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

Tamara Broderick
Associate Professor
MIT

http://www.tamarabroderick.com/tutorials.html
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
Marginal cluster assignments
Marginal cluster assignments

- Integrate out the frequencies
Marginal cluster assignments

- Integrate out the frequencies
  \[
  \rho_1 \sim \text{Beta}(a_1, a_2),
  z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)
  \]
Marginal cluster assignments

- Integrate out the frequencies
  \[ \rho_1 \sim \text{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2) \]
Marginal cluster assignments

- Integrate out the frequencies

\[ \rho_1 \sim \text{Beta}(a_1, a_2), \quad z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2) \]
Marginal cluster assignments

- Integrate out the frequencies

\[ \rho_1 \sim \text{Beta}(a_1, a_2), \quad z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2) \]

\[ p(z_n = 1|z_1, \ldots, z_{n-1}) = \frac{a_{1,n}}{a_{1,n} + a_{2,n}} \]
Marginal cluster assignments

- Integrate out the frequencies

\[ \rho_1 \sim \text{Beta}(a_1, a_2), \quad z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2) \]

\[ p(z_n = 1 | z_1, \ldots, z_{n-1}) = \frac{a_{1,n}}{a_{1,n} + a_{2,n}} \]

\[ a_{1,n} := a_1 + \sum_{m=1}^{n-1} 1\{z_m = 1\}, \quad a_{2,n} = a_2 + \sum_{m=1}^{n-1} 1\{z_m = 2\} \]
Marginal cluster assignments

• Integrate out the frequencies
  \( \rho_1 \sim \text{Beta}(a_1, a_2) \), \( z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2) \)
  
  \[
p(z_n = 1 | z_1, \ldots, z_{n-1}) = \frac{a_{1,n}}{a_{1,n} + a_{2,n}}
  \]

  \[
a_{1,n} := a_1 + \sum_{m=1}^{n-1} 1\{z_m = 1\}, \quad a_{2,n} = a_2 + \sum_{m=1}^{n-1} 1\{z_m = 2\}
  \]

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

- Integrate out the frequencies
  \[ \rho_1 \sim \text{Beta}(a_1, a_2), \quad z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2) \]
  \[ p(z_n = 1|z_1, \ldots, z_{n-1}) = \frac{a_{1,n}}{a_{1,n} + a_{2,n}} \]
  \[ a_{1,n} := a_1 + \sum_{m=1}^{n-1} 1\{z_m = 1\}, \quad a_{2,n} = a_2 + \sum_{m=1}^{n-1} 1\{z_m = 2\} \]

- 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

- Integrate out the frequencies

\[ \rho_1 \sim \text{Beta}(a_1, a_2), \quad z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2) \]

\[ p(z_n = 1|z_1, \ldots, z_{n-1}) = \frac{\rho_1^n}{\rho_1^n + \rho_2^n} \]

\[ a_{1,n} := a_1 + \sum_{m=1}^{n-1} 1\{z_m = 1\}, \quad a_{2,n} = a_2 + \sum_{m=1}^{n-1} 1\{z_m = 2\} \]

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

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

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

- Hoppe urn / Blackwell-MacQueen urn
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn
Marginal cluster assignments

• Hoppe urn / Blackwell-MacQueen urn
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
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
------|------
⚫️    | 🔧  🔧  🔧  🔧  🔧  🔧  🔧

[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

Step 0

Step 1

Step 2

Step 3

[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

Step 0

Step 1

Step 2

Step 3

Step 4

(#orange, #other)
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

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

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

\[
\begin{align*}
\text{Step 0} & : \quad \text{not orange: } (\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha) \\
\text{Step 1} & : \quad \text{not orange, green: } (\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha) \\
\text{Step 2} & : \quad \text{not orange, green: } (\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)
\end{align*}
\]
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

Step 0

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

Step 1

\[\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \]

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

Step 2

\[\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \]

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

Step 3

\[\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \]

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

Step 4

\[\bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \]

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

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

\[
\begin{align*}
\text{Step 0} & & \text{Step 1} & & \text{Step 2} & & \text{Step 3} & & \text{Step 4} \\
\text{(orange, other)} & = & \text{PolyaUrn}(1, \alpha) \\
\text{not orange: (green, other)} & = & \text{PolyaUrn}(1, \alpha) \\
\text{not orange, green: (red, other)} & = & \text{PolyaUrn}(1, \alpha) \\
\end{align*}
\]

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

\[
V_k \overset{iid}{\sim} \text{Beta}(1, \alpha)
\]
\[
\rho_1 = V_1
\]
\[
\rho_2 = (1 - V_1)V_2
\]

(#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

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

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

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

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

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

$\uparrow$

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

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

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 $\alpha$

\[
\phi_1
\]

\[
\mathcal{Z}_1 = \{1, 2\}, \mathcal{Z}_2 = \{3, 5\}, \mathcal{Z}_3 = \{6\}
\]
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 $\alpha$

$\phi_1$

$\phi_2$

Partition of $[8]$: set of mutually exclusive & exhaustive sets of $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$
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 $\alpha$

\[ \phi_1 \]
\[ \phi_2 \]
\[ \phi_3 \]
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 $\alpha$
- 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
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
  $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$
  $\Rightarrow \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}$
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
  
  $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$

  \[ \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\} \]

  \textit{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:
\[
\frac{\alpha}{\bar{\alpha}} \cdot \frac{1}{\alpha + 1}
\]
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:

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

- Probability of this seating:
\[
\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}
\]
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 \(k\)): 
  
  \[\begin{array}{cccc}
  1 & 7 & 6 & 3 \\
  2 & 5 & 4 & \text{---}
  \end{array}\]
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 \( k \)):

  \[
  \frac{1}{\alpha \cdots (\alpha + N - 1)}
  \]
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:
  \[
  \frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{3}{\alpha + 6} \cdot \frac{\alpha}{\alpha + 7}
  \]

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

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

• Prob doesn’t depend on customer order: exchangeable
Chinese restaurant process

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

- 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
\[
\mathbb{P}(\pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}) = \mathbb{P}(\pi_8 = \{\{2, 3, 8, 1\}, \{4, 6, 7\}, \{5\}\})
\]
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 \( C \)):
  \[
  \alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)! \left/ \alpha \cdot \cdots \cdot (\alpha + N - 1) \right. = P(\Pi_N = \pi_N)
  \]

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

• So:
  \[
p(\Pi_N | \Pi_{N,-n}) =
  \]
Probability of $N$ customers ($K_N$ tables, $\#C$ at table $C$):

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

So:

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

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

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

- 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 $C$):

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

• So:

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

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

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

• 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 $C$):
  \[ \alpha^{K_N} \frac{\prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N) \]
- So:
  \[ p(\Pi_N | \Pi_{N,-n}) = \begin{cases} \frac{\#C}{\alpha + N - 1} & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} & \text{if } n \text{ starts a new cluster} \end{cases} \]
- 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 $C$):
  \[
  \frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)
  \]

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

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

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

• Data $x_{1:N}$

• Generative model

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

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

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

• Data $x_1:N$

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

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

- Gibbs sampler:

  $p(\Pi_N | \Pi_{N,-n}, x) \propto \{ \ldots \}$
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:
  $$p(\Pi_N | \Pi_{N,-n}, x) \propto \begin{cases} \text{if } n \text{ joins cluster } C \end{cases}$$
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:

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

• Data $x_{1:N}$

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

• Want: posterior $p(\Pi_N|x_{1:N})$

• Gibbs sampler:

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

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

- **For completeness:** $p(x_{C\cup\{n\}}|x_C) =$
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
  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \mu_C \sim \mathcal{N}(\mu_0, \Sigma_0)$
  $\forall C \in \Pi_N, \forall n \in C, x_n \sim \mathcal{N}(\mu_C, \Sigma)$

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

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

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

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

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

- **Data**: $x_{1:N}$
- **Generative model**
  \[ \Pi_N \sim \text{CRP}(N, \alpha) \]
  \[ \forall C \in \Pi_N, \phi_C \overset{iid}{\sim} G_0 \]
  \[ \forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma) \]
- **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**:
  \[ p(x_{C \cup \{n\}}|x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma) \]
  \[ \tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (#C)\Sigma^{-1} \]
  \[ \tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right) \]

[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
  $\Pi_N \sim \text{CRP}(N, \alpha)$
  $\forall C \in \Pi_N, \phi_C \sim G_0 \text{ iid}$
  $\forall C \in \Pi_N, \forall n \in C, x_n \sim F(\phi_C)$

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

- Gibbs sampler:

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

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

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

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

[MacEachern 1994; Neal 1992; Neal 2000]
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

[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]

[ESO/ L. Calçada/ M. Kornmesser 2017]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]
[Del Pozzo et al 2017, 2018]

[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]

[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]

[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]

[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]

[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]
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
Exercises

- Can you find a formula for the expected number of clusters from a Hoppe-urn(α) after N data points? What happens as N → ∞?
- Code a CRP mixture model simulator.
- Derive the CRP mixture model Gibbs sampler in the slides.
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well.
- Review slice sampling [Neal 2003], variational [Bishop 2006].
Hierarchies

Dependencies

Coalescents/Diffusions/Trees
de Finetti

Power laws

Feature allocations

Networks/graphs

Poisson processes

Here be Dragons
Clustering

Document 1                              Arts | Econ | Sports | Health | Technology
Document 2                              Arts | Econ | Sports | Health | Technology
Document 3                              Arts | Econ | Sports | Health | Technology
Document 4                              Arts | Econ | Sports | Health | Technology
Document 5                              Arts | Econ | Sports | Health | Technology
Document 6                              Arts | Econ | Sports | Health | Technology
Document 7                              Arts | Econ | Sports | Health | Technology
# 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></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>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Document 2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Document 3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Document 4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Document 5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Document 6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Document 7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

- Indian buffet process
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

[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>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

[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>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

[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

### Feature allocation

<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document 2</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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document 3</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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document 4</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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document 5</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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document 6</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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Document 7</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>
</tbody>
</table>

- Indian buffet process
- Beta process

## Trait 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>30</td>
<td></td>
<td></td>
<td>15</td>
<td></td>
</tr>
<tr>
<td><strong>Document 2</strong></td>
<td>5</td>
<td></td>
<td>12</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td><strong>Document 3</strong></td>
<td>3</td>
<td>9</td>
<td></td>
<td>19</td>
<td>28</td>
</tr>
<tr>
<td><strong>Document 4</strong></td>
<td></td>
<td>4</td>
<td>21</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td><strong>Document 5</strong></td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td><strong>Document 6</strong></td>
<td></td>
<td></td>
<td>9</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td><strong>Document 7</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experimental design under a budget

[Masoero, Camerlenghi, Favaro, Broderick 2019]
Experimental design under a budget

Individual sequence: GAT AAAT CT GGT CTT TATT TCC
Reference genome: GAT AAAT CTT GGT CTT TAT TCC
Binary encoding of variation: 0 0 1 0 1 0

[Masoero, Camerlenghi, Favaro, Broderick 2019]
Experimental design under a budget

- State-of-the-art performance under constant experimental conditions

[Masoero, Camerlenghi, Favaro, Broderick 2019]
Experimental design under a budget

- State-of-the-art performance under constant experimental conditions
- Can estimate # new variants when experimental conditions change

[Masoero, Camerlenghi, Favaro, Broderick 2019]
Experimental design under a budget

- State-of-the-art performance under constant experimental conditions
- Can estimate # new variants when experimental conditions change

[Figure: A subpopulation of the gnomAD data]

[Masoero, Camerlenghi, Favaro, Broderick 2019]
Experimental design under a budget

• State-of-the-art performance under constant experimental conditions
• Can estimate # new variants when experimental conditions change
• Can be used to optimize under a budget

[Masoero, Camerlenghi, Favaro, Broderick 2019]
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

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

\[ p(\cdot) \]

- Interpretable, flexible, 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

- Interpretable, flexible, coherent uncertainties, prior info
- Example models:

  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

\[ p(\cdot) \]

- Interpretable, flexible, coherent uncertainties, prior info
- Example models:
  - Stochastic block model

[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

Interpretable, flexible, coherent uncertainties, prior info

Example models:
- Stochastic block model
- Mixed membership stochastic block model

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

Interpretable, flexible, coherent uncertainties, prior info

Example models:
- Stochastic block model
- Mixed membership stochastic block model
- Infinite relational model

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

\[ p(\cdot) \]

- Interpretable, flexible, coherent uncertainties, prior info
- Example models:
  - Stochastic block model
  - Mixed membership stochastic block model
  - Infinite relational model
  - Latent space model
  - Eigenmodel
  - Latent feature relational model
  - Infinite latent attribute model
  - Sparse matrix-variate Gaussian process block model
  - Random function model

[References: Holland et al. 1983; Hoff 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
Probabilistic models for graphs

\[ p(\cdot) \]

- Interpretable, flexible, 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

$p(\cdot)$

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

- Interpretable, flexible, 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*)

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

\[ p(\ ) \]

- Interpretable, flexible, 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

$p(\quad)$

- Interpretable, flexible, 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**: an alternative framework for sparse graphs

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

\[ p(\cdot) \]

- Interpretable, flexible, 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 (\textit{projectivity})
- \textbf{Problem}: model misspecification, dense graphs
- \textbf{Solution}: an alternative framework for sparse graphs
  - Don’t miss independent graphs work by Crane & Dempseyey

[Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,16a,b; Cai, Campbell, Broderick 2016; Campbell, Cai, Broderick 2016; Campbell, Cai, Broderick 2018; Williamson 2016]

social: Facebook, Twitter, email
biological: ecological, protein, gene
transportation: roads, railways
DP or not DP, that is the question

- GEM:
- Compare to:
  - Finite (small $K$) mixture model
  - Finite (large $K$) mixture model
  - Time series
Local Exchangeability

Recall

• GEM:

• vs time series

[Campbell, Syed, Yang, Jordan, Broderick 2019]
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.org]

[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/ L. Calçada/ M. Kornmesser 2017]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
[Prabhakaran, Azizi, Carr, Pe’er 2016]
[Del Pozzo et al 2017, 2018]
[ESO/L. Calçada/M. Kornmesser 2017]
[MIT xPRO]
References (1/7)


- *NeurIPS 2015 Workshop on Networks in the Social & Information Sciences*.
- *NeurIPS 2015 Workshop, Bayesian Nonparametrics: The Next Generation*.


- *NeurIPS 2016 Workshop on Practical Bayesian Nonparametrics*. 


References (3/7)


References (5/7)


References (6/7)


ESO/L. Calçada/M. Kornmesser. 16 October 2017, 16:00:00. Obtained in 2018 from: https://commons.wikimedia.org/wiki/File:Artist%E2%80%99s_impression_of_merging_neutron_stars.jpg || Source: https://www.eso.org/public/images/eso1733a/ (Creative Commons Attribution 4.0 International License)
