step 3} & & \text{Step 4} \\
\bullet & & \bullet & & \bullet & & \bullet & & \bullet \\
\end{align*}
\]

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

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

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

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

Step 1

Step 2

Step 3

Step 4

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

Step 0
Step 1
Step 2
Step 3
Step 4

(#orange, #other) = PolyaUrn(1, \alpha)

• not orange: (#green, #other) = PolyaUrn(1, \alpha)
• not orange, green: (#red, #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

\[
\begin{align*}
\text{Step 0} & \quad \text{Step 1} & \quad \text{Step 2} & \quad \text{Step 3} & \quad \text{Step 4} \\
\black & \quad \bullet & \quad \bullet \bullet & \quad \bullet \bullet \bullet & \quad \bullet \bullet \bullet \\
\end{align*}
\]

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

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

- 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

Step 0 | Step 1 | Step 2 | Step 3 | Step 4
--- | --- | --- | --- | ---
• | • | • | • | •

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

$\rho_1 = V_1$

$(\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- 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

\[
\begin{align*}
(V_k)_{iid} & \sim \text{Beta}(1, \alpha) \\
\rho_1 &= V_1 \\
\rho_2 &= (1 - V_1)V_2
\end{align*}
\]

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

Step 0

Step 1

Step 2

Step 3

Step 4

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

\( \rho_1 = V_1 \)

\( \rho_2 = (1 - V_1)V_2 \)

\( \rho_3 = \prod_{k=1}^{2} (1 - V_k) V_3 \)

(#orange, #other) = PolyaUrn(1, \alpha)

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

- not orange, green: (#red, #other) = PolyaUrn(1, \alpha)
Chinese restaurant process

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

\[ \{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\} \]

$\mathcal{Z} = \{1, \ldots, 8\}$

\[ \uparrow \]

\[ \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\} \]
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 $\alpha$
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

- 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 $\alpha$

\[\phi_1\]

\[\mathbb{P}(\mathcal{Z}) = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}\]
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 $\alpha$

\[ \phi_1 \]

1

Partition of 8: set of mutually exclusive & exhaustive sets

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

\[ [8] = \{1, \ldots, 8\} \]
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 $\alpha$

\[\phi_1, \phi_2, \phi_3, \phi_4\]
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 $\alpha$

\[ [8] = \{\{1,2,7,8\}, \{3,5,6\}, \{4\}\} \]

\[ [\phi_1, \phi_2, \phi_3] = [1,2,7,8] \]

[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
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
  
  $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$
• 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
  $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$
  $\Rightarrow \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}$
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

$$\begin{align*}
z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3 \\
\Rightarrow \Pi_8 &= \left\{ \{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\} \right\}
\end{align*}$$

- *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:
  \[
  \frac{\alpha}{k_n}\]

\begin{align*}
\phi_1 & \quad 7 & \quad \phi_2 & \quad 6 & \quad \phi_3 & \quad 4 \\
1 & \quad 7 & \quad 2 & \quad 6 & \quad 3 & \quad 4 \\
2 & \quad 5 & \quad 5 & \quad 5 & \quad 5 & \quad 5 \\
8 & \quad 5 & \quad 8 & \quad 8 & \quad 8 & \quad 8 \\
\end{align*}
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:
\[
\frac{1}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4}
\]
Chinese restaurant process

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

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

- Probability of this seating:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
  \]
• Probability of 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 \( k \)):
  \[
  \frac{1}{\alpha \cdots (\alpha + N - 1)}
  \]
• 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 (\(K\) 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 \( k \)):
  \[
  \frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}
  \]
• Probability of this seating:
\[
\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{3}{\alpha + 6} \cdot \frac{\alpha + 7}{\alpha + 8}
\]

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

- Prob doesn’t depend on customer order: exchangeable
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 \( C \)):
  \[
  \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (#C - 1)!}{\alpha \cdot \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]

- Prob doesn’t depend on customer order: exchangeable
  \[
  \mathbb{P}(\Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}) = \mathbb{P}(\Pi_8 = \{\{2, 3, 8, 1\}, \{4, 6, 7\}, \{5\}\})
  \]
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 \(C\)):
  \[
  \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]

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

- 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
  \[
  p(\Pi_N | \Pi_{N,-n})
  \]
  - 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 $C$):
  \[
  \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdot \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]
- So:
  \[
p(\Pi_N | \Pi_{N,-n}) = \]

\[
\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)! \quad \alpha \cdot \cdots (\alpha + N - 1)
\]
Chinese restaurant process

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

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

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

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

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

- So:
  \[
  p(\Pi_N | \Pi_{N,-n}) = \begin{cases} 
  \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\
  \alpha^{K_{N-1}} \frac{\prod_{C \in \Pi_{N-1}} (\#C - 1)!}{\alpha \cdots (\alpha + N - 2)} & \text{if } n \text{ starts a new cluster}
  \end{cases}
  \]
Chinese restaurant process

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

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

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

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

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

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

• 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: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$, $v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$, $v_3^{(t)} \sim p(v_3 | v_1^{(t)}, v_2^{(t)})$
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
  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0)$

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

- Data $x_1:N$

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

- Data $x_{1:N}$

- Generative model

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

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

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

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

• Want: posterior $p(\Pi_N|x_{1:N})$

• Gibbs sampler:
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**:

  \[ p(\Pi_N | \Pi_{N,-n}, x) \]
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:
  \[ p(\Pi_N|\Pi_{N,-n}, x) \propto \left\{ \right\} \]
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:**
  
  $p(\Pi_N|\Pi_{N,-n}, x) \propto \left\{ \begin{array}{ll}
  \text{if } n \text{ joins cluster } C
  \end{array} \right.$
CRP mixture model: inference

• Data \( x_{1:N} \)

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

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

• Gibbs sampler:

\[
p(\Pi_N | \Pi_{N,-n}, x) \propto \begin{cases} & \text{if } n \text{ joins cluster } C \\
& \text{if } n \text{ starts a new cluster} 
\end{cases}
\]
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:**

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

- Data $x_1:N$
- Generative model
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0) \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \overset{\text{indep}}{\sim} \mathcal{N}(\mu_C, \Sigma) \]

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

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

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

• 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:

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

- 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:

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

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

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

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

• 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
  \[
  \Pi_N \sim \text{CRP}(N, \alpha) \\
  \forall C \in \Pi_N, \phi_C \sim G_0^{\text{iid}} \\
  \forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma)
  \]

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

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

- 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

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

  $\forall C \in \Pi_N, \phi_C \sim G_0$

  $\forall C \in \Pi_N, \forall n \in C, x_n \sim F(\phi_C)$

• Want: posterior $p(\Pi_N|x_{1:N})$

• Gibbs sampler:

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

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

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

- **Gibbs sampler:**

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

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

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

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

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

- 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

- 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 $p(\theta | x)$

$p(\theta | x)$

$q^*(\theta)$
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

$p(\theta|x)$
$q^*(\theta)$
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
Exercises

- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive \( p(x_{C \cup \{n\}} \mid x_C) \) 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; Kall, 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
Power laws
Hierarchies
Dependencies
Coalescents/Diffusions/Trees
de Finetti
Feature allocations
Networks/graphs
Poisson processes
Here be Dragons
## Clustering

<table>
<thead>
<tr>
<th>Document 1</th>
<th>Arts</th>
<th>Econ</th>
<th>Sports</th>
<th>Health</th>
<th>Technology</th>
</tr>
</thead>
<tbody>
<tr>
<td></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>
# Feature allocation

<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>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>
## Feature allocation

<table>
<thead>
<tr>
<th>Document 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts</td>
<td>E</td>
<td>n</td>
<td>o</td>
<td></td>
</tr>
<tr>
<td>Economics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Indian buffet process
### Feature allocation

- Indian buffet process

<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>
## 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>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Document 2</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Document 3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Document 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Document 5</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Document 6</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Document 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005]
# 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>

- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005, Hjort 1990]
### 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
- Beta process

Power laws
Power laws

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

$K_N \sim \alpha \log N$ w.p. 1,

$K_N \sim N \alpha$ w.p. 1,

$\# j \sim C(j)$, $j \rightarrow 1$ w.p. 1

Power laws

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

\[ K_N \sim \alpha \log N \text{ w.p. } 1, \quad \implies \quad \frac{1}{j} \sim C(j), \quad j \to \infty, \text{ w.p. } 1\]
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

$K_N \sim \alpha \log N \text{ w.p. 1}$
$\implies # j \sim C(j), j \to \infty, \text{ w.p. 1}$

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

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: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps’ law, Herdan’s law, etc
- Pitman-Yor process:

\[ [\text{Gnedin, et al 2007, Pitman, Yor 1997}] \]
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:
  $K_N \sim S_\alpha N^\sigma \text{ w.p. 1}$

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

- $K_N := \#$ clusters occupied by $N$ 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)

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)
- 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, Rodríguez et al 2008]
Hierarchies

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

[Teh et al 2006]
Genealogy, trees, beyond trees

[Diagram of a genealogy tree with labels 1 to 9 and branches labeled $T_2, T_3, T_4$.]
Genealogy, trees, beyond trees

- Kingman coalescent

[Diagram of a genealogy tree with nodes labeled 1 to 9, and branches labeled $T_2$, $T_3$, $T_4$.]

[Note from Wakeley 2008]
Genealogy, trees, beyond trees

- Kingman coalescent

[Diagram showing a tree structure with labels 1 to 9, and references to Wakeley 2008 and Kingman 1982]
Genealogy, trees, beyond trees

- Kingman coalescent
- Fragmentation
- Coagulation

[Wakeley 2008]

[Kingman 1982]
Genealogy, trees, beyond trees

- Kingman coalescent
- Fragmentation
- Coagulation

Genealogy, trees, beyond trees

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

[Wakeley 2008]

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

Conjugacy & Poisson point processes

- Beta process, Bernoulli process (Indian buffet)

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


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)
- 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

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

\[ p(\cdot) \]

social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

$p(\cdot)$

- Rich relationships, coherent uncertainties, prior info

social: Facebook, Twitter, email

biological: ecological, protein, gene

transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more

social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008]
Probabilistic models for graphs

$p(\cdot)$

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership, stochastic block, infinite relational, and many more

Social: Facebook, Twitter, email
Biological: ecological, protein, gene
Transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

\( p(\cdot) \)

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn’t change distribution of earlier data (projectivity)

Social: Facebook, Twitter, email
Biological: ecological, protein, gene
Transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

\[ p( ) \]

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn’t change distribution of earlier data (*projectivity*)
- **Problem**: model misspecification, dense graphs

**social**: Facebook, Twitter, email

**biological**: ecological, protein, gene

**transportation**: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

\[ p(\cdot) \]

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn’t change distribution of earlier data (projectivity)
- **Problem**: model misspecification, dense graphs
- **Solution**: a new framework for sparse graphs

social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn’t change distribution of earlier data (*projectivity*)
- **Problem**: model misspecification, dense graphs
- **Solution**: a new framework for sparse graphs

social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

\[ p(\cdot) \]

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn’t change distribution of earlier data (*projectivity*)
- **Problem**: model misspecification, dense graphs
- **Solution**: a *new framework* for sparse graphs

social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

\[ p(\cdot) \]

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn’t change distribution of earlier data (*projectivity*)
- **Problem**: model misspecification, dense graphs
- **Solution**: a new framework for *sparse graphs*

- social: *Facebook, Twitter, email*
- biological: *ecological, protein, gene*
- transportation: *roads, railways*

[Holland et al 1983; Kemp et al 2006; Xu et al 2007; Airoldi et al 2008; Lloyd et al 2012]
Probabilistic models for graphs

\[ p(\cdot) \]

- Rich relationships, coherent uncertainties, prior info
- Example models: Stochastic block, mixed membership stochastic block, infinite relational, and many more
- Assume: Adding more data doesn’t change distribution of earlier data (projectivity)
- **Problem**: model misspecification, dense graphs
- **Solution**: a new framework for sparse graphs
  - Concurrent & independent graphs work by Crane & Dempsey

social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways
Sequence of graphs
Sequence of graphs

\[ G_1 \]

\[ G \]
Sequence of graphs

\[ G_1 \quad | \quad G_2 \quad | \quad G \]
Sequence of graphs

$G_1$

$G_2$

$G_3$

$G$
Sequence of graphs

$G_1$  $G_2$  $G_3$  $G_4$  $G$
Sequence of graphs

\( G_1 \)

\( G_2 \)

\( G_3 \)

\( G_4 \)

\( \ldots \)
Sequence of graphs

- Dense graph sequence \( \#\text{edges}(G_n) \geq c \cdot [\#\text{nodes}(G_n)]^2 \)
Sequence of graphs

- Dense graph sequence \( \#\text{edges}(G_n) \geq c \cdot [\#\text{nodes}(G_n)]^2 \)
- Sparse graph sequence \( \#\text{edges}(G_n) \in o([\#\text{nodes}(G_n)]^2) \)
The Old Way: Nodes

$G_1$
The Old Way: Nodes

$G_1$
The Old Way: Nodes

\[ G_1 \]

\[ G_2 \]
The Old Way: Nodes

$G_1$

$G_2$

$G_3$
The Old Way: Nodes

\[ G_1 \]

\[ G_2 \]

\[ G_3 \]

\[ G_4 \]

[Hoover 1979, Aldous 1981]
Exchangeability

$G_1$

$G_2$

$G_3$

$G_4$

[Hoover 1979, Aldous 1981]
Exchangeability

$G_1$ | $G_2$ | $G_3$ | $G_4$

[Hoover 1979, Aldous 1981]
Exchangeability

\[ G_1 \]
\[ G_2 \]
\[ G_3 \]
\[ G_4 \]

\[ p( ) \]
Exchangeability

\[ p( G_1 ) = p( G_2 ) = p( G_3 ) = p( G_4 ) \]

[Hoover 1979, Aldous 1981]
Node exchangeability

\[ p(1, 2, 3, 4) = p(2, 3, 1, 4) \]

\[ G_1 \] \hspace{2cm} \[ G_2 \] \hspace{2cm} \[ G_3 \] \hspace{2cm} \[ G_4 \]

[Hoover 1979, Aldous 1981]
The Old Way: Node exchangeability

\[ p(G_1) = p(G_2) = p(G_3) = p(G_4) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ \begin{array}{|c|c|}
\hline
0 & 1 \\
\hline
\end{array} \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]
Aldous-Hoover

$W(x, y)$

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]

\[ [\text{Hoover 1979, Aldous 1981}] \]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Aldous-Hoover

\[ W(x, y) \]

[Hoover 1979, Aldous 1981]
Every node-exchangeable graph has a graphon rep

[Hoover 1979, Aldous 1981]
Every node-exchangeable graph has a graphon rep

\[ \mathbb{E}[\#\text{edges}(G_n)] \]

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]
Every node-exchangeable graph has a graphon rep

\[
\mathbb{E}[\# \text{edges}(G_n)] = \mathbb{E} \left[ \frac{n}{2} \int_0^1 \int_0^1 W(x, y) \, dx \, dy \right]
\]

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]
Aldous-Hoover

Every node-exchangeable graph has a graphon rep

\[ \mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E} \left[ \binom{n}{2} \frac{1}{2} \int_0^1 \int_0^1 W(x, y) \, dx \, dy \right] \]

\[ \sim cn^2 \]

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]
Every node-exchangeable graph has a graphon rep

\[ E[\#\text{edges}(G_n)] = E \left[ \binom{n}{2} \frac{1}{2} \int_0^1 \int_0^1 W(x, y) \, dx \, dy \right] \]

\[ \sim cn^2 = c \cdot [\#\text{nodes}(G_n)]^2 \]
Every node-exchangeable graph has a graphon rep

\[ \mathbb{E}[\# \text{edges}(G_n)] = \mathbb{E} \left[ \binom{n}{2} \frac{1}{2} \int_0^1 \int_0^1 W(x, y) \, dx \, dy \right] \]

\[ \sim cn^2 = c \cdot [\# \text{nodes}(G_n)]^2 \]

Every node-exch graph sequence is dense (or empty)

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015]
Every node-exchangeable graph has a graphon rep

\[ \mathbb{E}[\#\text{edges}(G_n)] = \mathbb{E} \left[ \binom{n}{2} \frac{1}{2} \int_0^1 \int_0^1 W(x, y) \, dx \, dy \right] \]

\[ \sim cn^2 = c \cdot [\#\text{nodes}(G_n)]^2 \]

Every node-exch graph sequence is dense (or empty)

[Hoover 1979, Aldous 1981, Orbanz, Roy 2015; Caron, Fox 2014; Veitch, Roy 2015; Borgs, Chayes, Cohn, Holden 2016; Broderick, Cai 2015; Crane, Dempsey 2015; Crane, Dempsey 2016; Cai, Campbell, Broderick 2016]
A New Way: Edges

\[ G_1 \]
A New Way: Edges

$G_1$
A New Way: Edges

$G_1$  $G_2$
A New Way: Edges

$G_1$  

$G_2$  

$G_3$
A New Way: Edges

$G_1$  

$G_2$  

$G_3$  

$G_4$
Edge exchangeability

$G_1$

$G_2$

$G_3$

$G_4$
Edge exchangeability

$G_1$

$G_2$

$G_3$

$G_4$
Edge exchangeability

\[ p(G_1) = p(G_2) = p(G_3) = p(G_4) \]
Edge exchangeability

\[ p(1, 2, 3, 4) = p(2, 4, 1, 3) \]
Edge exchangeability

Thm. A wide range of edge-exchangeable graph sequences are sparse

\[ p(1 \rightarrow 2 \rightarrow 3) = p(2 \rightarrow 4 \rightarrow 1) \]
Nonparametric Bayes

- Bayesian methods that are 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]

[Ed Bowlby, NOAA]

[Prabhakaran, Azizi, Carr, Pe’er 2016]

[Escobar, West 1995; Ghosal, et al 1999]

[Saria et al 2010]

[Fox, et al 2014]

[Ed Bowlby, NOAA]

[Lloyd et al 2012; Miller et al, 2010]

[Ewens, 1972; Hartl, Clark 2003]

[Sudderth, Jordan 2009]
References (page 1 of 4)


