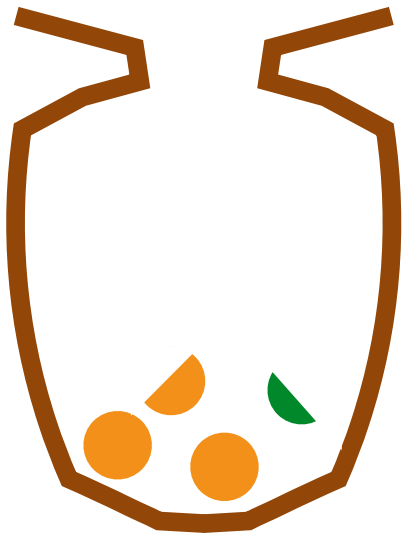


# Nonparametric Bayesian Methods: Models, Algorithms, and Applications (Part 3)

Tamara Broderick

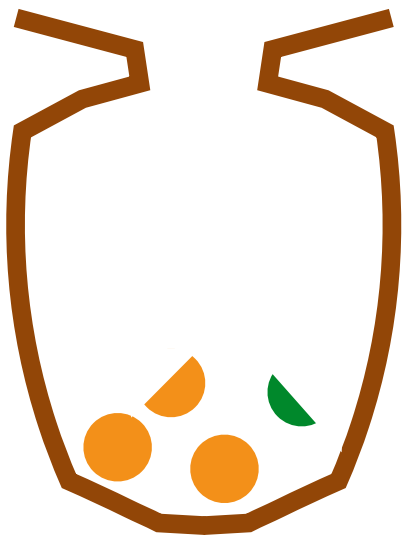
ITT Career Development Assistant Professor  
Electrical Engineering & Computer Science  
MIT

# Marginal cluster assignments



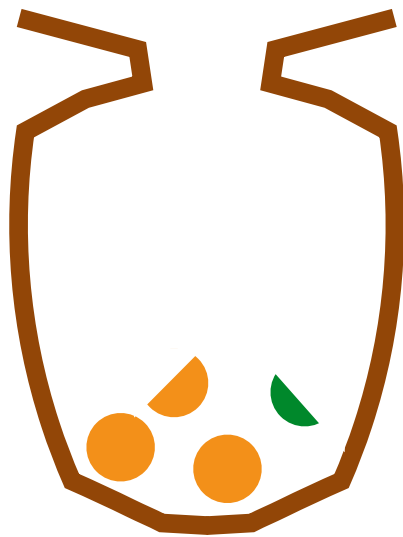
# Marginal cluster assignments

- Pólya urn



# Marginal cluster assignments

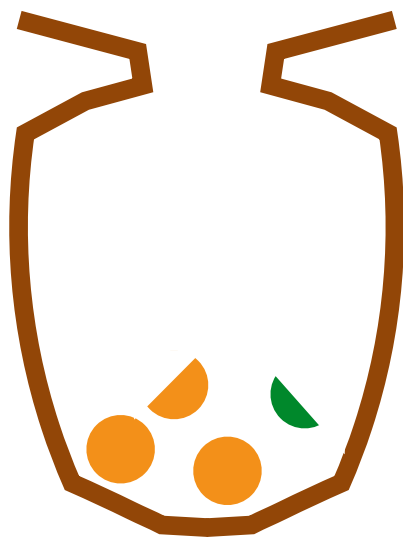
- Pólya urn
  - Choose any ball with prob proportional to its mass
  - Replace and add ball of same color



# Marginal cluster assignments

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

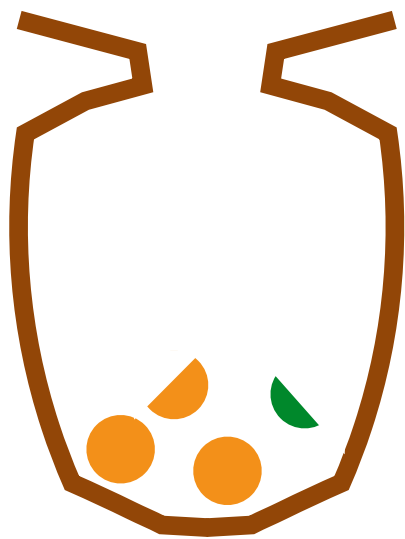
$\text{PolyaUrn}(a_{\text{orange}}, a_{\text{green}})$



# Marginal cluster assignments

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

$\text{PolyaUrn}(a_{\text{orange}}, a_{\text{green}})$



$$\lim_{n \rightarrow \infty} \frac{\# \text{ orange}}{\# \text{ total}} = \rho_{\text{orange}} \stackrel{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

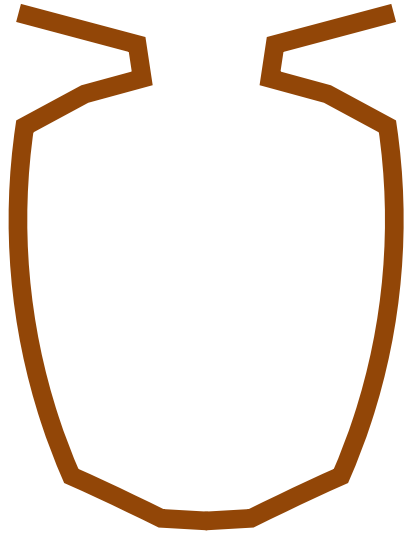
# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



# Marginal cluster assignments

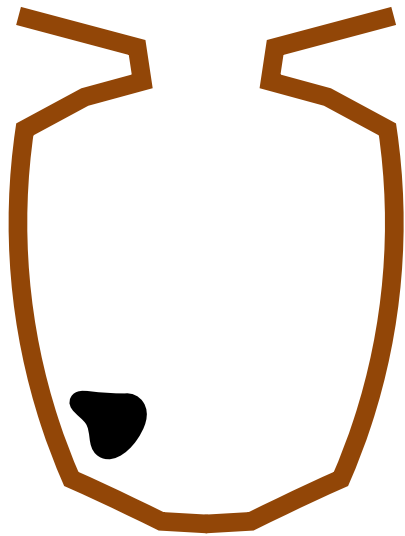
- Hoppe urn / Blackwell-MacQueen urn





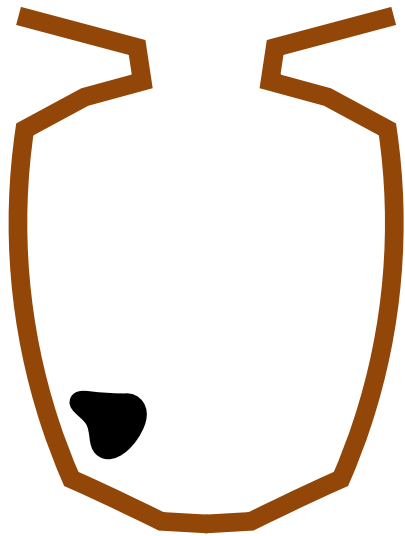
# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



# Marginal cluster assignments

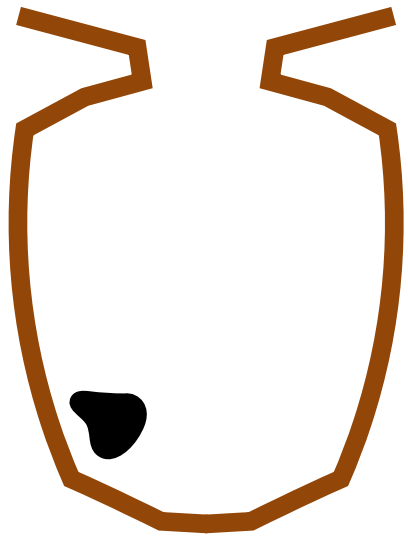
- Hoppe urn / Blackwell-MacQueen urn



- Choose ball with prob proportional to its mass

# Marginal cluster assignments

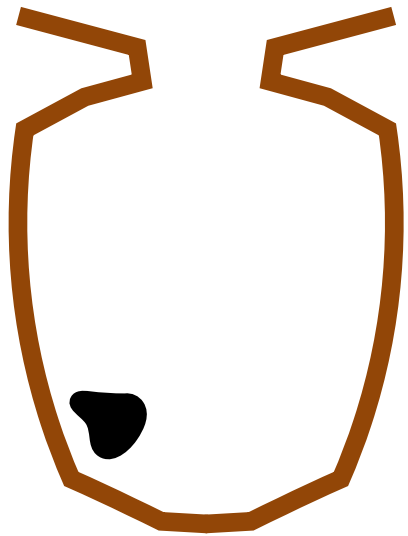
- Hoppe urn / Blackwell-MacQueen urn



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

# Marginal cluster assignments

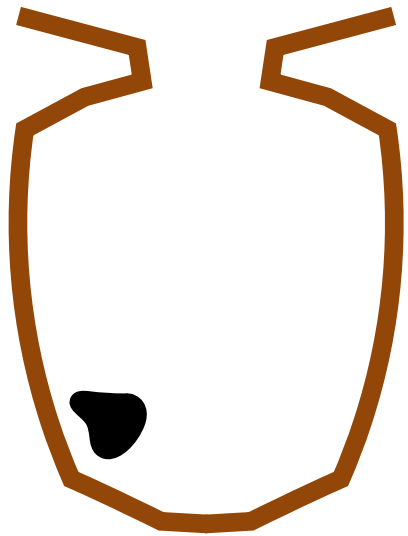
- Hoppe urn / Blackwell-MacQueen urn



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

# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



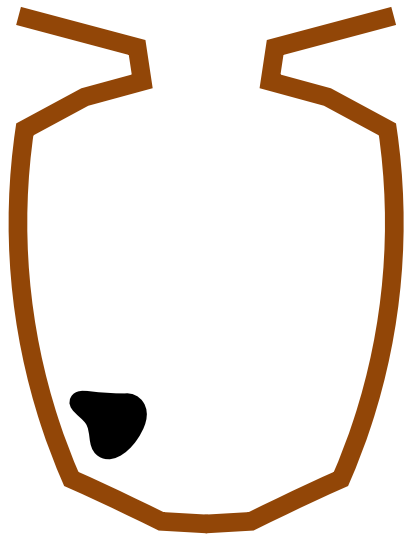
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

Step 0

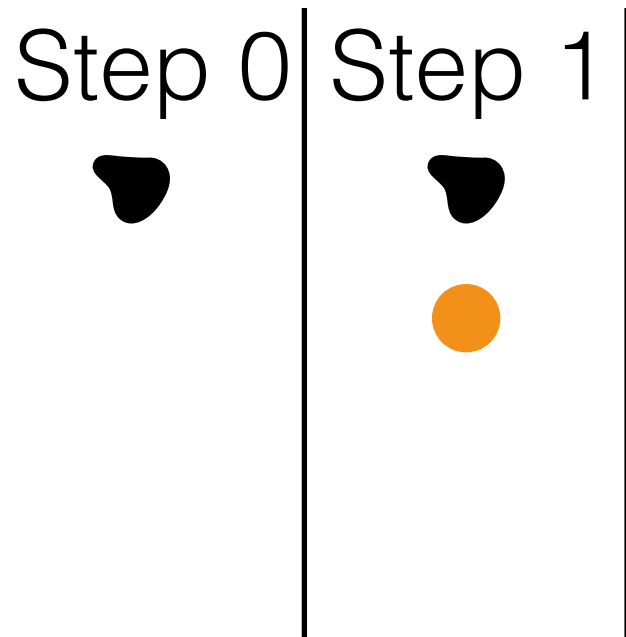


# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

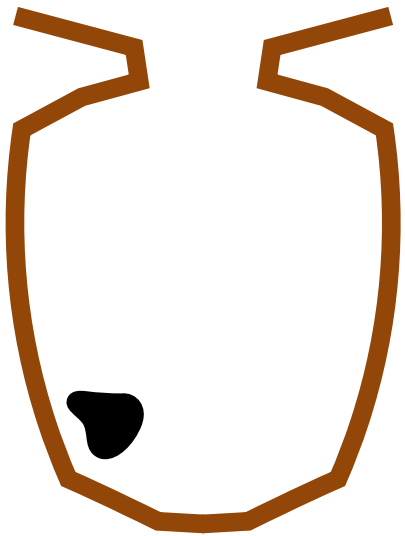


- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

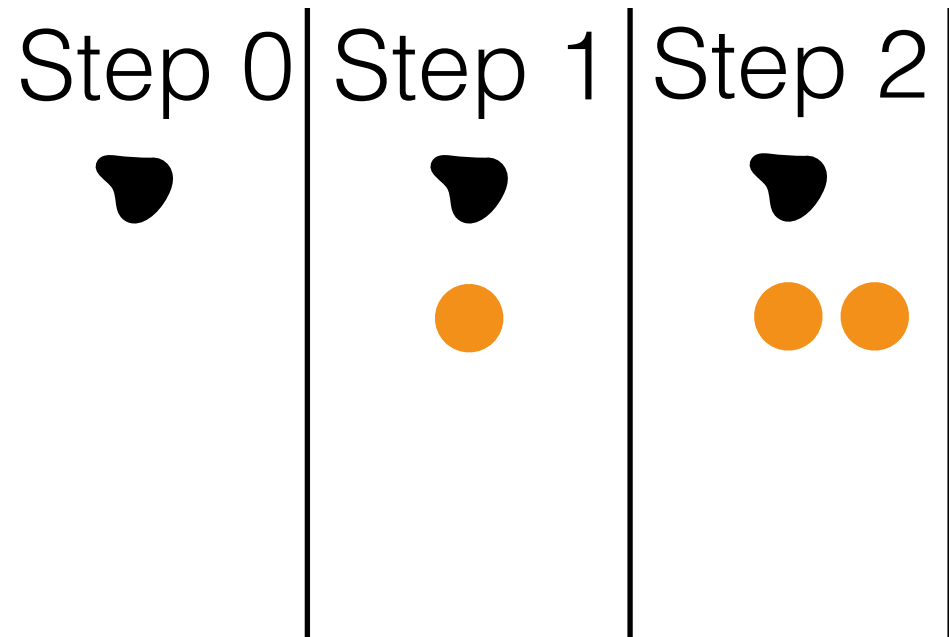


# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



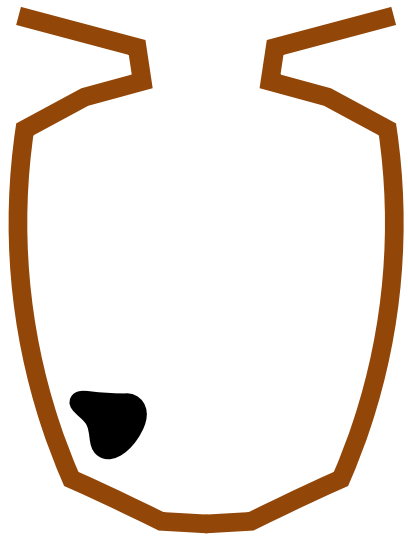
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



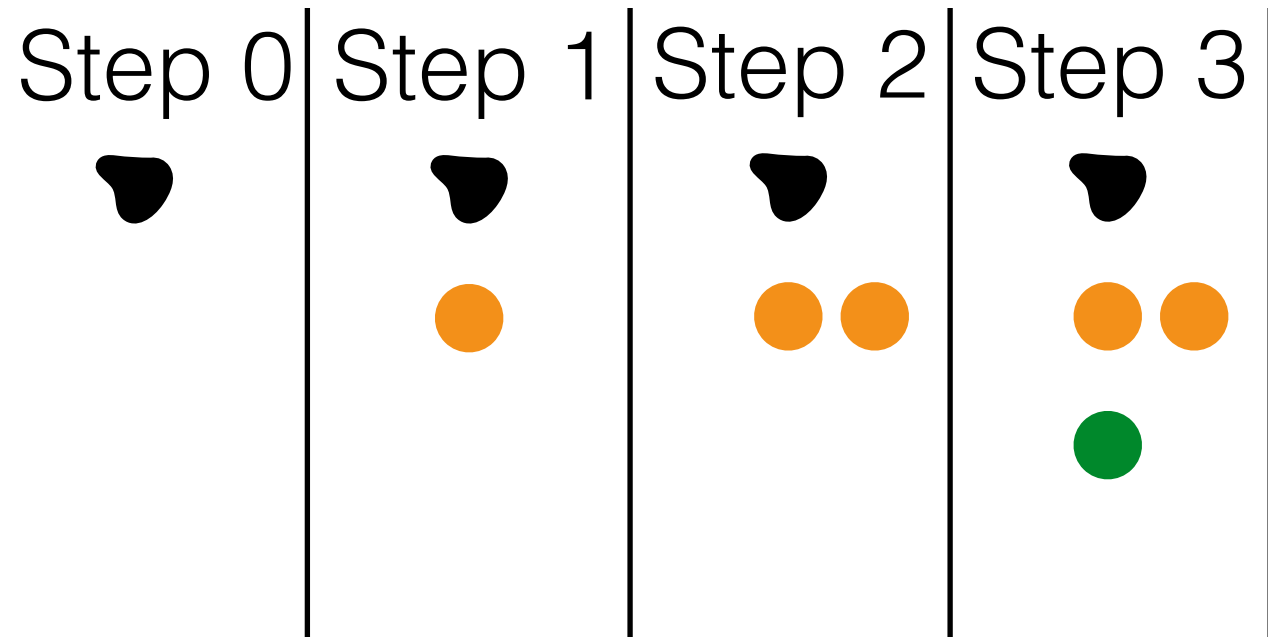


# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



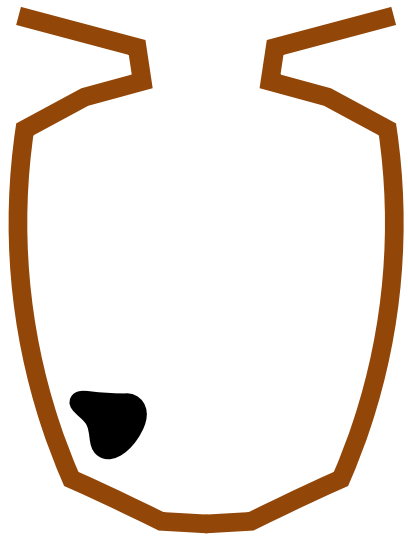
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



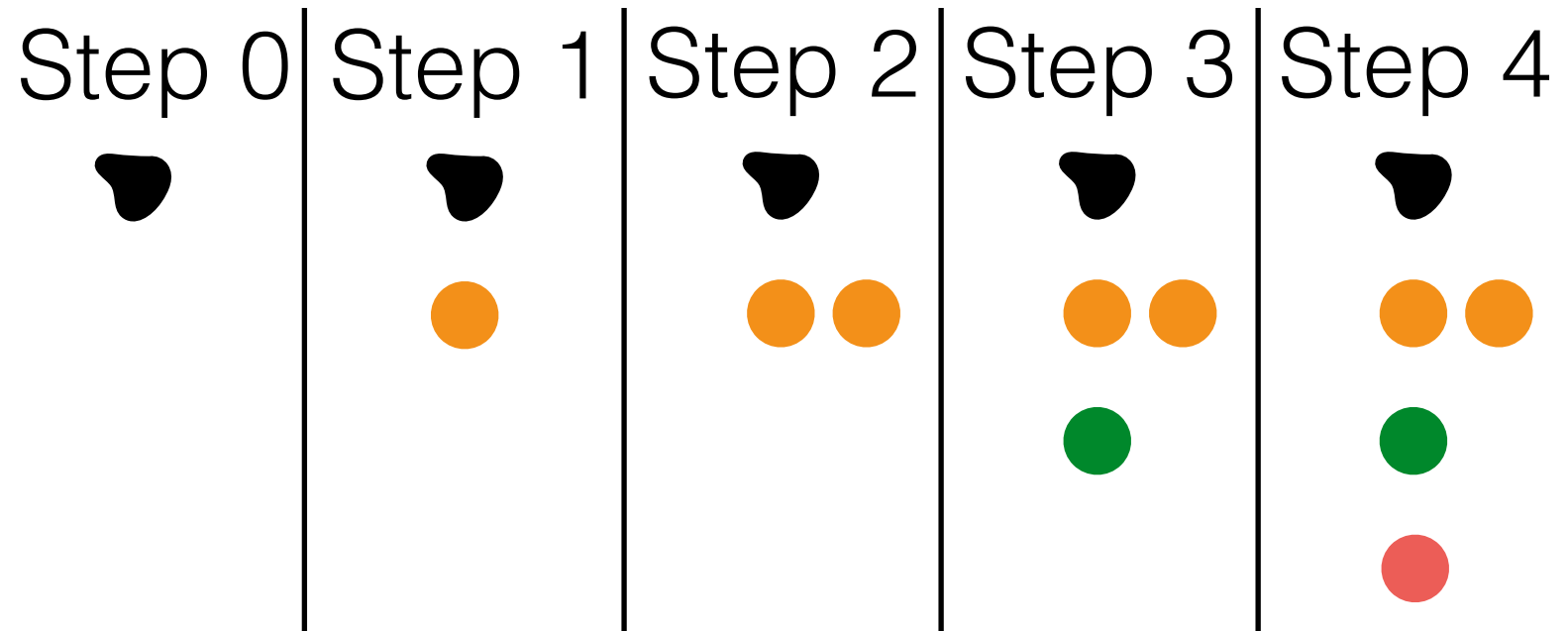


# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

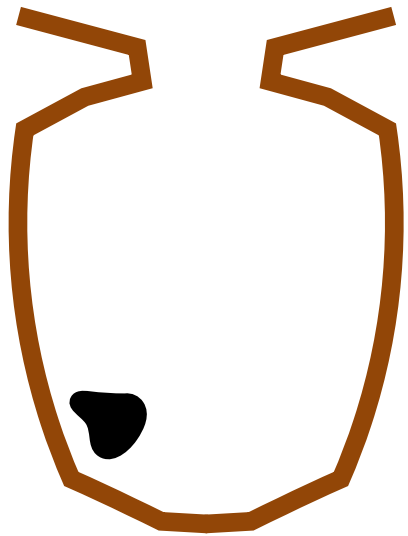


- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

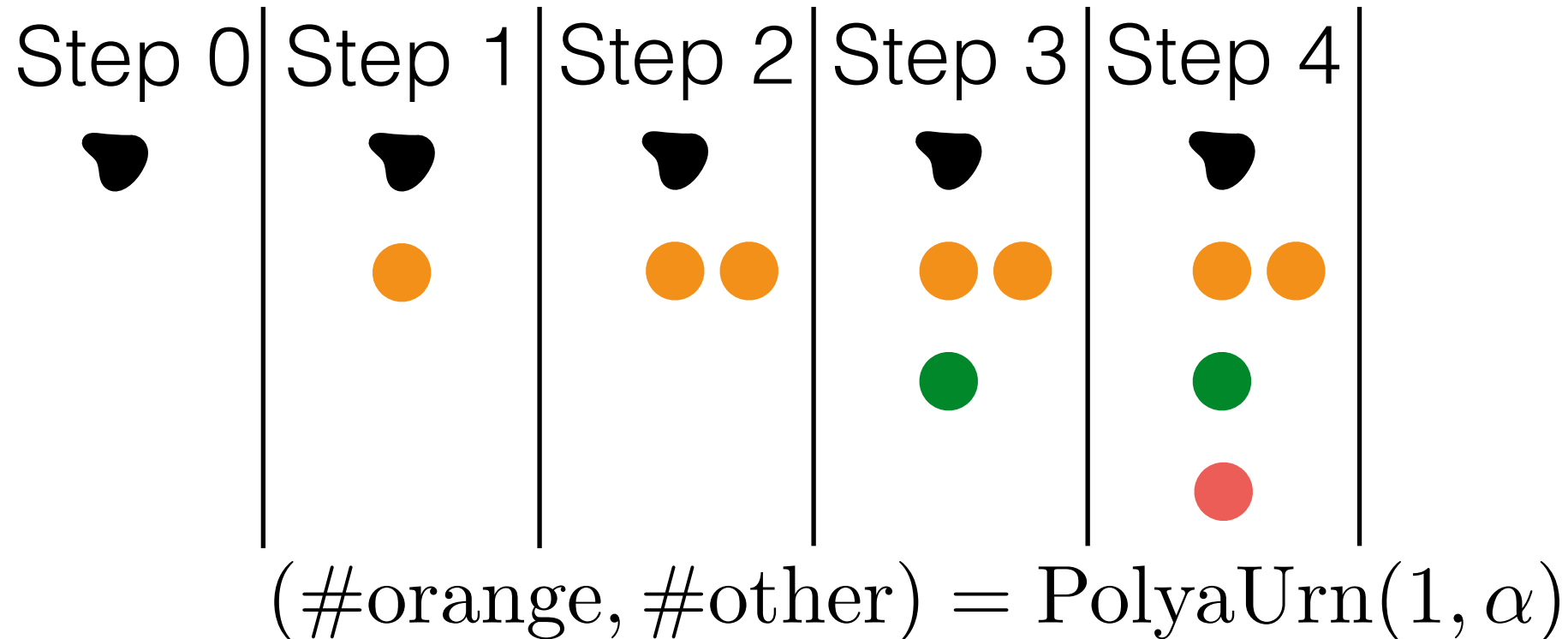


# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

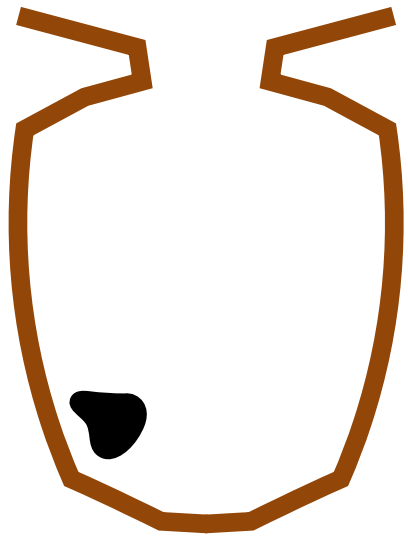


- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

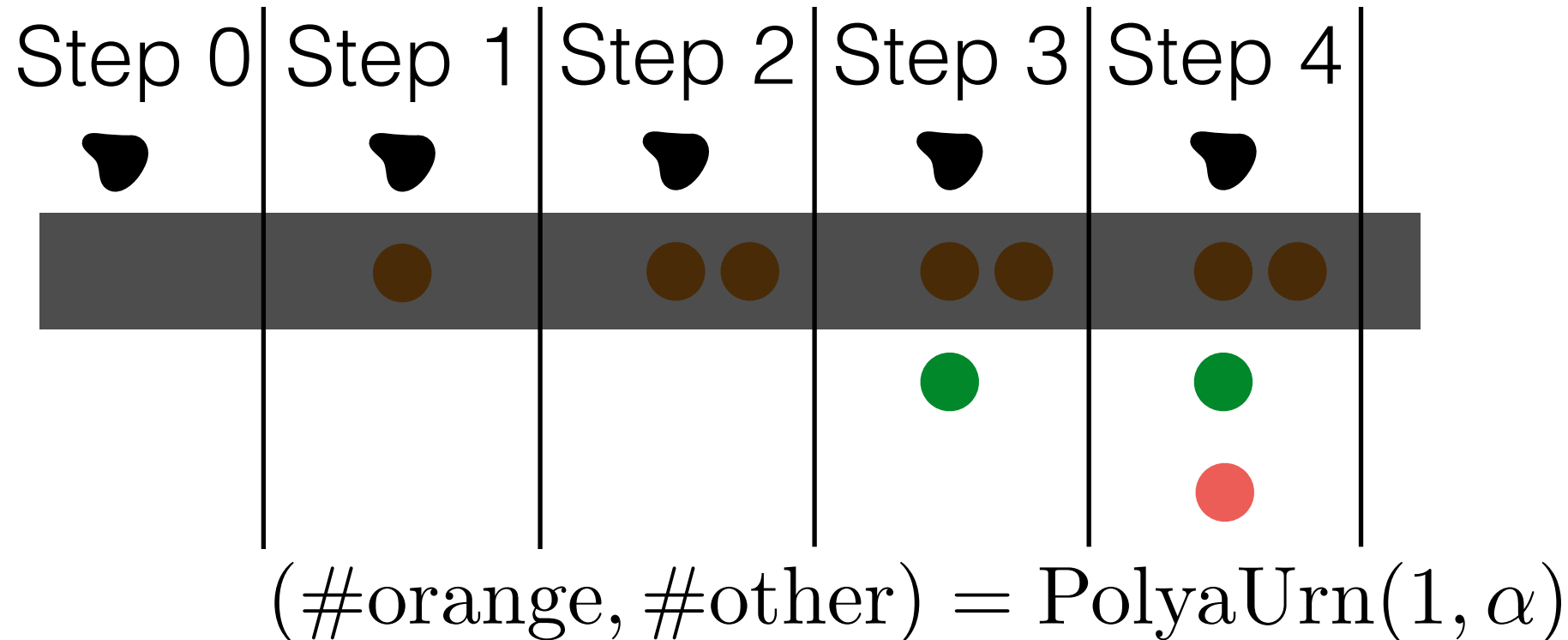


# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

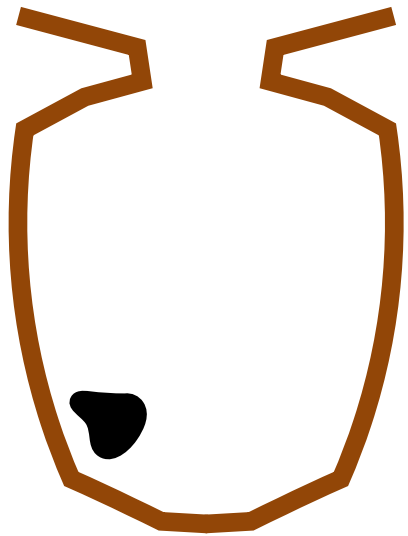


- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

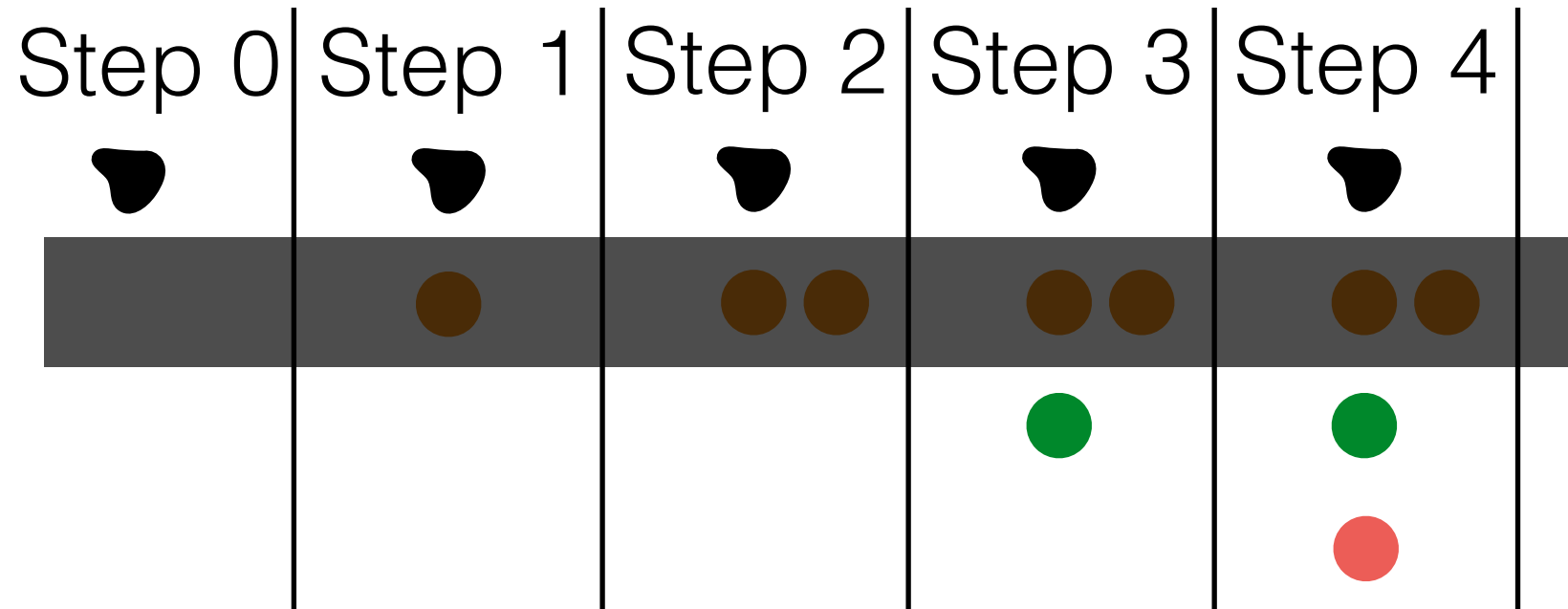


# Marginal cluster assignments

- 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

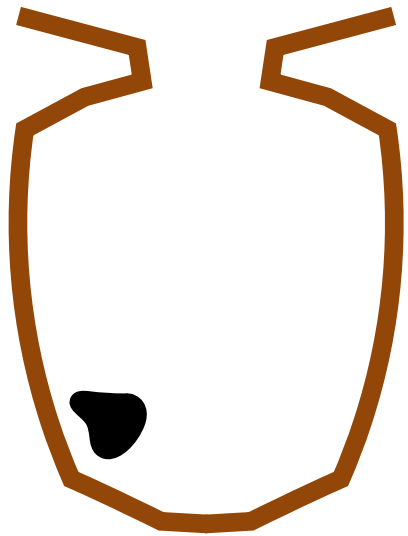


$$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$$

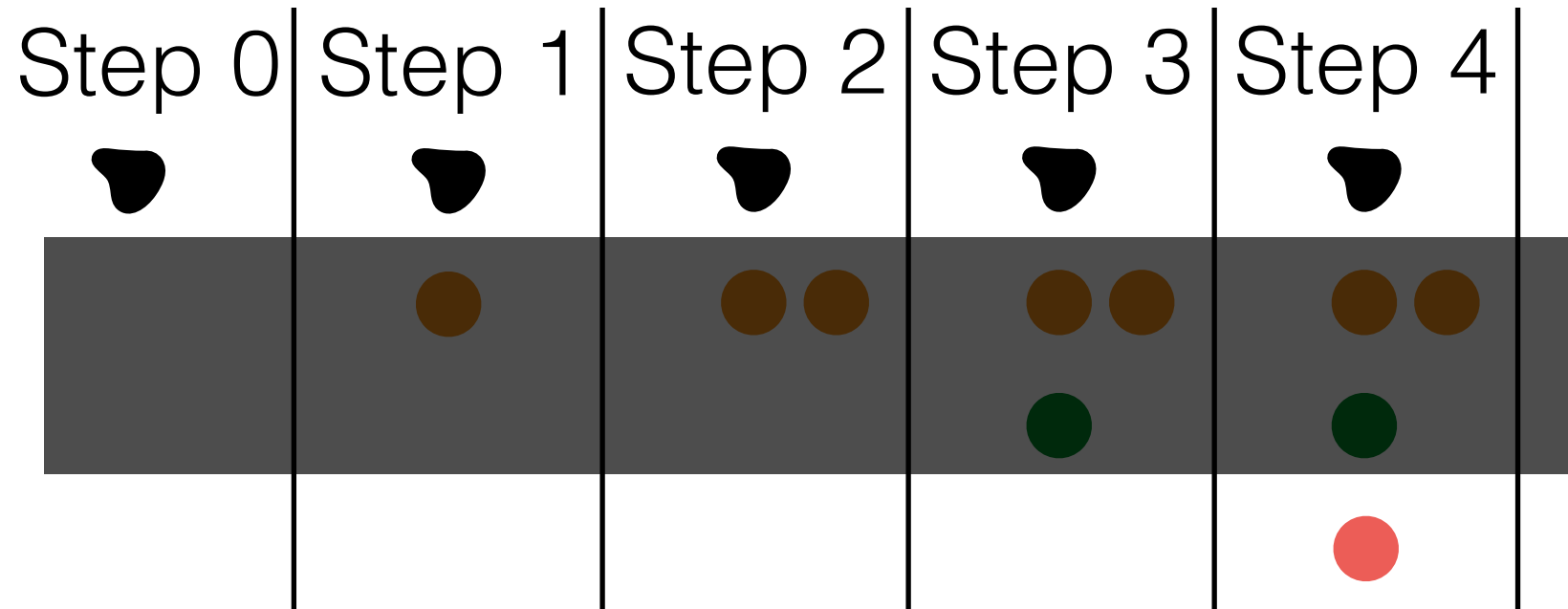
- not orange:  $(\# \text{green}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

# 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

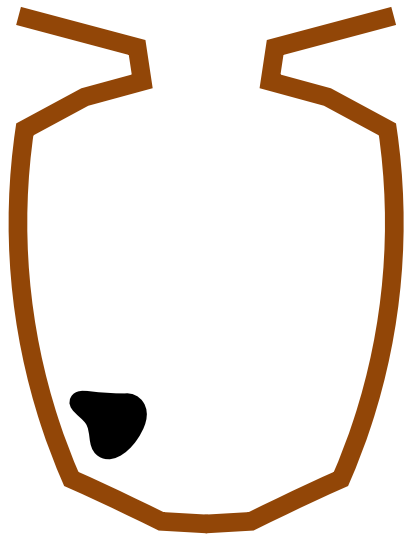


$$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$$

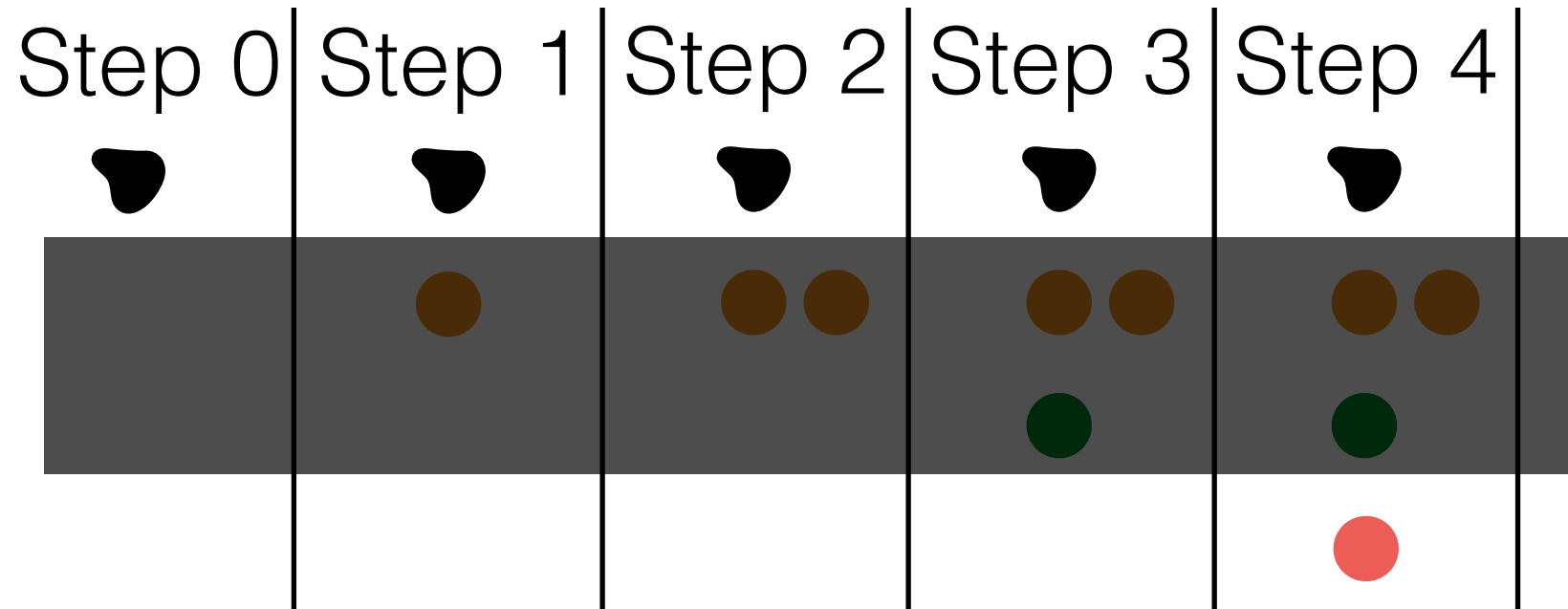
- not orange:  $(\# \text{green}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

# 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

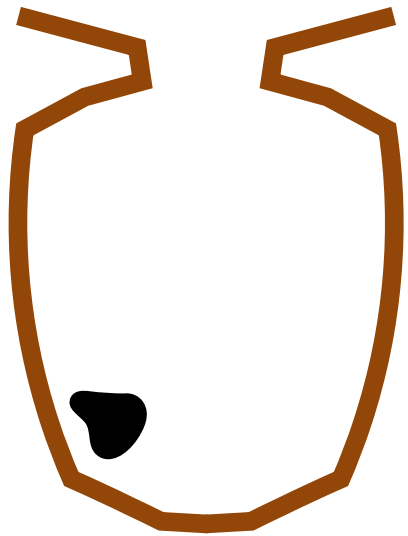


$$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$$

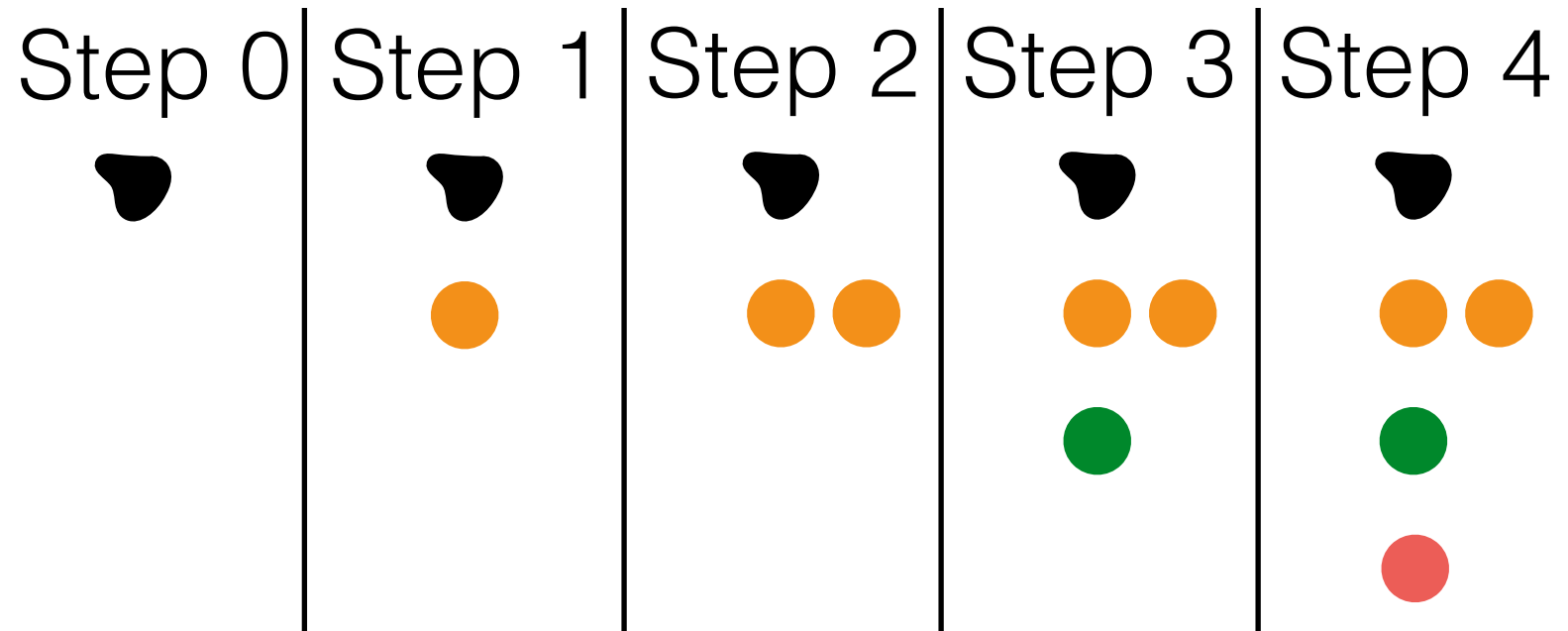
- not orange:  $(\# \text{green}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green:  $(\# \text{red}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

# 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



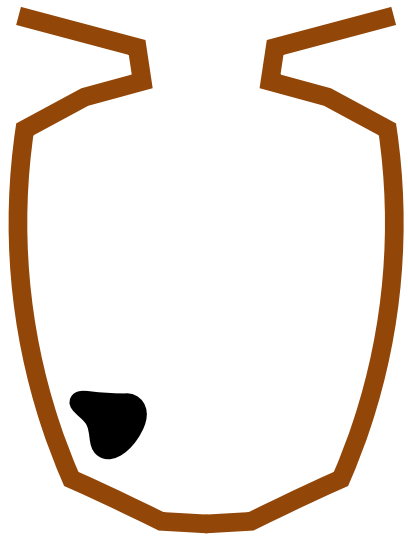
$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

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

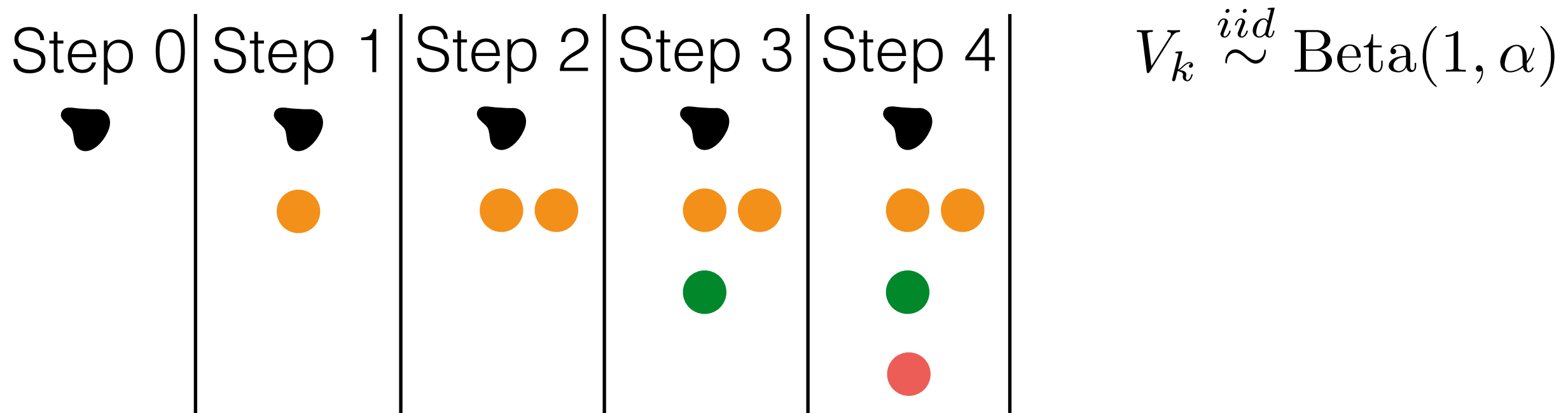


# Marginal cluster assignments

- 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



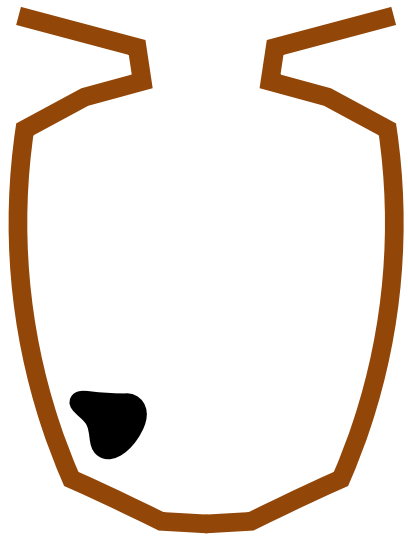
$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

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

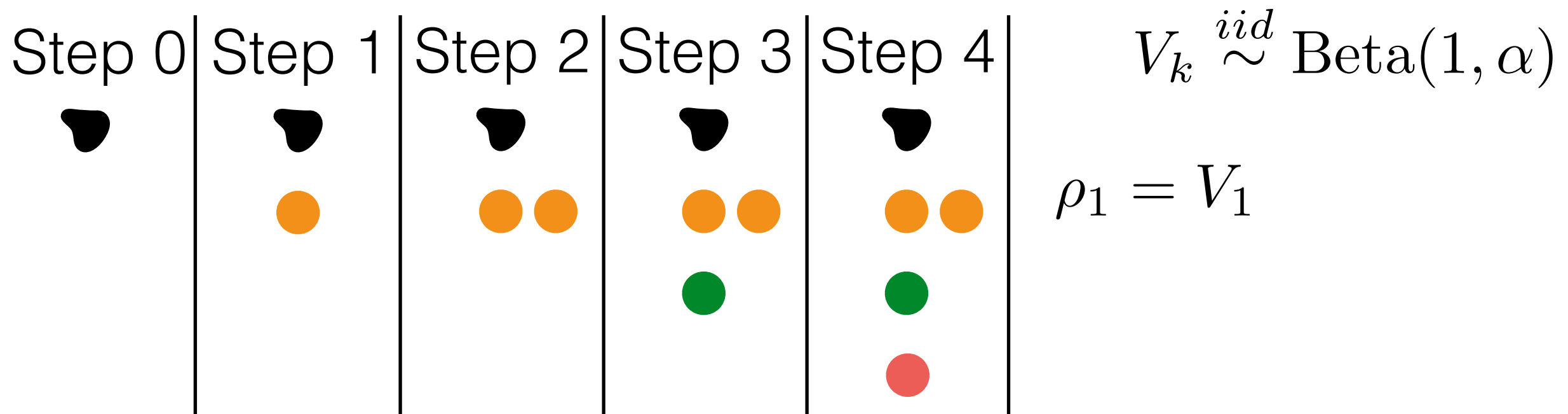


# Marginal cluster assignments

- 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

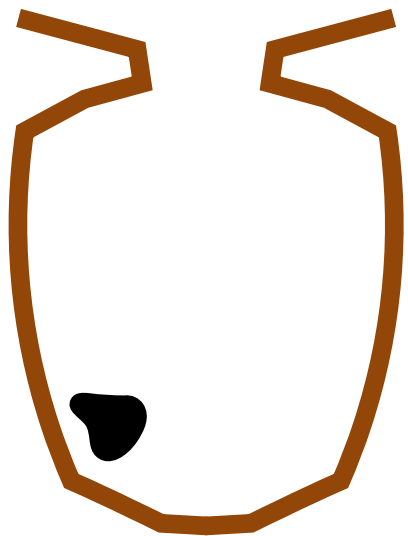


$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

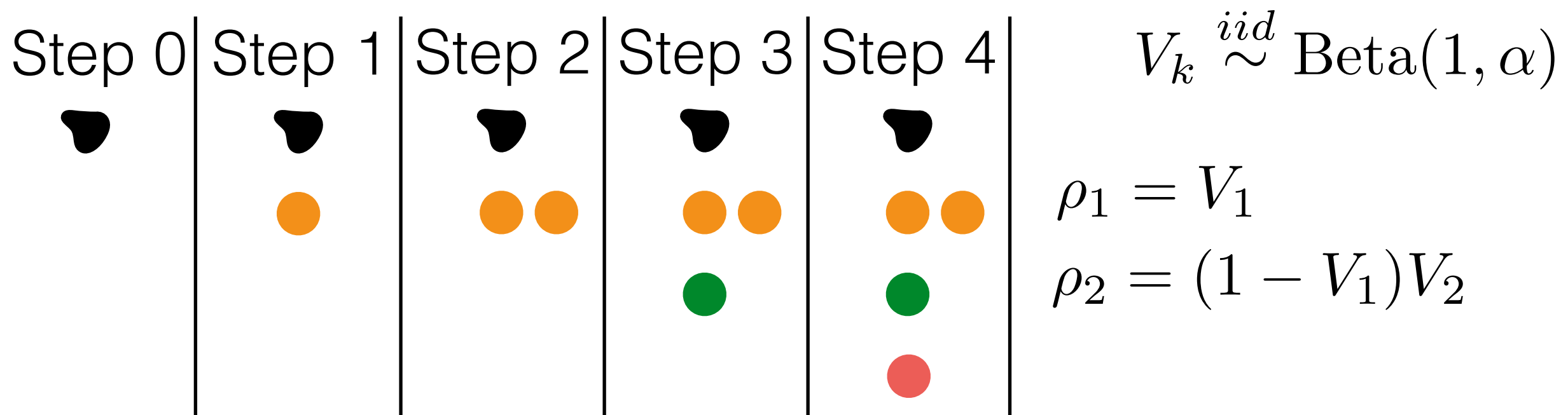
- not orange:  $(\# \text{green}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green:  $(\# \text{red}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

# 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

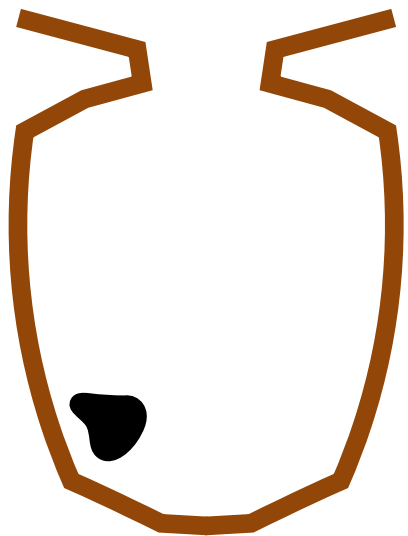


$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

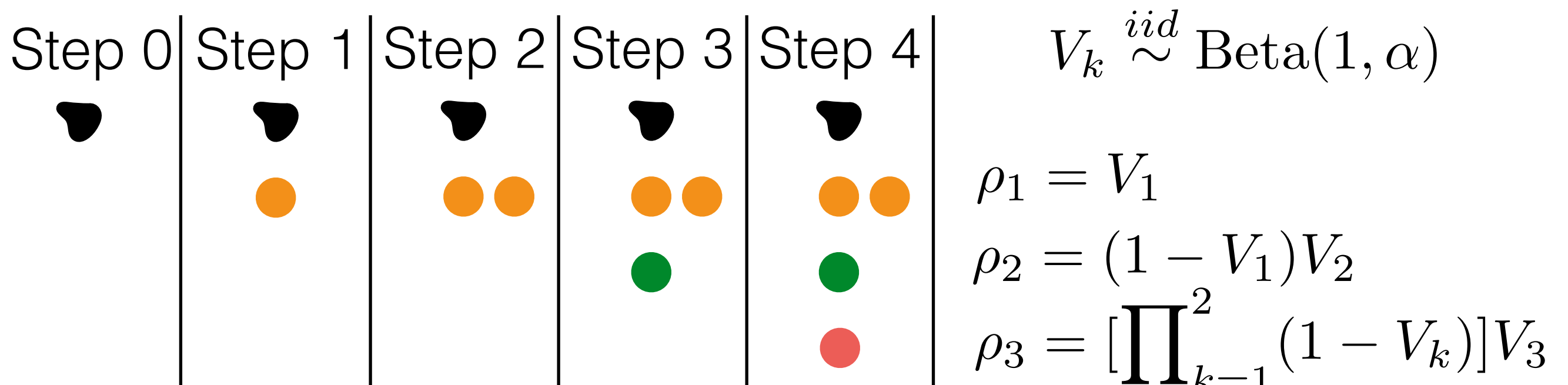
- not orange:  $(\# \text{green}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green:  $(\# \text{red}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

# Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



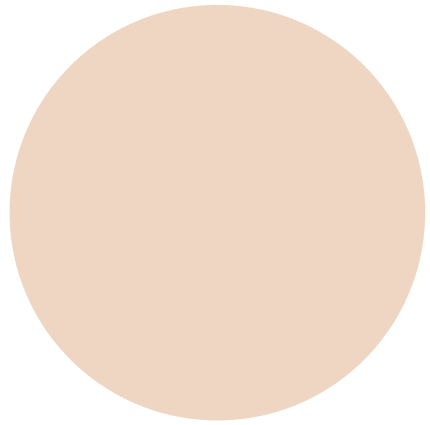
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



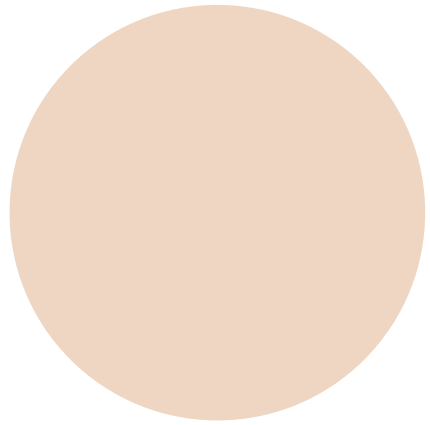
$(\# \text{orange}, \# \text{other}) = \text{PolyaUrn}(1, \alpha)$

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

# Chinese restaurant process

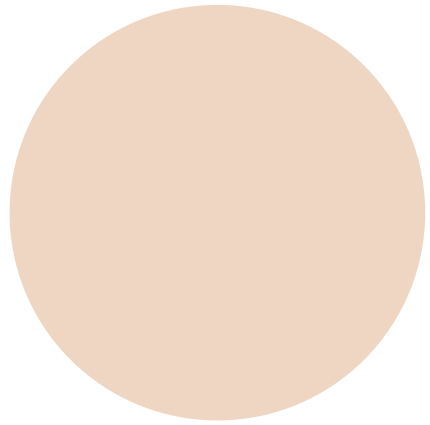


# Chinese restaurant process



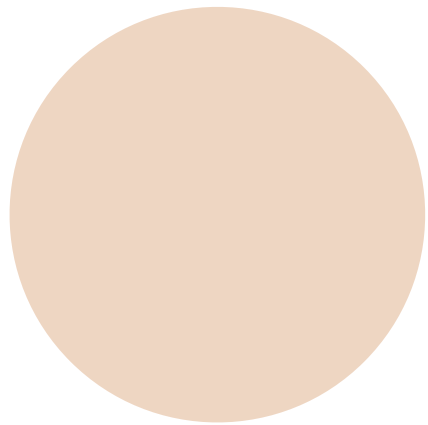
- Same thing we just did

# Chinese restaurant process



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

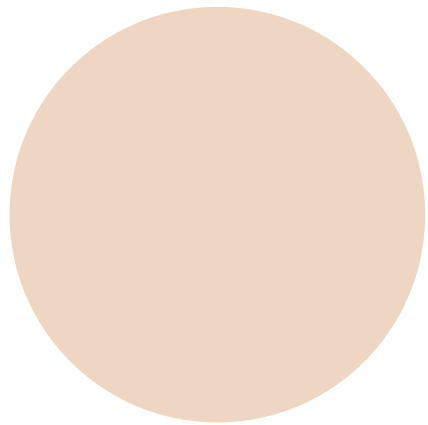
# Chinese restaurant process



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



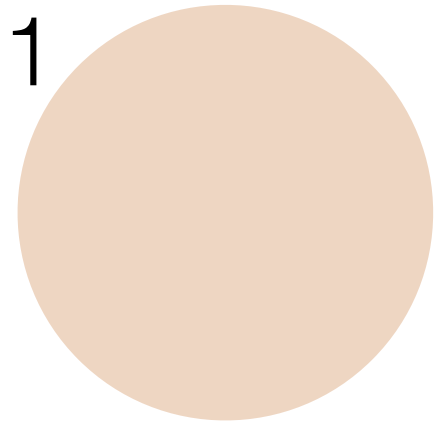
# Chinese restaurant process



- 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  $\alpha$

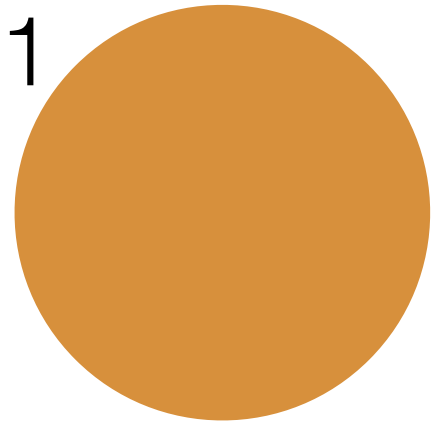


# Chinese restaurant process



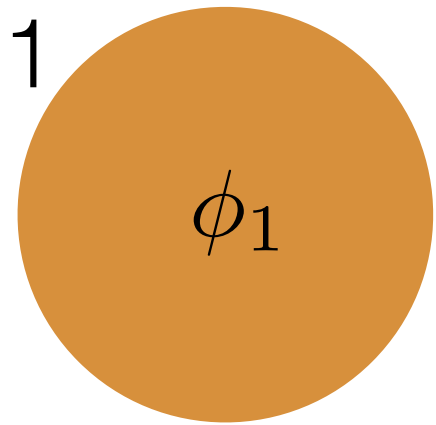
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



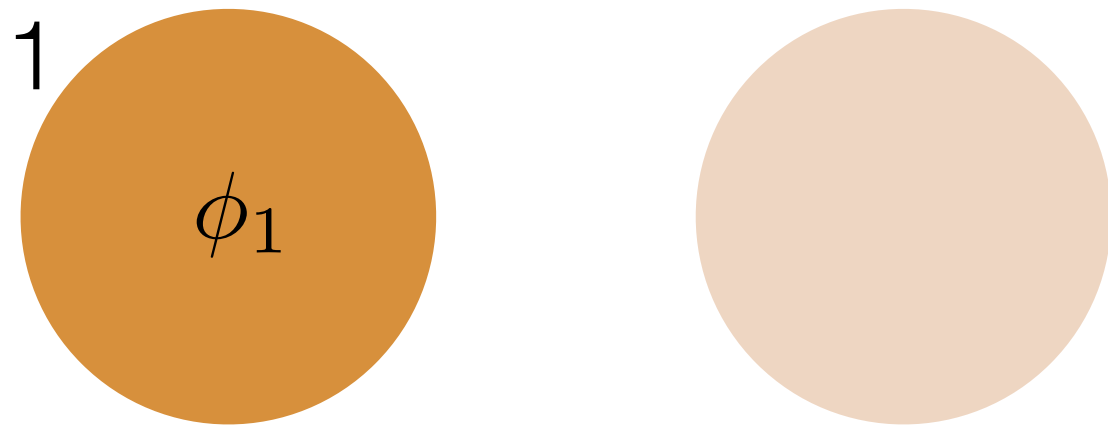
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



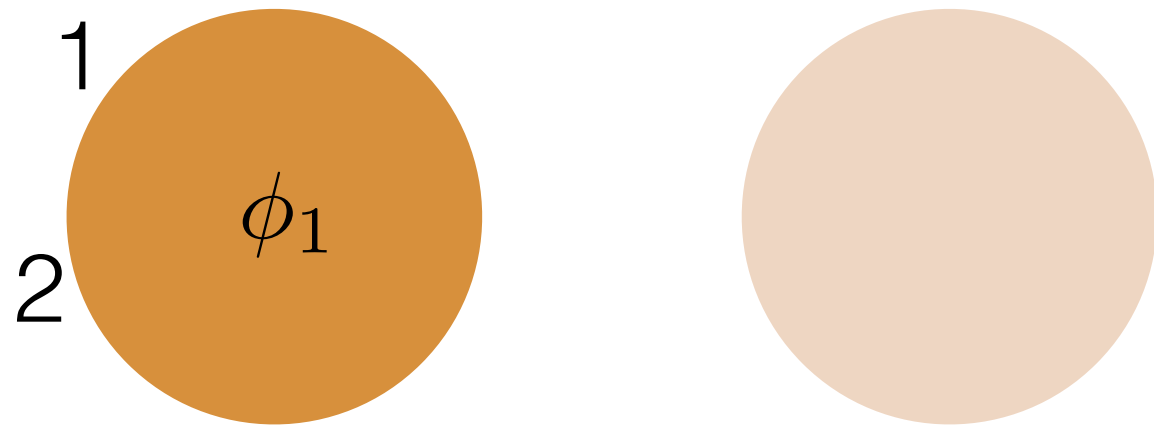
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



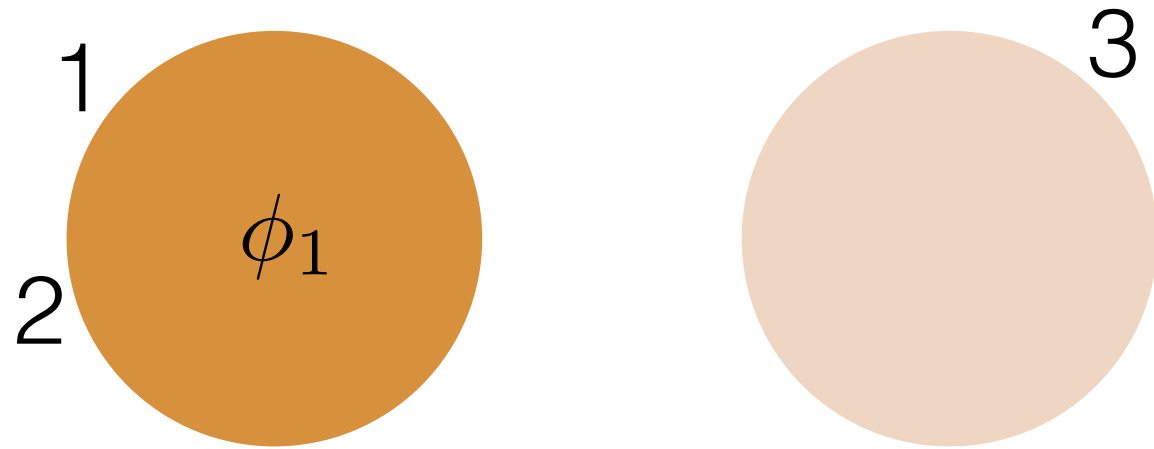
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



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

# Chinese restaurant process



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

# Chinese restaurant process



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



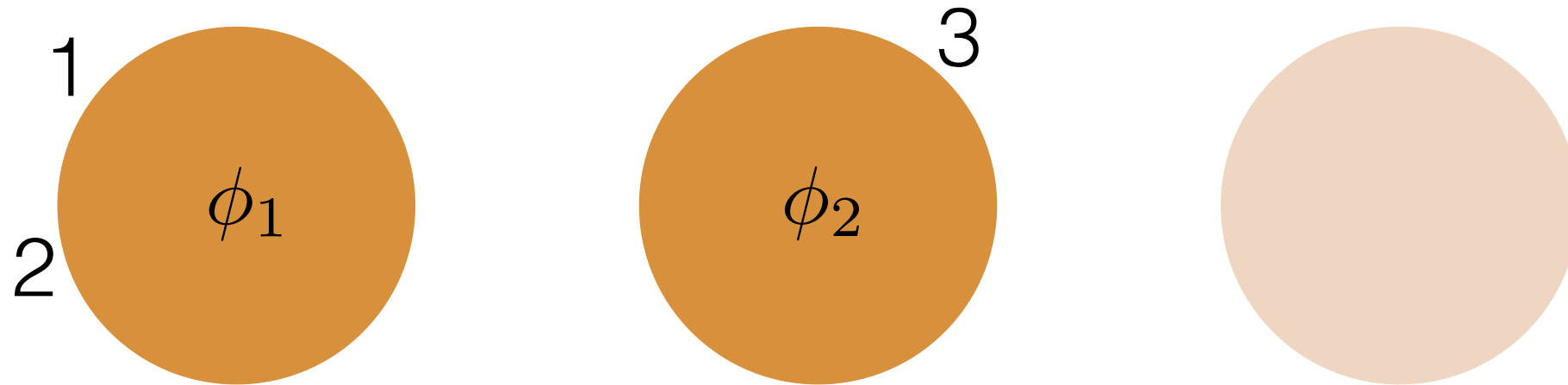
# Chinese restaurant process



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

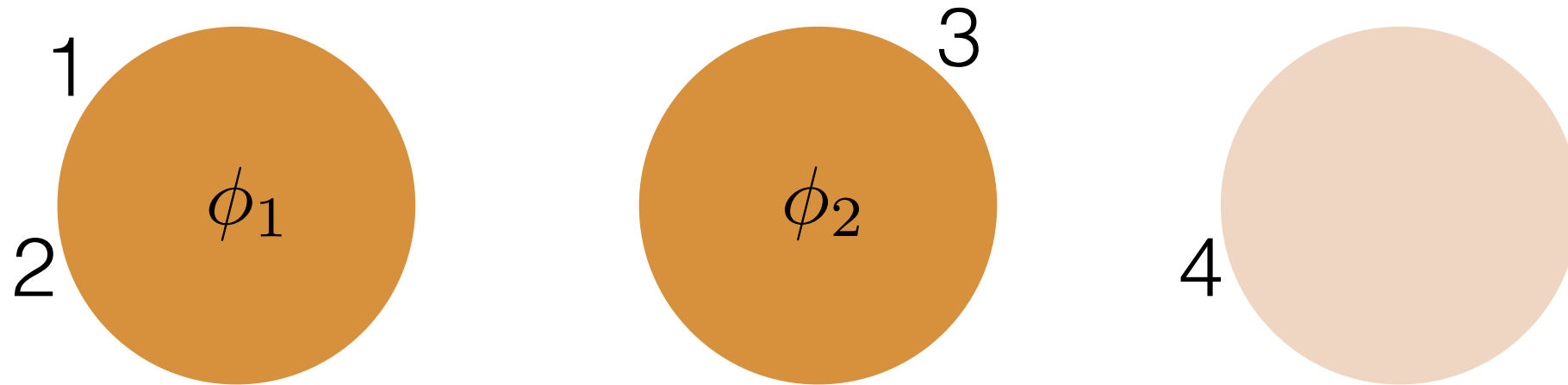


# Chinese restaurant process



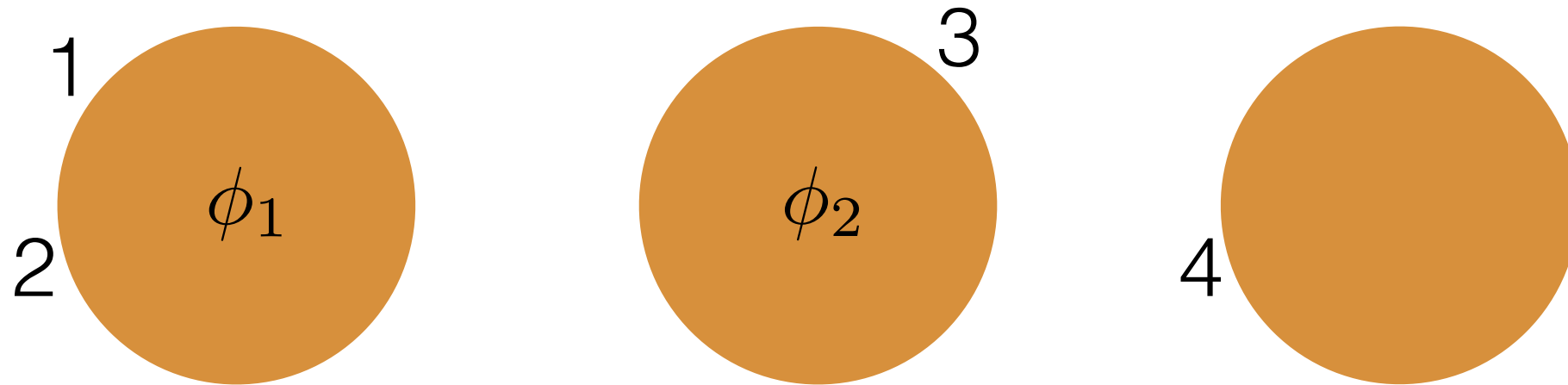
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



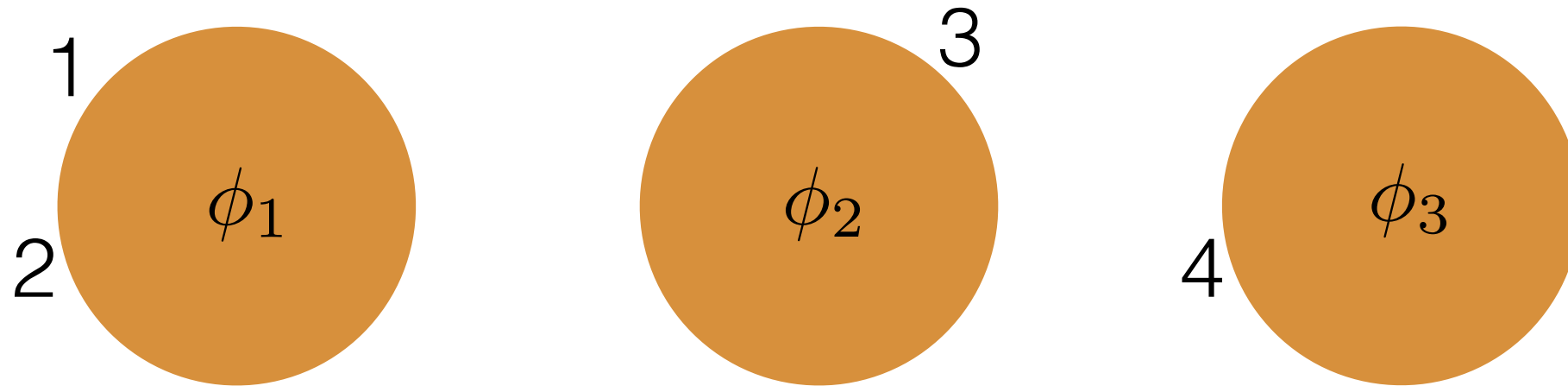
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



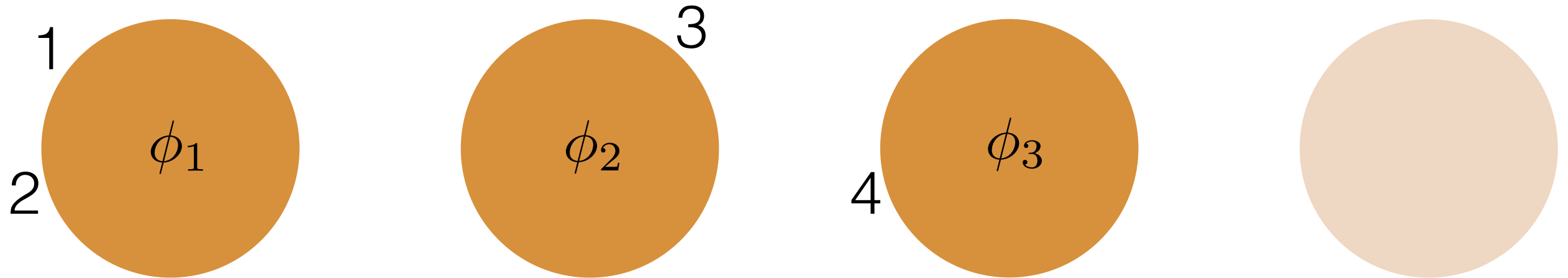
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



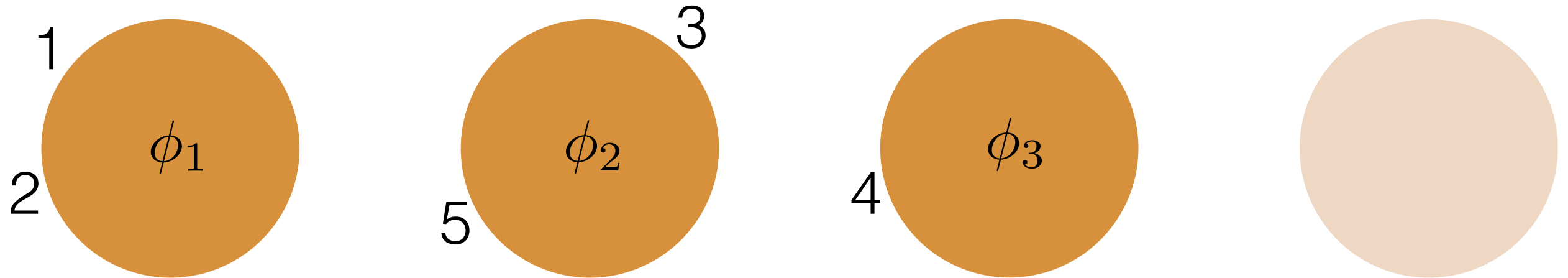
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



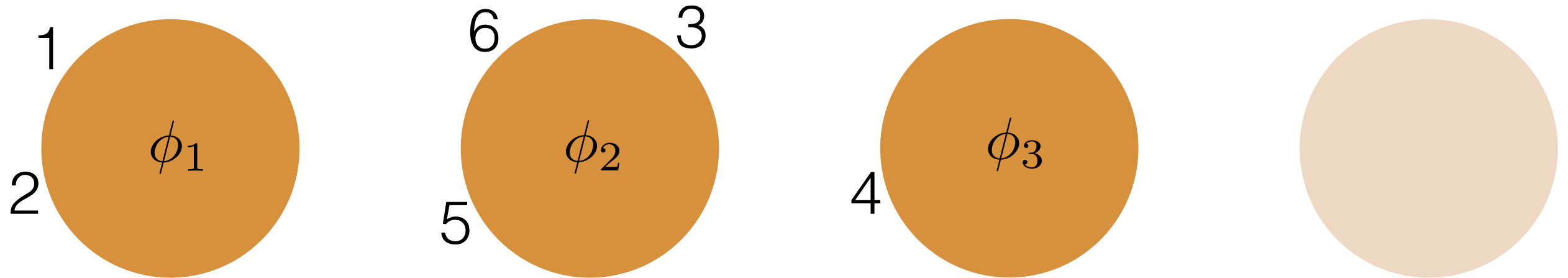
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



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

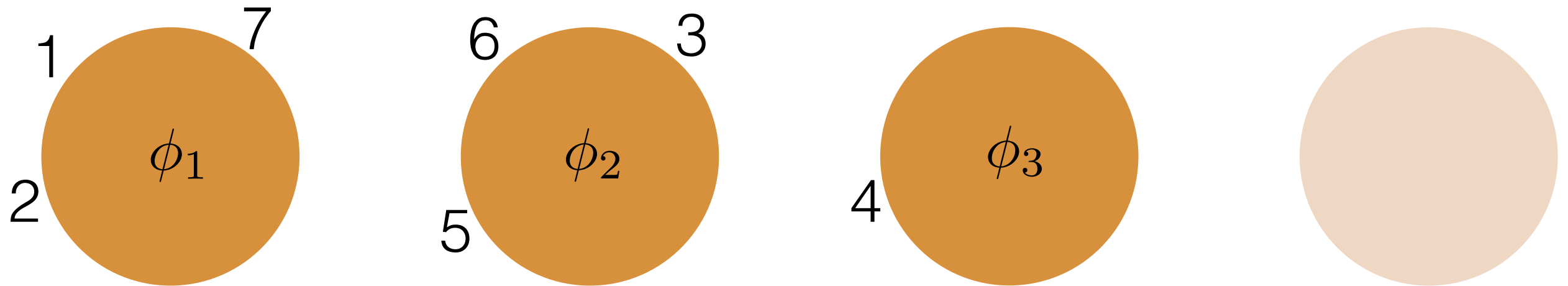
# Chinese restaurant process



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



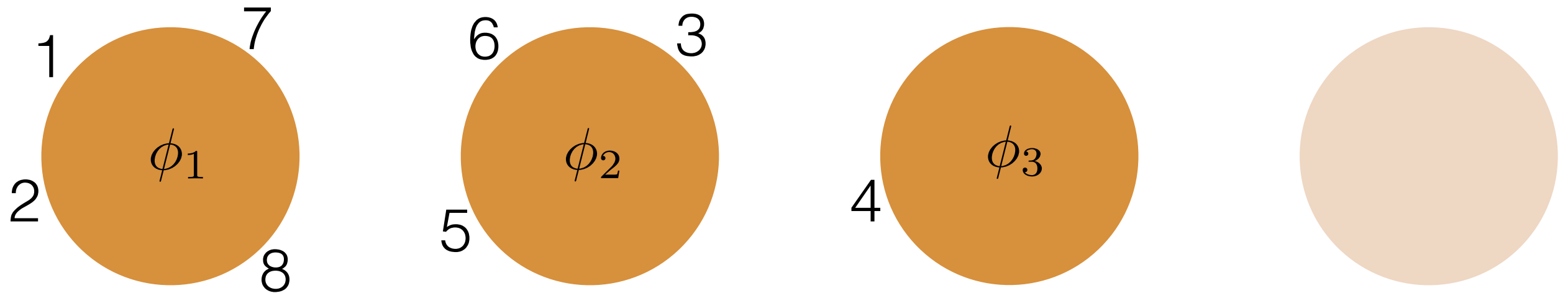
# Chinese restaurant process



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

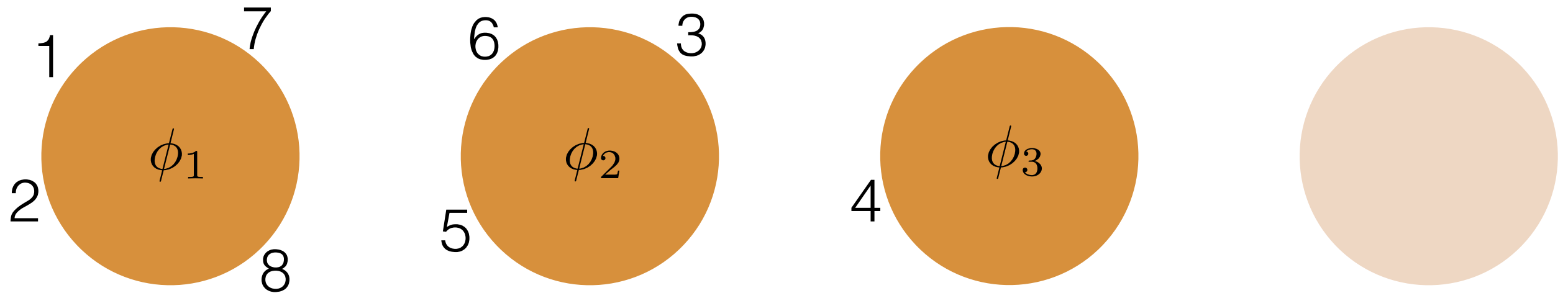


# Chinese restaurant process



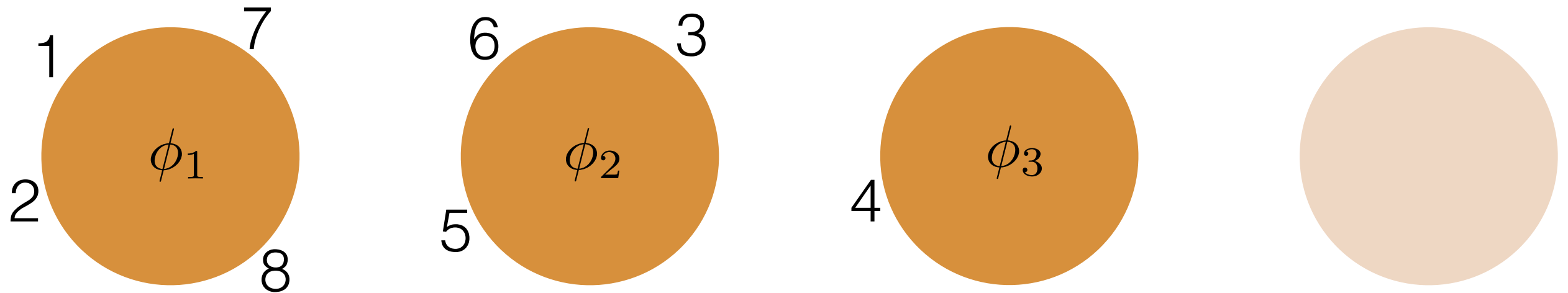
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# Chinese restaurant process



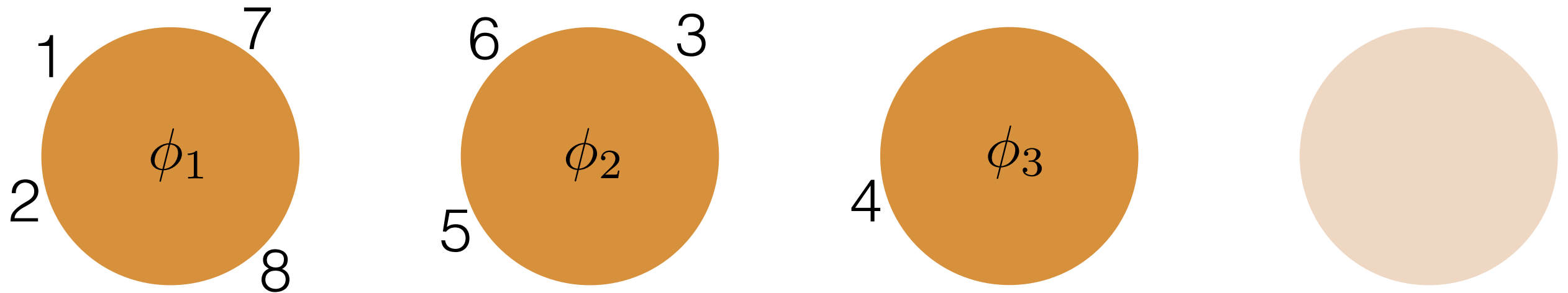
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$

# 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

# Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes
- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?

# Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes
- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?

# Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes
- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?



# Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes
- Big questions
  - Why NPBayes?
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?



# Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes
- Big questions
  - Why NPBayes? Learn more as acquire more data
  - What does an infinite/growing number of parameters really mean (in NPBayes)?
  - Why is NPBayes challenging but practical?

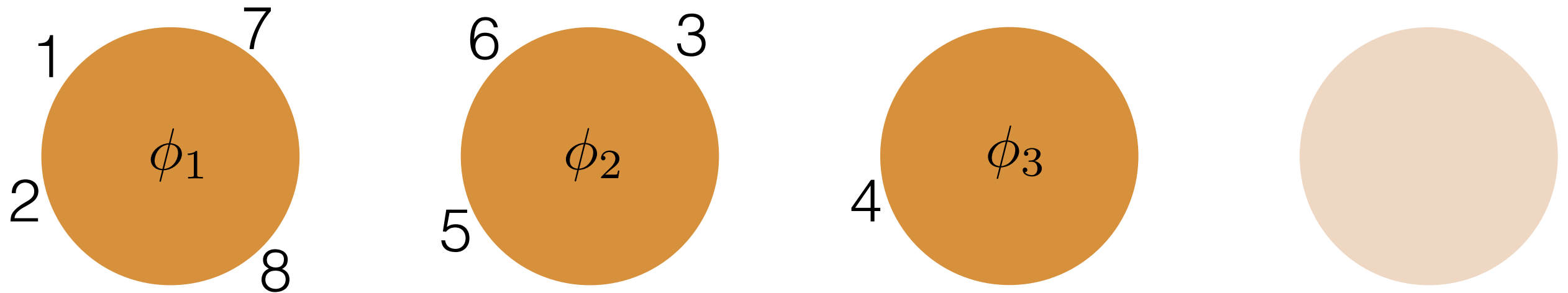
# Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes
- Big questions
  - Why NPBayes? Learn more as acquire more data
  - What does an infinite/growing number of parameters really mean (in NPBayes)? Components vs. clusters; latent vs. realized
  - Why is NPBayes challenging but practical?

# Roadmap

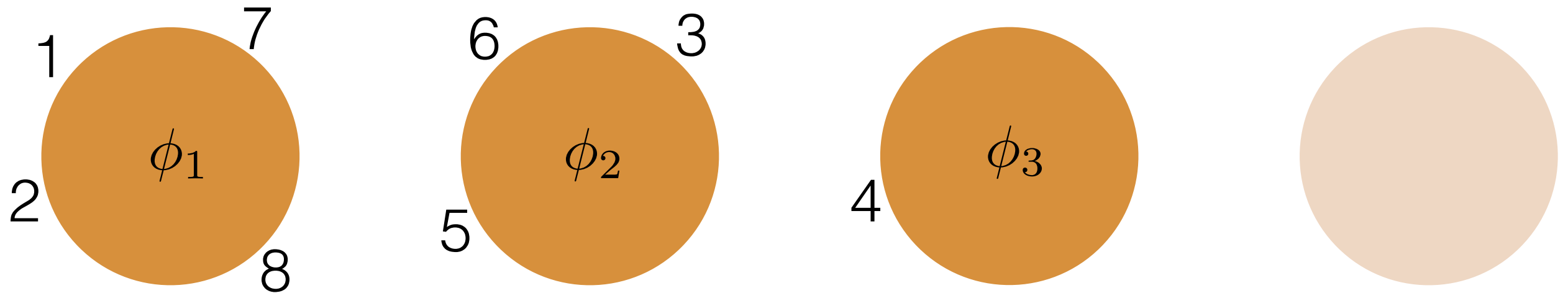
- Example problem: clustering
- Example NPBayes model: Dirichlet process
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayes
- Big questions
  - Why NPBayes? Learn more as acquire more data
  - What does an infinite/growing number of parameters really mean (in NPBayes)? Components vs. clusters; latent vs. realized
  - Why is NPBayes challenging but practical? Infinite dimensional parameter, but finitely many parameters realized

# Chinese restaurant process



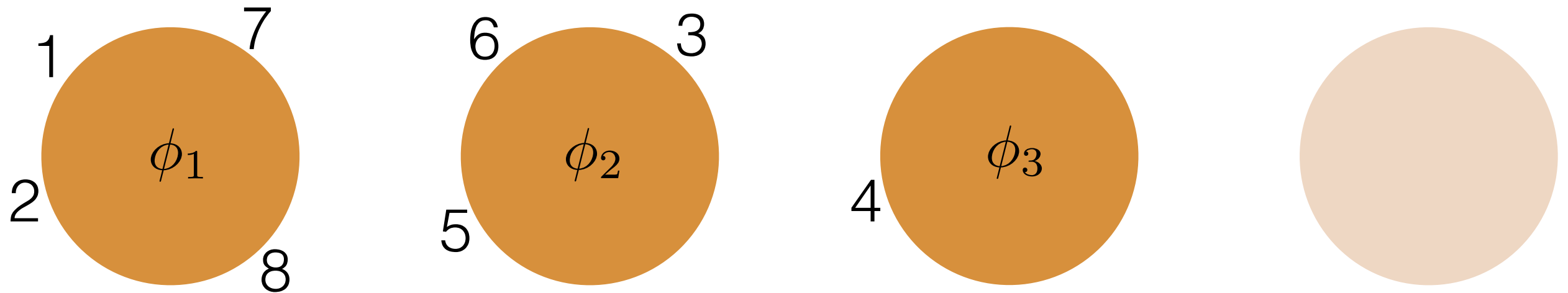
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$
- Marginal for the Categorical likelihood with GEM prior

# Chinese restaurant process



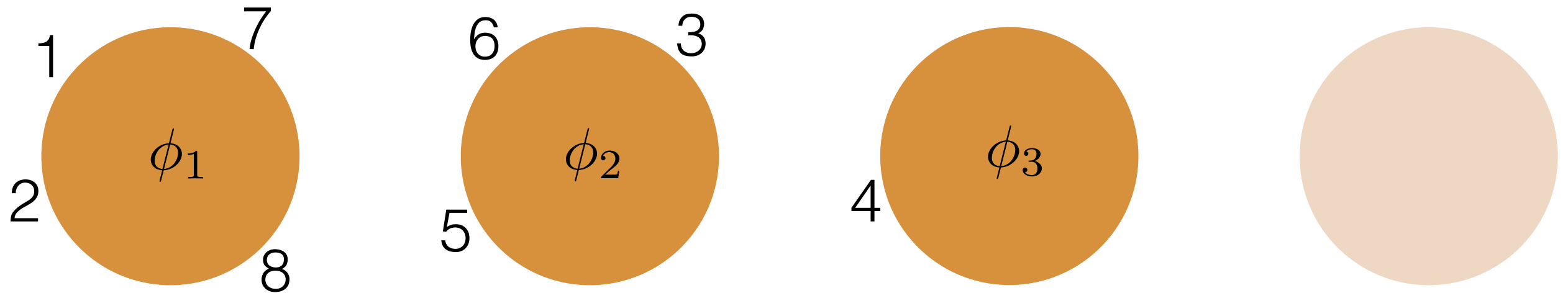
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$
- Marginal for the Categorical likelihood with GEM prior  
 $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$

# Chinese restaurant process



- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to  $\alpha$
- Marginal for the Categorical likelihood with GEM prior
$$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$$
$$\Rightarrow \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}$$

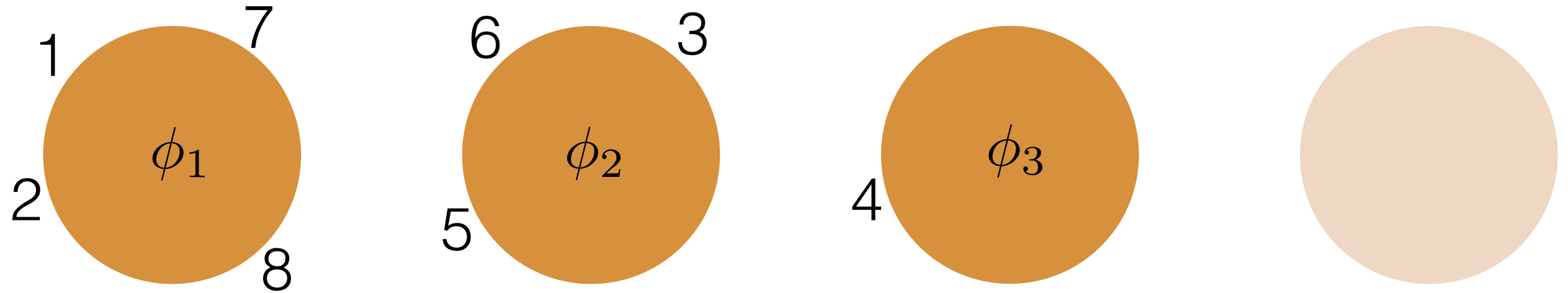
# Chinese restaurant process



- 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
$$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$$
$$\Rightarrow \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}$$
- *Partition of [8]*: set of mutually exclusive & exhaustive sets of  $[8] := \{1, \dots, 8\}$



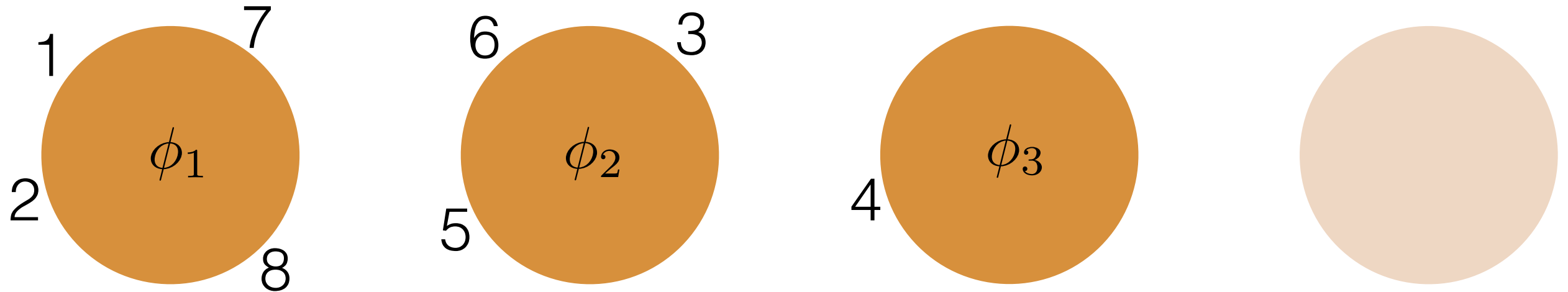
# Chinese restaurant process



- Probability of this seating:

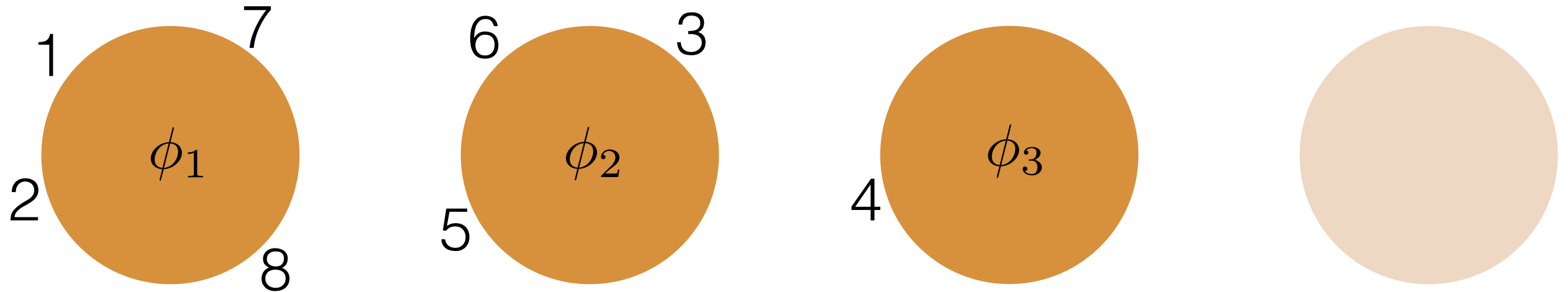


# Chinese restaurant process



- Probability of this seating:  
 $\frac{\alpha}{\alpha}$

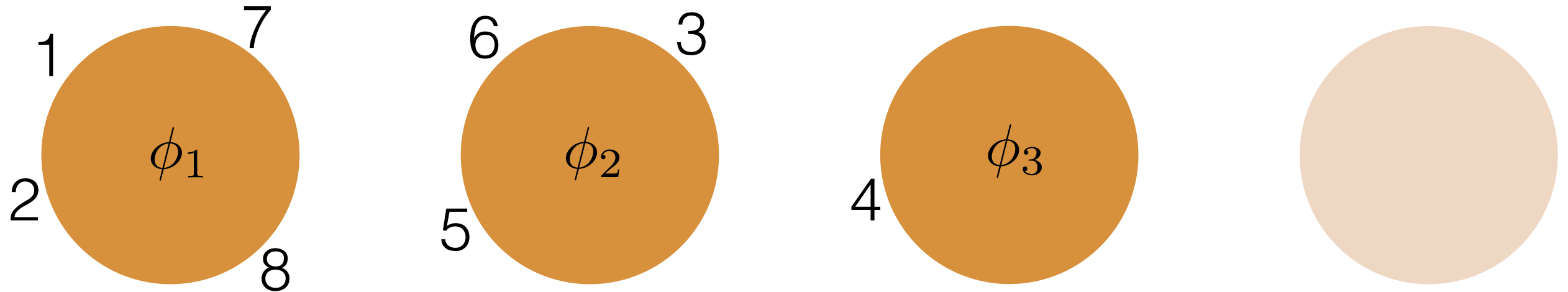
# Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1}$$

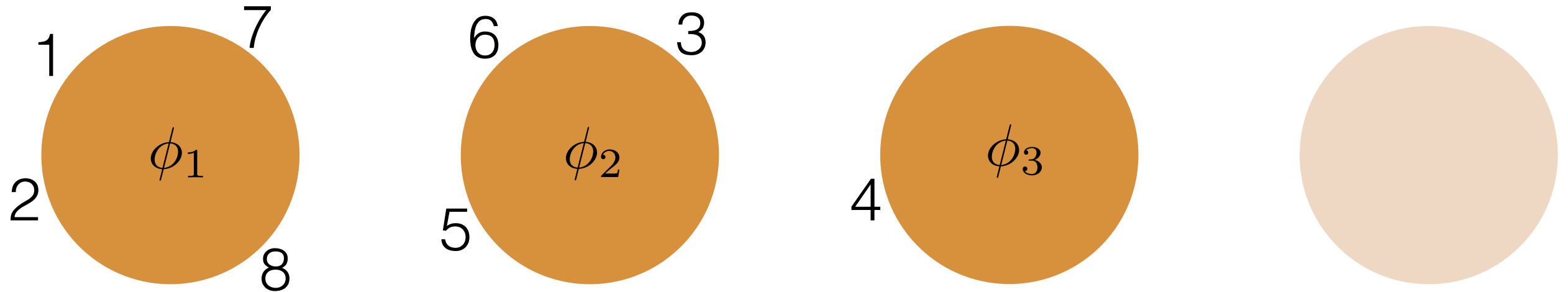
# Chinese restaurant process



- Probability of this seating:

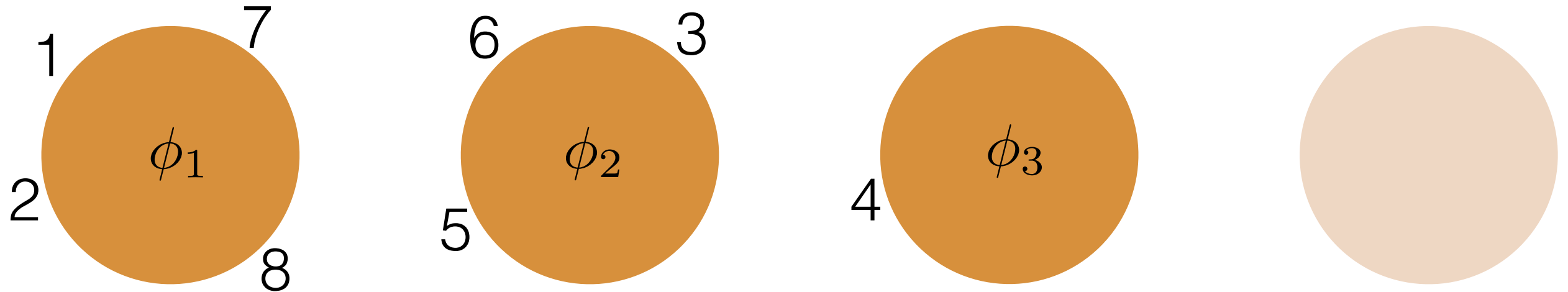
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2}$$

# Chinese restaurant process



- Probability of this seating:  
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3}$$

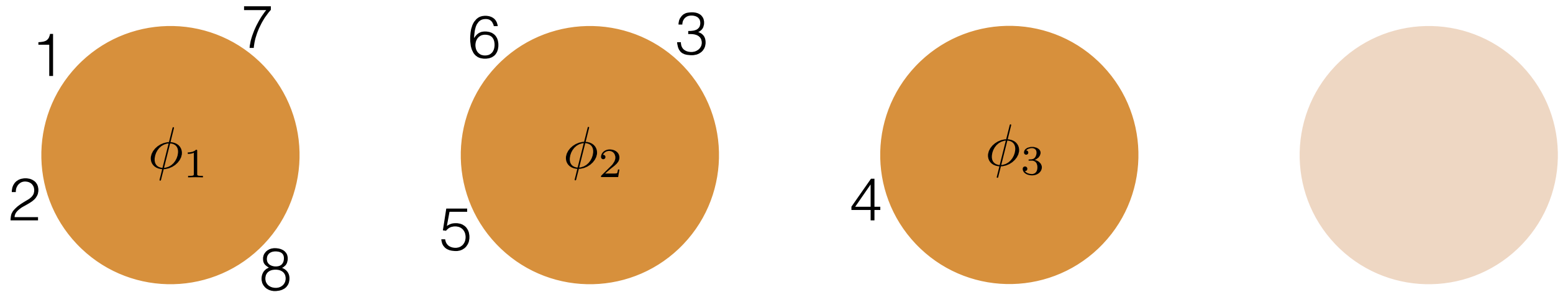
# Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4}$$

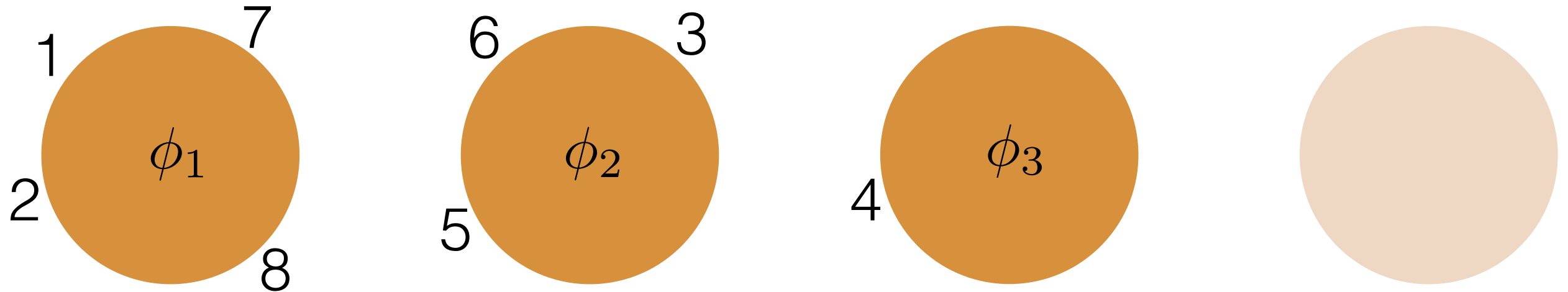
# Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5}$$

# Chinese restaurant process

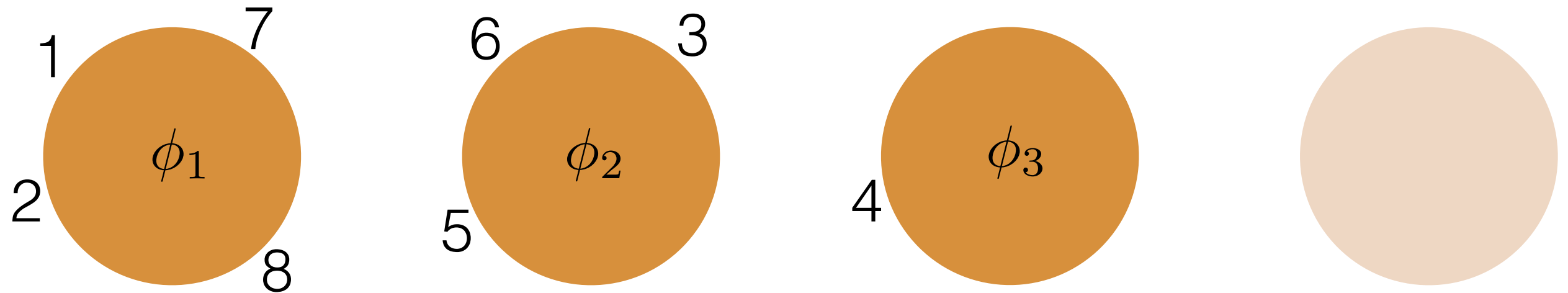


- Probability of this seating:

$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6}$$



# Chinese restaurant process

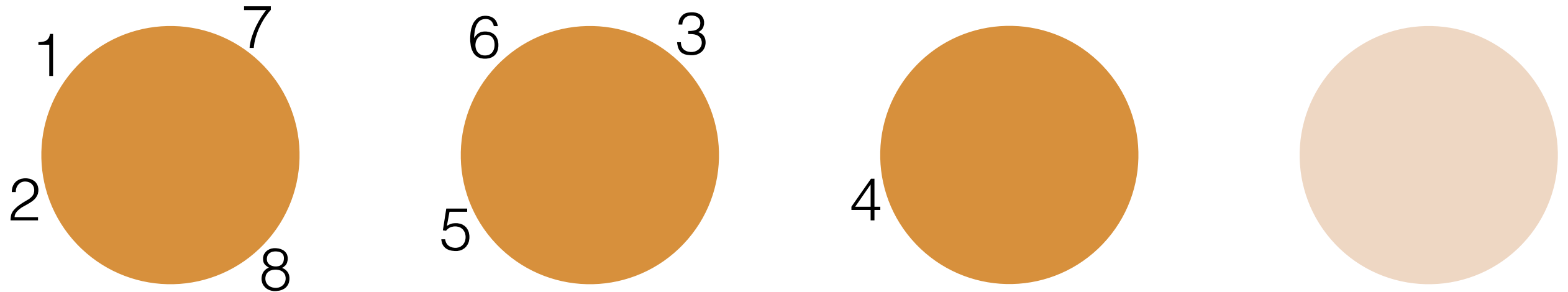


- Probability of this seating:

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



# Chinese restaurant process

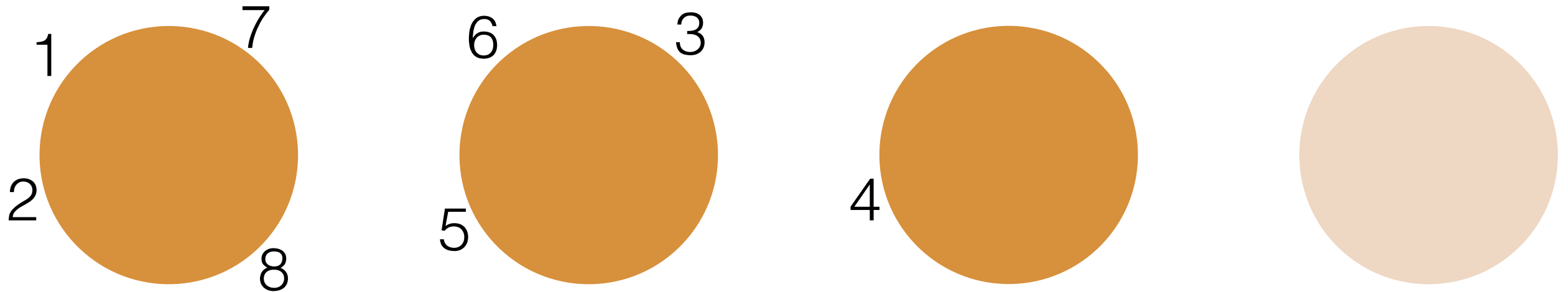


- Probability of this seating:

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

- Probability of  $N$  customers ( $K_N$  tables,  $n_k$  at table  $k$ ):

# Chinese restaurant process



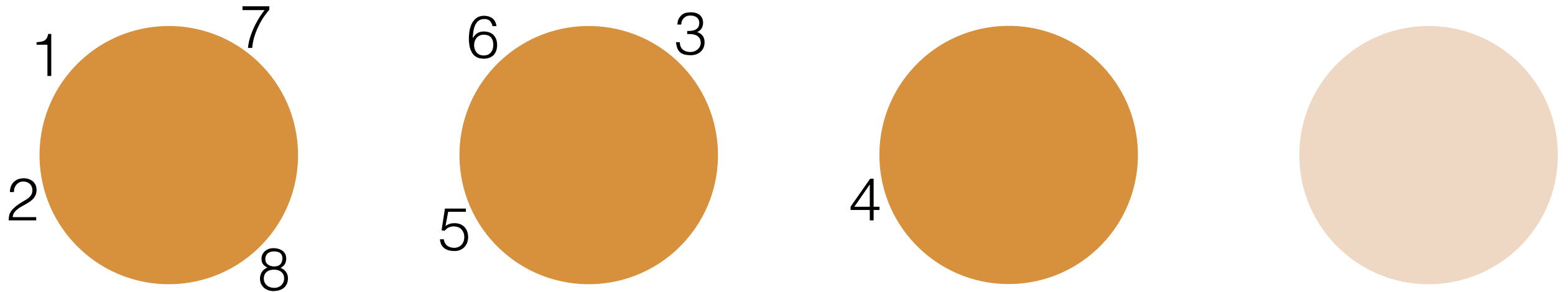
- Probability of this seating:

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

- Probability of  $N$  customers ( $K_N$  tables,  $n_k$  at table  $k$ ):

---

# Chinese restaurant process



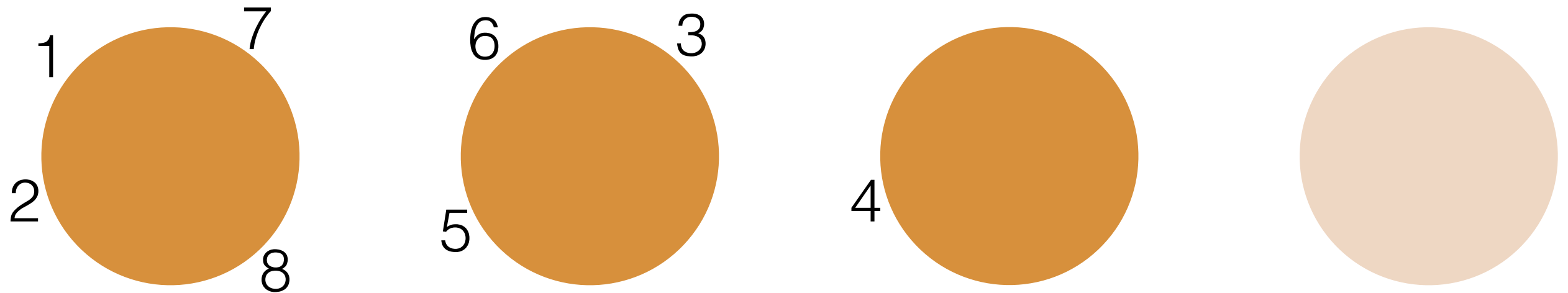
- Probability of this seating:

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

- Probability of  $N$  customers ( $K_N$  tables,  $n_k$  at table  $k$ ):

$$\frac{1}{\alpha \cdots (\alpha + N - 1)}$$

# Chinese restaurant process



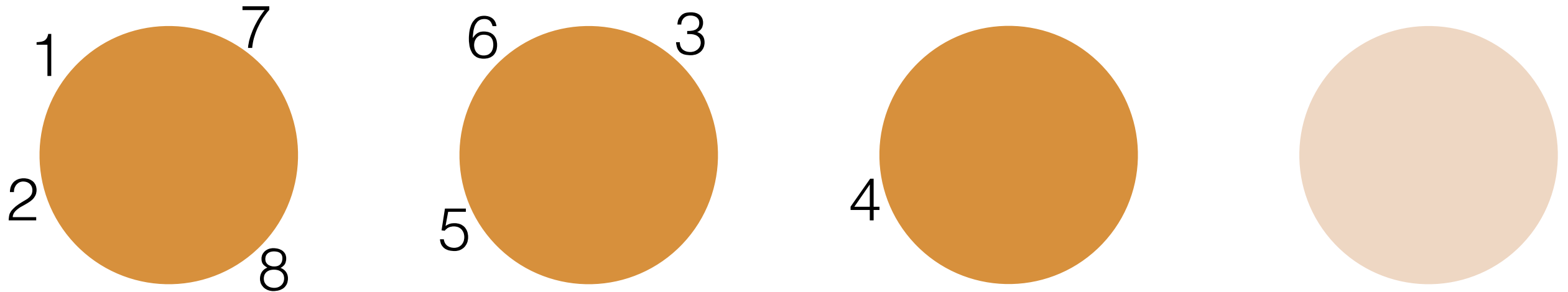
- Probability of this seating:

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

- Probability of  $N$  customers ( $K_N$  tables,  $n_k$  at table  $k$ ):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$

# Chinese restaurant process



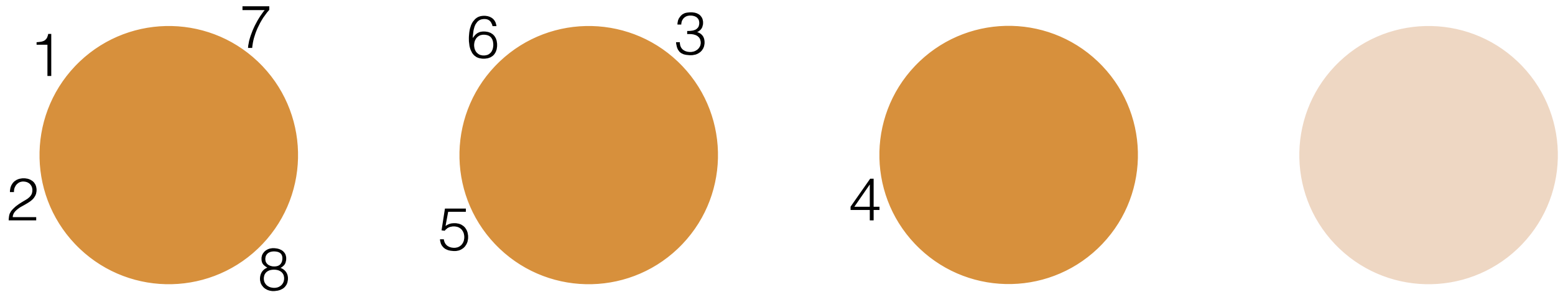
- Probability of this seating:

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

- Probability of  $N$  customers ( $K_N$  tables,  $n_k$  at table  $k$ ):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$

# Chinese restaurant process



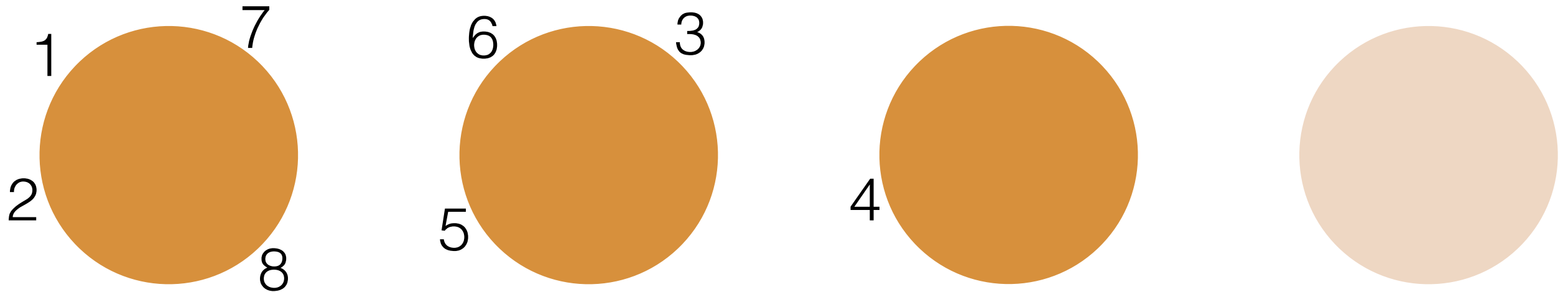
- Probability of this seating:

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

- 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:

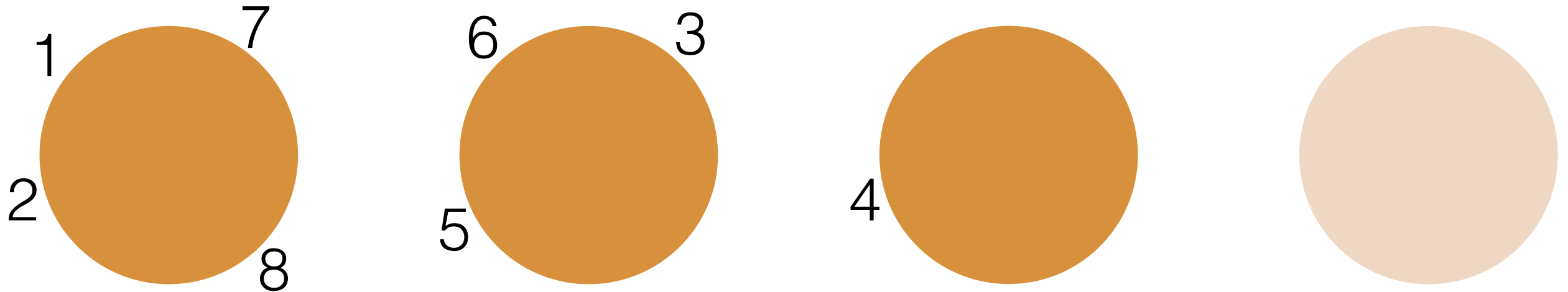
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}$$

- 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:

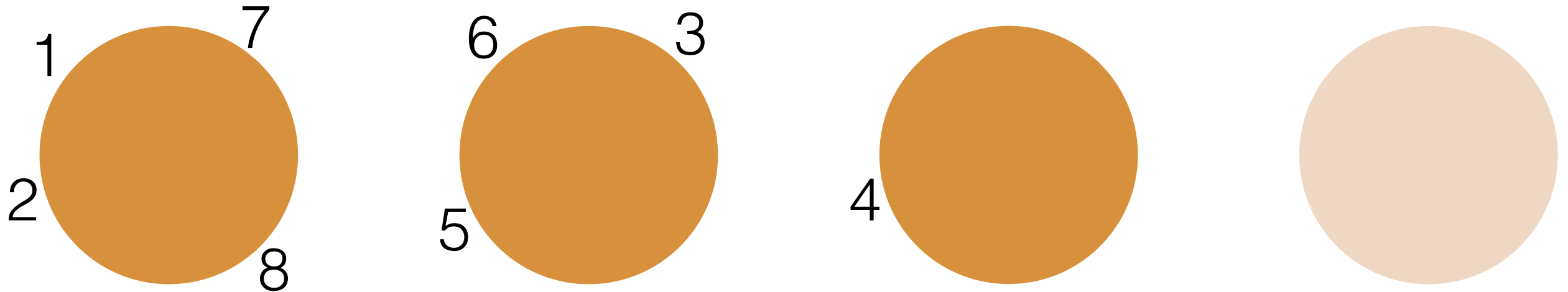
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}$$

- Probability of  $N$  customers ( $K_N$  tables,  $n_k$  at table  $k$ ):

$$\frac{\alpha^{K_N}}{\alpha \cdots (\alpha + N - 1)}$$



# Chinese restaurant process



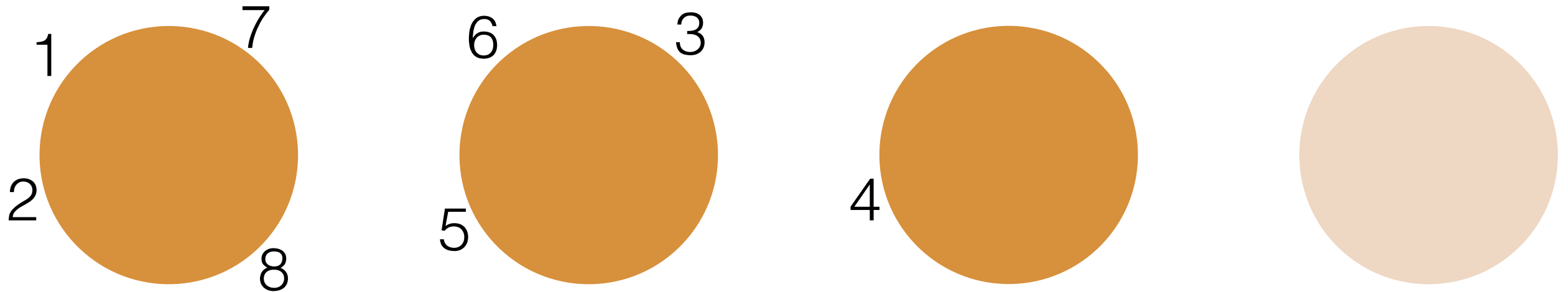
- Probability of this seating:

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

- Probability of  $N$  customers ( $K_N$  tables,  $n_k$  at table  $k$ ):

$$\frac{\alpha^{K_N} \prod_{k=1}^{K_N} (n_k - 1)!}{\alpha \cdots (\alpha + N - 1)}$$

# Chinese restaurant process



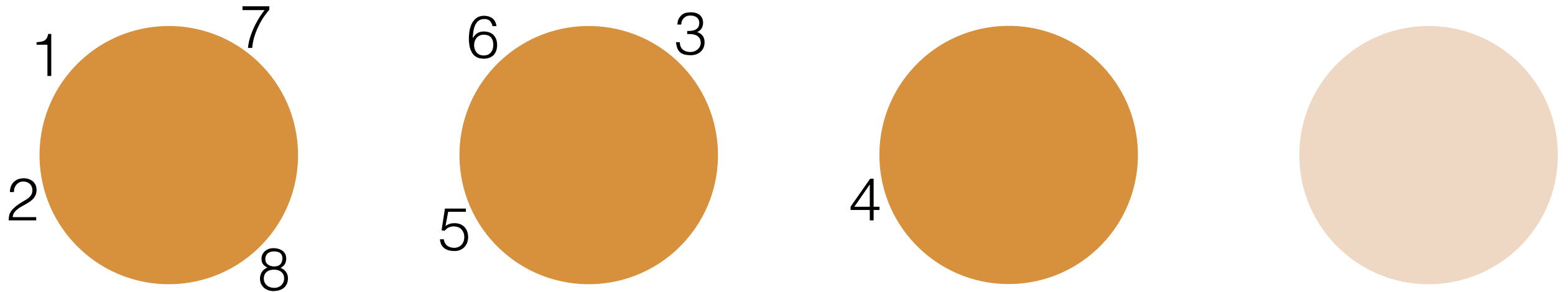
- Probability of this seating:

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

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

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

# Chinese restaurant process



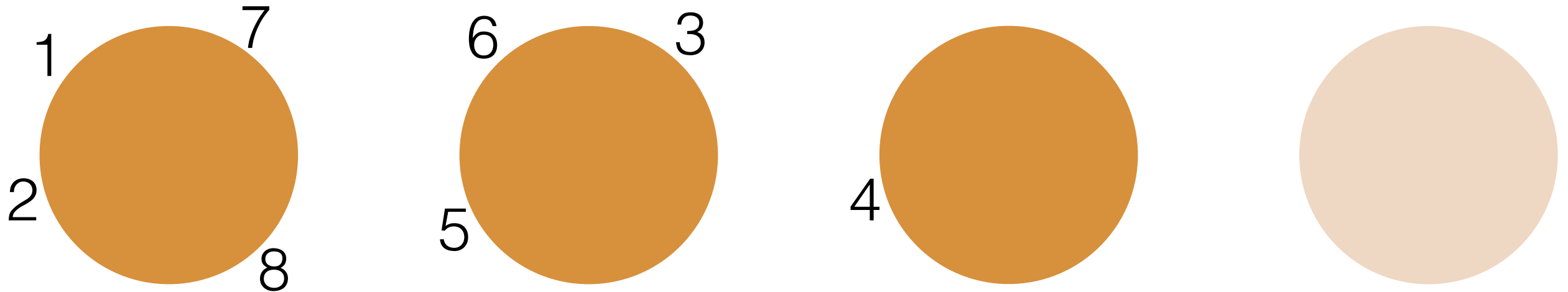
- Probability of this seating:

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

- 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:

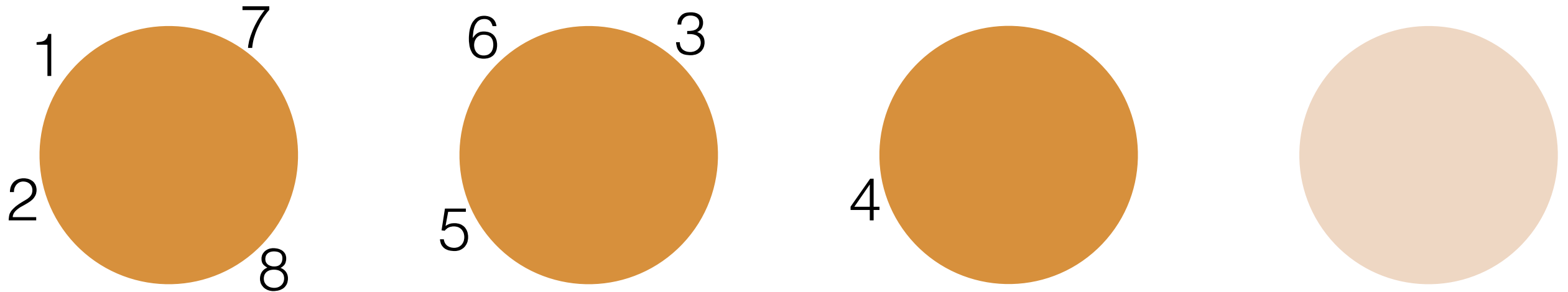
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}$$

- 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:

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

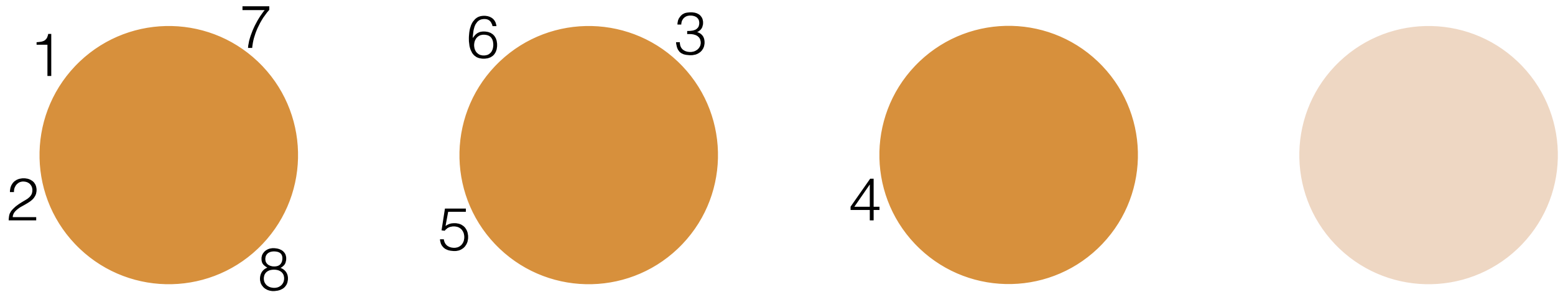
- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):

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

- 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  $C$ ):

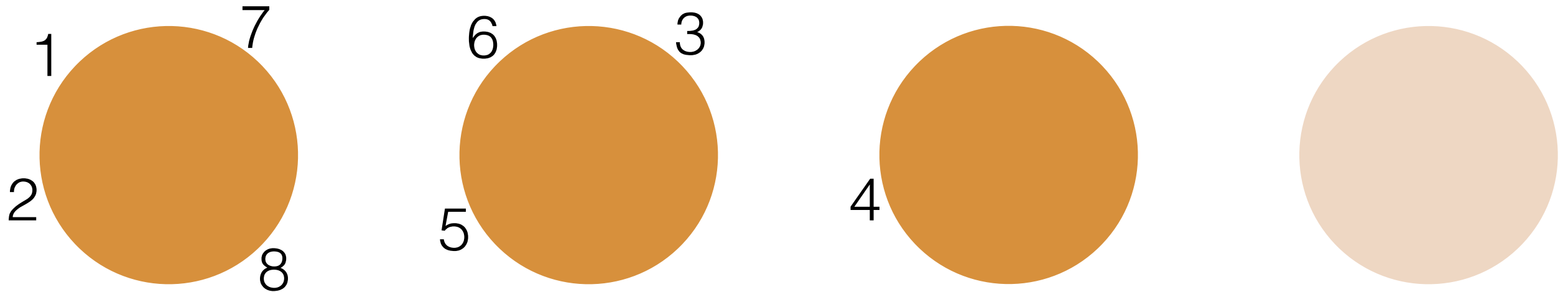
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- 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  $C$ ):

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

- Prob doesn't depend on customer order: *exchangeable*

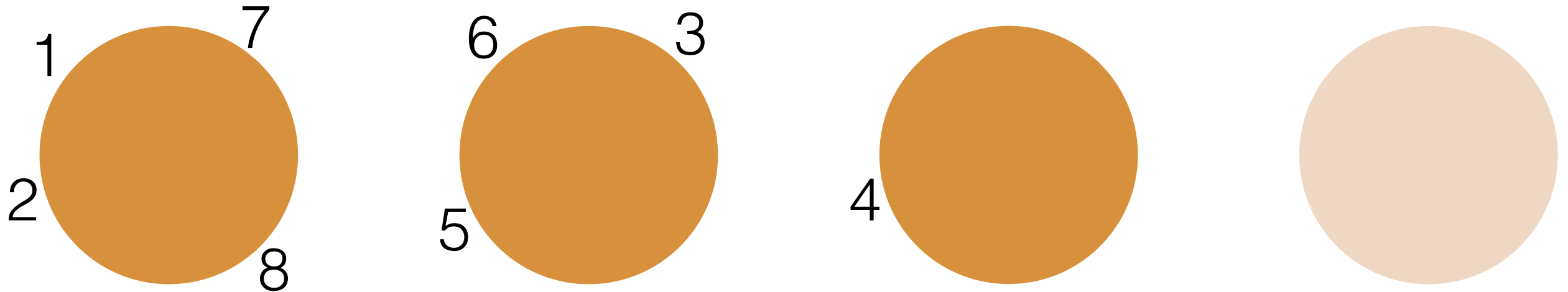
$$\mathbb{P}(\Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}) = \mathbb{P}(\Pi_8 = \{\{2, 3, 8, 1\}, \{4, 6, 7\}, \{5\}\})$$

- Can always pretend  $n$  is the last customer and calculate  $p(\Pi_N | \Pi_{N, -n})$

- e.g.  $\Pi_{8, -5} = \{\{1, 2, 7, 8\}, \{3, 6\}, \{4\}\}$



# Chinese restaurant process



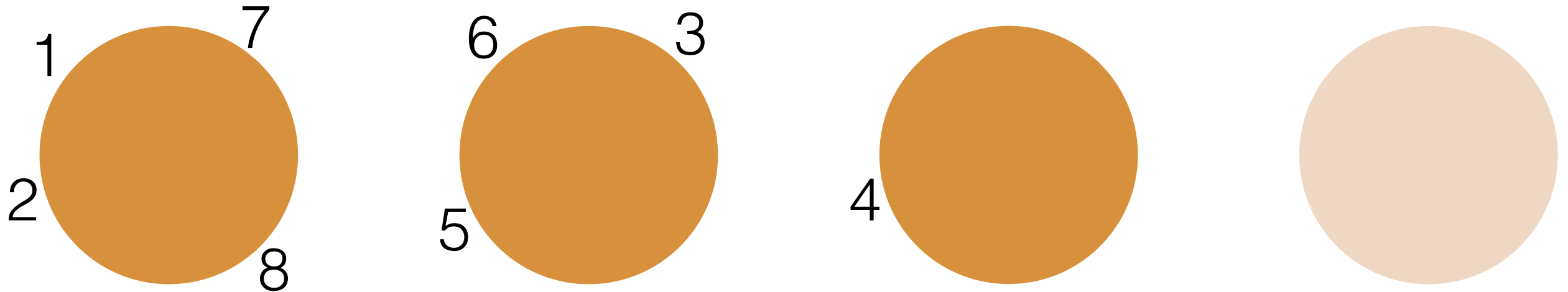
- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):

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

- So:  $p(\Pi_N | \Pi_{N,-n}) =$



# Chinese restaurant process

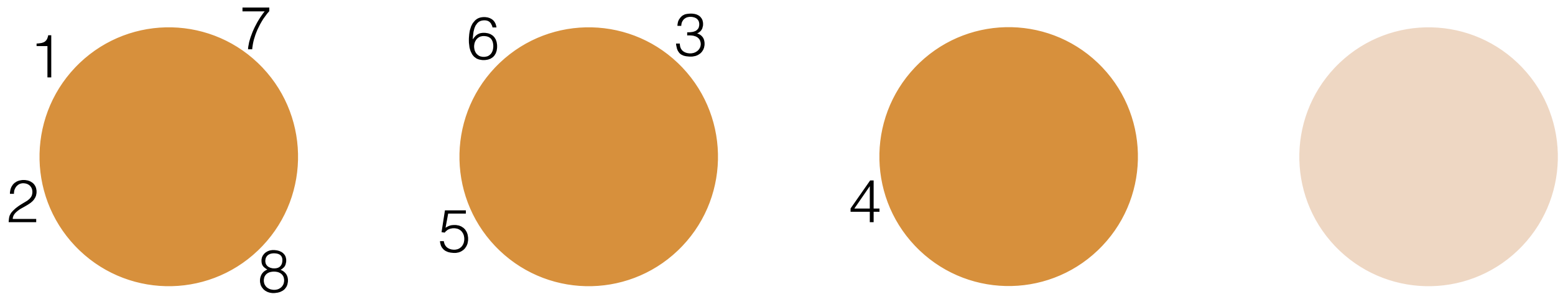


- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):

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

- So:  $p(\Pi_N | \Pi_{N,-n}) = \left\{ \right.$

# Chinese restaurant process

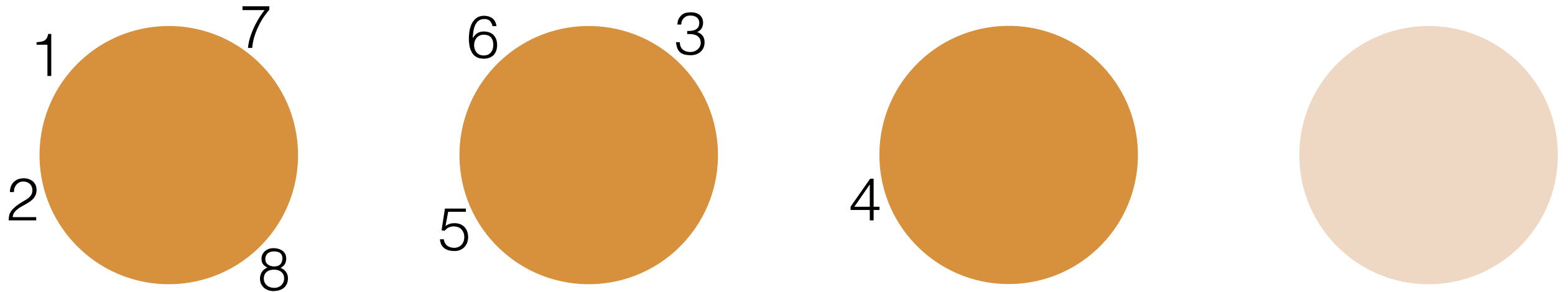


- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):

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

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

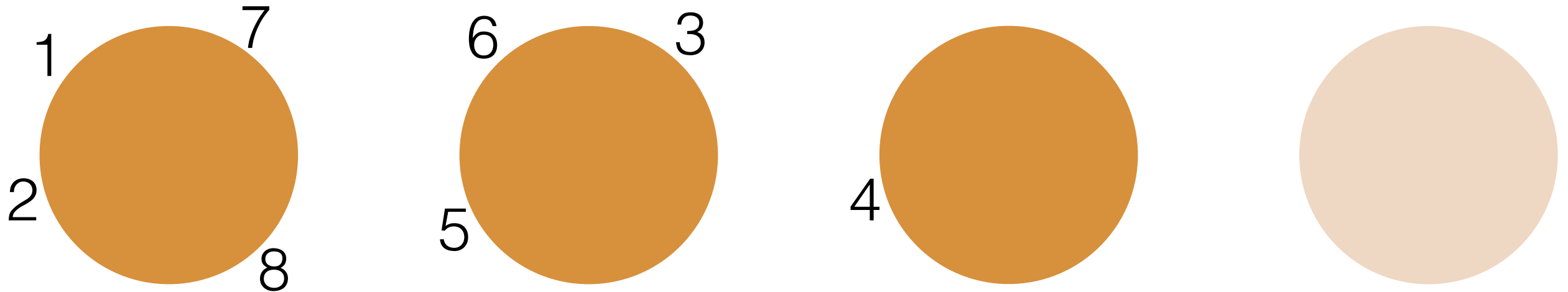
# Chinese restaurant process



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

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

# Chinese restaurant process

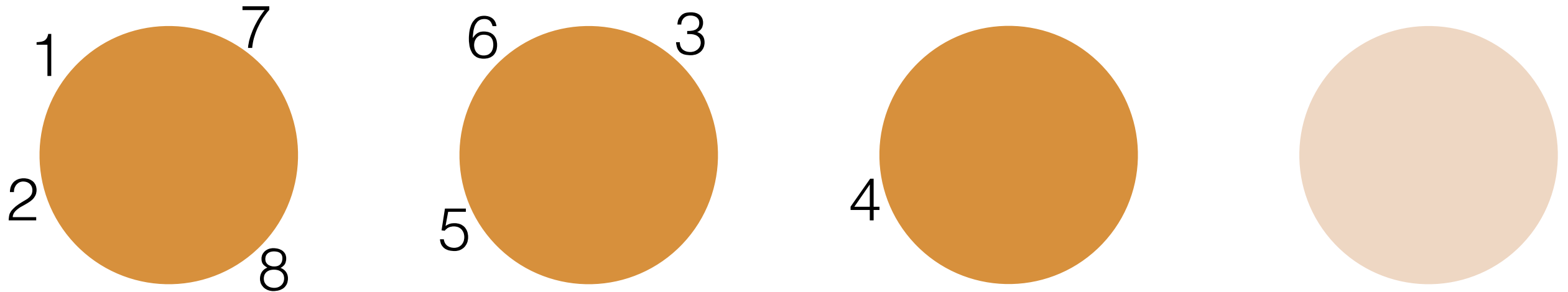


- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):

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

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

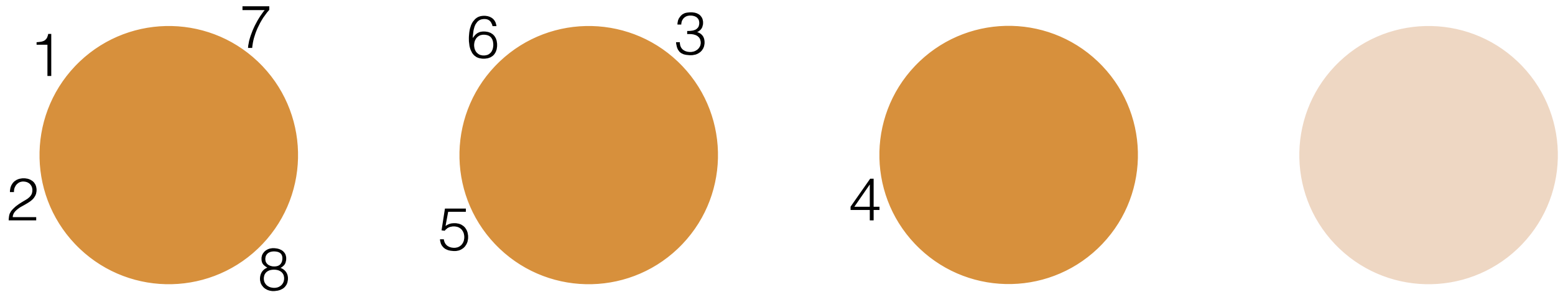
# Chinese restaurant process



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

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

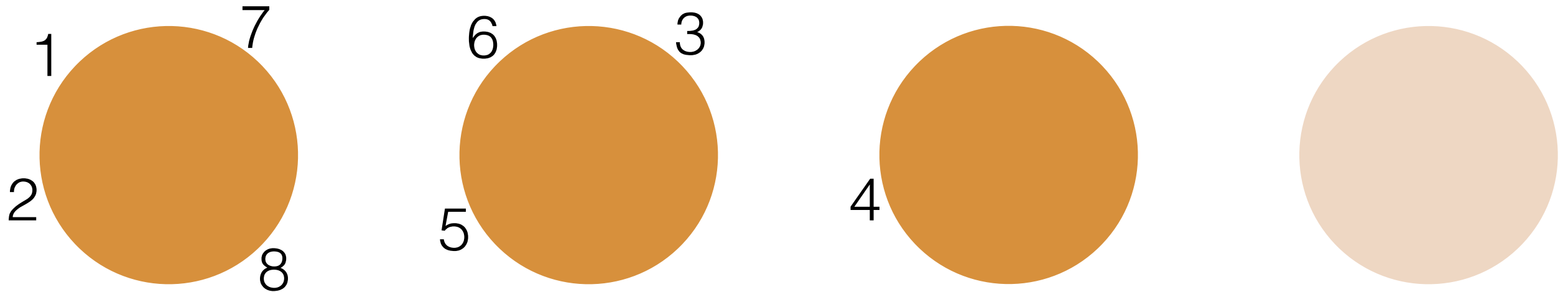
# Chinese restaurant process



- 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

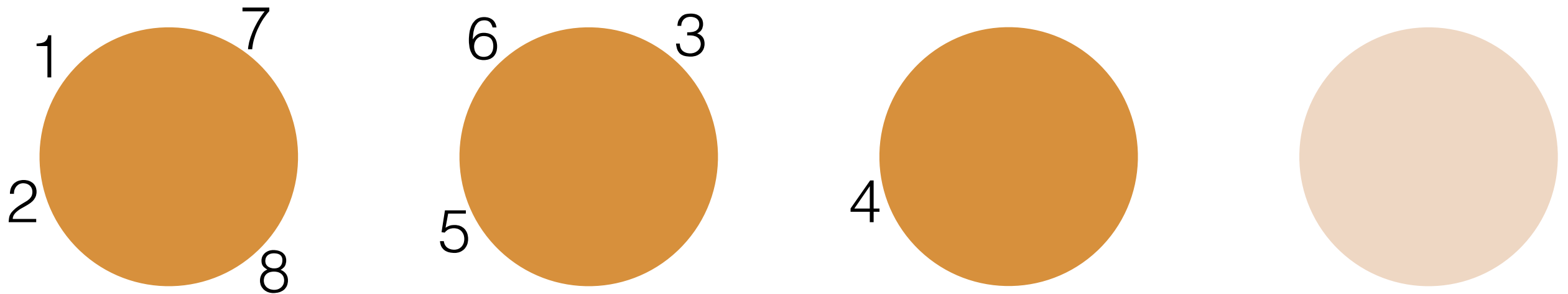


- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):  

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

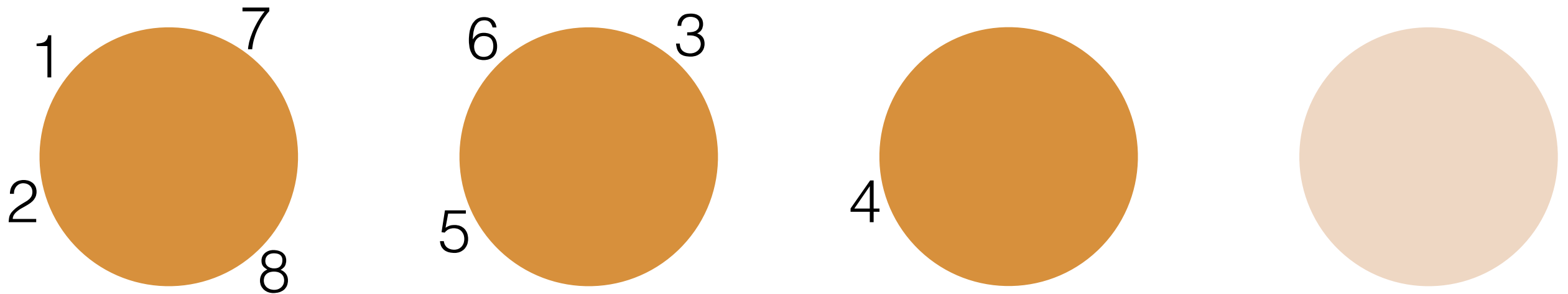


- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):  

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$
- 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:  $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - $t^{\text{th}}$  step:  $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$



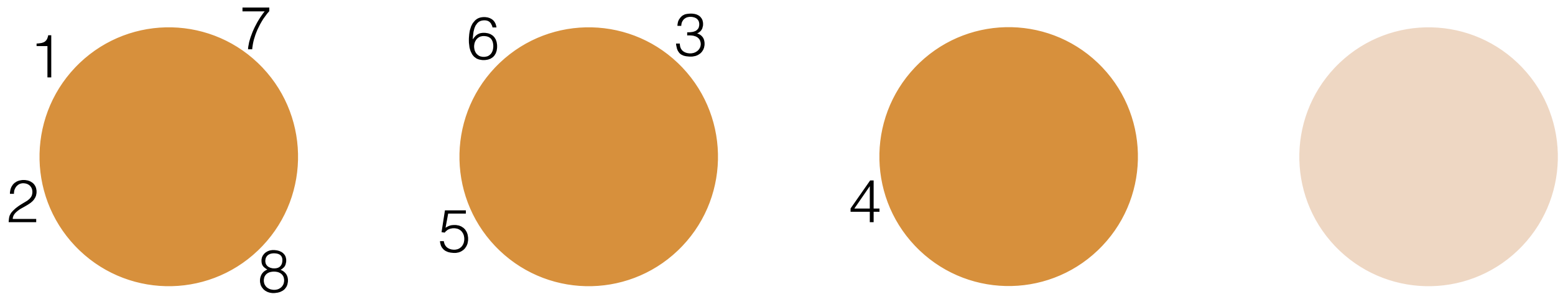
# Chinese restaurant process



- 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:  $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$   $v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$
  - $t^{\text{th}}$  step:  $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

# Chinese restaurant process



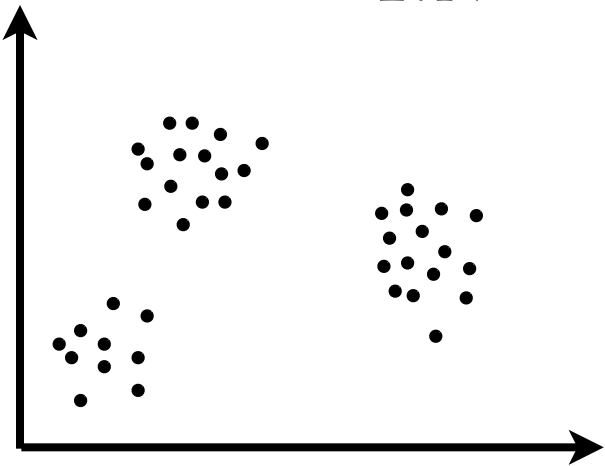
- Probability of  $N$  customers ( $K_N$  tables,  $\#C$  at table  $C$ ):  

$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$
- 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:  $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$   $v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$
  - $t^{\text{th}}$  step:  $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$   $v_3^{(t)} \sim p(v_3 | v_1^{(t)}, v_2^{(t)})$

# CRP mixture model: inference

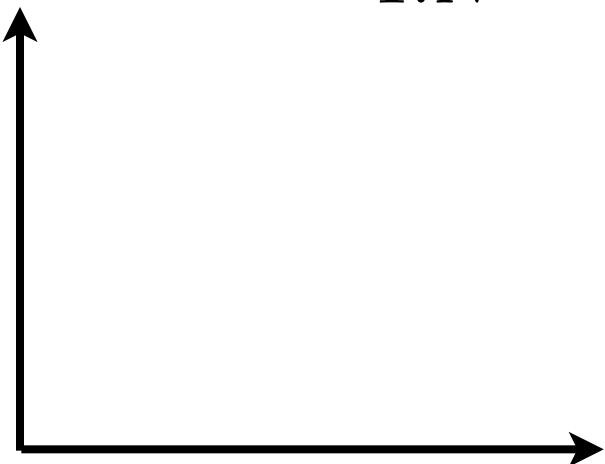
# CRP mixture model: inference

- Data  $x_{1:N}$



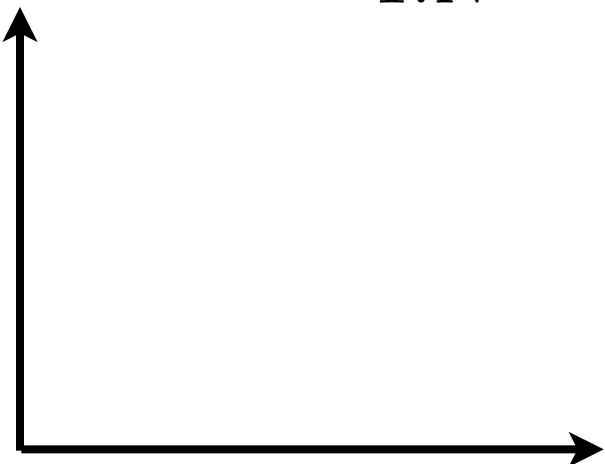
# CRP mixture model: inference

- Data  $x_{1:N}$



# CRP mixture model: inference

- Data  $x_{1:N}$
- Generative model



# CRP mixture model: inference

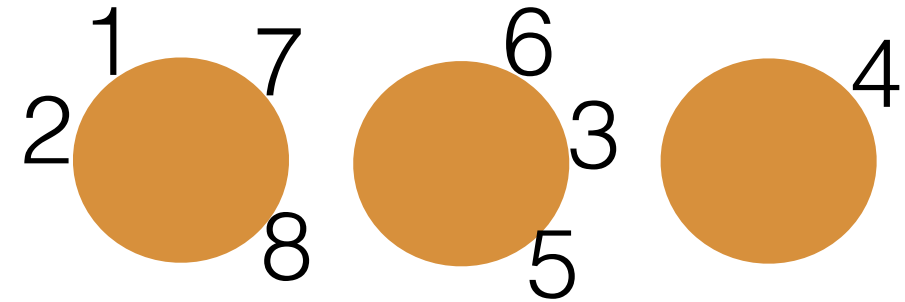
- Data  $x_{1:N}$
- Generative model  
 $\Pi_N \sim \text{CRP}(N, \alpha)$





# CRP mixture model: inference

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

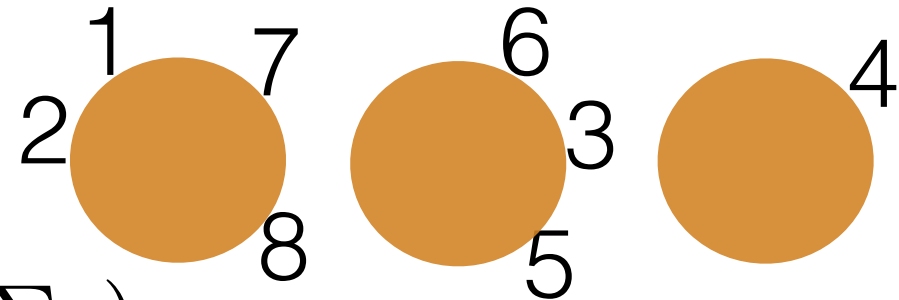




# CRP mixture model: inference

- Data  $x_{1:N}$
- 

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



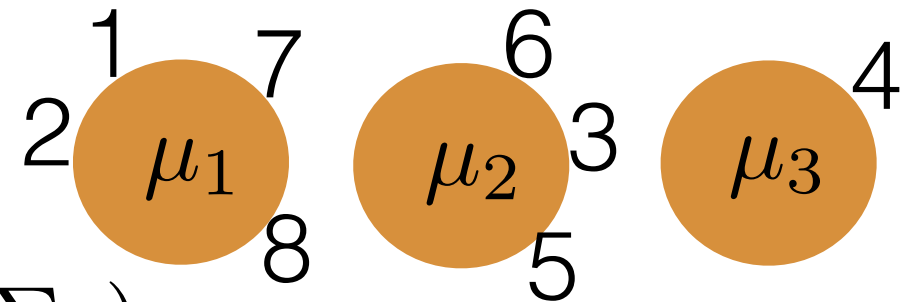
# CRP mixture model: inference

- Data  $x_{1:N}$

- Generative model

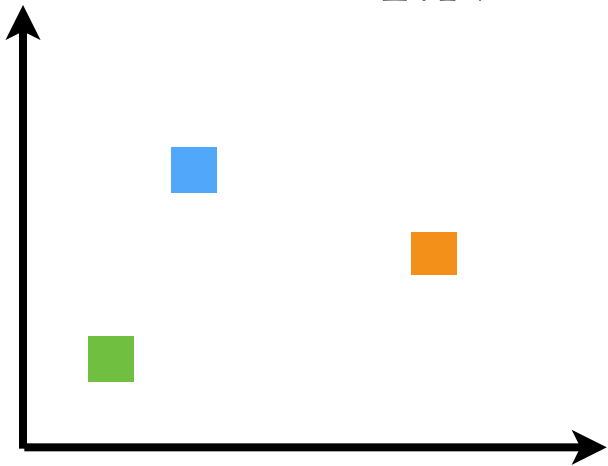
$$\Pi_N \sim \text{CRP}(N, \alpha)$$

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



# CRP mixture model: inference

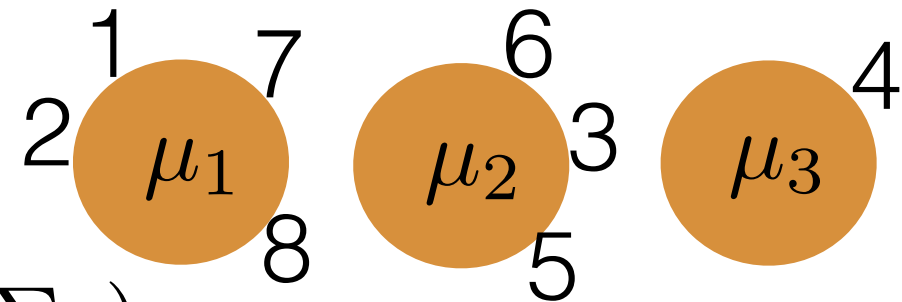
- Data  $x_{1:N}$



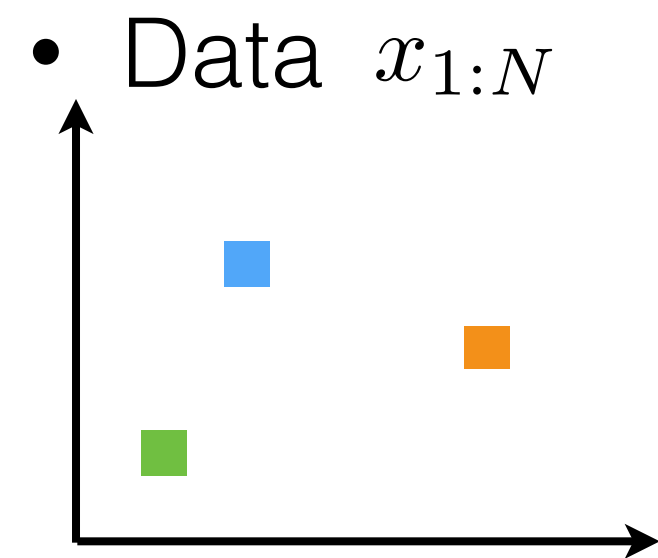
- Generative model

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

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



# CRP mixture model: inference

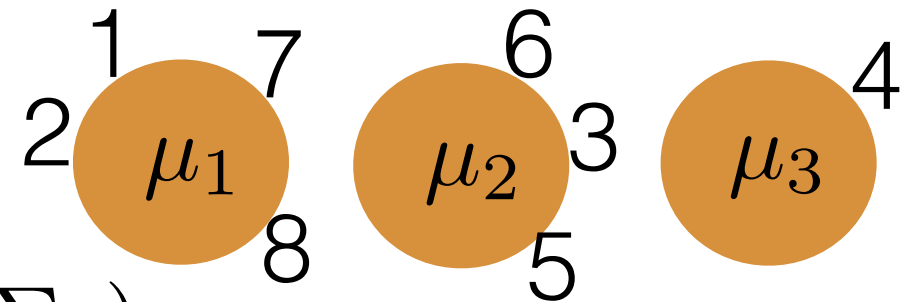


- Generative model

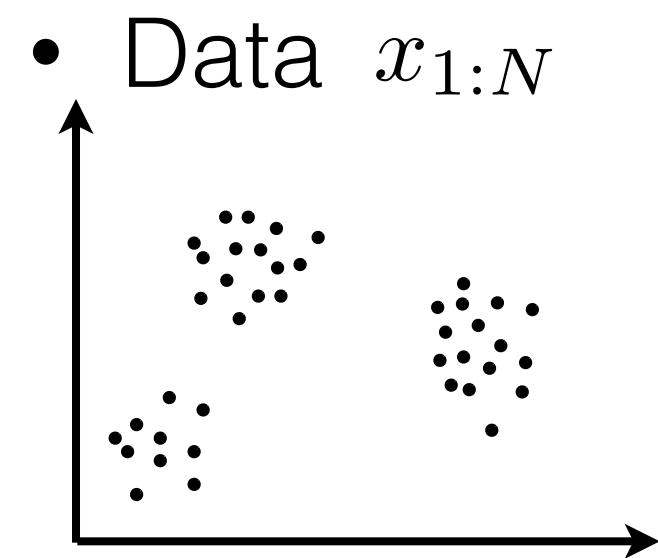
$$\Pi_N \sim \text{CRP}(N, \alpha)$$

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

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



# CRP mixture model: inference

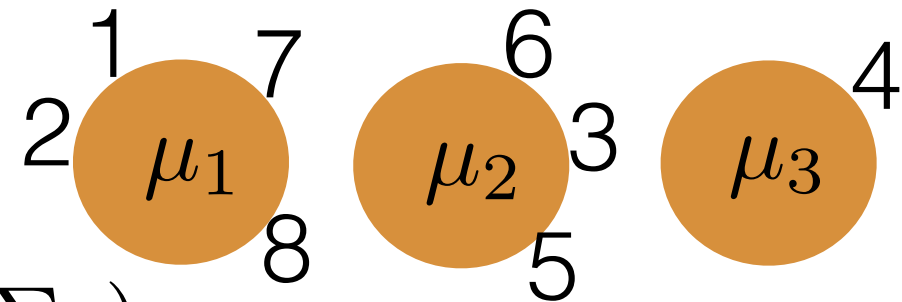


- Generative model

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

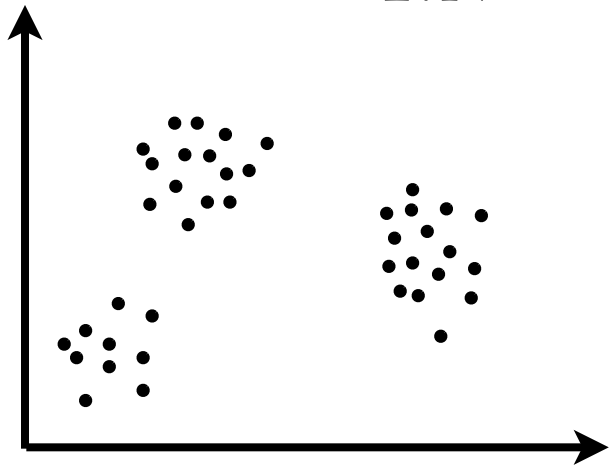
$$\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

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



# CRP mixture model: inference

- Data  $x_{1:N}$

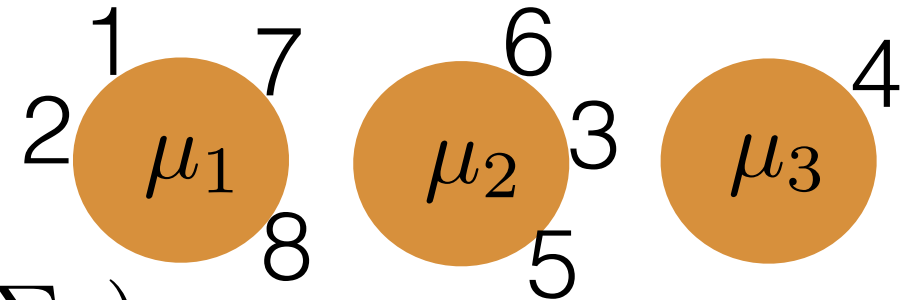


- Generative model

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

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

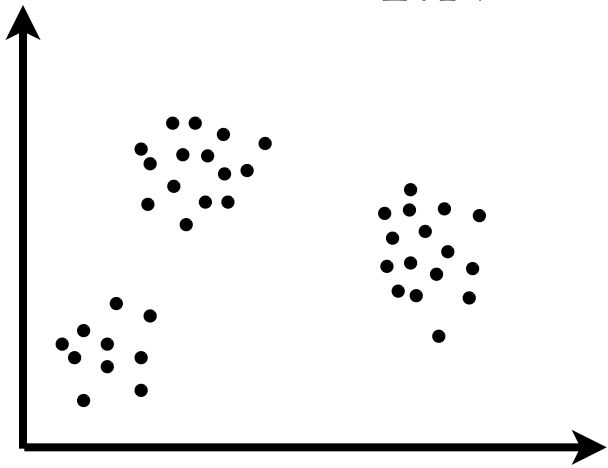
$$\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior

# CRP mixture model: inference

- Data  $x_{1:N}$

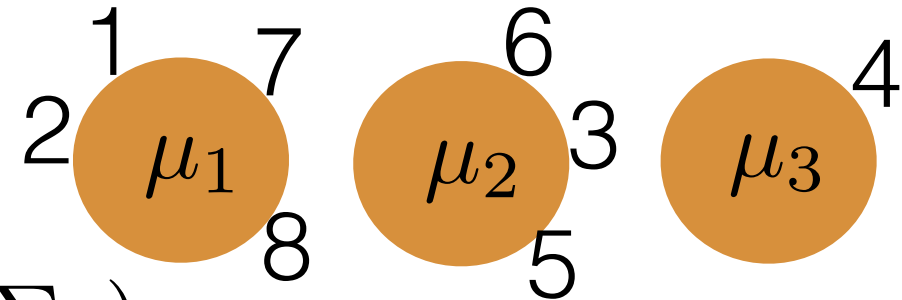


- Generative model

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

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

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

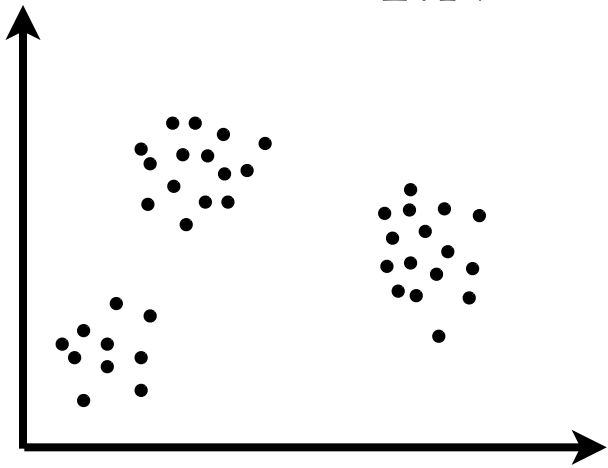


- Want: posterior  $p(\Pi_N | x_{1:N})$



# CRP mixture model: inference

- Data  $x_{1:N}$

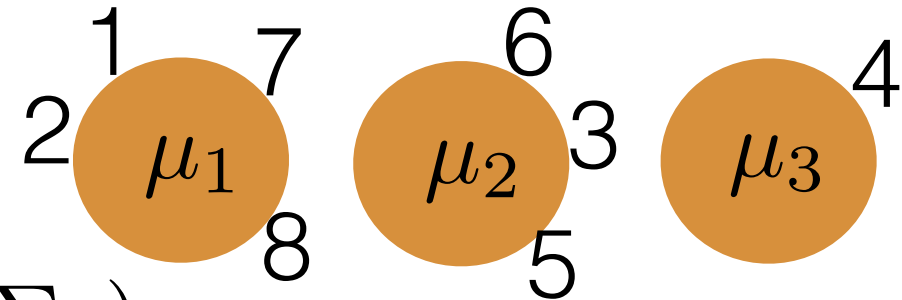


- Generative model

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

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

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

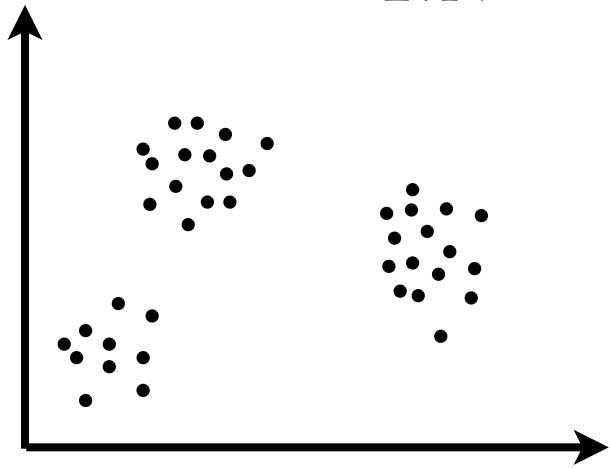


- Want: posterior  $p(\Pi_N | x_{1:N})$
- Gibbs sampler:



# CRP mixture model: inference

- Data  $x_{1:N}$

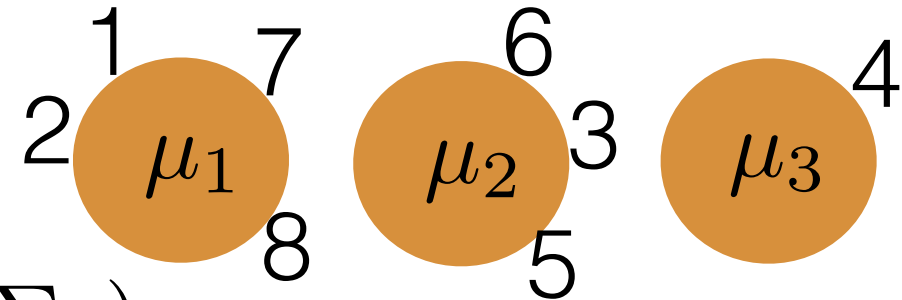


- Generative model

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

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

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



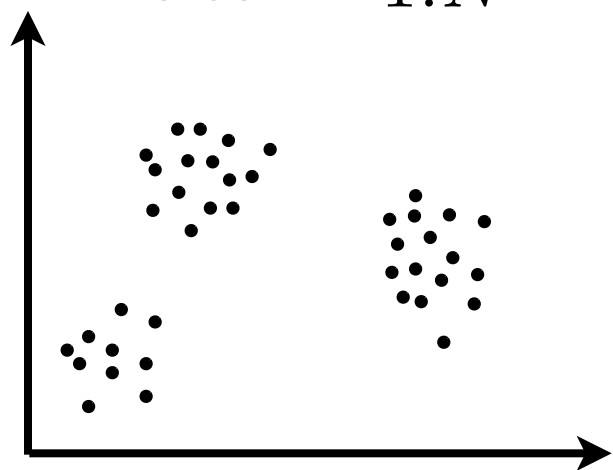
- Want: posterior  $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,-n}, x)$$

# CRP mixture model: inference

- Data  $x_{1:N}$

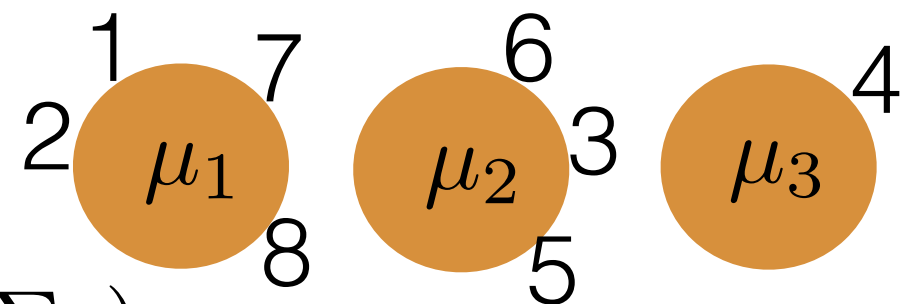


- Generative model

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

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

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



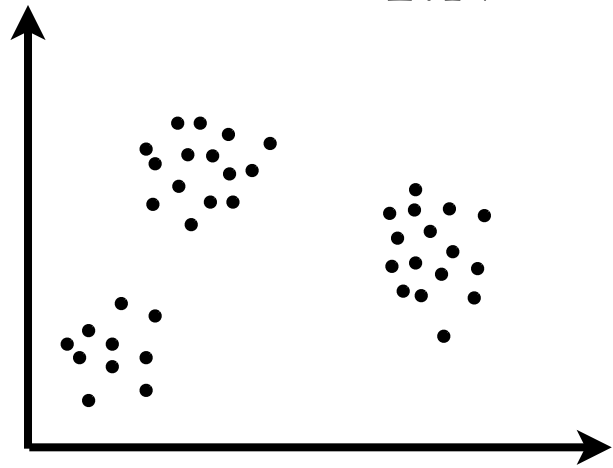
- Want: posterior  $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,-n}, x) \propto \left\{ \right.$$

# CRP mixture model: inference

- Data  $x_{1:N}$

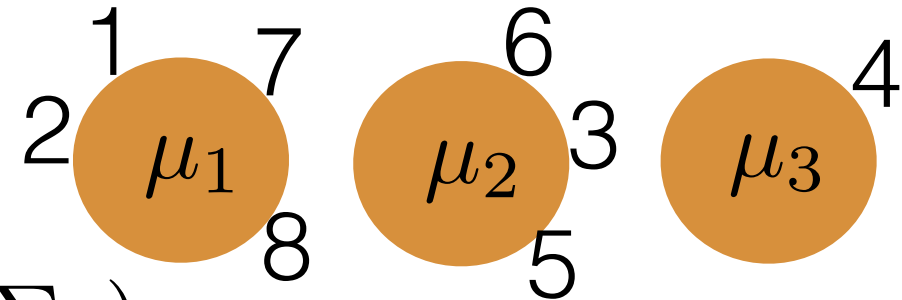


- Generative model

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

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

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



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

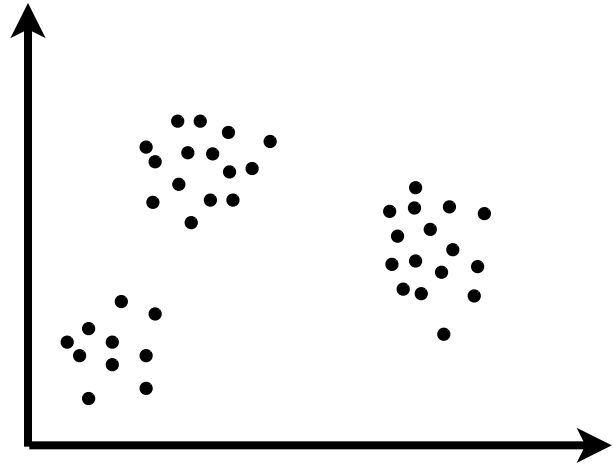
- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,-n}, x) \propto \left\{ \begin{array}{l} \end{array} \right.$$

if  $n$  joins cluster  $C$

# CRP mixture model: inference

- Data  $x_{1:N}$

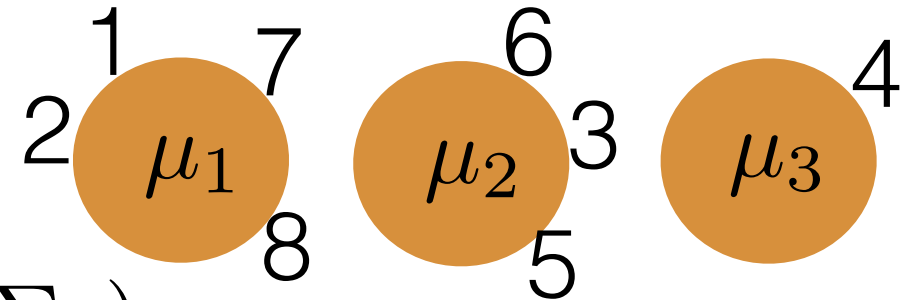


- Generative model

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

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

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



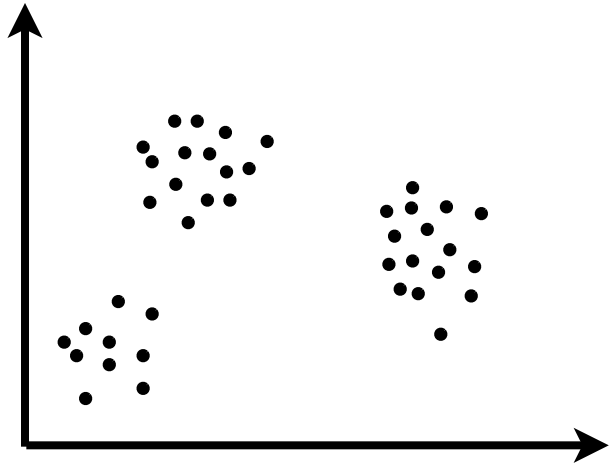
- Want: posterior  $p(\Pi_N | x_{1:N})$

- 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  $x_{1:N}$

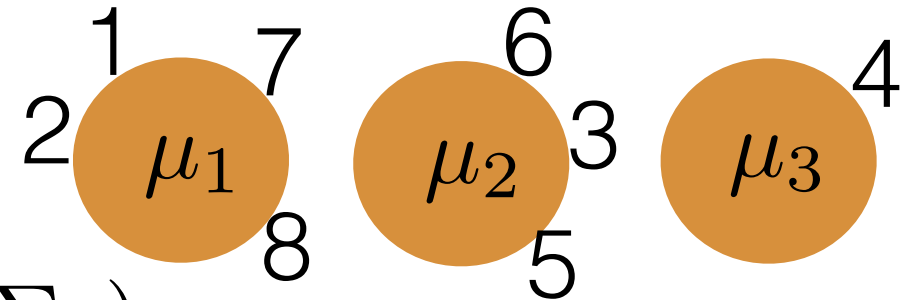


- Generative model

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

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

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



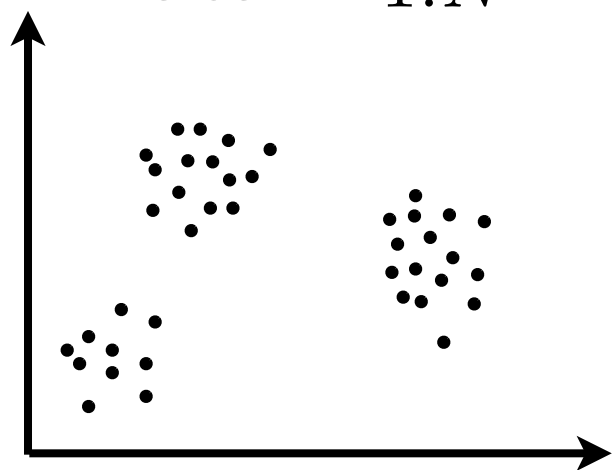
- Want: posterior  $p(\Pi_N | x_{1:N})$

- 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  $x_{1:N}$

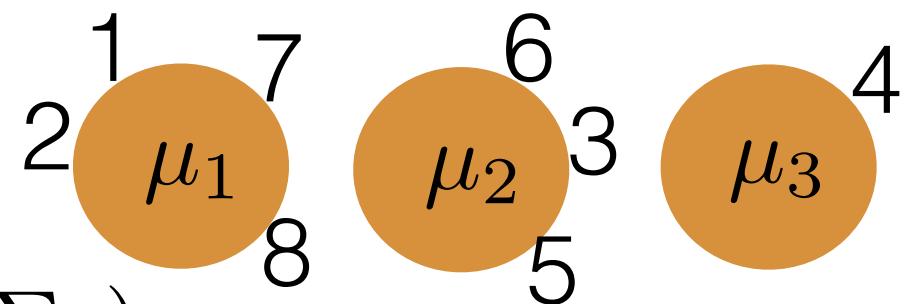


- Generative model

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

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

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



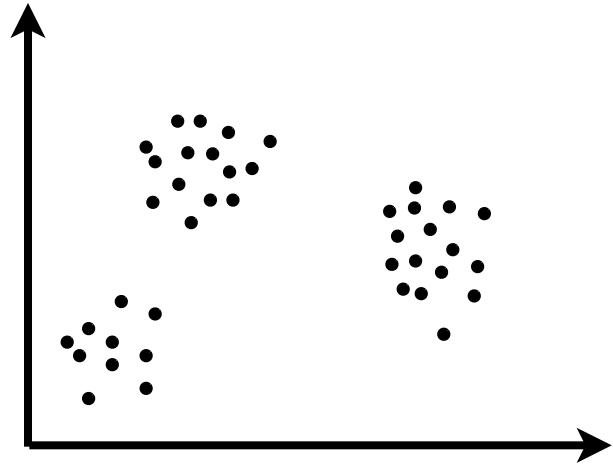
- Want: posterior  $p(\Pi_N | x_{1:N})$

- 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  $x_{1:N}$

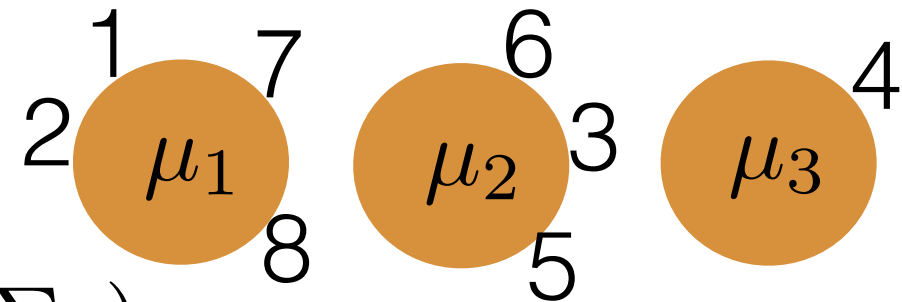


- Generative model

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

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

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



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

- Gibbs sampler:

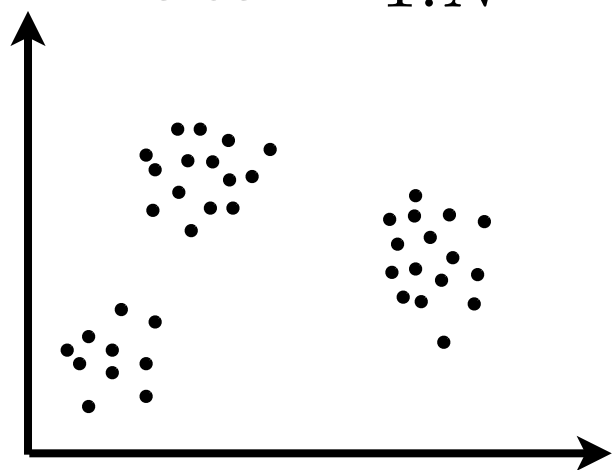
$$p(\Pi_N | \Pi_{N,-n}, x) \propto \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

- For completeness:  $p(x_{C \cup \{n\}} | x_C) =$



# CRP mixture model: inference

- Data  $x_{1:N}$

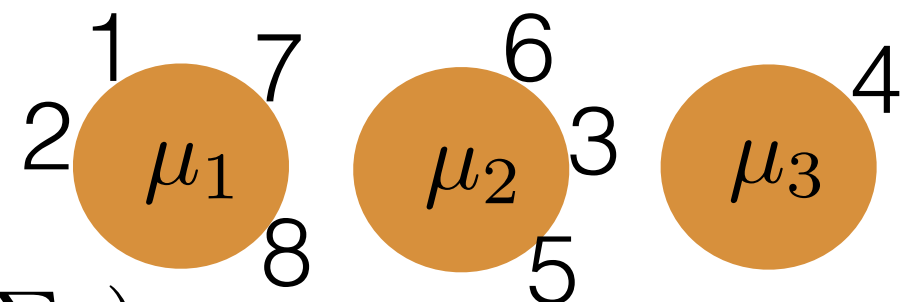


- Generative model

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

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

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



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

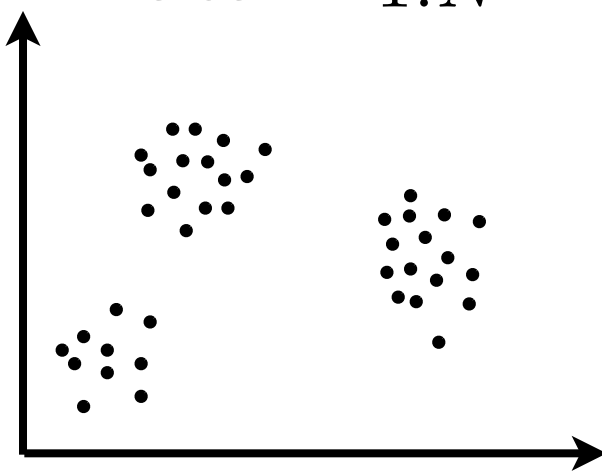
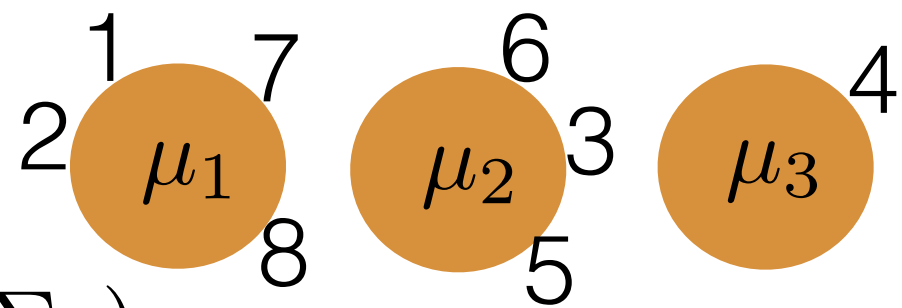
- Gibbs sampler:

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

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



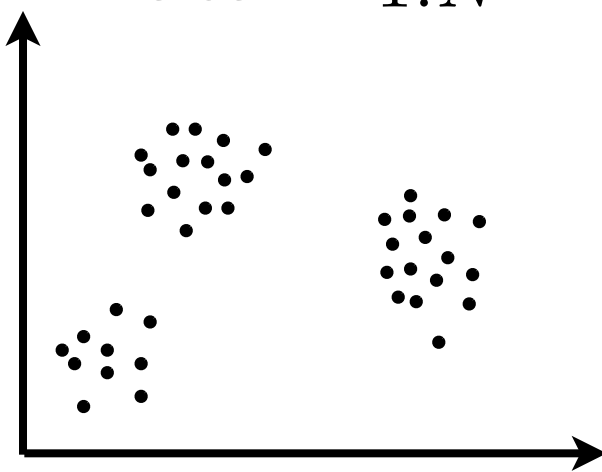
# CRP mixture model: inference

- Data  $x_{1:N}$ 

  - Generative model
    - $\Pi_N \sim \text{CRP}(N, \alpha)$
    - $\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$
    - $\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$
- 
- 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:  $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$ 

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

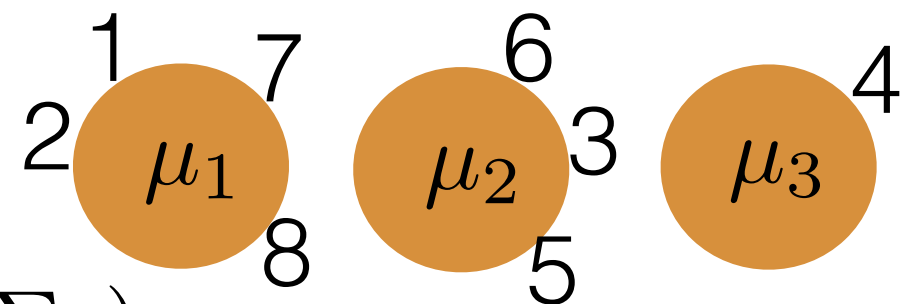
$$\tilde{m} := \tilde{\Sigma} \left( \Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right)$$

# CRP mixture model: inference

- Data  $x_{1:N}$ 

- Generative model
  - $\Pi_N \sim \text{CRP}(N, \alpha)$
  - $\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$
  - $\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$
- 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:  $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$ 

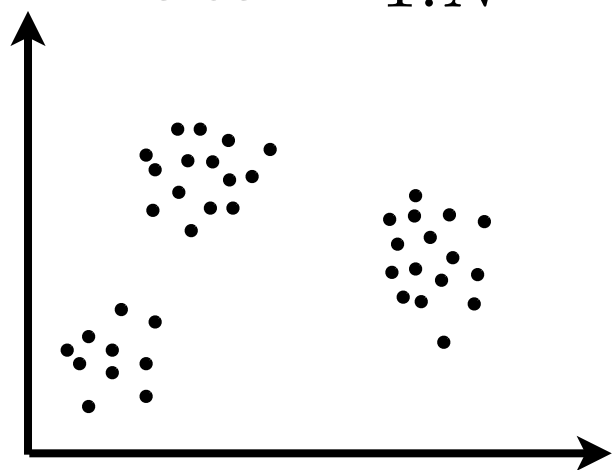
$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$

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



# CRP mixture model: inference

- Data  $x_{1:N}$

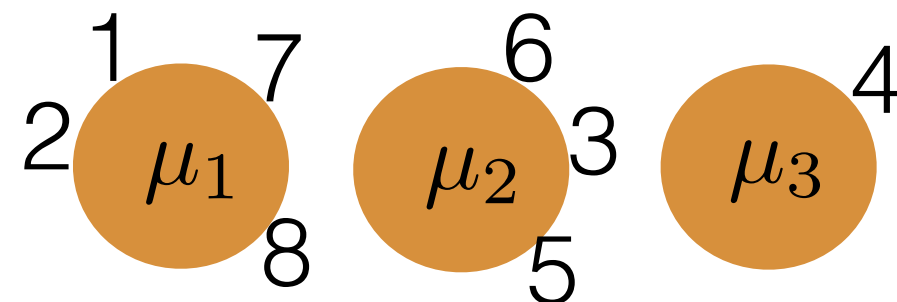


- Generative model

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

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

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



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

- Gibbs sampler:

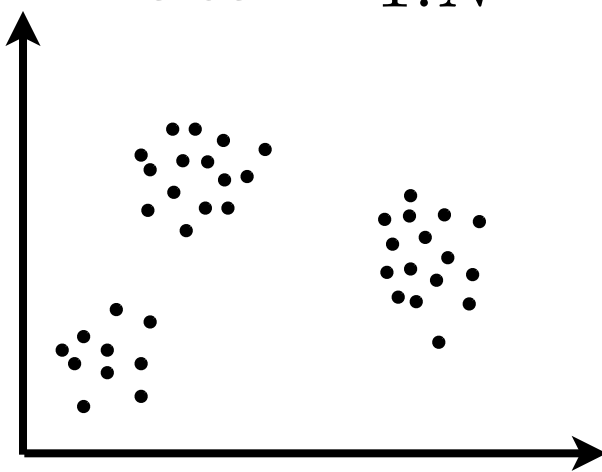
$$p(\Pi_N | \Pi_{N,-n}, x) \propto \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

- 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

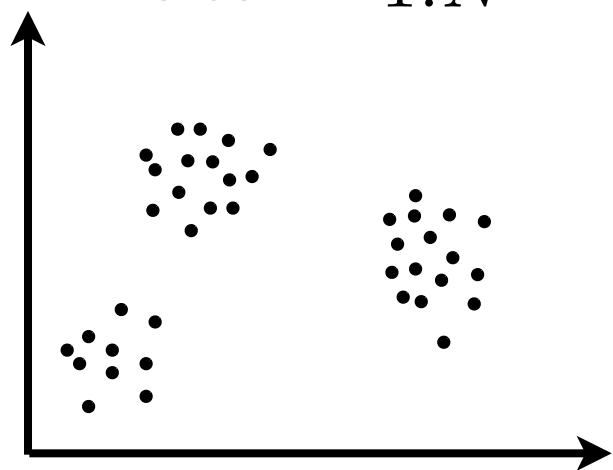
- Data  $x_{1:N}$ 

  - Generative model
    - $\Pi_N \sim \text{CRP}(N, \alpha)$
    - $\forall C \in \Pi_N, \phi_C \stackrel{iid}{\sim} G_0$
    - $\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} F(\phi_C)$
- Diagram illustrating three clusters of data points, each associated with a cluster mean  $\mu_1, \mu_2, \mu_3$ . The points are numbered 1 through 8, showing their assignment to the clusters.
- 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  $x_{1:N}$

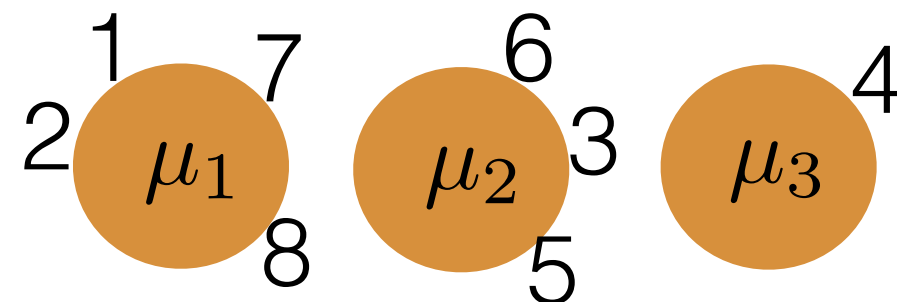


- Generative model

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

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

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

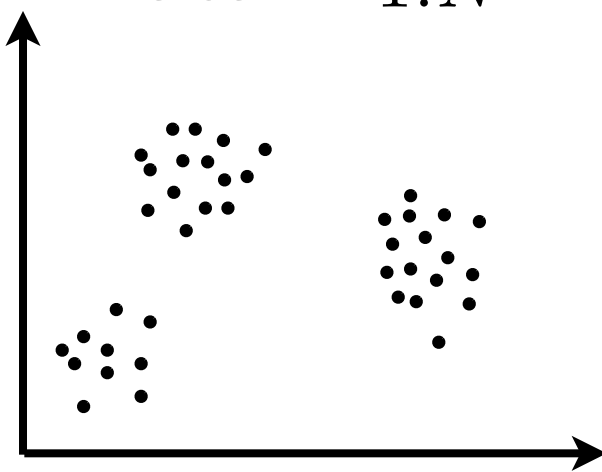
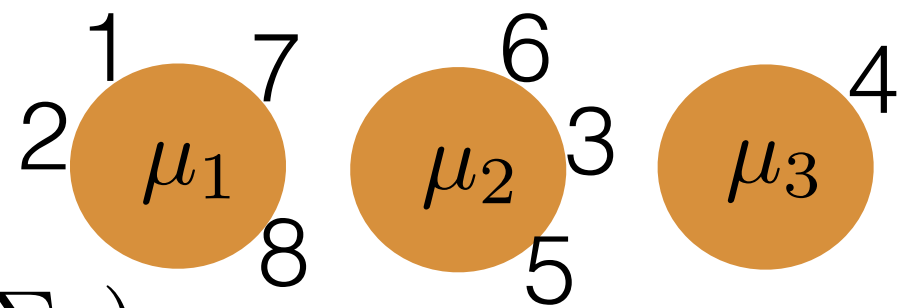


- Want: posterior  $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

# CRP mixture model: inference

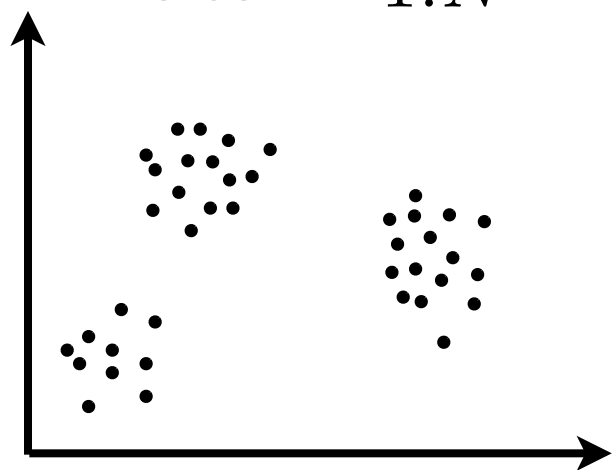
- Data  $x_{1:N}$ 

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

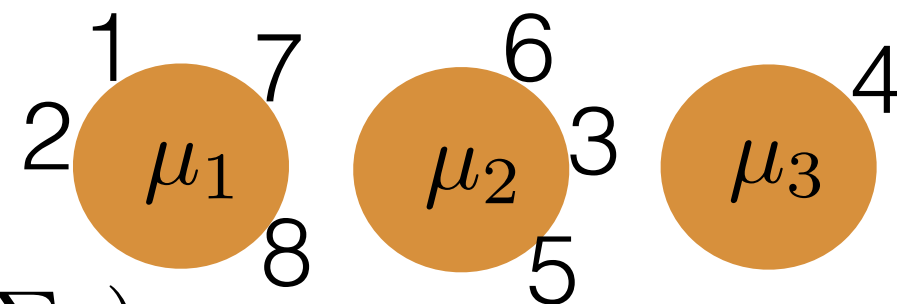


- Generative model

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

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

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



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

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

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



# More Markov Chain Monte Carlo



# More Markov Chain Monte Carlo

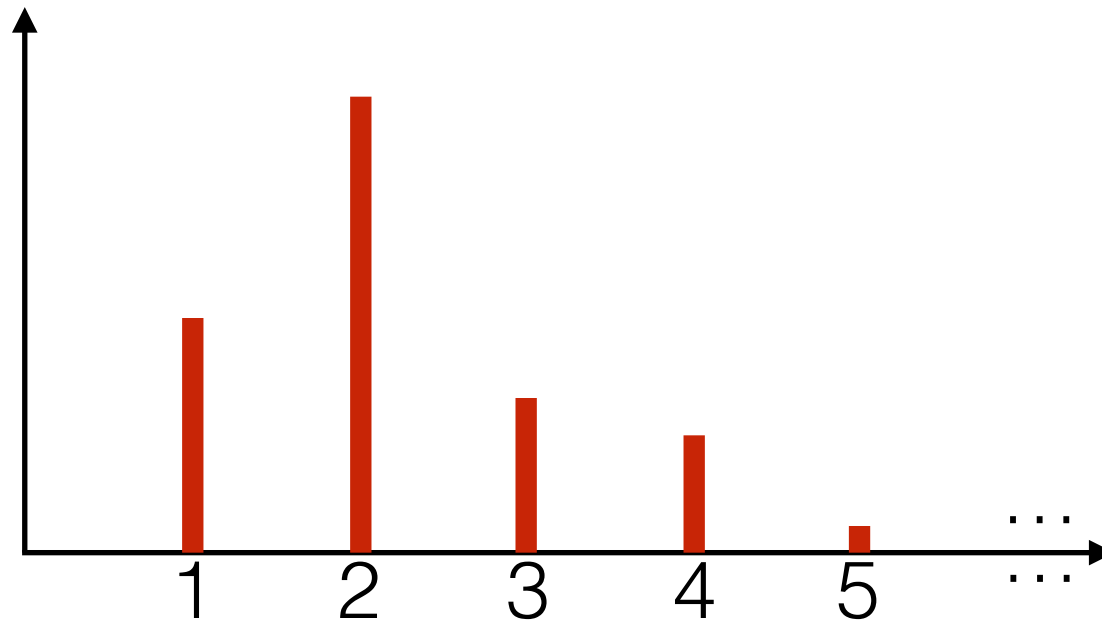
- Slice sampling

# More Markov Chain Monte Carlo

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

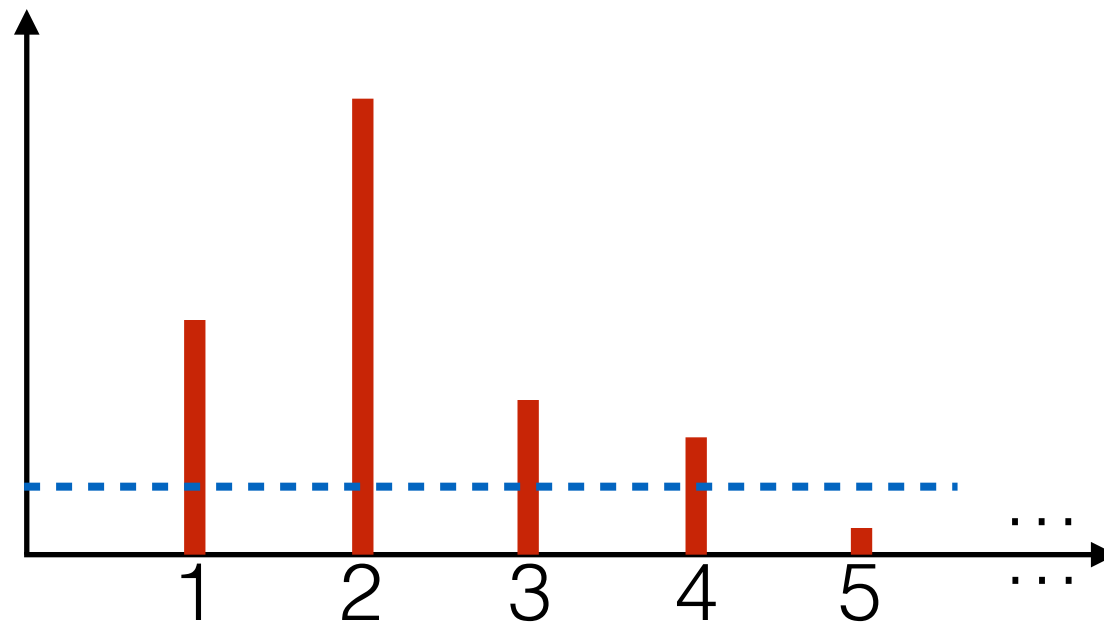
# More Markov Chain Monte Carlo

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



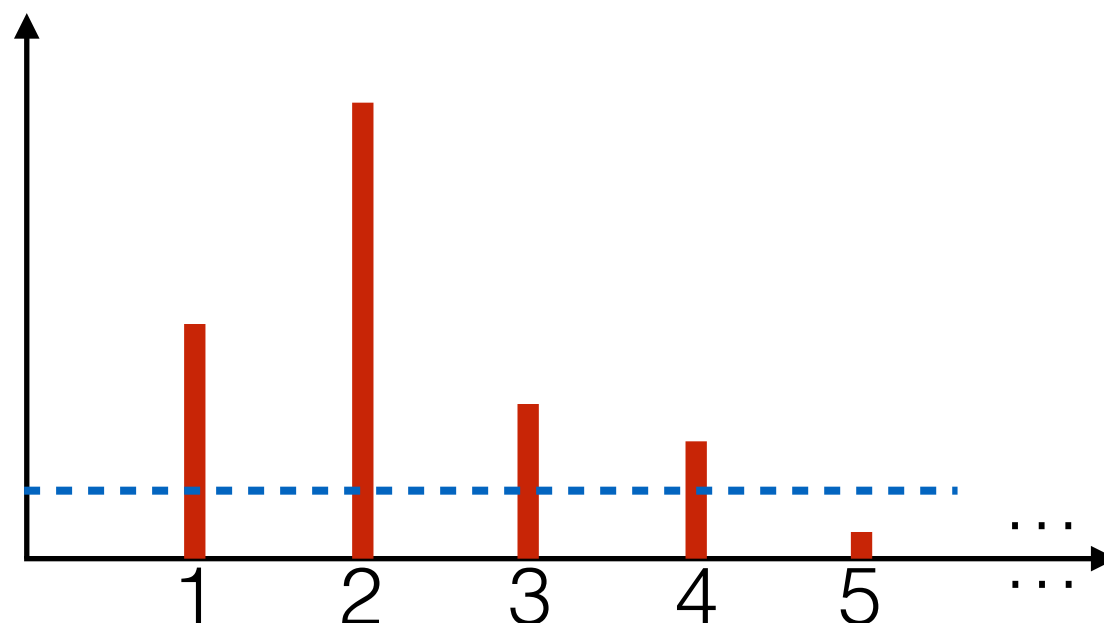
# More Markov Chain Monte Carlo

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



# More Markov Chain Monte Carlo

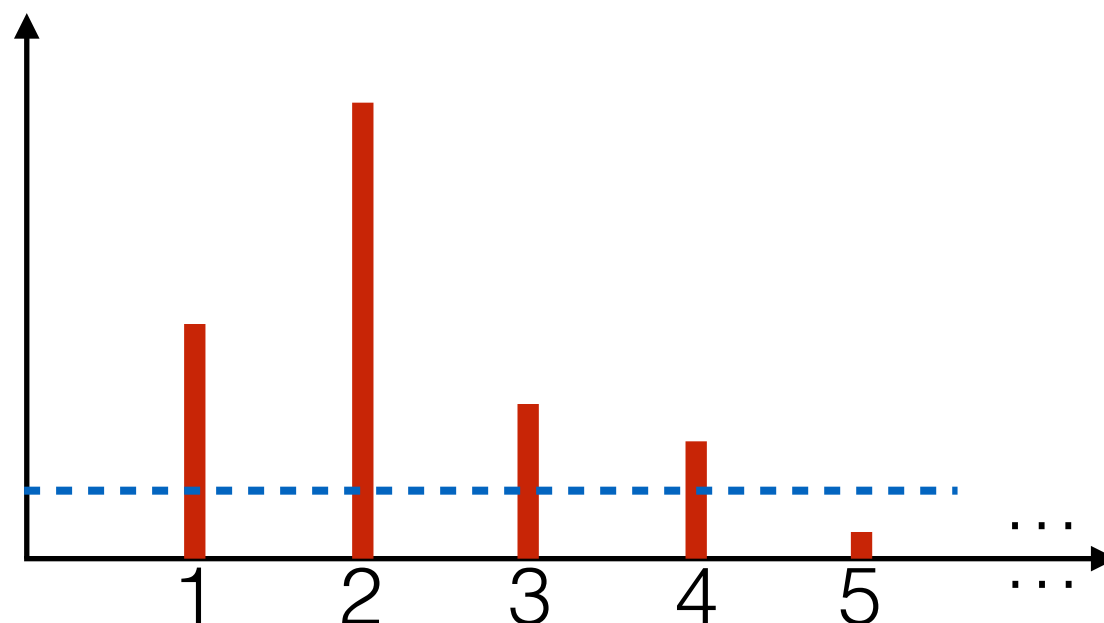
- Slice sampling
  - auxiliary variable  $\rightarrow$  finite conditionals



- Approximate with truncated distribution

# More Markov Chain Monte Carlo

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



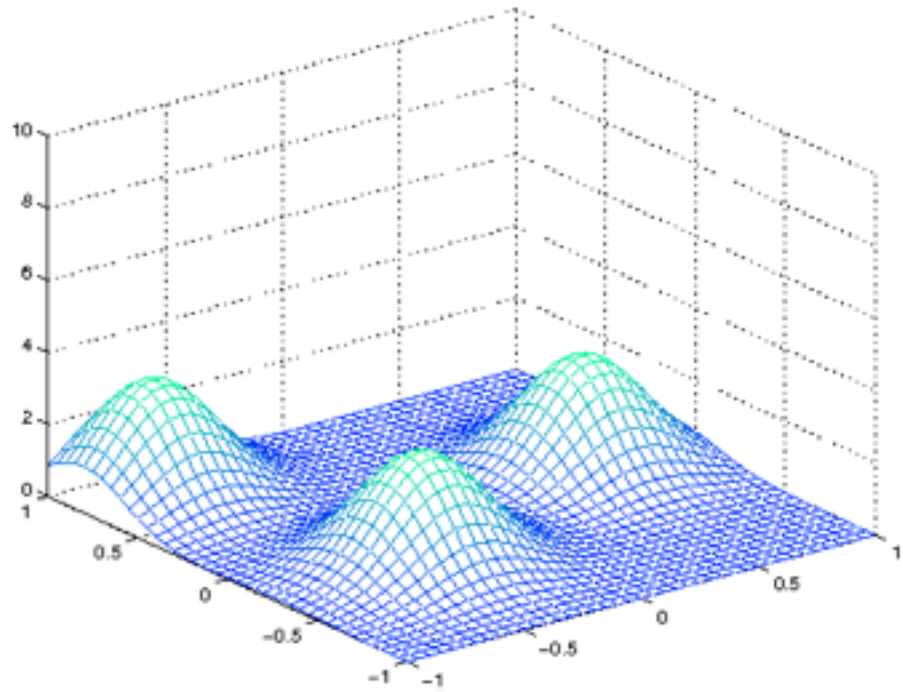
- Approximate with truncated distribution
  - E.g., Hamiltonian Monte Carlo

# Variational Bayes



# Variational Bayes

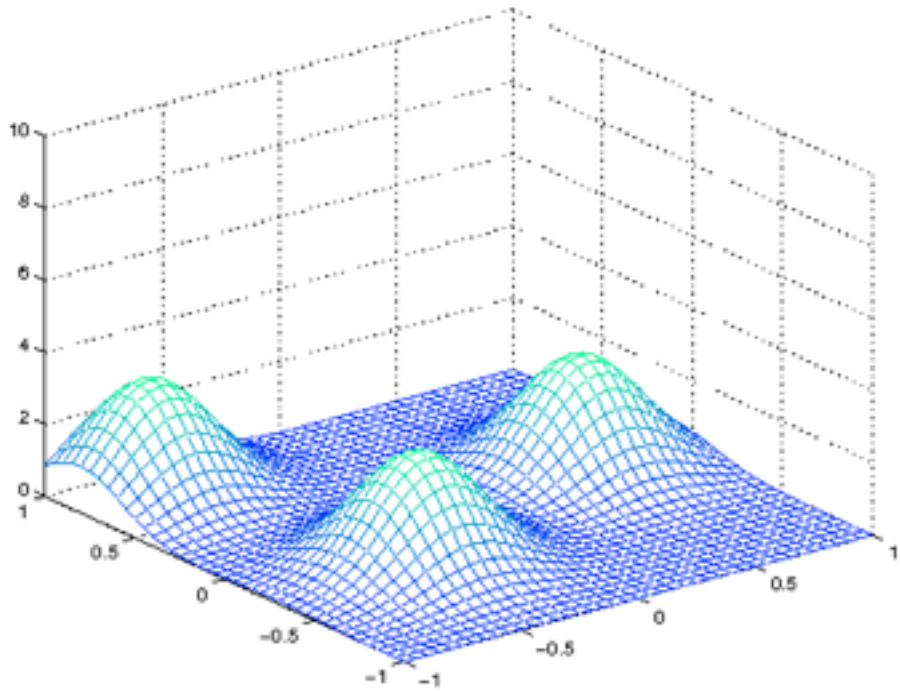
- Variational Bayes (VB)





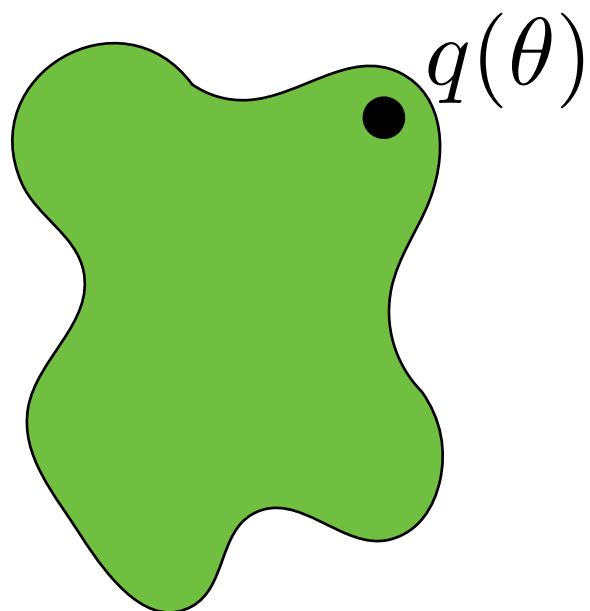
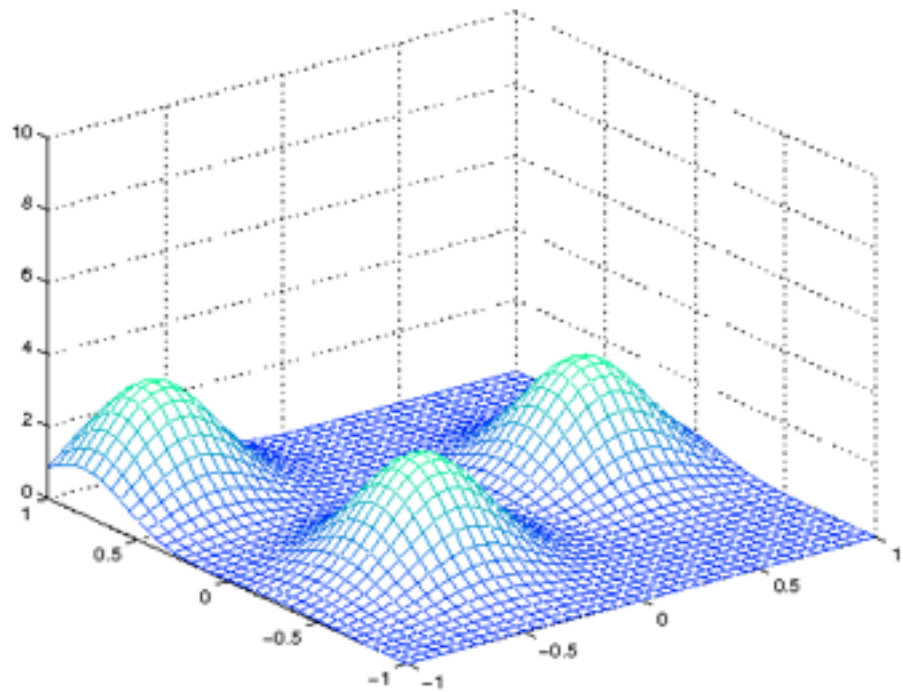
# Variational Bayes

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



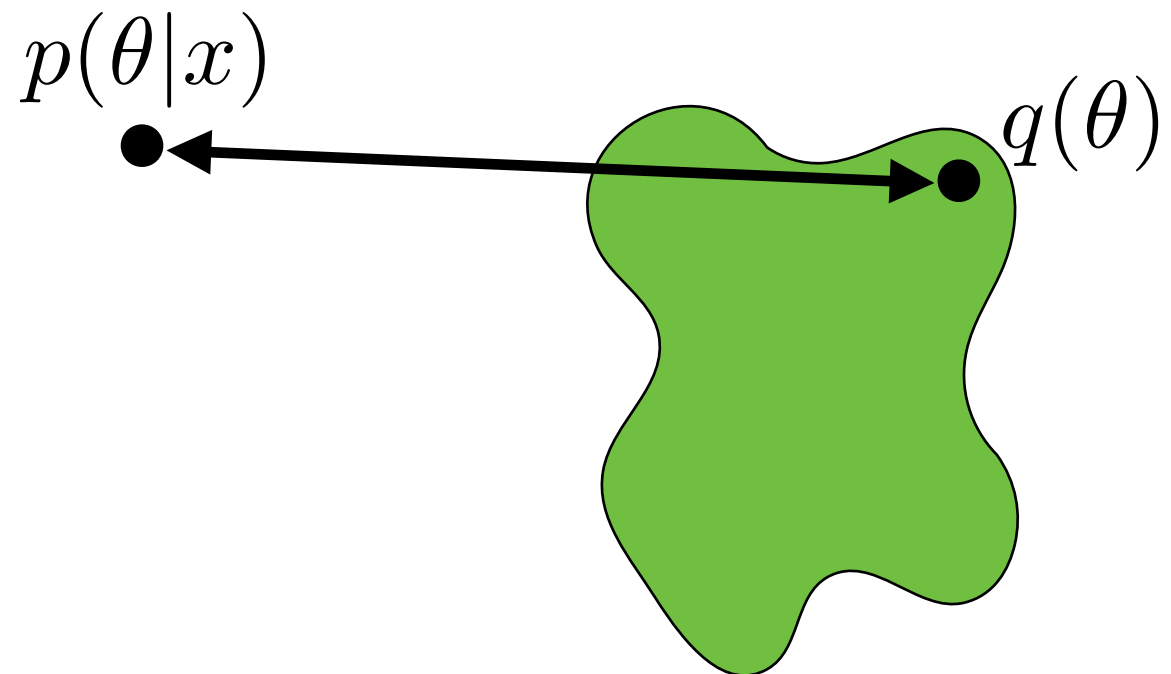
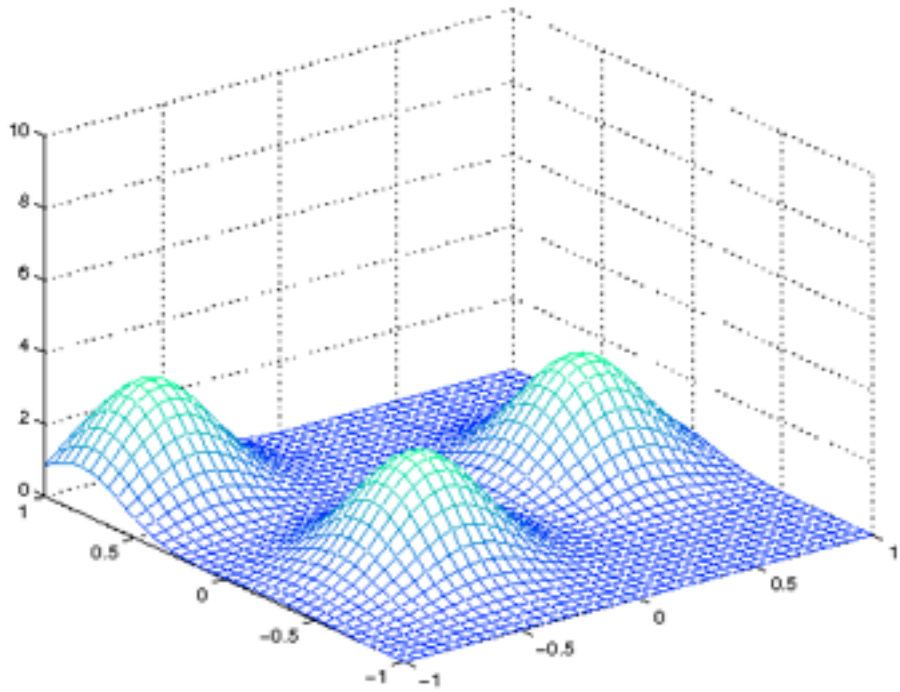
# Variational Bayes

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



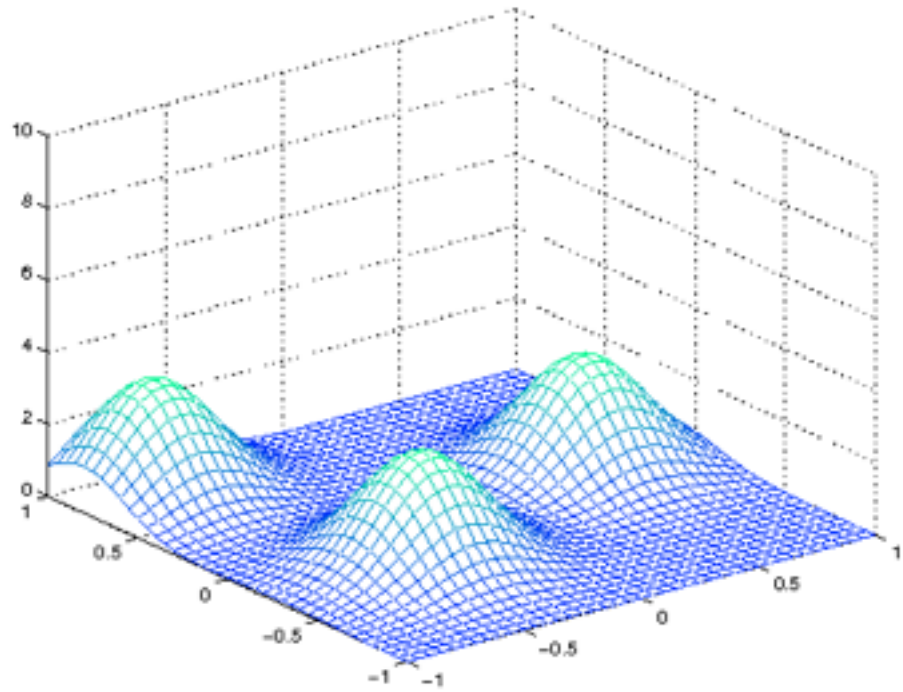
# Variational Bayes

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

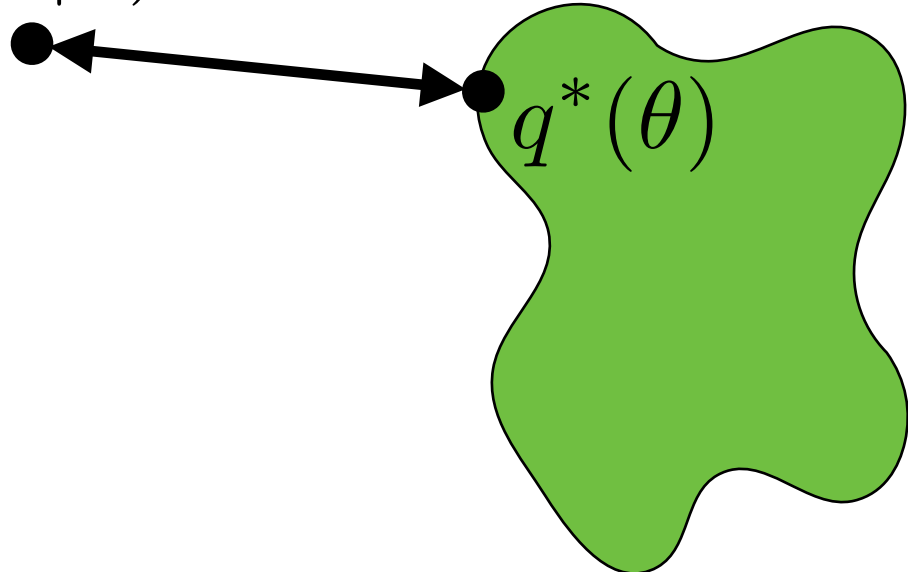


# Variational Bayes

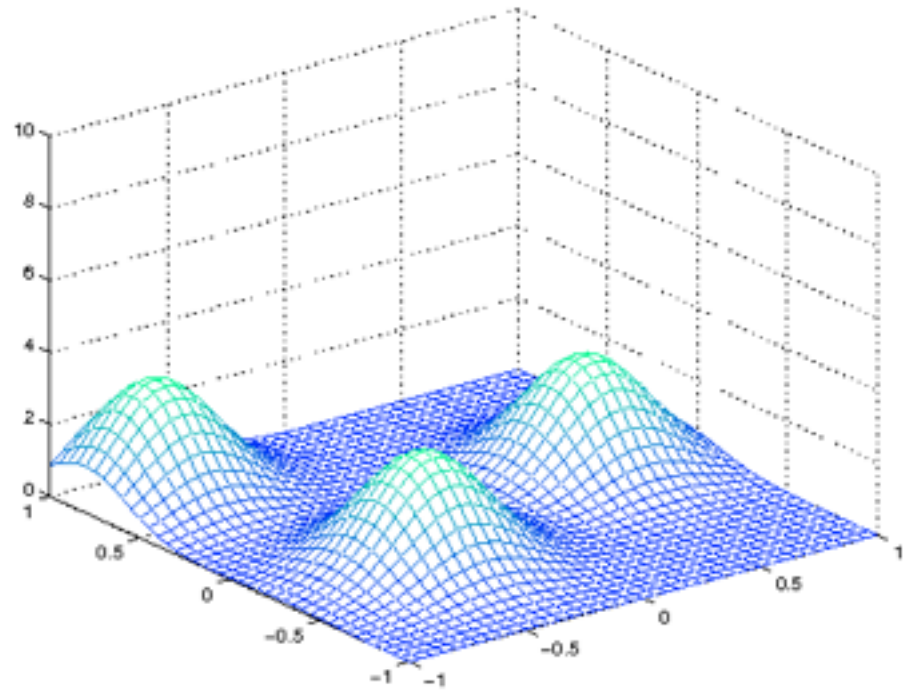
- Variational Bayes (VB)
  - Approximation  $q^*(\theta)$  for posterior  $p(\theta|x)$



$p(\theta|x)$

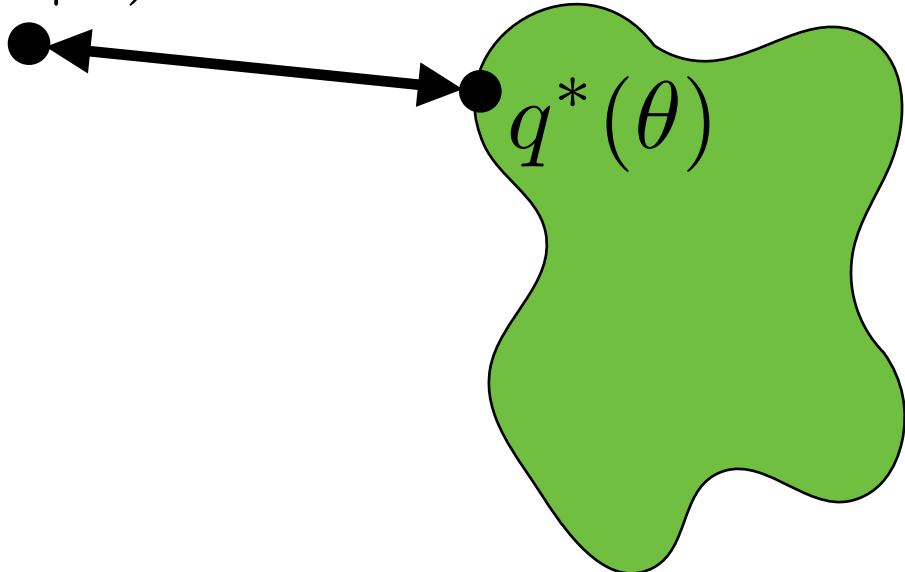


# Variational Bayes

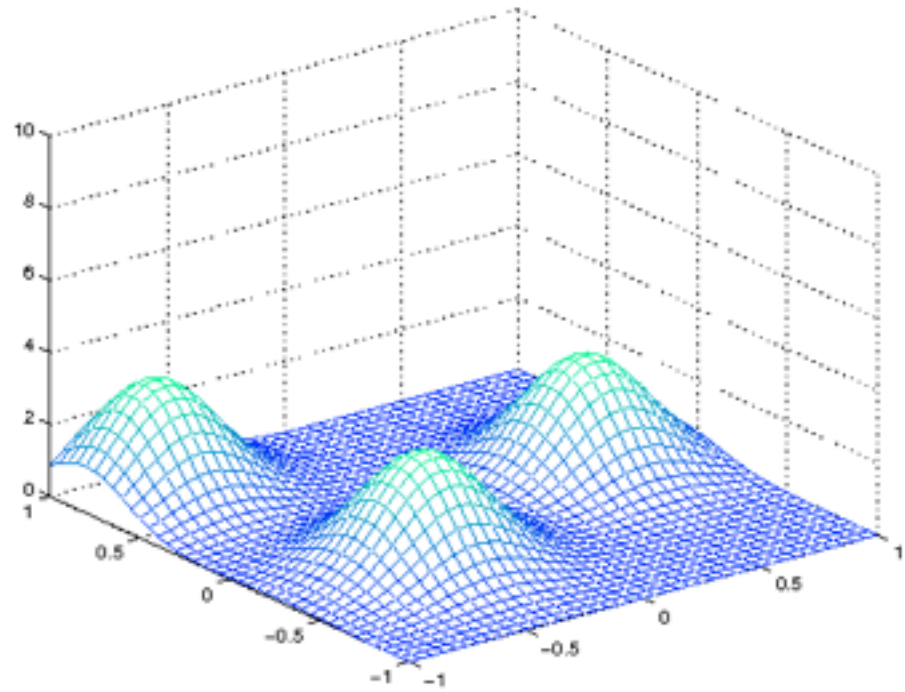


- Variational Bayes (VB)
  - Approximation  $q^*(\theta)$  for posterior  $p(\theta|x)$
  - “Close”: Minimize Kullback-Liebler (KL) divergence:
$$KL(q||p(\cdot|x))$$

$p(\theta|x)$

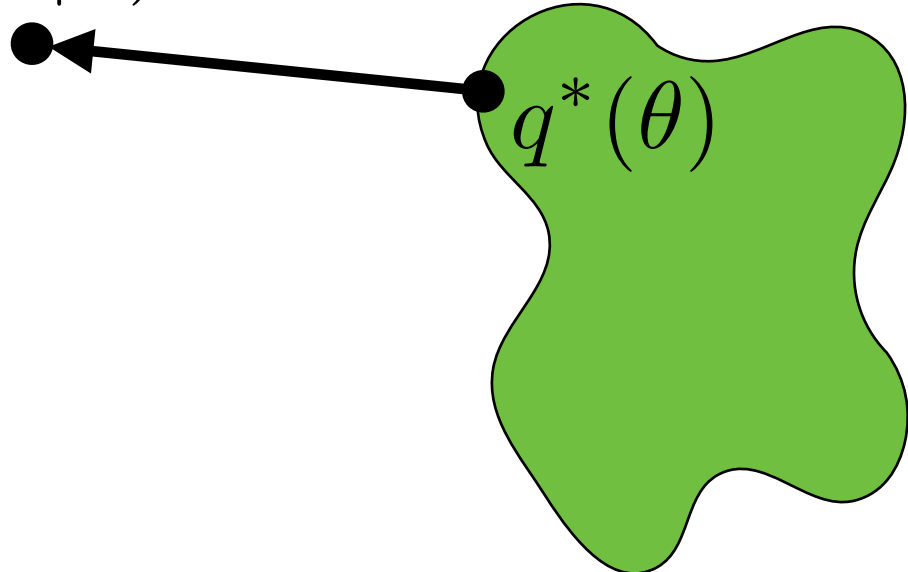


# Variational Bayes



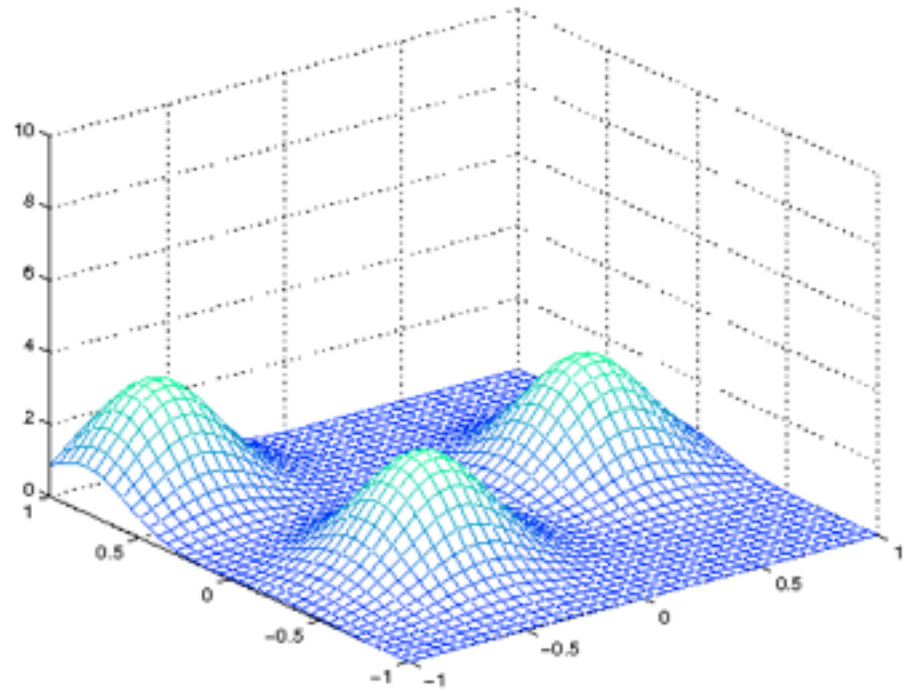
- Variational Bayes (VB)
  - Approximation  $q^*(\theta)$  for posterior  $p(\theta|x)$
  - “Close”: Minimize Kullback-Liebler (KL) divergence:
$$KL(q||p(\cdot|x))$$

$p(\theta|x)$



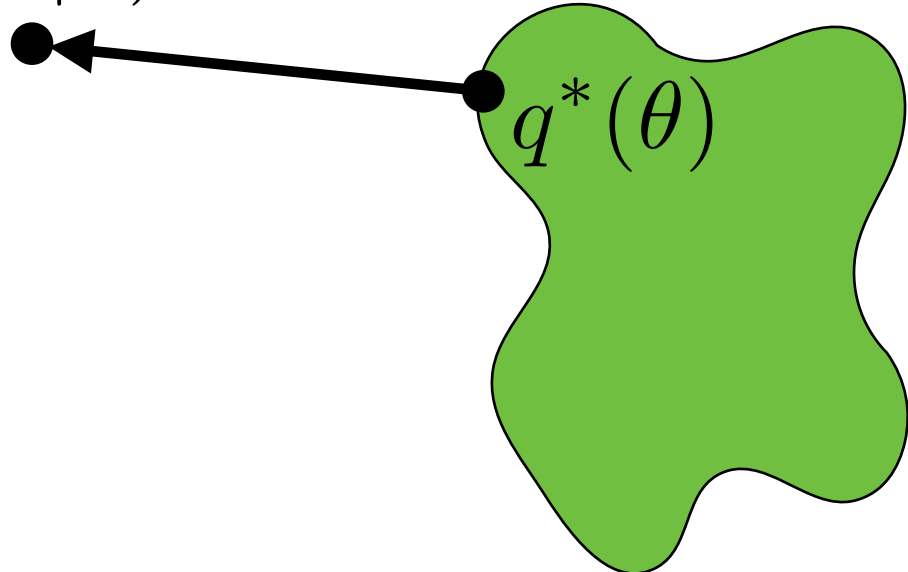


# Variational Bayes

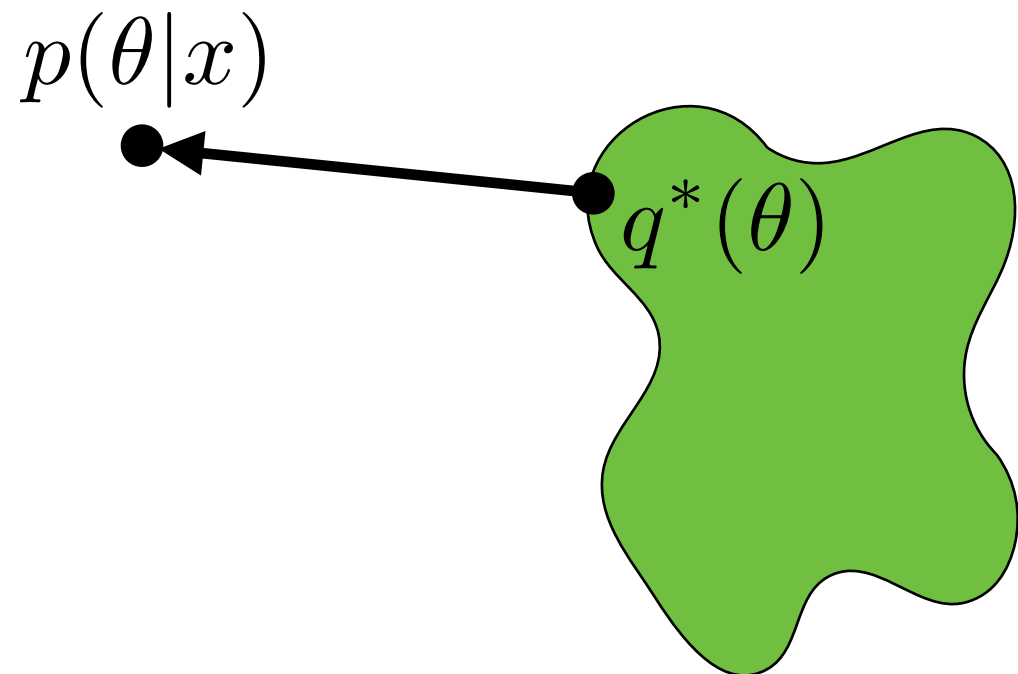
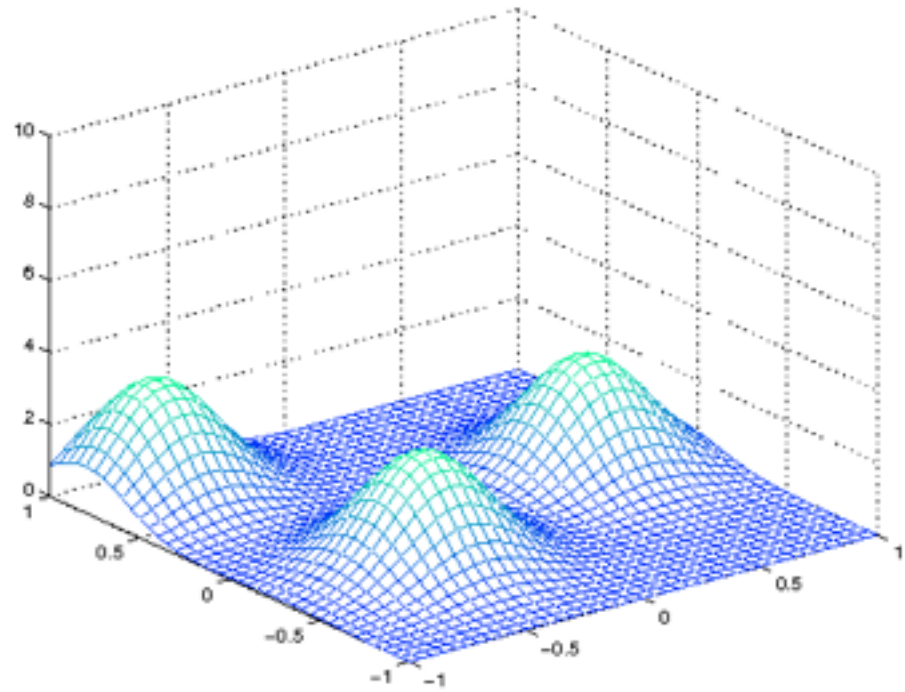


- Variational Bayes (VB)
  - Approximation  $q^*(\theta)$  for posterior  $p(\theta|x)$
  - “Close”: Minimize Kullback-Liebler (KL) divergence:
$$KL(q||p(\cdot|x))$$
  - “Nice”: factorizes, exponential family, truncation

$p(\theta|x)$



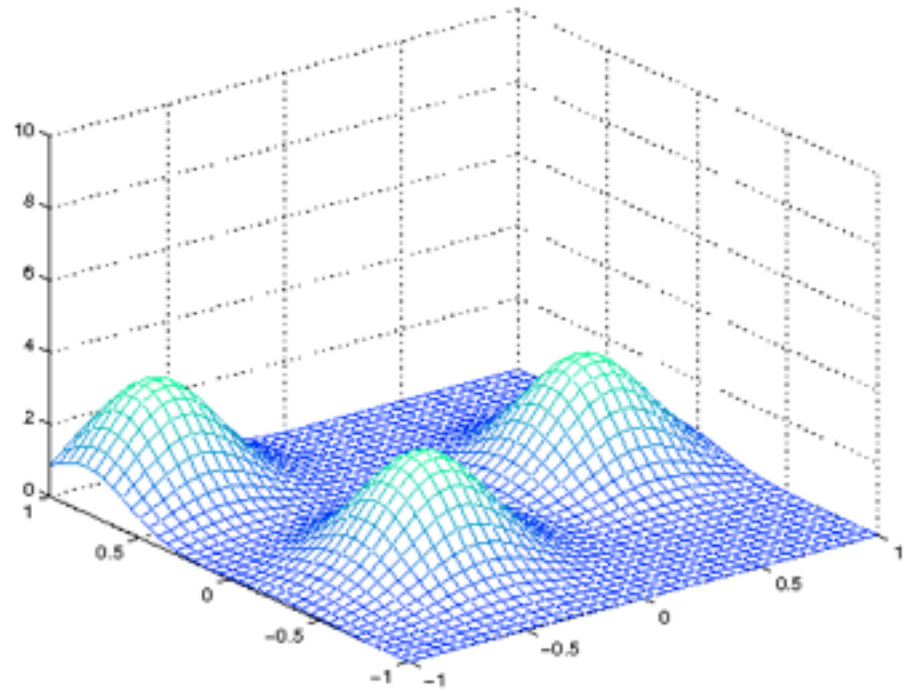
# Variational Bayes



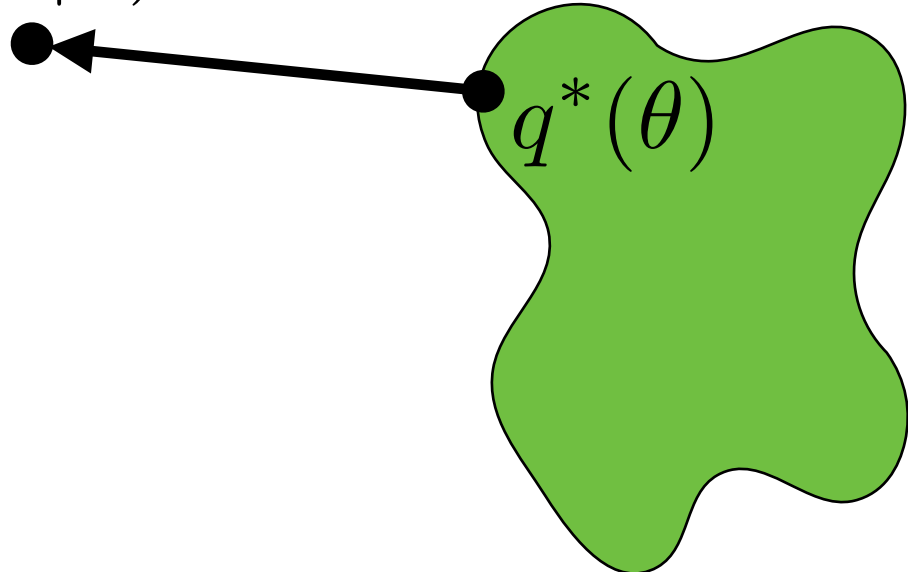
- Variational Bayes (VB)
  - Approximation  $q^*(\theta)$  for posterior  $p(\theta|x)$
  - “Close”: Minimize Kullback-Liebler (KL) divergence:
$$KL(q||p(\cdot|x))$$
  - “Nice”: factorizes, exponential family, truncation
- VB practical success



# Variational Bayes

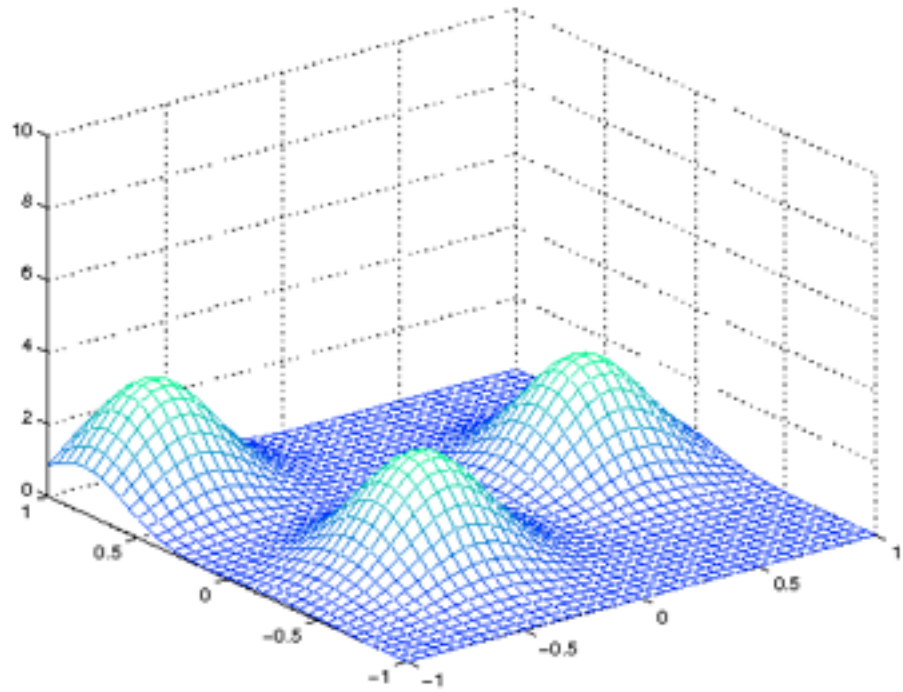


$p(\theta|x)$

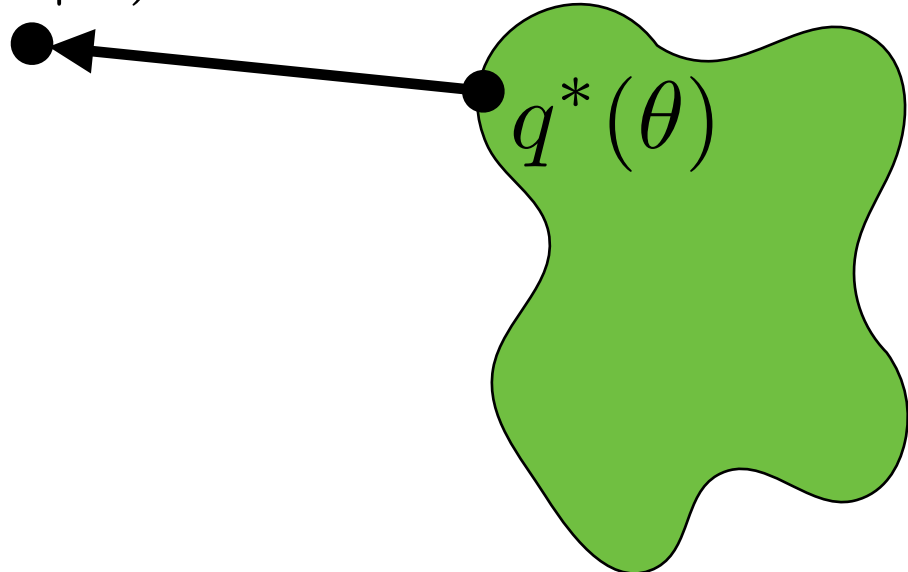


- Variational Bayes (VB)
  - Approximation  $q^*(\theta)$  for posterior  $p(\theta|x)$
  - “Close”: Minimize Kullback-Liebler (KL) divergence:
$$KL(q||p(\cdot|x))$$
  - “Nice”: factorizes, exponential family, truncation
- VB practical success
  - point estimates and prediction

# Variational Bayes

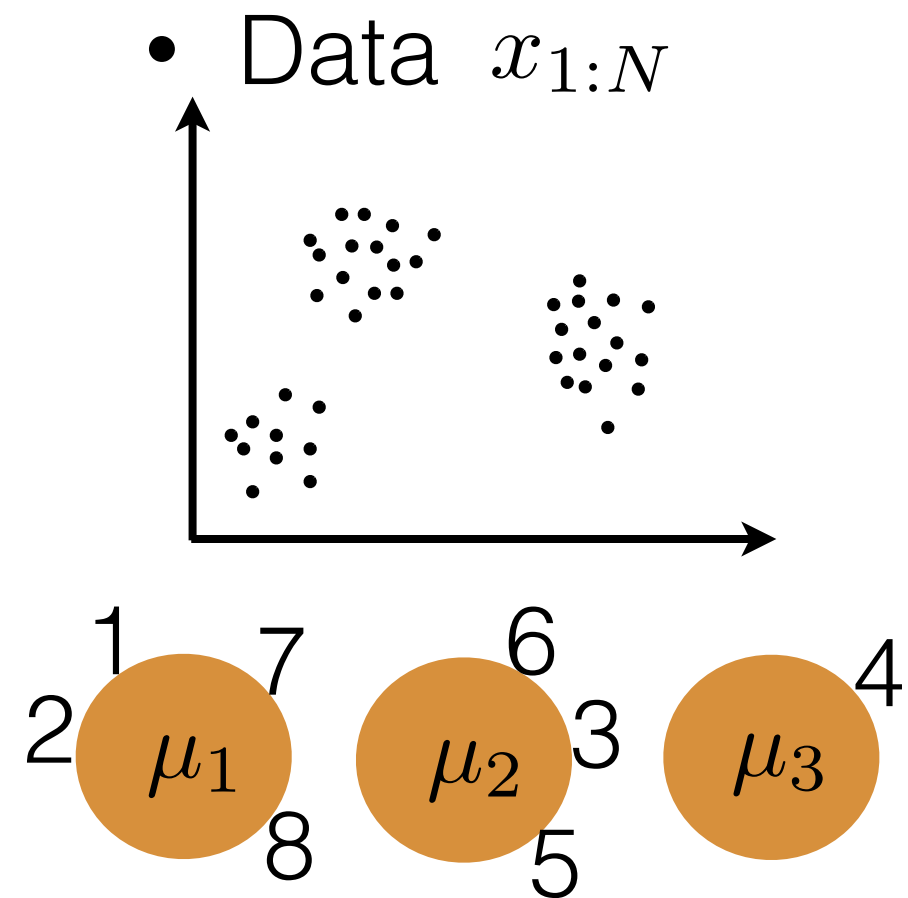


$p(\theta|x)$



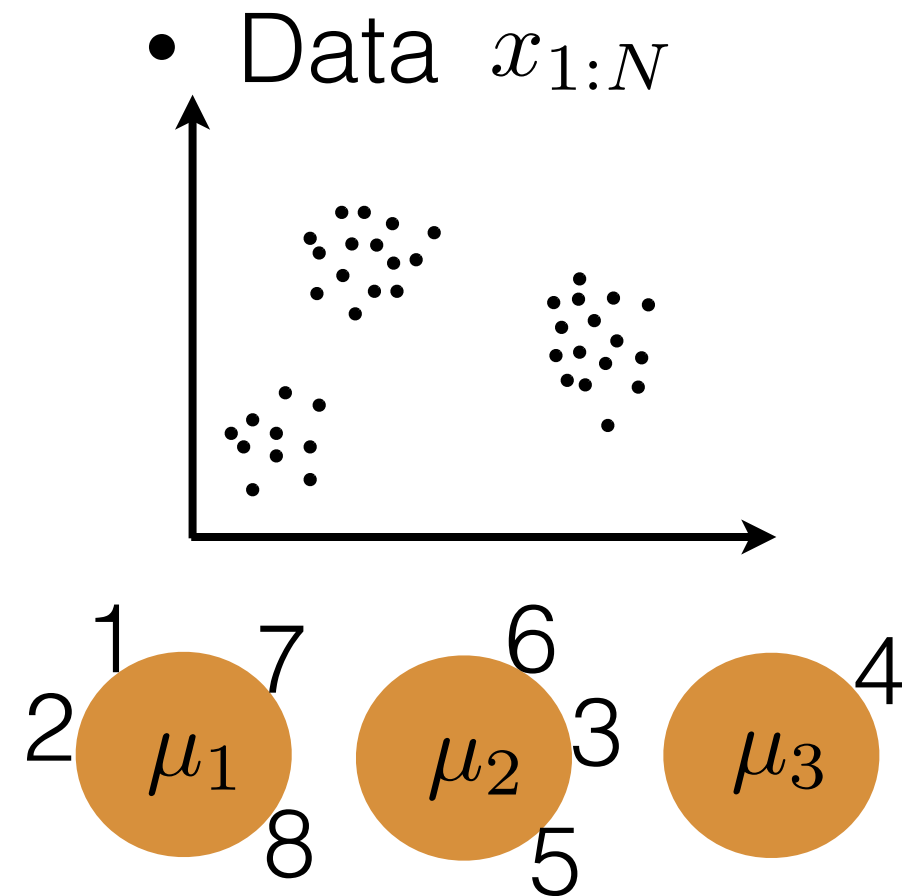
- Variational Bayes (VB)
  - Approximation  $q^*(\theta)$  for posterior  $p(\theta|x)$
  - “Close”: Minimize Kullback-Liebler (KL) divergence:
$$KL(q||p(\cdot|x))$$
  - “Nice”: factorizes, exponential family, truncation
- VB practical success
  - point estimates and prediction
  - fast, streaming, distributed

# Exercises



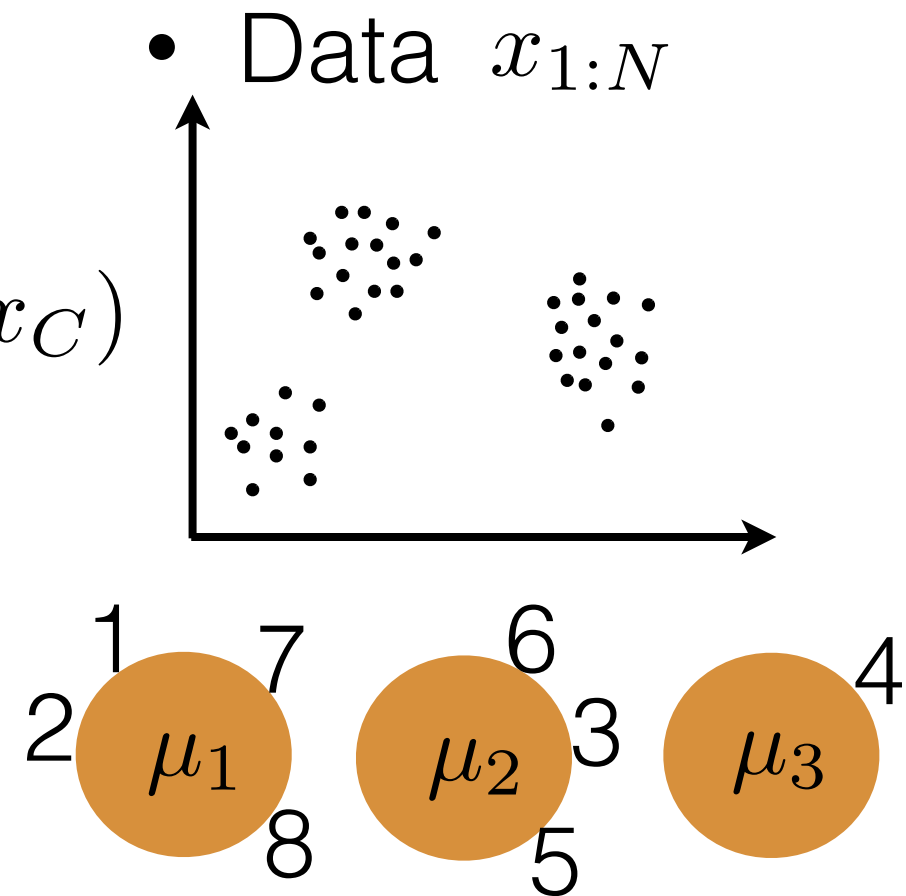
# Exercises

- Code a CRP mixture model simulator



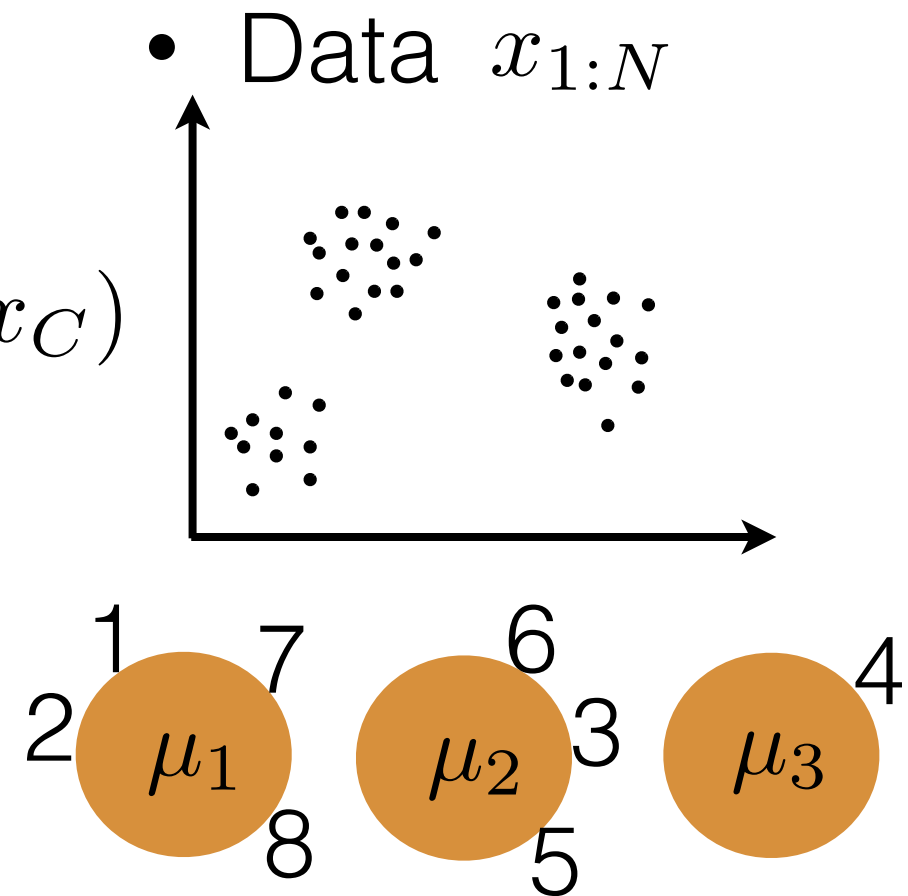
# Exercises

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



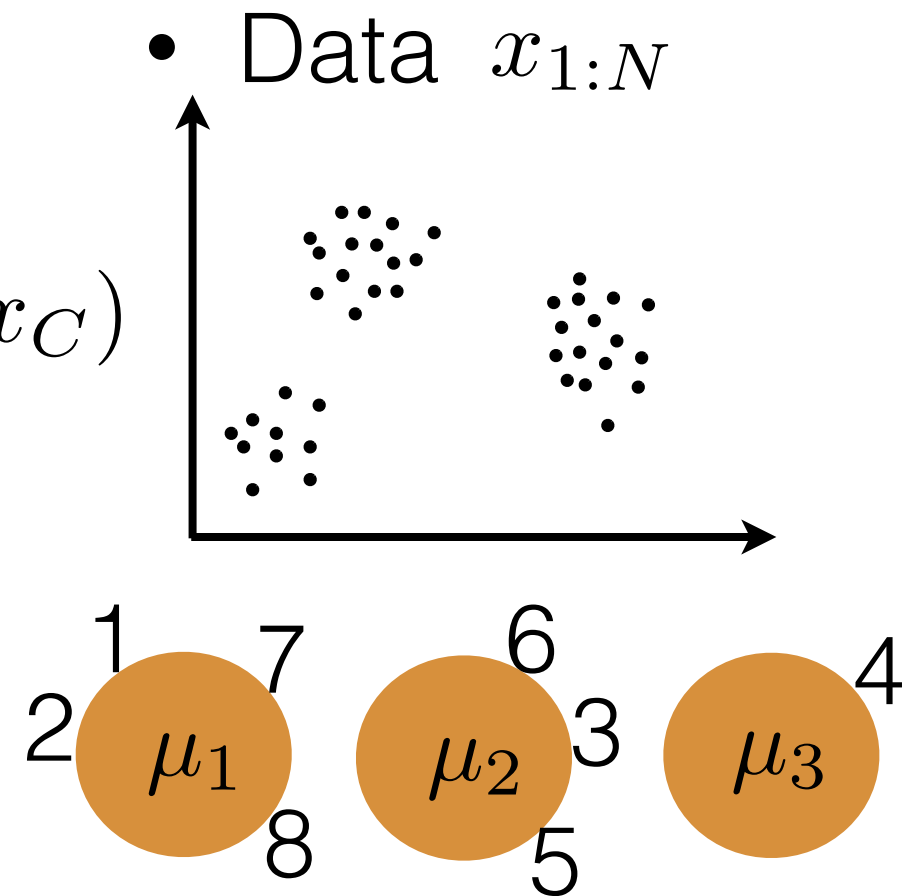
# Exercises

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



# Exercises

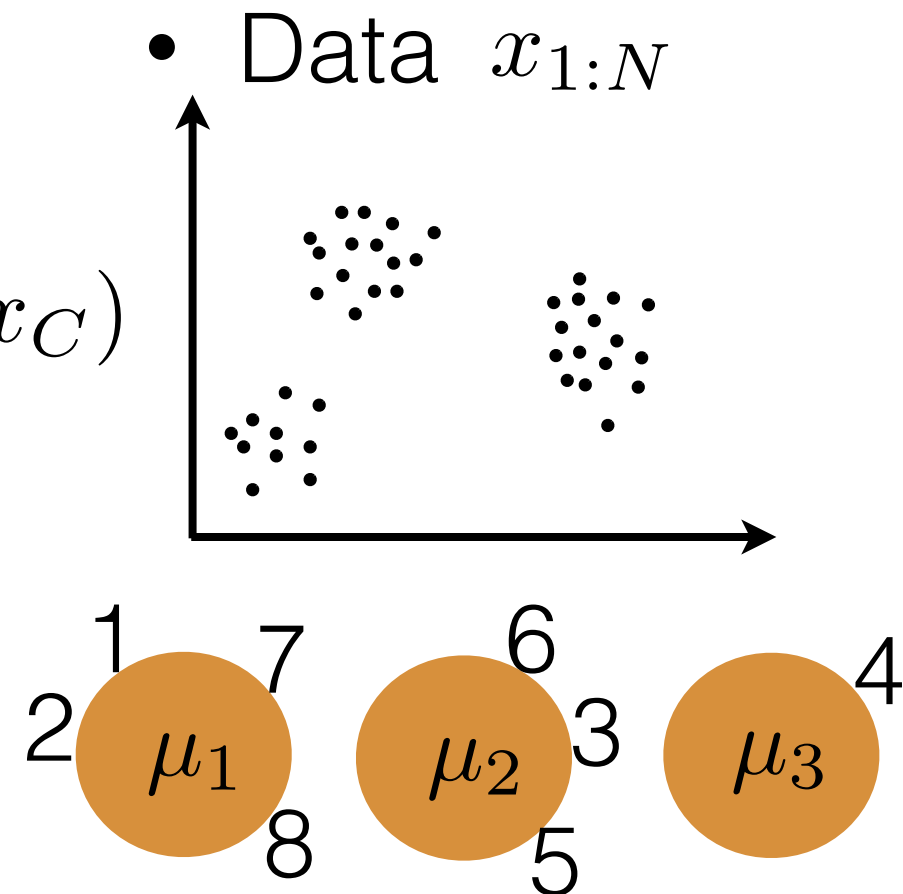
- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive  $p(x_{C \cup \{n\}} | x_C)$  explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
- Review Gibbs sampling,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





# Exercises

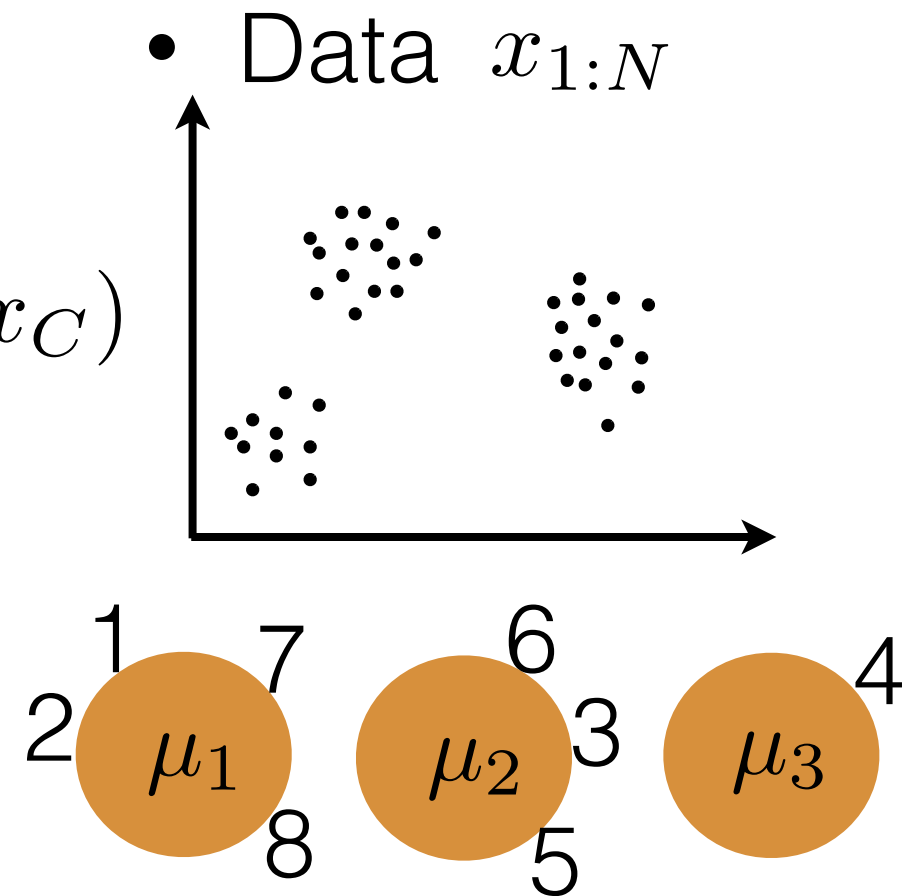
- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive  $p(x_{C \cup \{n\}} | x_C)$  explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well
- Review Gibbs sampling,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





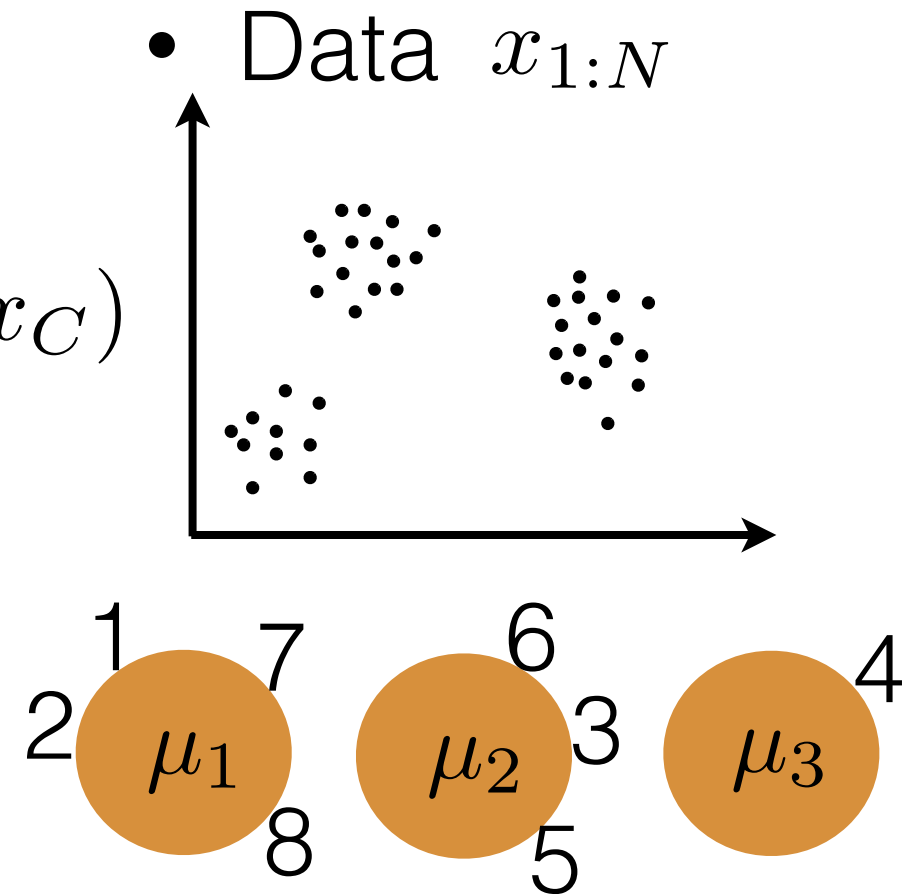
# Exercises

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



# Exercises

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







Hierarchies

Power laws

Dependencies

Feature allocations

Coalescents/  
Diffusions/Trees

Networks/graphs

de Finetti

Poisson processes

*Here be Dragons*



# Clustering

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

# Feature allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

# Feature allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

- Indian buffet process

# Feature allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

- Indian buffet process

# Feature allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

- Indian buffet process
- Beta process + Bernoulli likelihood



# Feature allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

- Indian buffet process
- Beta process + Bernoulli likelihood

# Feature allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

- Indian buffet process
- Beta process + Bernoulli likelihood

# Feature allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

- Indian buffet process
- Beta process + Bernoulli likelihood

# Trait allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 |      |      |        |        |            |
| Document 2 |      |      |        |        |            |
| Document 3 |      |      |        |        |            |
| Document 4 |      |      |        |        |            |
| Document 5 |      |      |        |        |            |
| Document 6 |      |      |        |        |            |
| Document 7 |      |      |        |        |            |

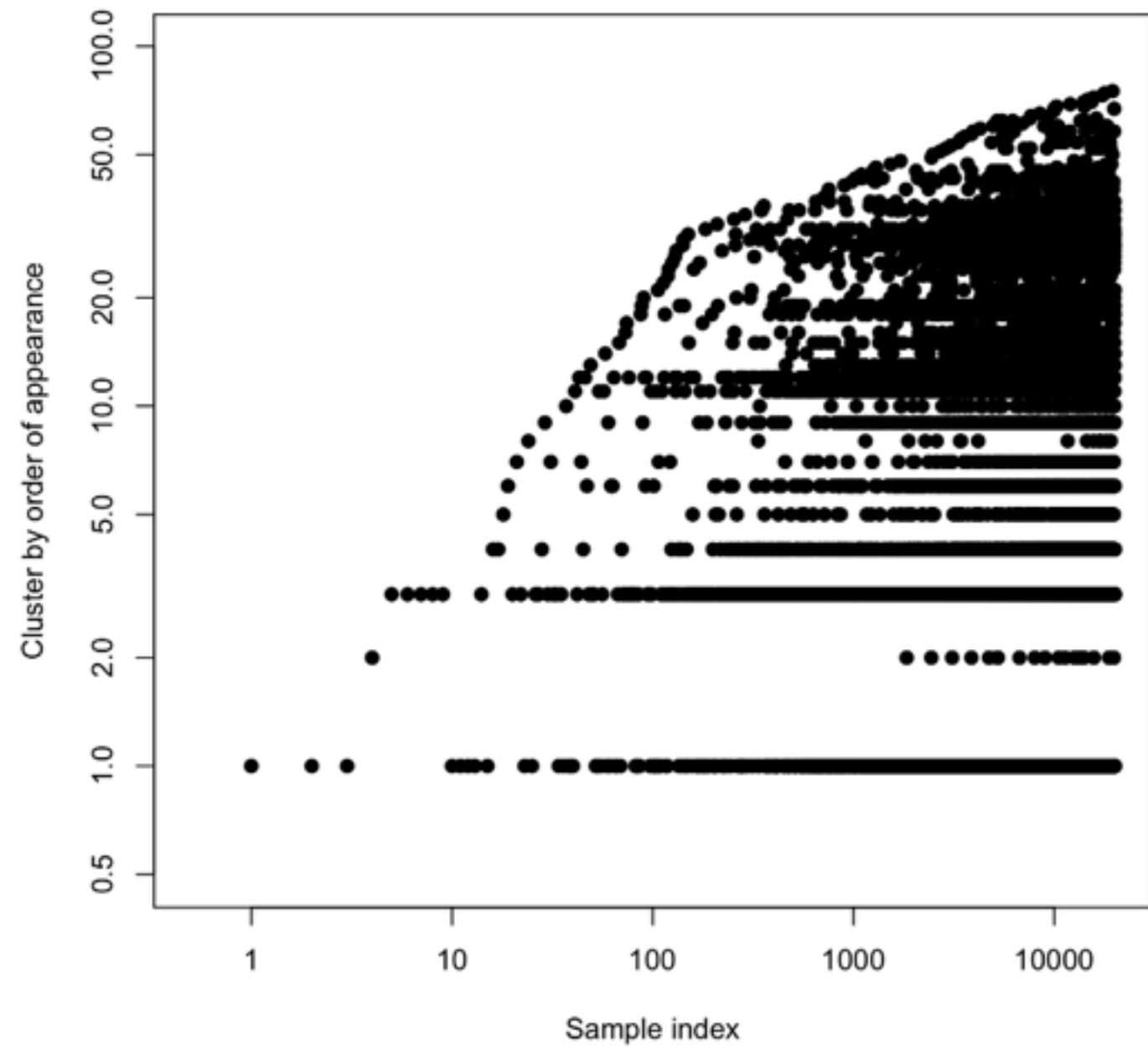
# Trait allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 | 25   |      |        |        | 42         |
| Document 2 | 33   |      |        | 15     | 8          |
| Document 3 | 9    | 28   |        | 14     | 18         |
| Document 4 |      |      | 9      | 21     | 10         |
| Document 5 |      | 11   |        |        | 32         |
| Document 6 |      |      |        | 8      | 27         |
| Document 7 |      |      |        |        |            |

# Trait allocation

|            | Arts | Econ | Sports | Health | Technology |
|------------|------|------|--------|--------|------------|
| Document 1 | 25   |      |        |        | 42         |
| Document 2 | 33   |      |        | 15     | 8          |
| Document 3 | 9    | 28   |        | 14     | 18         |
| Document 4 |      |      | 9      | 21     | 10         |
| Document 5 |      | 11   |        |        | 32         |
| Document 6 |      |      |        | 8      | 27         |
| Document 7 |      |      |        |        |            |

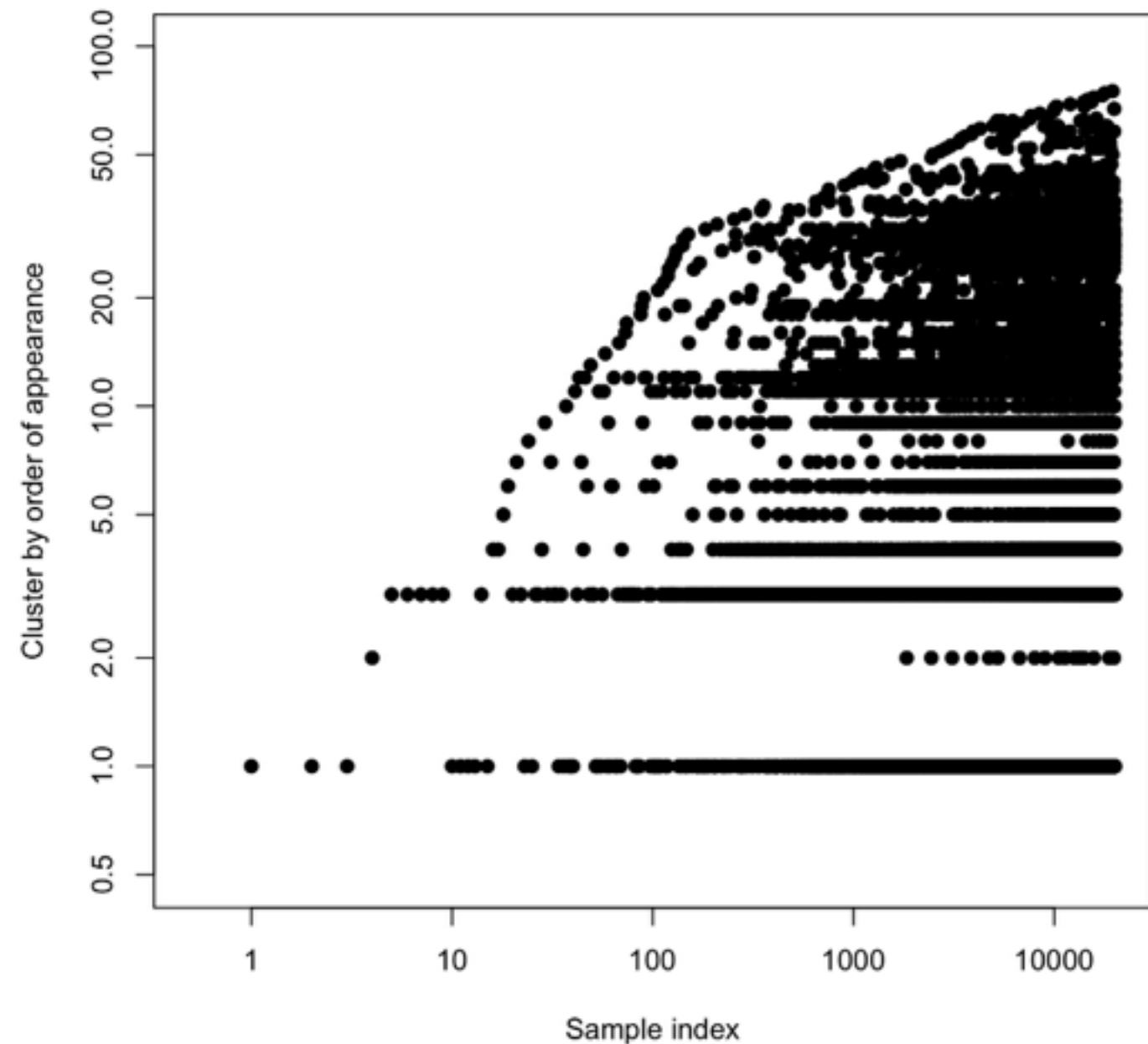
# Power laws





# Power laws

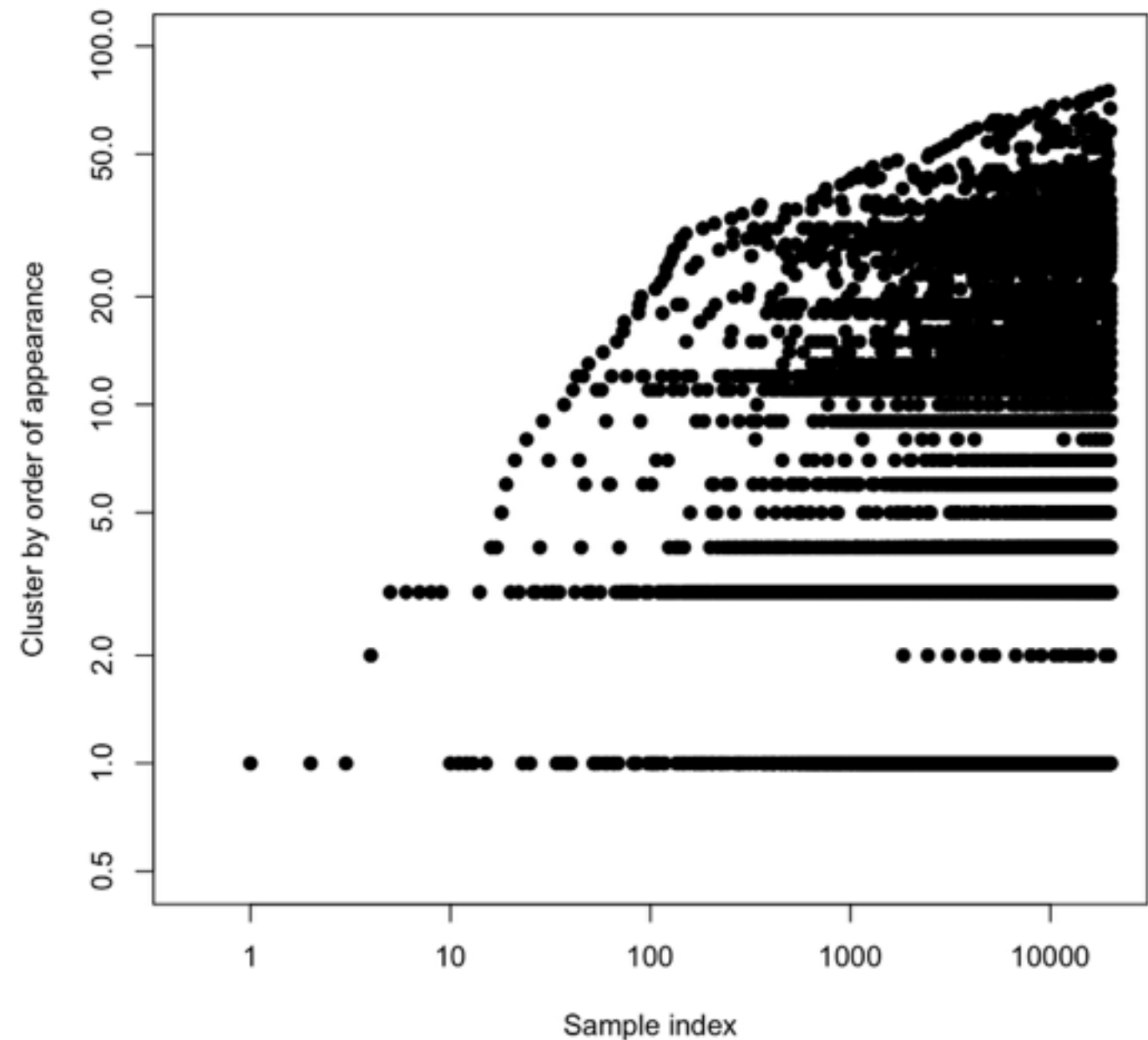
- $K_N := \#$  clusters occupied by  $N$  data points





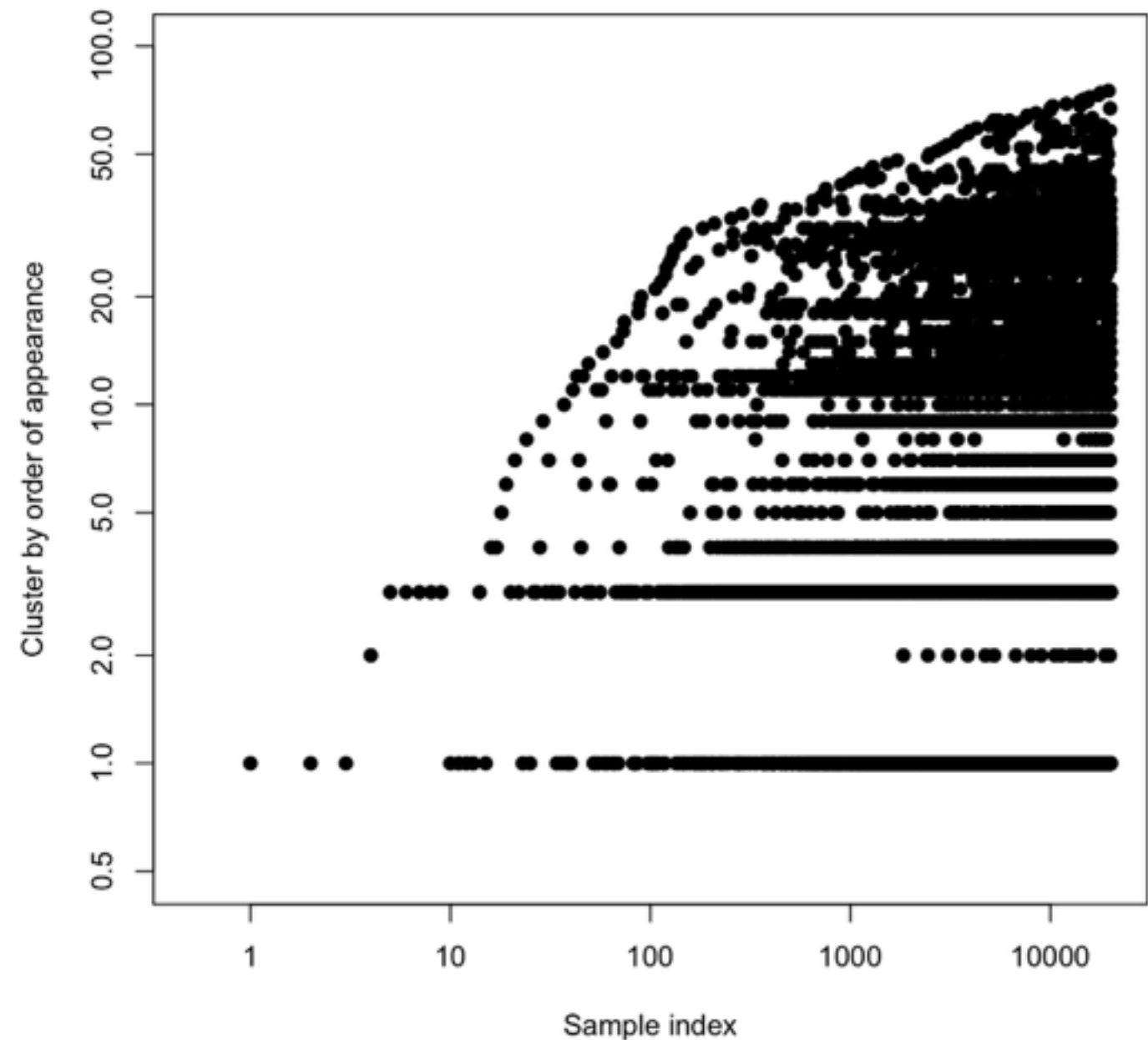
# Power laws

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



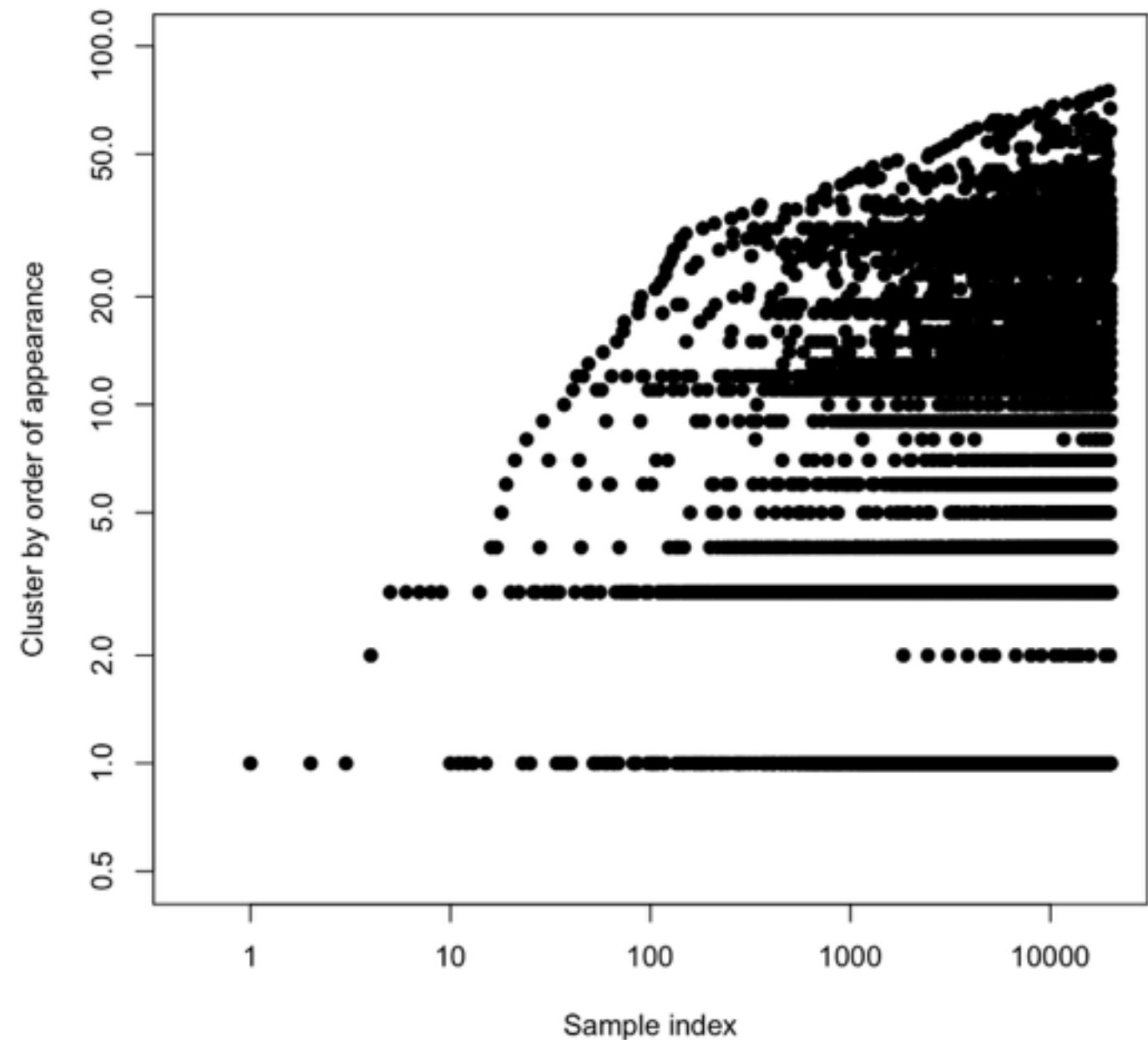
# Power laws

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



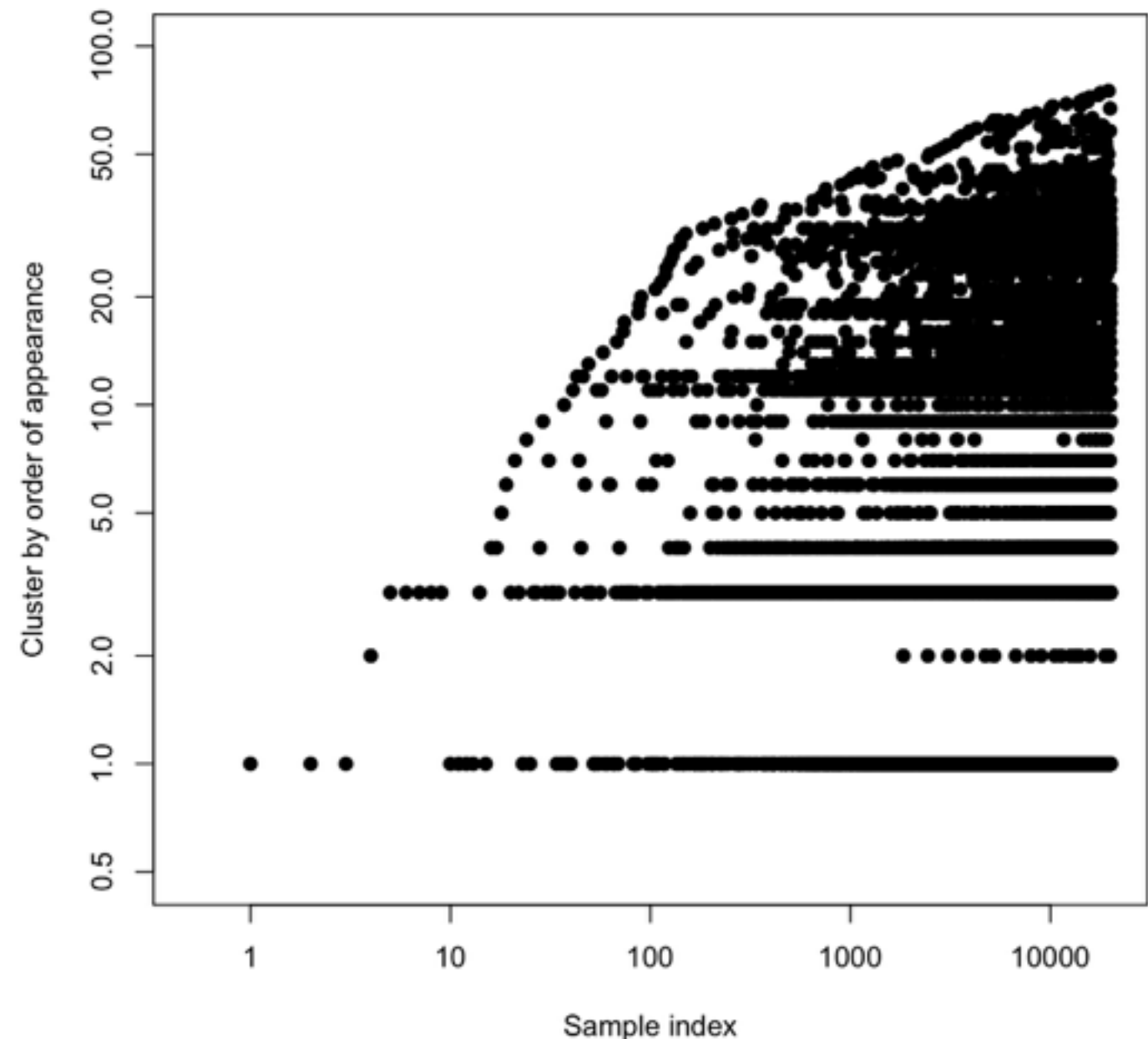
# Power laws

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



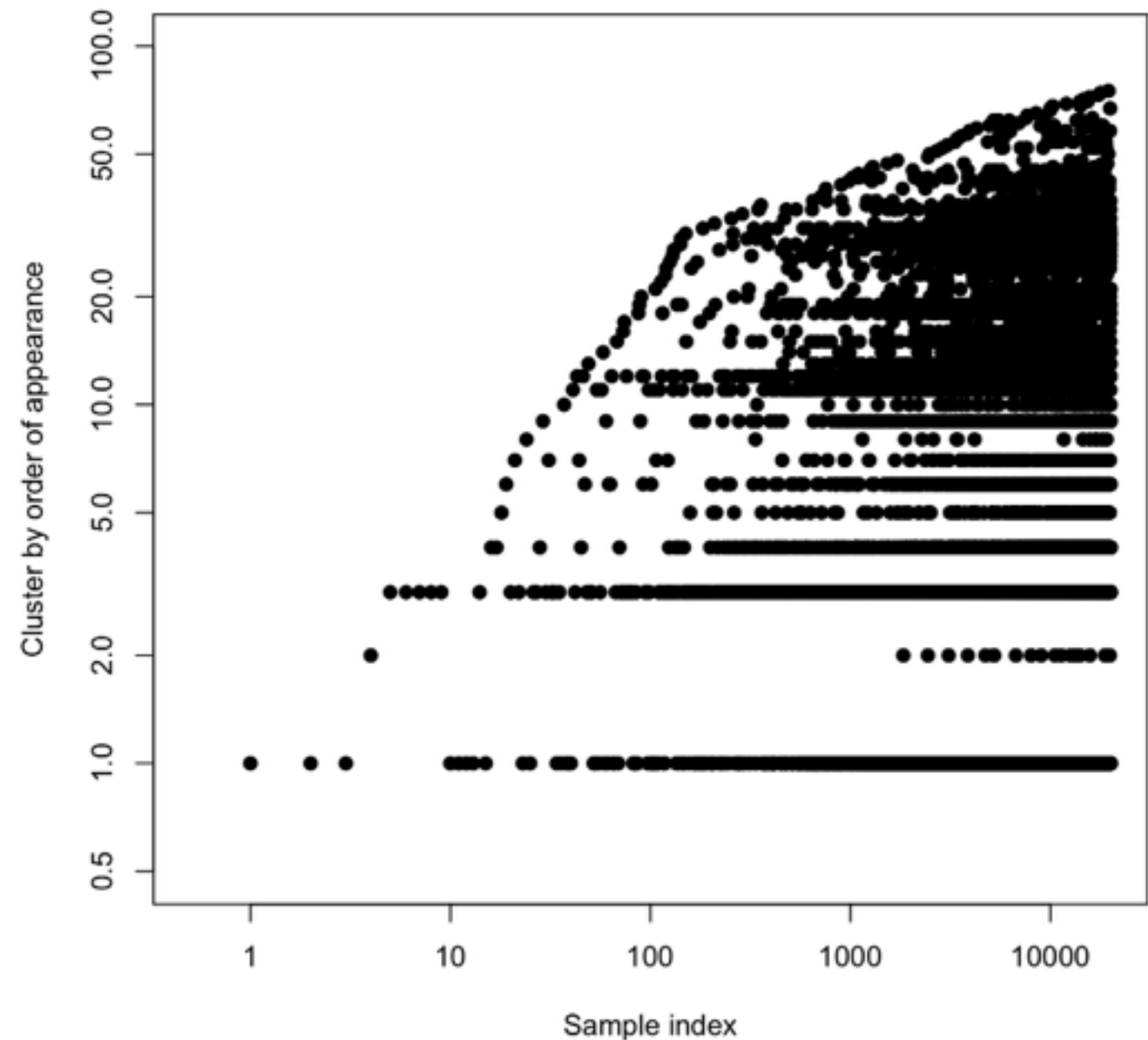
# Power laws

- $K_N := \#$  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:



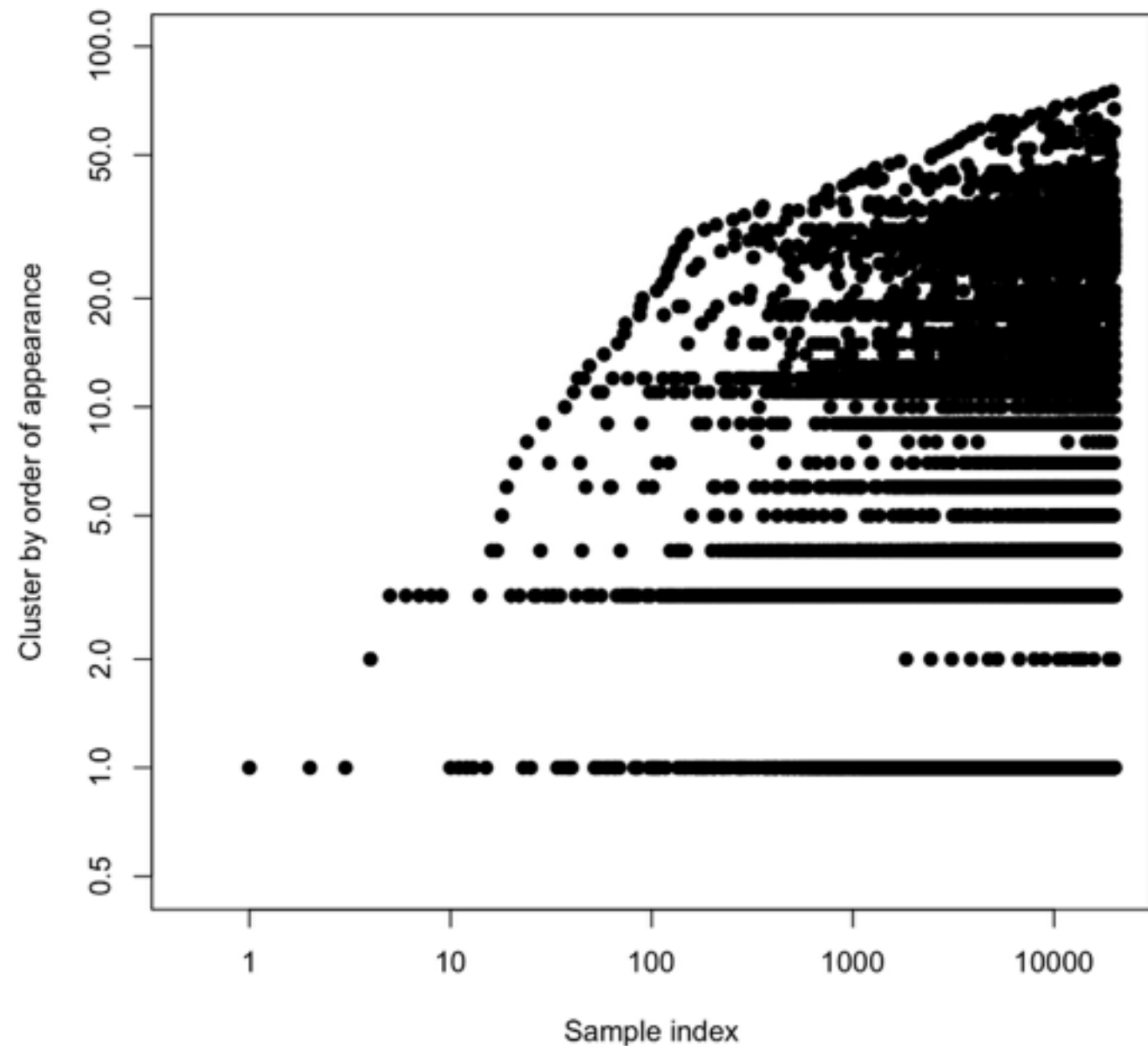
# Power laws

- $K_N := \#$  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:



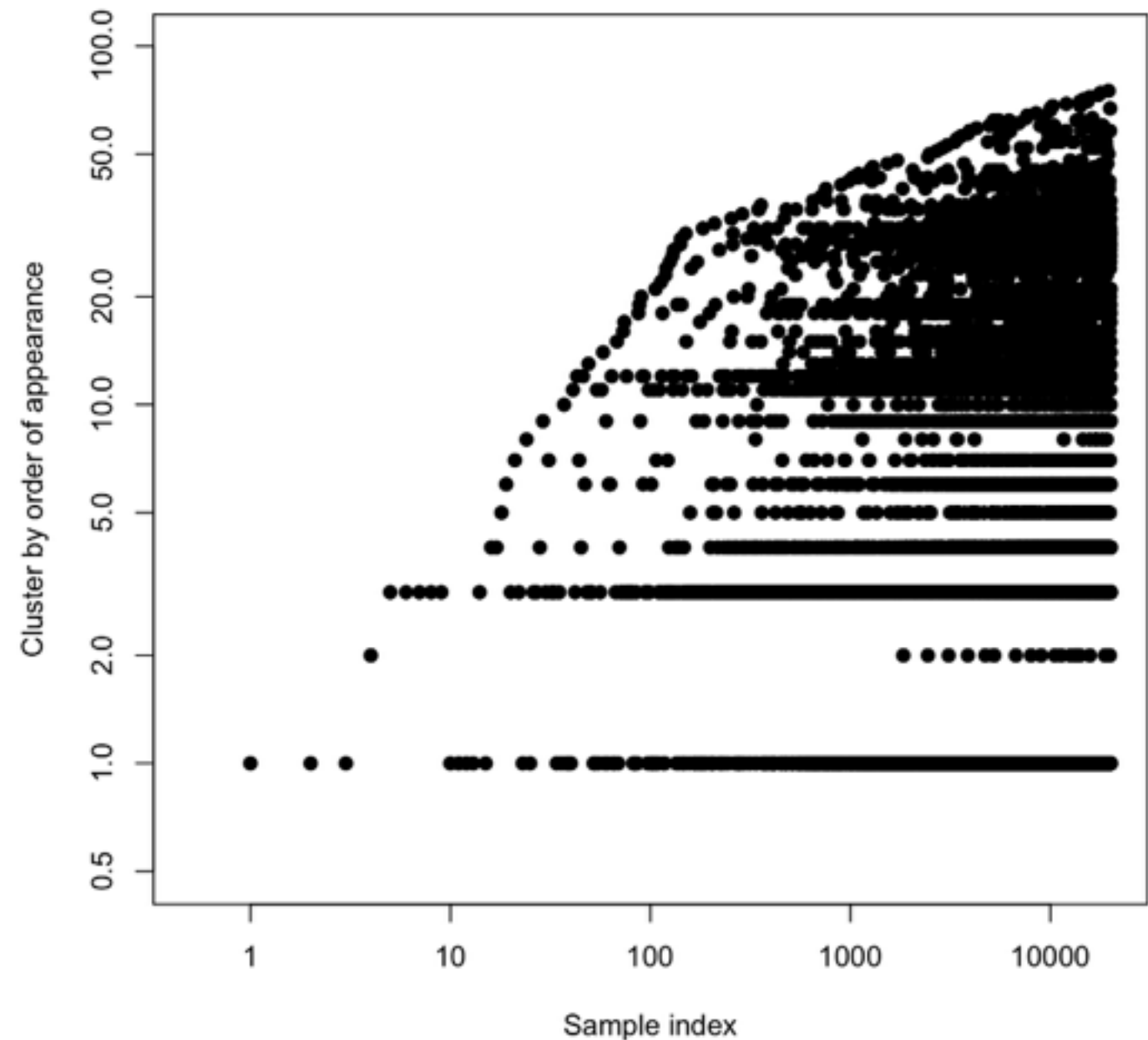
# Power laws

- $K_N := \#$  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



# Power laws

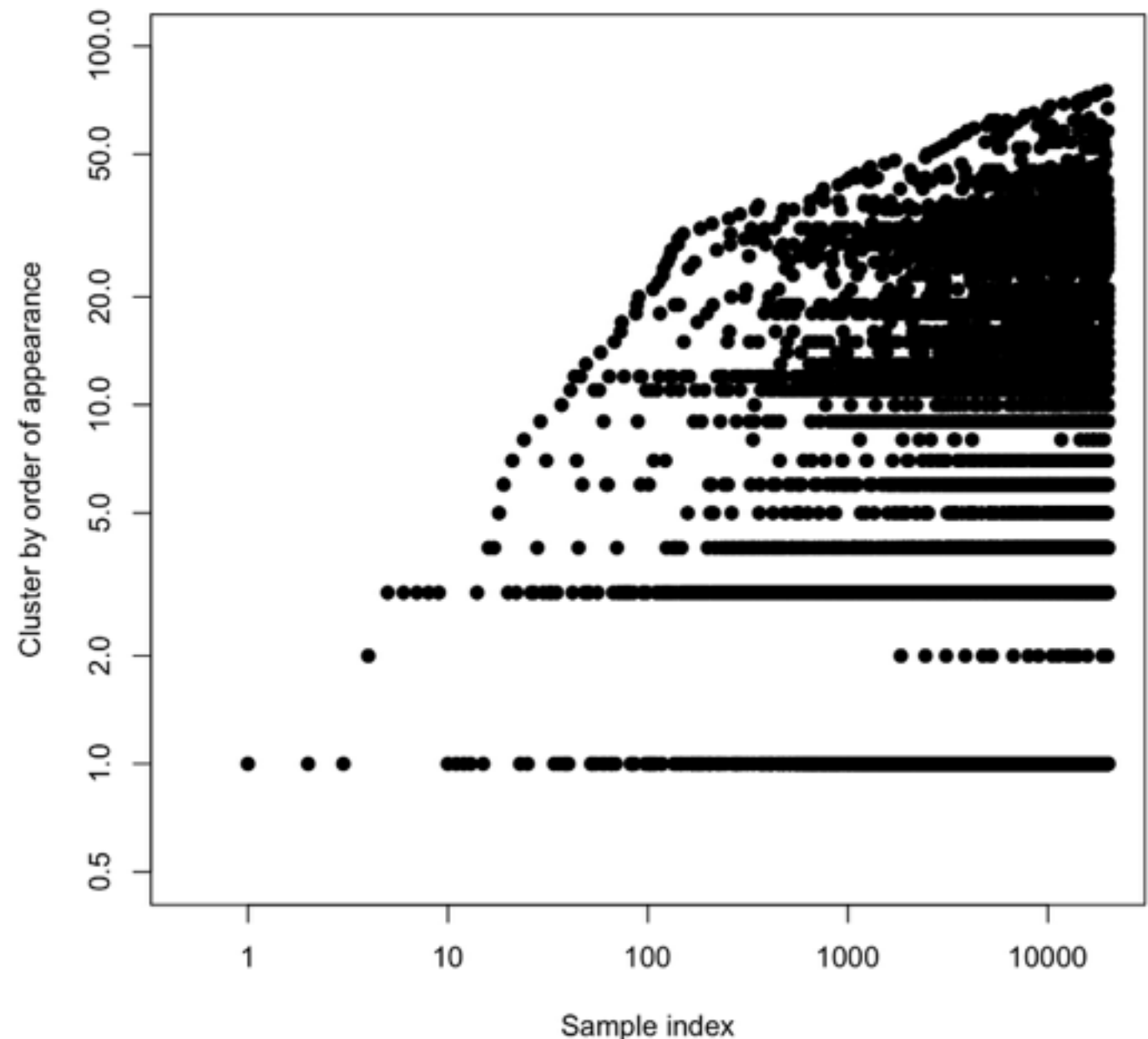
- $K_N := \#$  clusters occupied by  $N$  data points
- CRP:  $K_N \sim \alpha \log N$  w.p. 1
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:  
 $K_N \sim S_\alpha N^\sigma$  w.p. 1





# Power laws

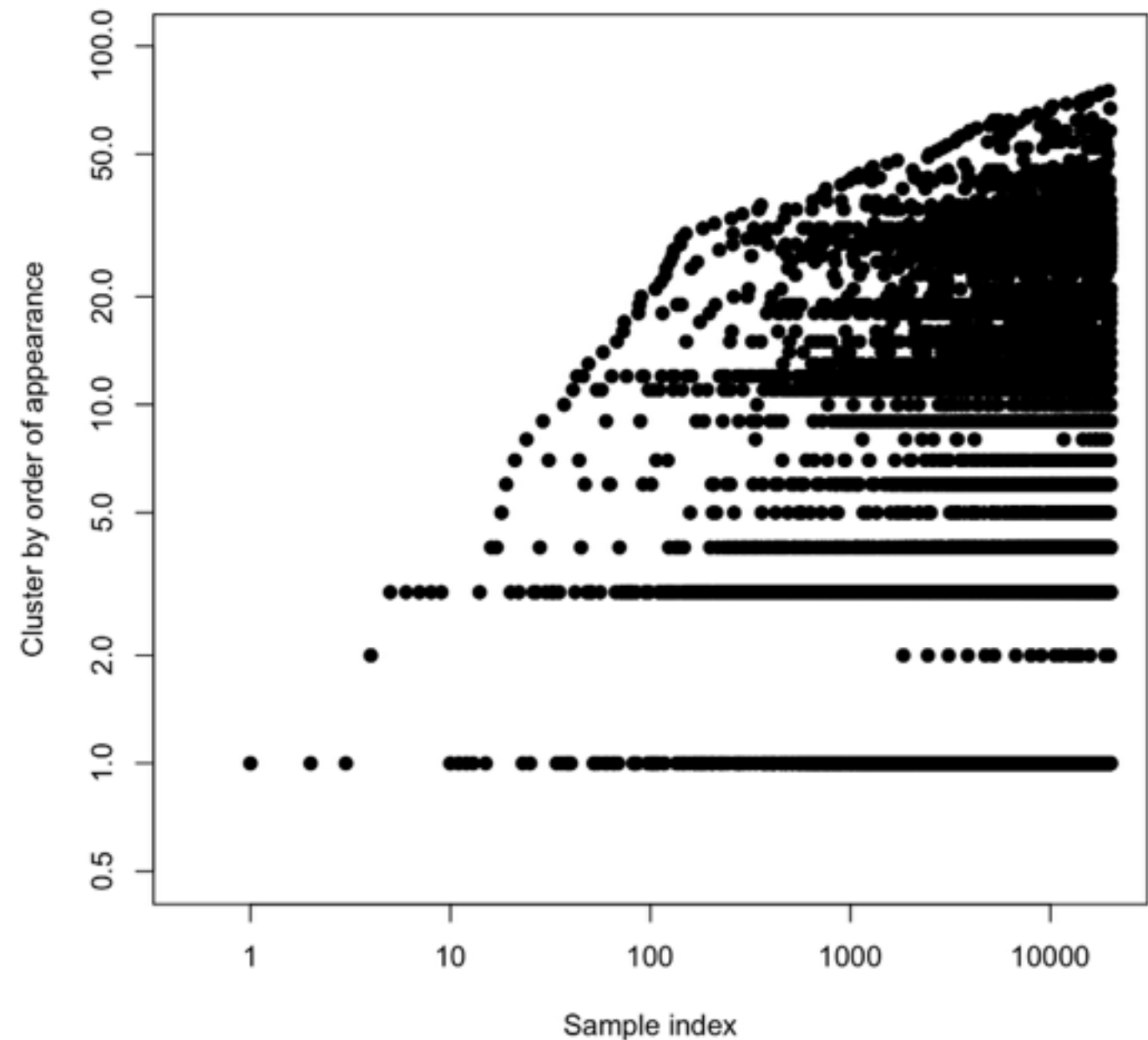
- $K_N := \#$  clusters occupied by  $N$  data points
- CRP:  $K_N \sim \alpha \log N$  w.p. 1
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:
$$K_N \sim S_\alpha N^\sigma \text{ w.p. } 1$$
  - related to Zipf's law (ranked frequencies)





# 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:
$$K_N \sim S_\alpha N^\sigma \text{ w.p. } 1$$
  - related to Zipf's law (ranked frequencies)
- Not just clusters



# Hierarchies

# Hierarchies

- Hierarchical  
Dirichlet process

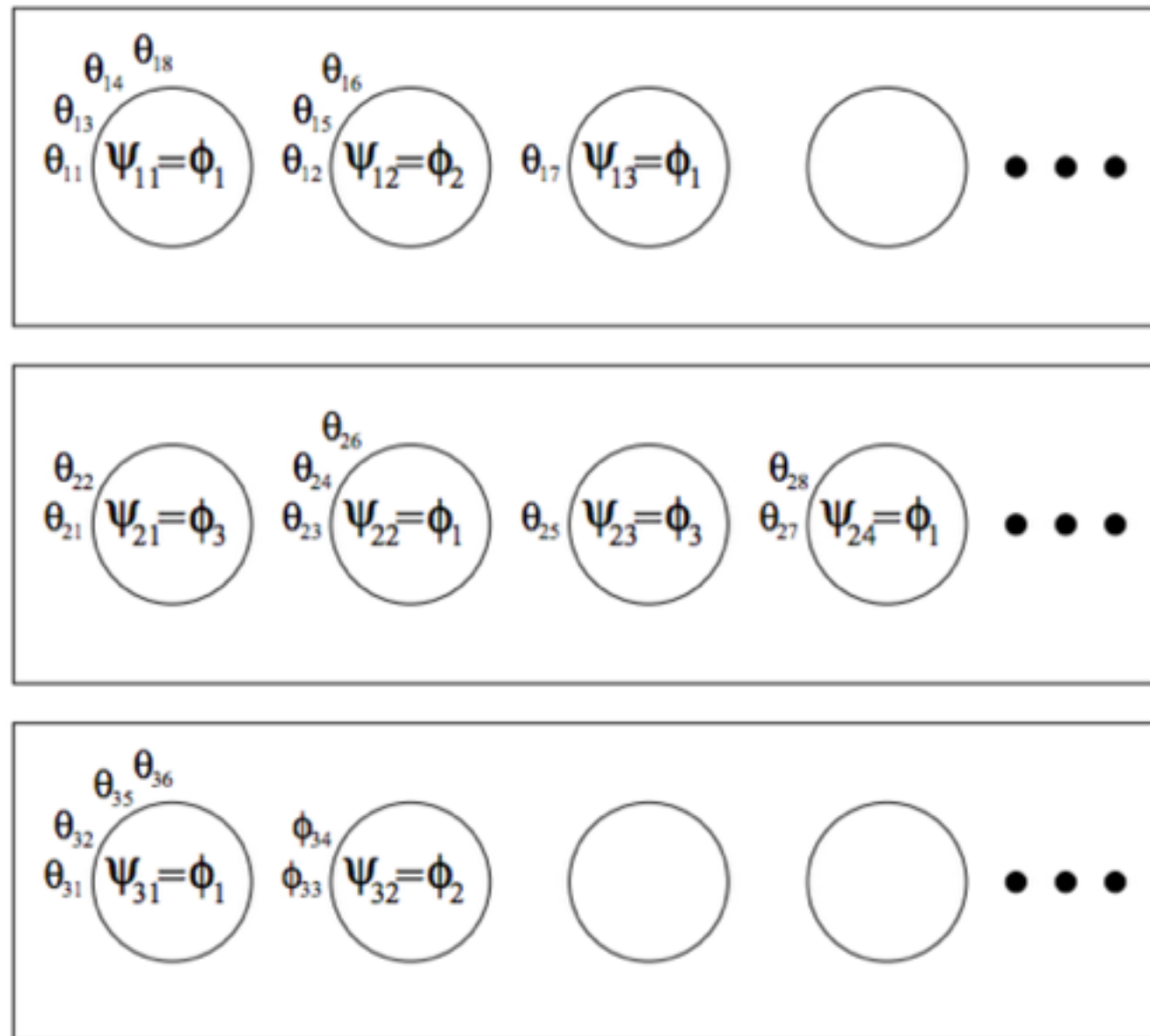
# Hierarchies

- Hierarchical Dirichlet process

# Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise

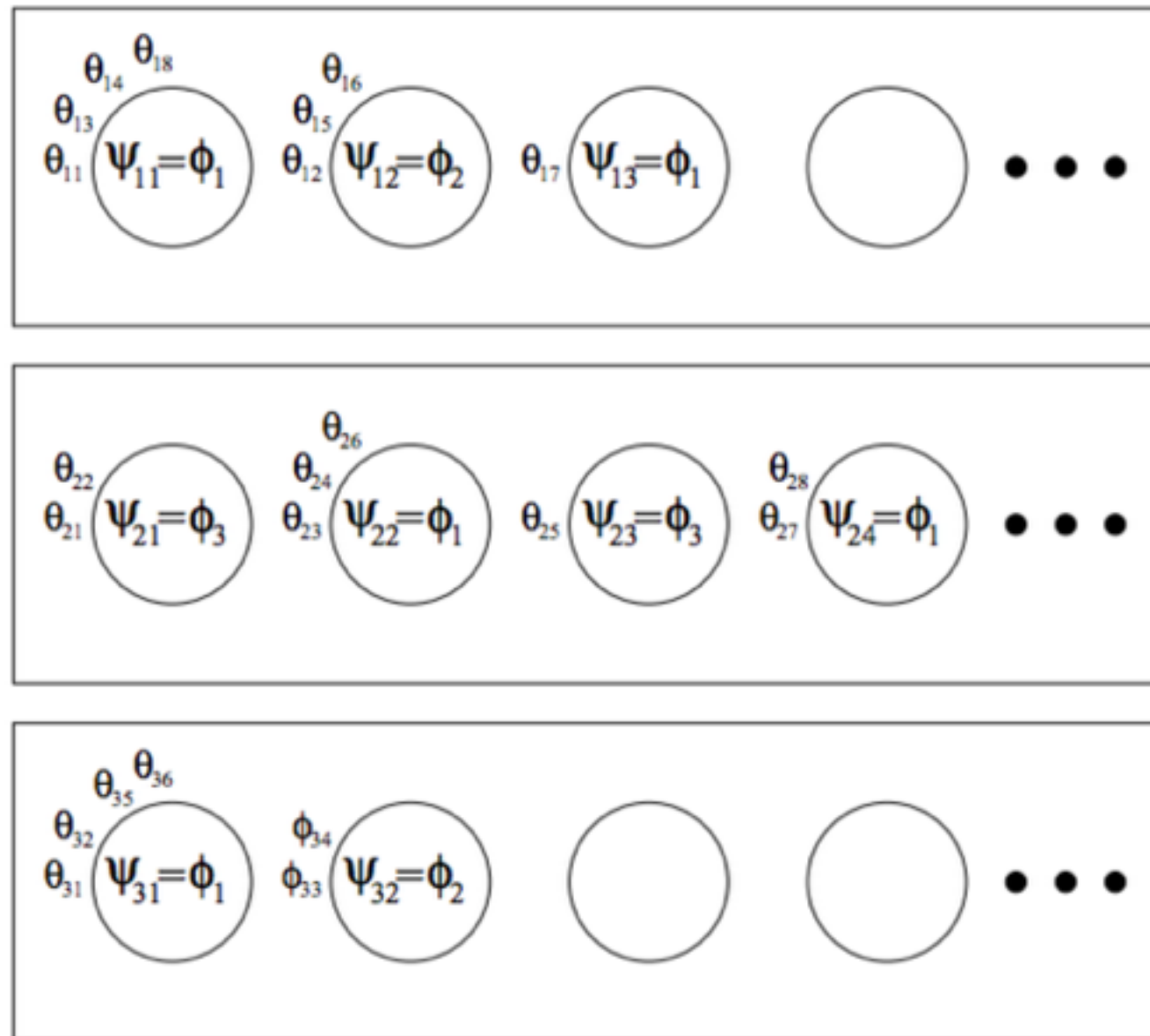
# Hierarchies



- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006]

# Hierarchies

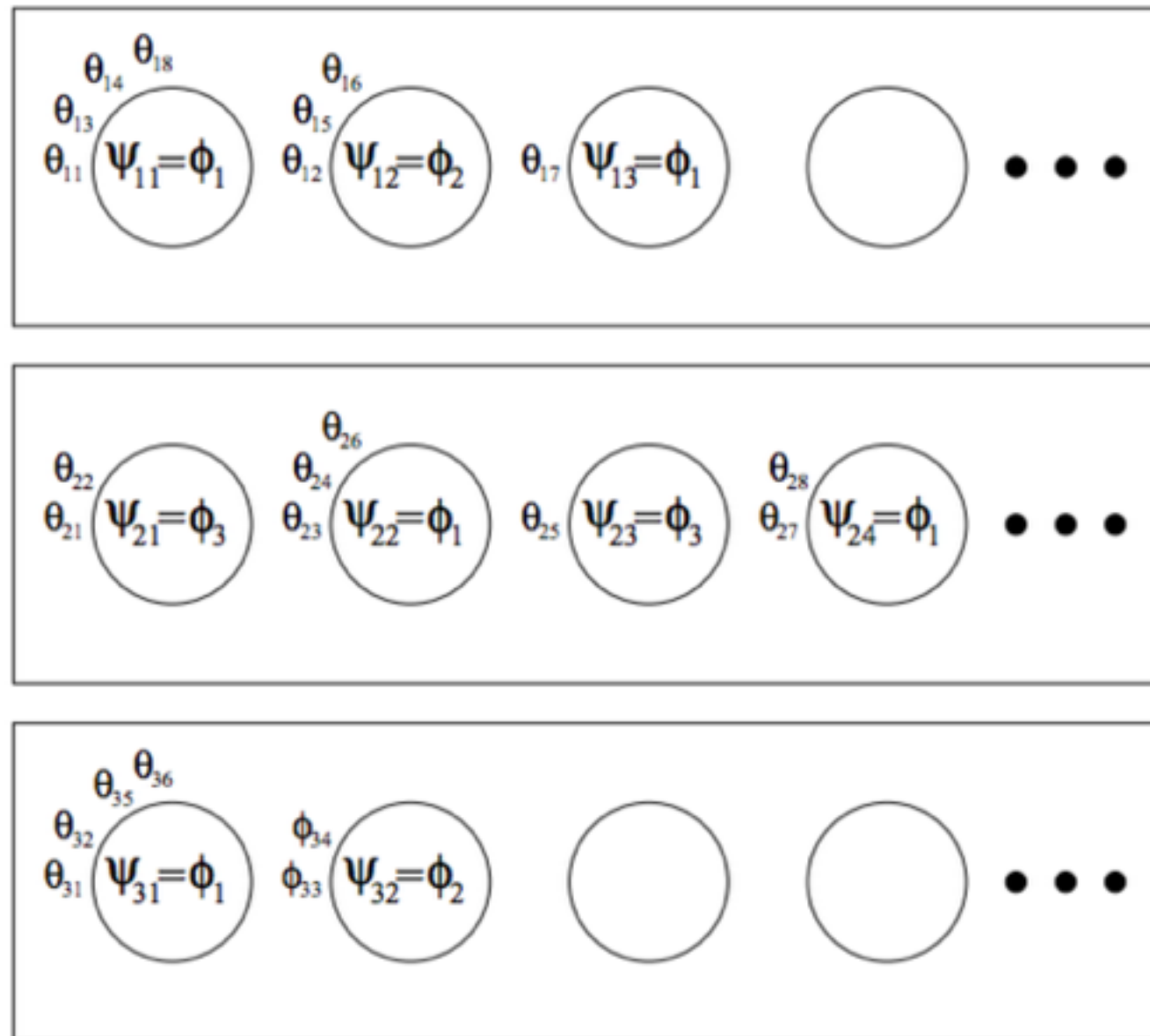


- Hierarchical Dirichlet process
- Chinese restaurant franchise
- Hierarchical beta process

[Teh et al 2006]



# Hierarchies

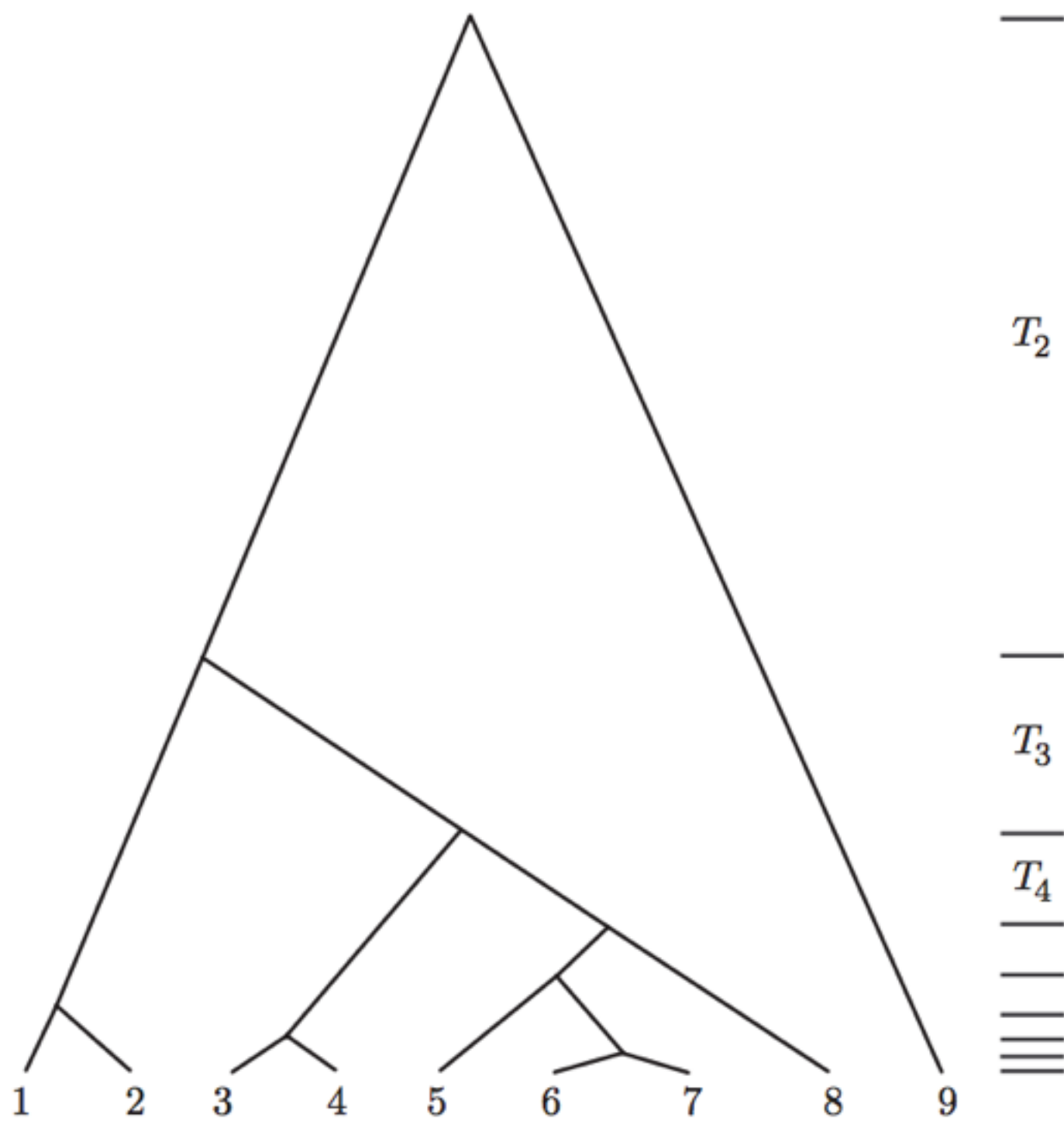


- Hierarchical Dirichlet process
- Chinese restaurant franchise
- Hierarchical beta process

[Teh et al 2006]

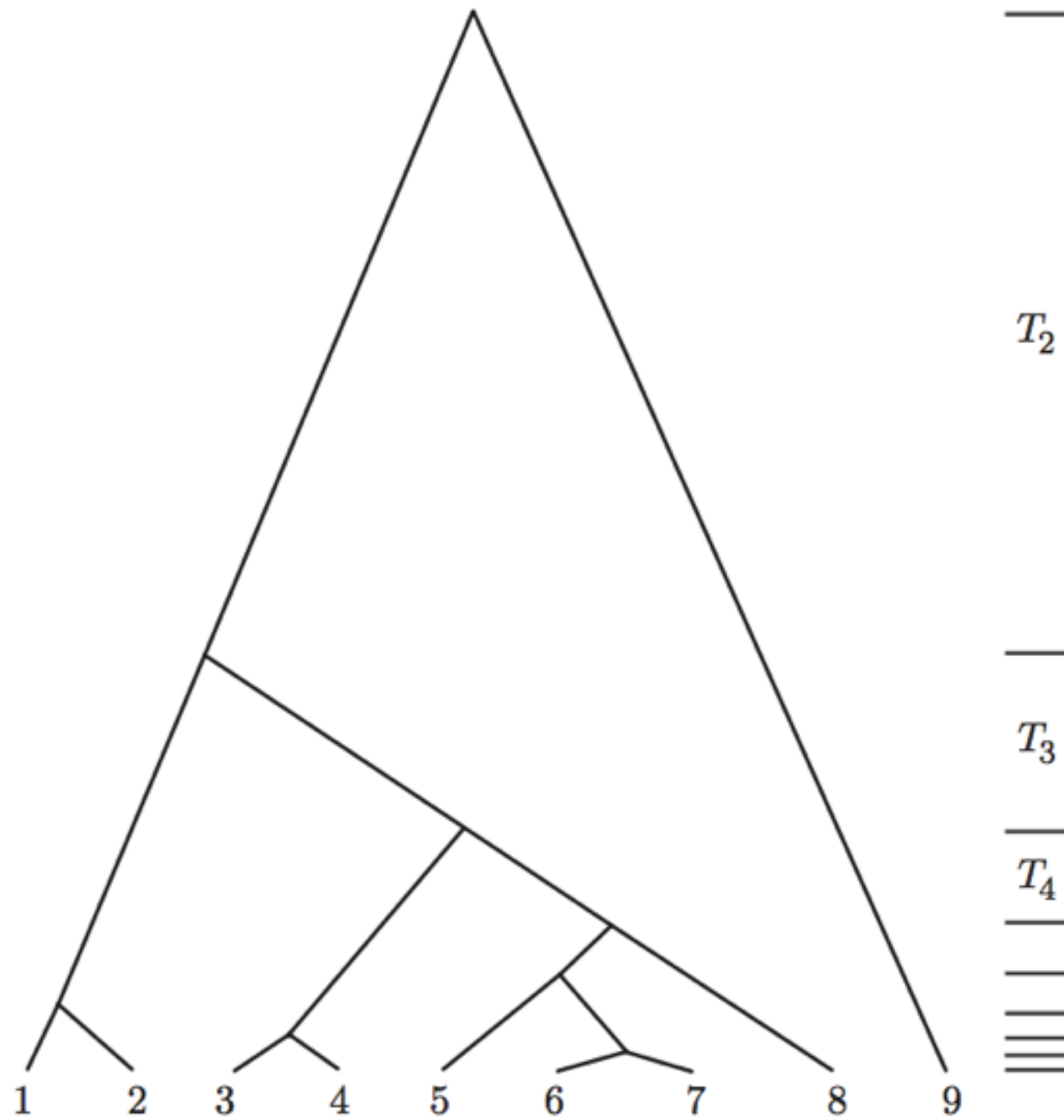


# Genealogy, trees, beyond trees



[Wakeley 2008]

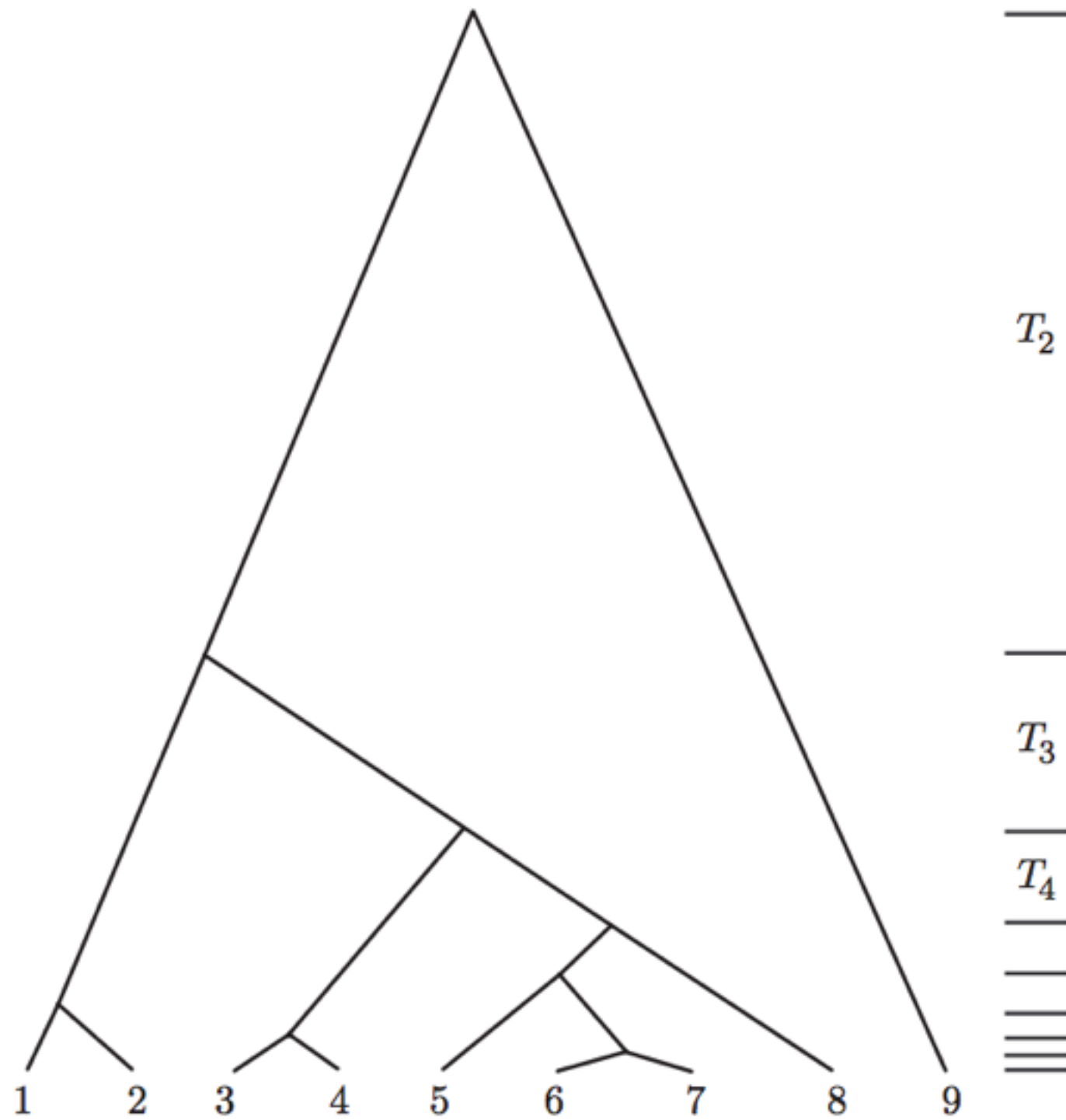
# Genealogy, trees, beyond trees



- Kingman coalescent

[Wakeley 2008]

# Genealogy, trees, beyond trees

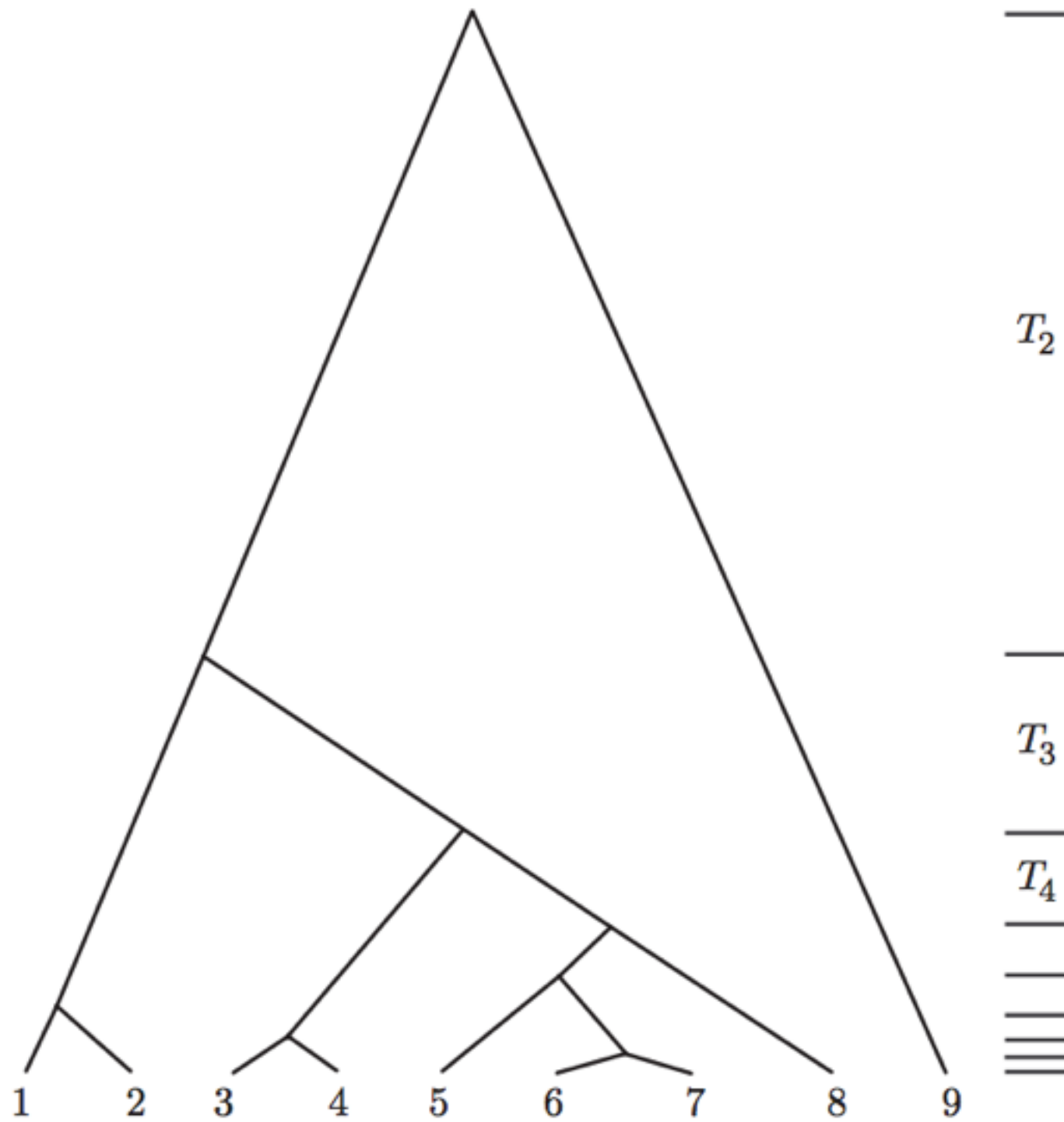


- Kingman coalescent

[Wakeley 2008]

[Kingman 1982]

# Genealogy, trees, beyond trees



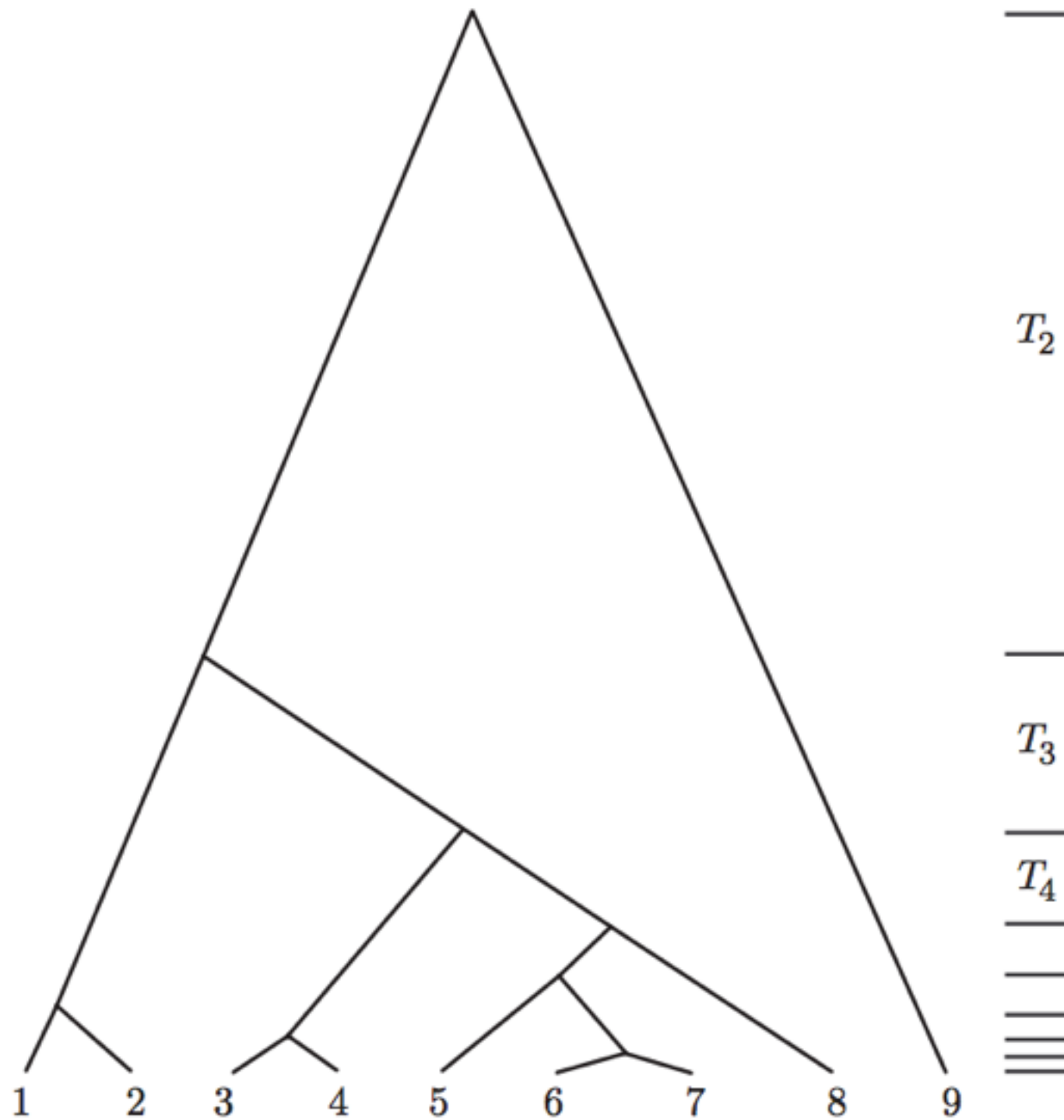
[Wakeley 2008]

[Kingman 1982]



- Kingman coalescent
- Fragmentation
- Coagulation

# Genealogy, trees, beyond trees

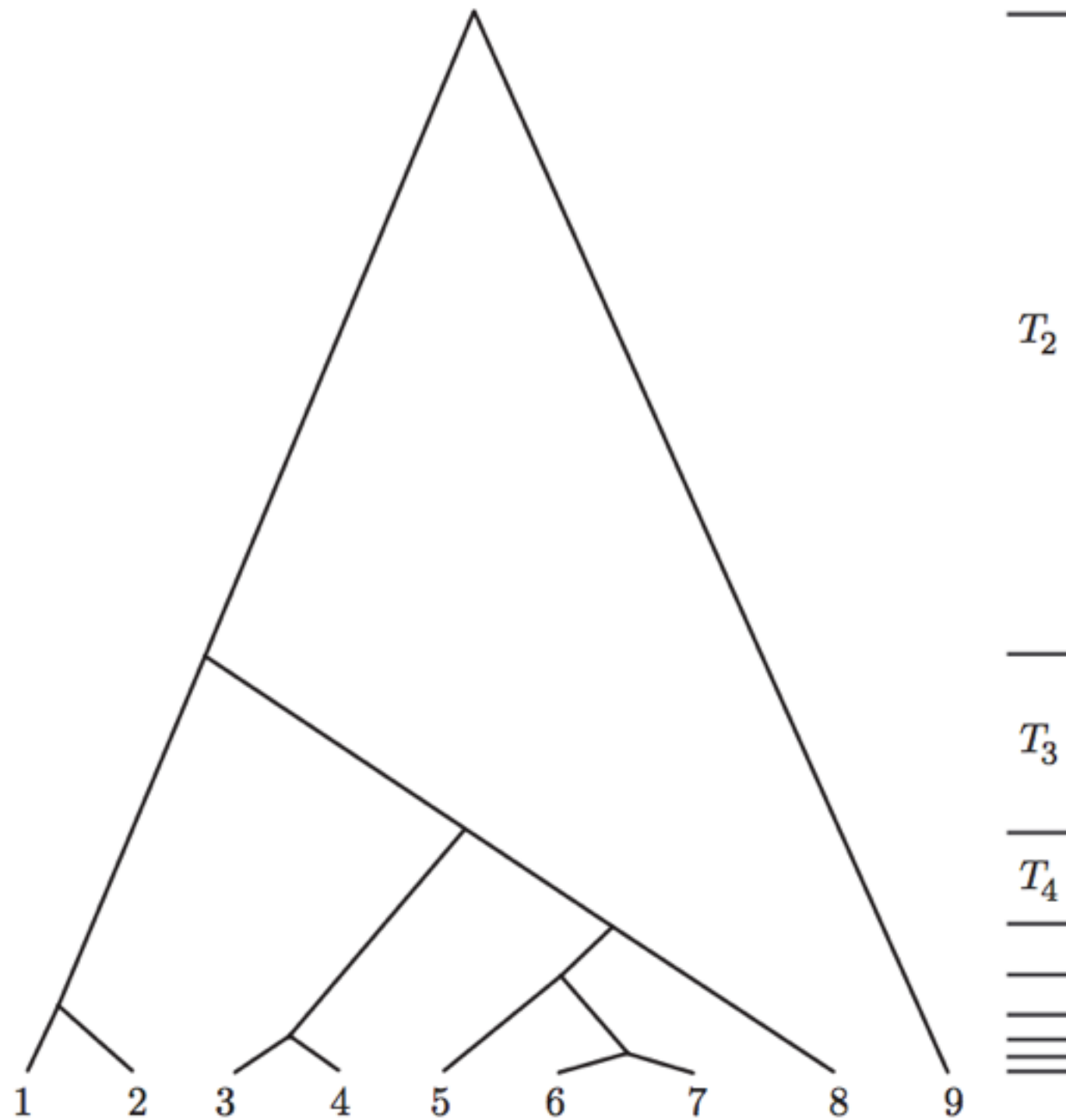


- Kingman coalescent
- Fragmentation
- Coagulation

[Wakeley 2008]

[Kingman 1982, Bertoin 2006, Teh et al 2011]

# Genealogy, trees, beyond trees

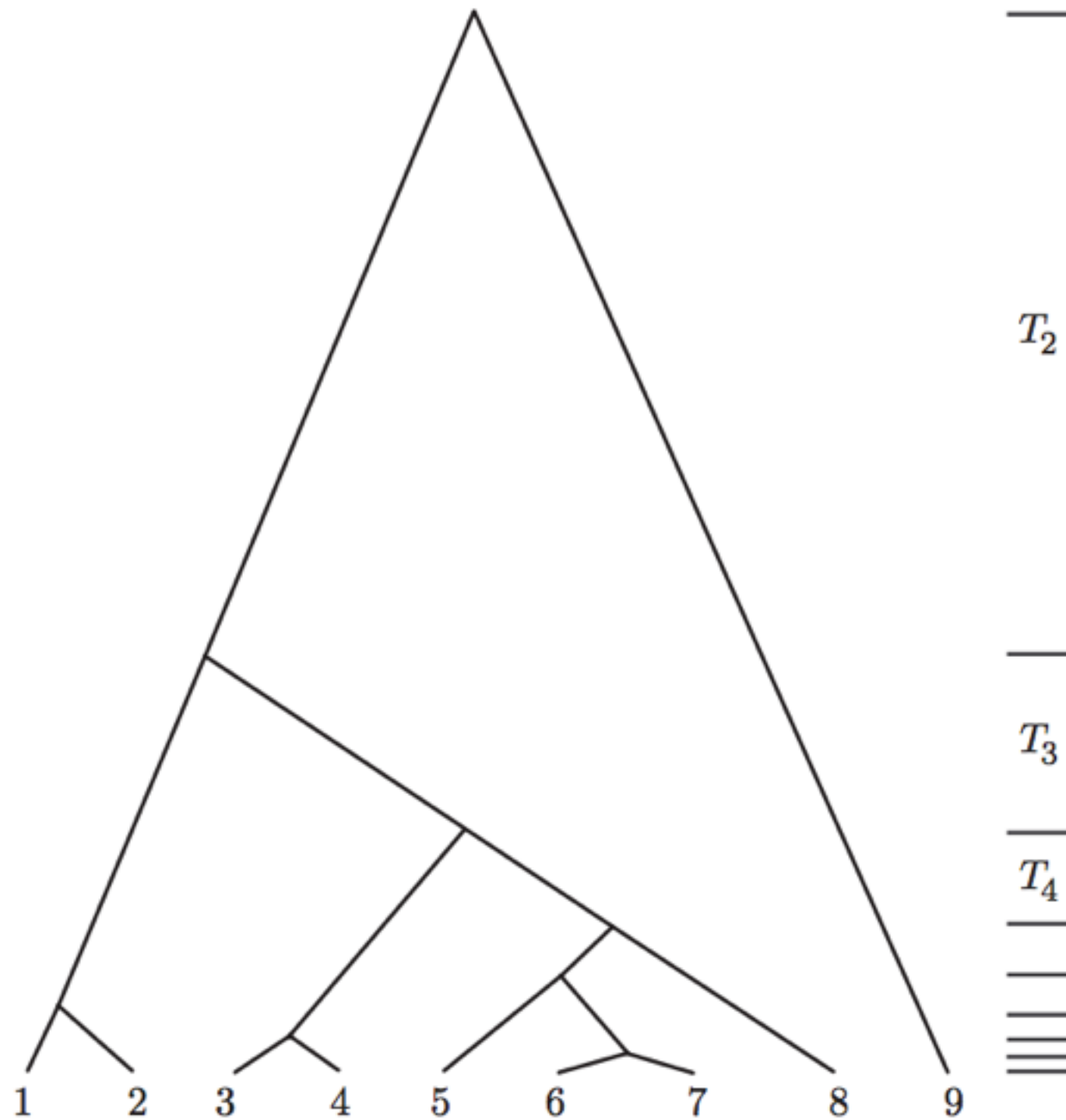


[Wakeley 2008]

[Kingman 1982, Bertoin 2006, Teh et al 2011]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

# Genealogy, trees, beyond trees



[Wakeley 2008]

[Kingman 1982, Bertoin 2006, Teh et al 2011, Neal 2003]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree



# Conjugacy & Poisson point processes



# Conjugacy & Poisson point processes

- Beta process, Bernoulli process (Indian buffet)

# Conjugacy & Poisson point processes

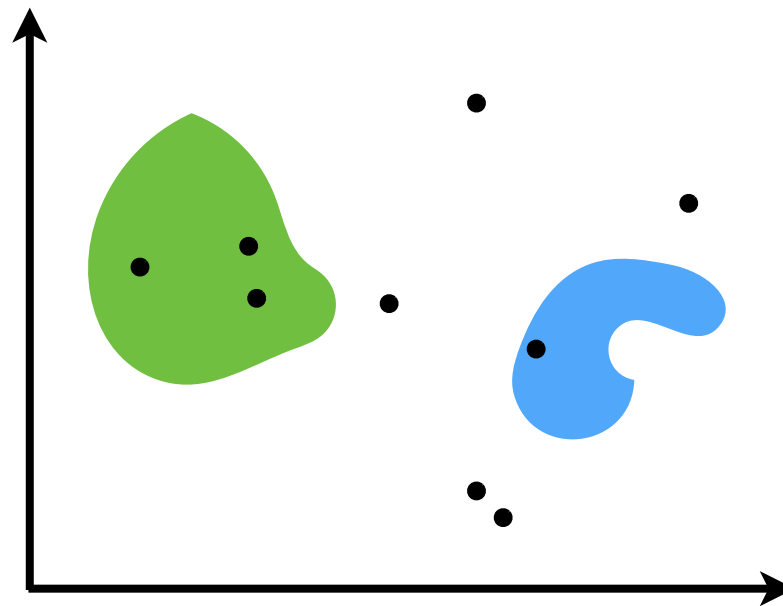
- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)

# Conjugacy & Poisson point processes

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

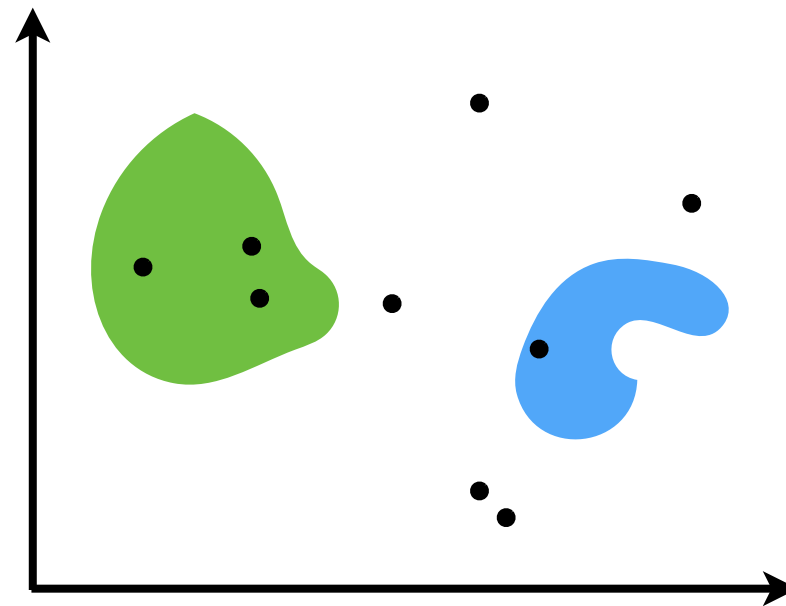
# Conjugacy & Poisson point processes

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



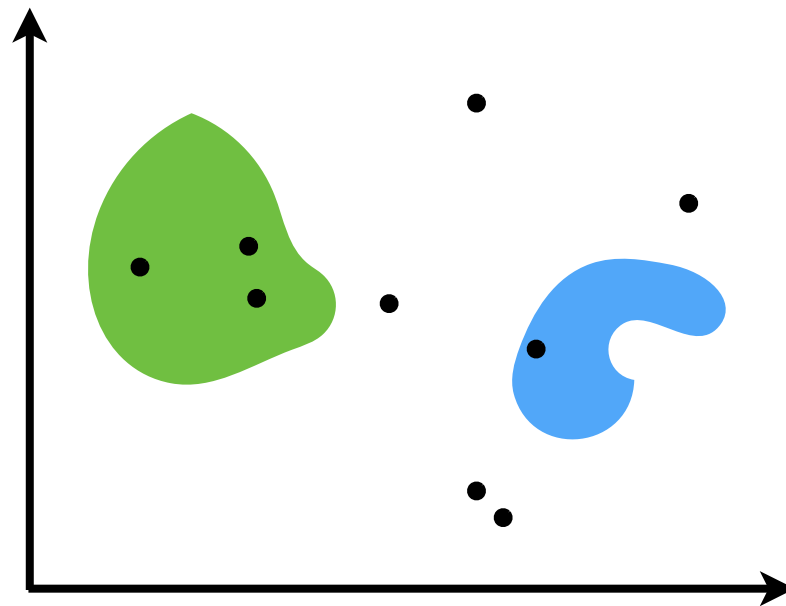
# Conjugacy & Poisson point processes

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



# Conjugacy & Poisson point processes

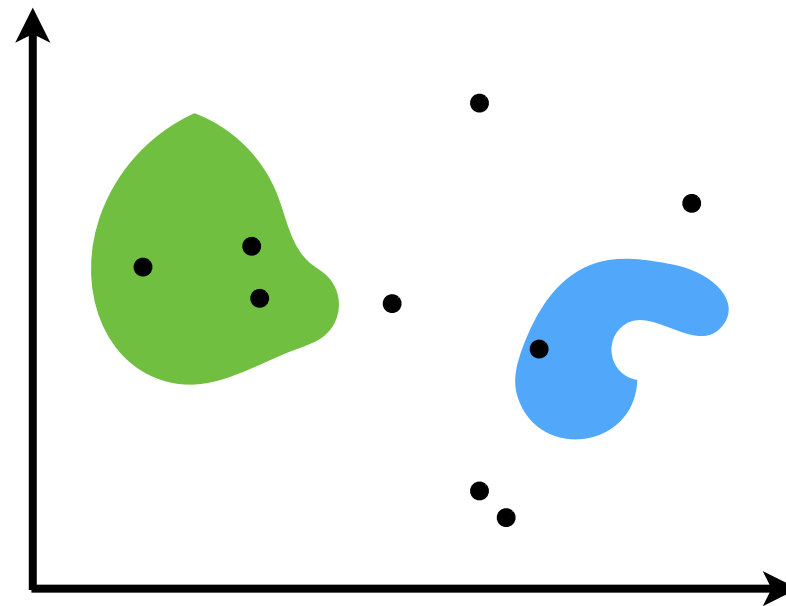
- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process



- 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

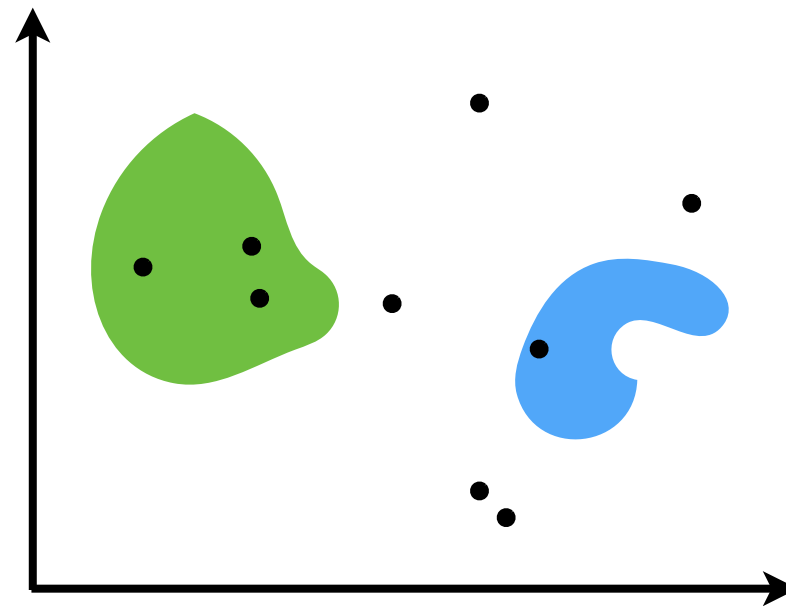


- Posteriors, conjugacy, and exponential families for completely random measures



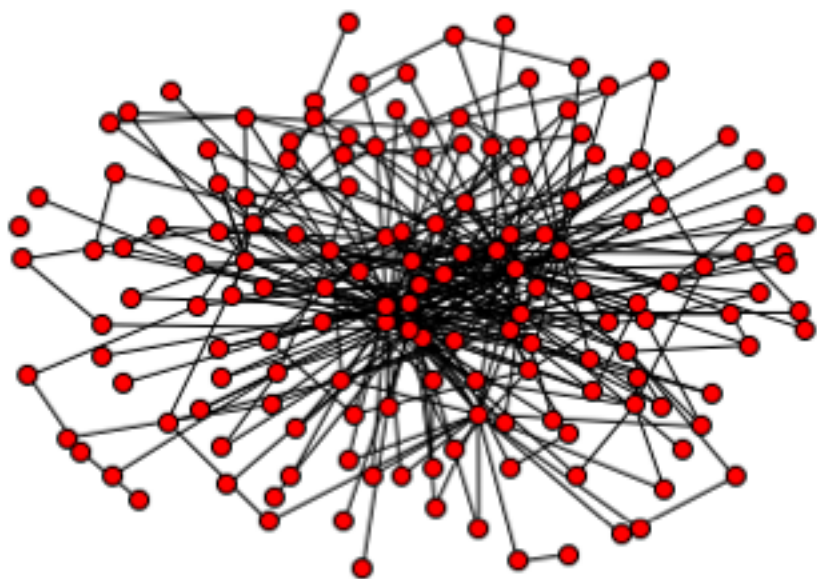
# Conjugacy & Poisson point processes

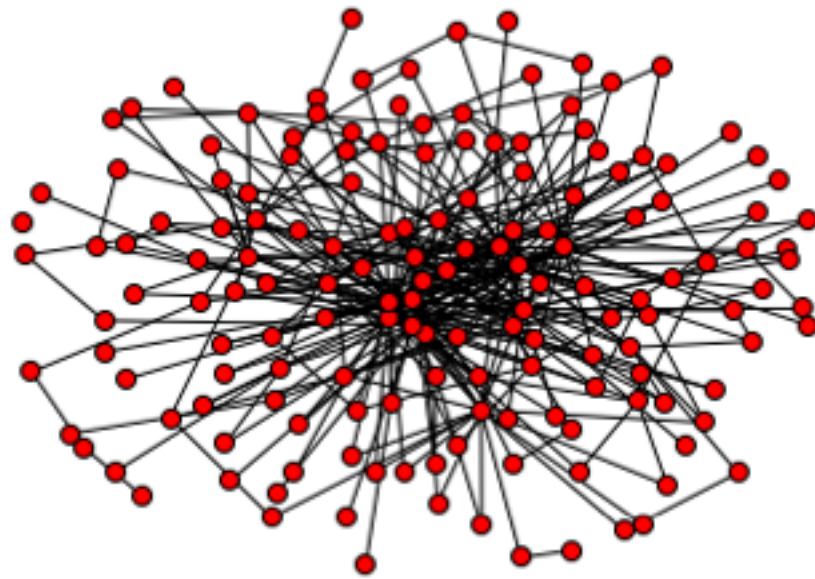
- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process



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

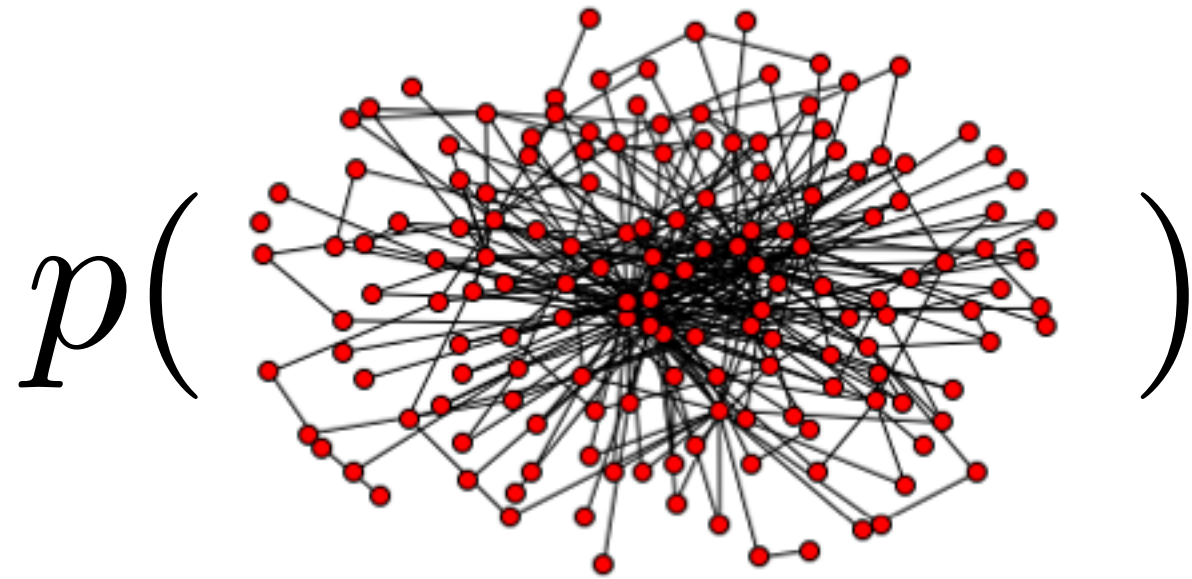






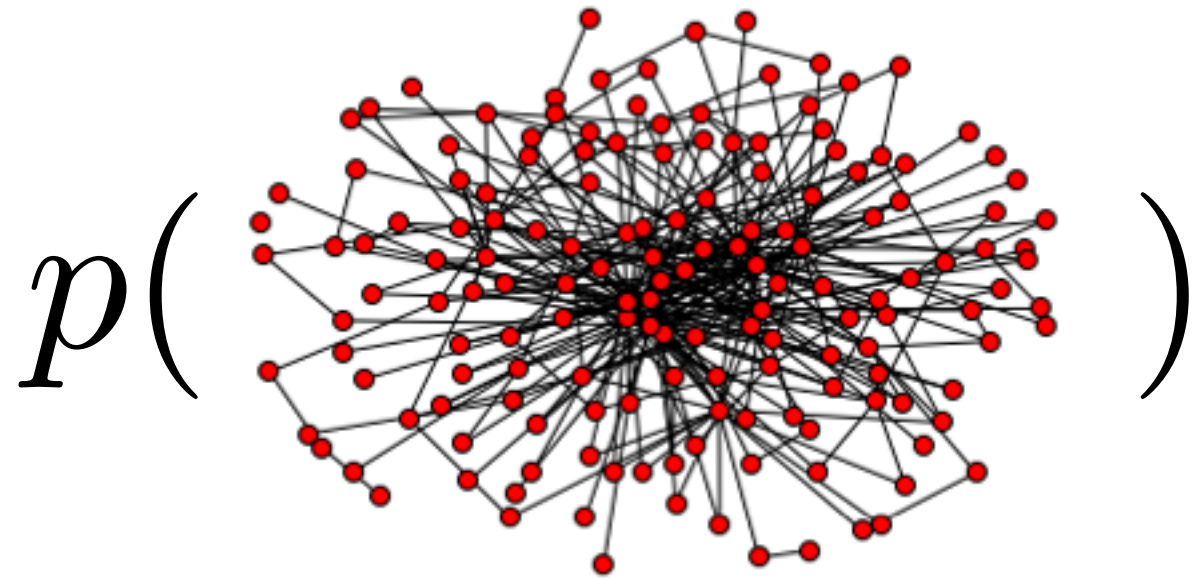
social: *Facebook, Twitter, email*  
biological: *ecological, protein, gene*  
transportation: *roads, railways*

# Probabilistic models for graphs



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

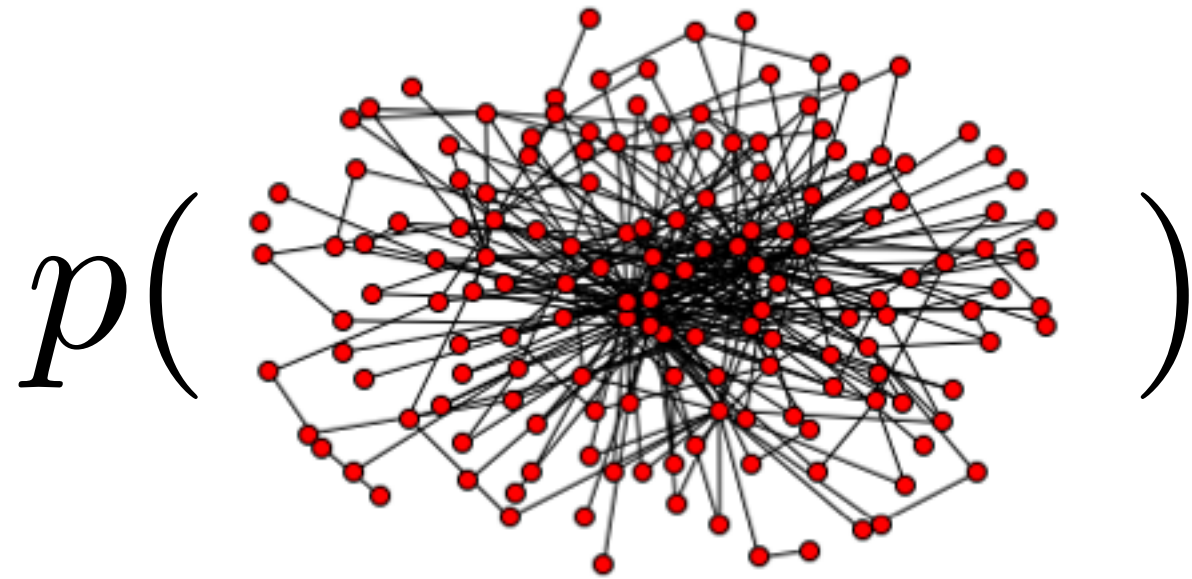
# Probabilistic models for graphs



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

- Rich relationships, coherent uncertainties, prior info

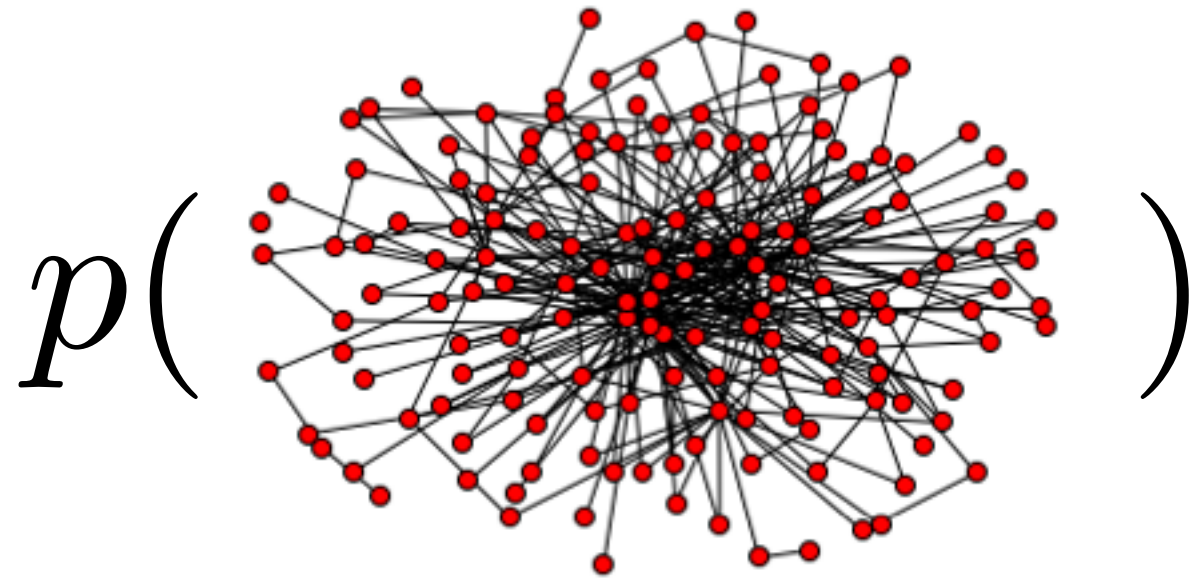
# Probabilistic models for graphs



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

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

# Probabilistic models for graphs

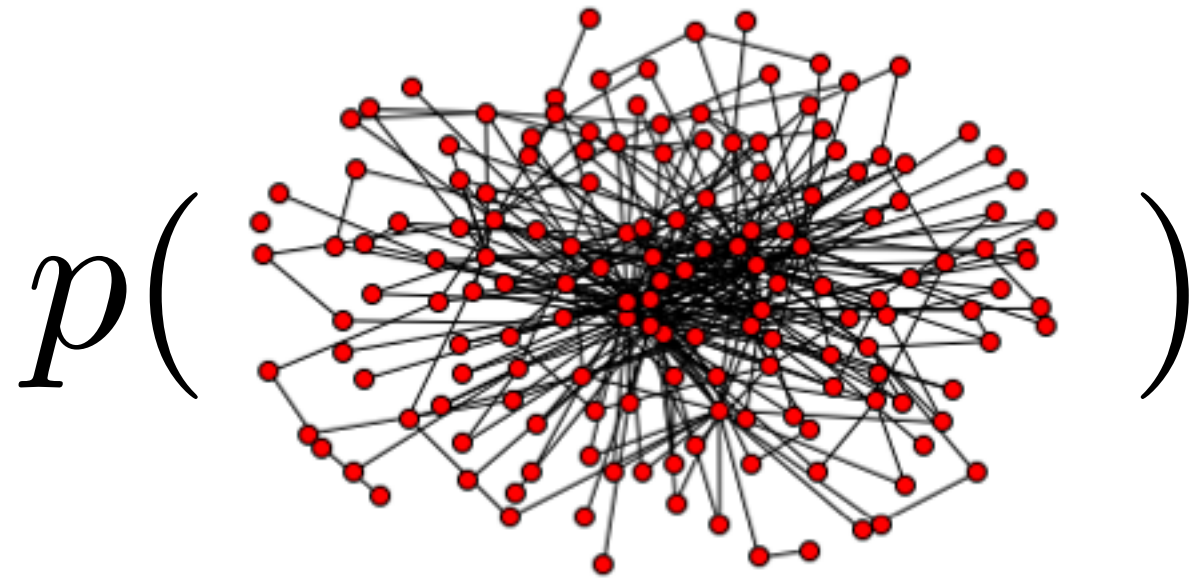


*social: Facebook, Twitter, email*  
*biological: ecological, protein, gene*  
*transportation: roads, railways*

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



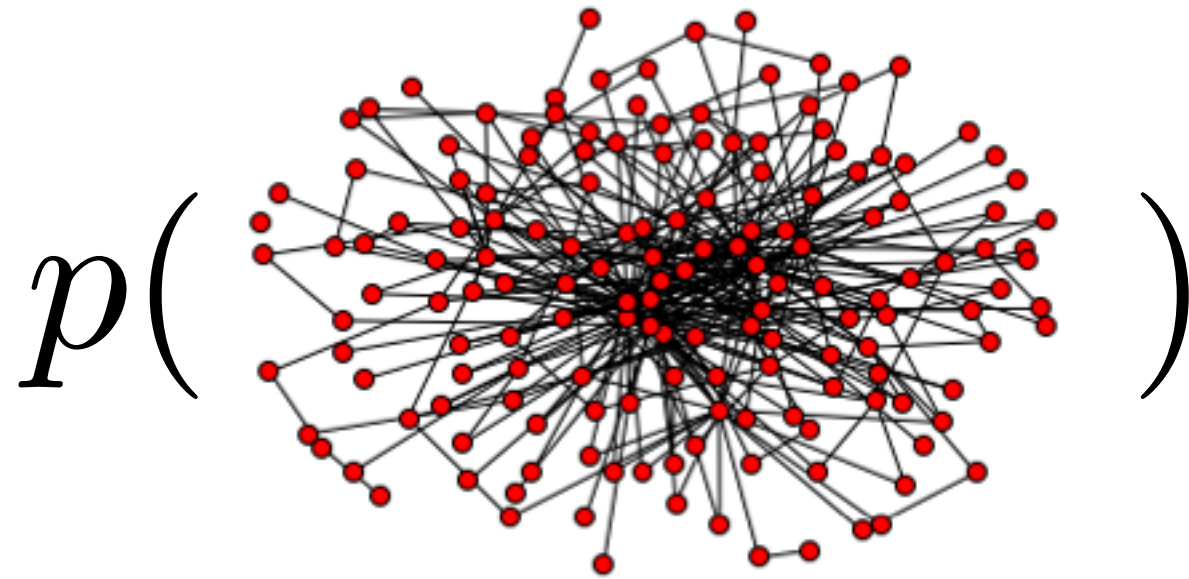
# Probabilistic models for graphs



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

- 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*)

# Probabilistic models for graphs

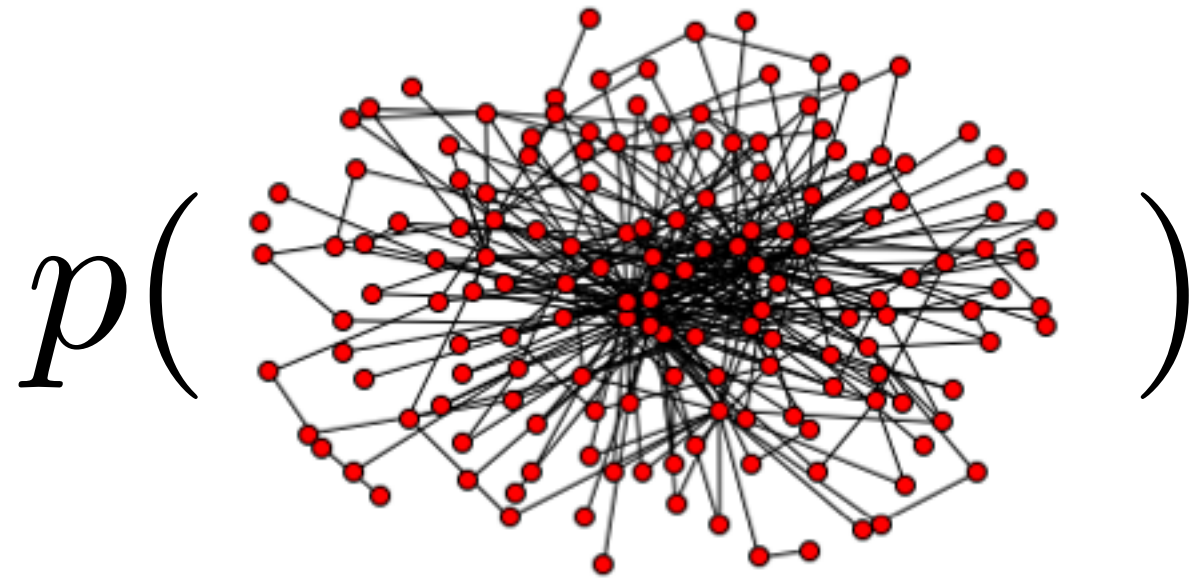


*social: Facebook, Twitter, email*  
*biological: ecological, protein, gene*  
*transportation: roads, railways*

- 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



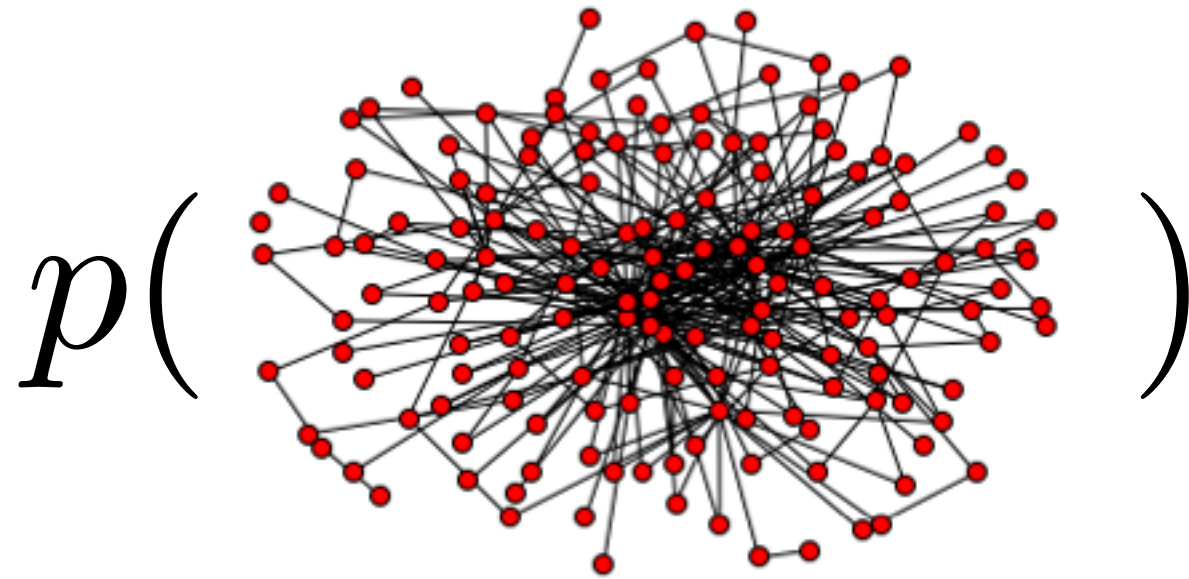
# Probabilistic models for graphs



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

- 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

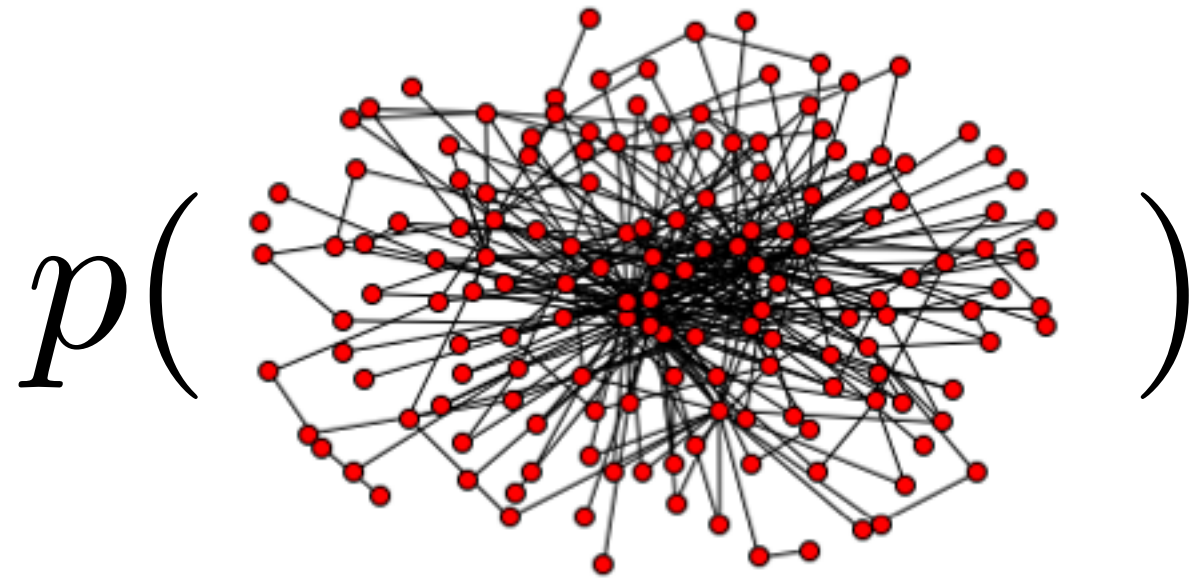
# Probabilistic models for graphs



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

- 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

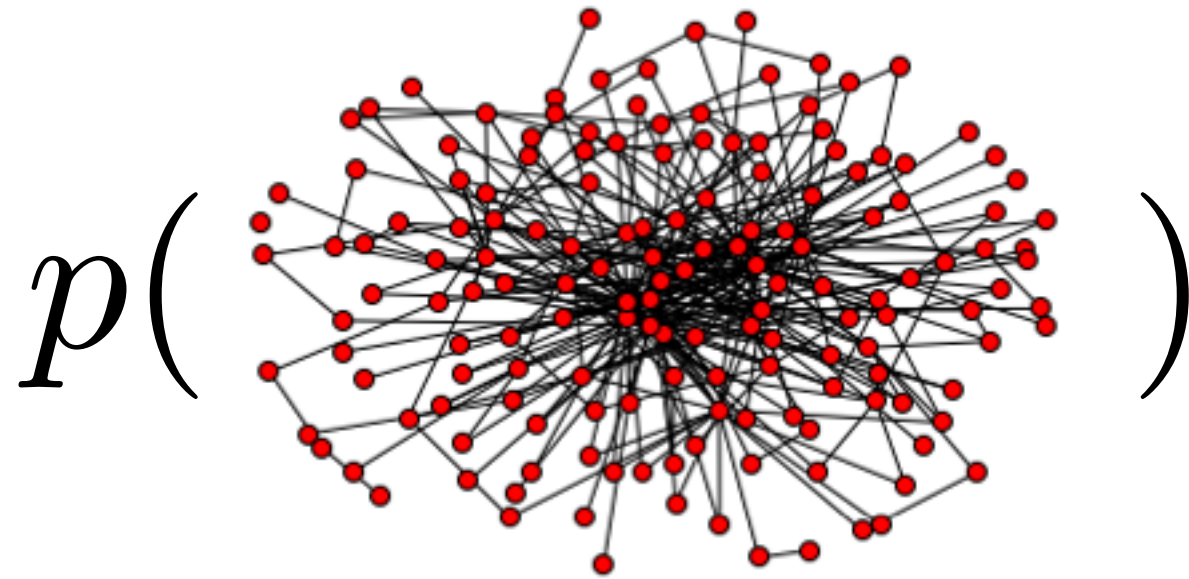
# Probabilistic models for graphs



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

- 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

# Probabilistic models for graphs

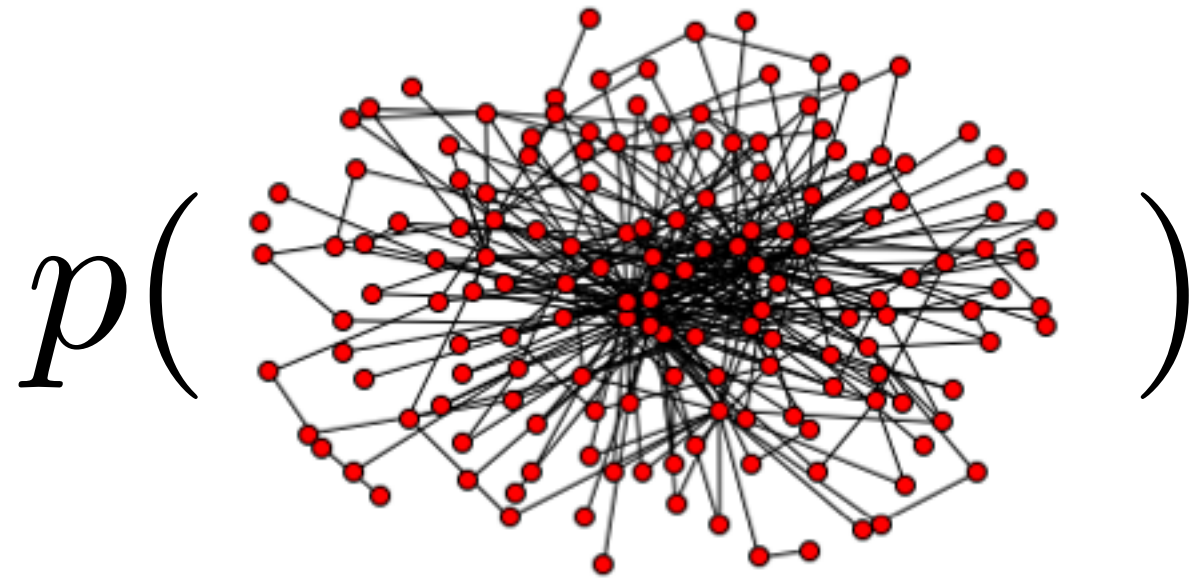


*social: Facebook, Twitter, email*  
*biological: ecological, protein, gene*  
*transportation: roads, railways*

- 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**



# Probabilistic models for graphs



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

- 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
  - “Edge exchangeability”

[Holland et al 1983;

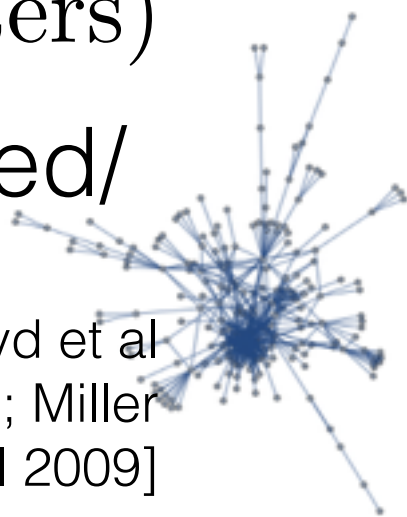
Kemp et al 2006; Xu et al 2007; Airol di et al 2008; Lloyd et al 2012; Broderick, Cai 2015; Cai, Broderick 2015a,b; Crane, Dempsey 2015a,b,2016a,b; Cai, Campbell, Broderick 2016; Campbell, Cai, Broderick 2016]

# Nonparametric Bayes

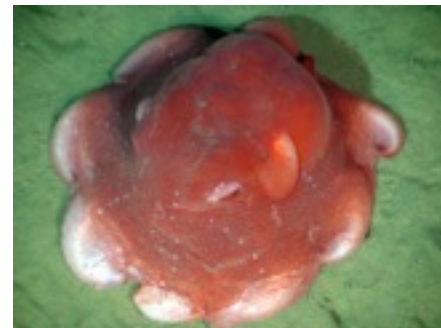
- Bayesian methods that are not parametric
- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- Not parametric (i.e. not finite parameter, unbounded/ growing/infinite number of parameters)



[Lloyd et al 2012; Miller et al 2009]



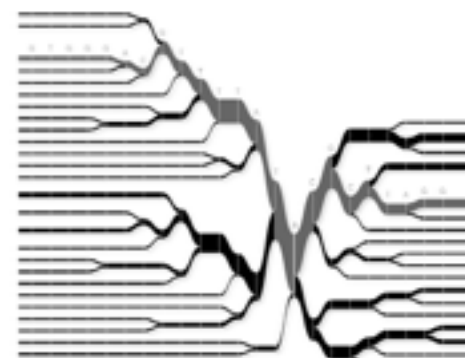
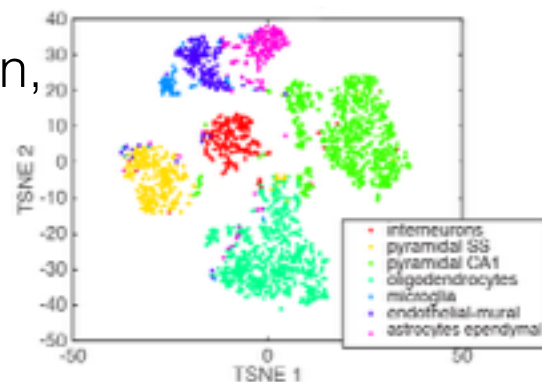
[Ed Bowlby, NOAA]



[Fox et al 2014]



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



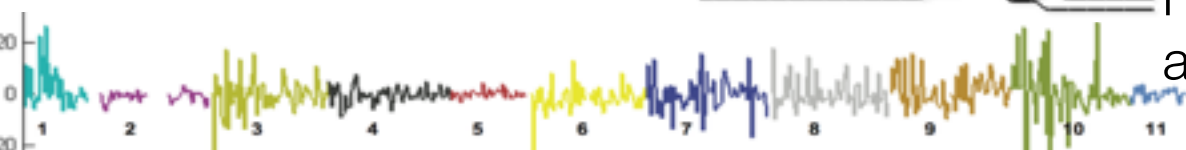
[Ewens 1972; Hartl, Clark 2003; Harris et al 2017]



[ESO/ L. Calçada/ M. Kornmesser 2017]

[Del Pozzo et al 2017, 2018]

[Saria et al 2010]



[Xu et al 2015]

[Cassidy et al 2015]

# References (1/6)

DJ Aldous. *Exchangeability and related topics*. Springer, 1983.

CE Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*, 1974.

E Arjas and D Gasbarra. Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, 1994.

J Bertoin. *Random Fragmentation and Coagulation Processes*. Cambridge University Press, 2006.

D Blackwell and JB MacQueen. Ferguson distributions via Pólya urn schemes. *The Annals of Statistics*, 1973.

T Broderick, MI Jordan, and J Pitman. Beta processes, stick-breaking, and power laws. *Bayesian Analysis*, 2012.

T Broderick, MI Jordan, and J Pitman. Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 2013.

T Broderick, J Pitman, and MI Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 2013.

T Broderick, AC Wilson, and MI Jordan. Posteriors, conjugacy, and exponential families for completely random measures. *Bernoulli*, 2018.

D Cai, T Campbell, and T Broderick. Edge-exchangeable graphs and sparsity. *NIPS*, 2016.

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

T Campbell, D Cai, and Broderick T. Exchangeable trait allocations. Submitted. ArXiv:1609.09147.

- *NIPS 2016 Workshop on Adaptive & Scalable Nonparametric Methods in ML*.
- *NIPS 2016 Workshop on Practical Bayesian Nonparametrics*.

# References (2/6)

T Campbell, D Cai, and T Broderick. Exchangeable trait allocations. arXiv preprint arXiv:1609.09147, 2016.

T Campbell, JH Huggins, JP How, and T Broderick. Truncated Random Measures. *Bernoulli*, to appear, 2018. <https://arxiv.org/abs/1603.00861>

H Crane and W Dempsey. Atypical scaling behavior persists in real world interaction networks. arXiv 1509.08184, 2015.

H Crane and W Dempsey. A framework for statistical network modeling. arXiv 1509.08185, 2015.

H Crane and W Dempsey. Edge exchangeable models for network data. arXiv 1603.04571, 2016.

H Crane and W Dempsey. Relational exchangeability. arXiv 1607.06762, 2016.

W Del Pozzo, TGF Li, and C Messenger. Cosmological inference using only gravitational wave observations of binary neutron stars. *Physical Review D*, 2017.

W Del Pozzo, CPL Berry, A Ghosh, TSF Haines, LP Singer, and A Vecchio. Dirichlet Process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations. *Monthly Notices of the Royal Astronomical Society*, 2018.

S Engen. A note on the geometric series as a species frequency model. *Biometrika*, 1975.

MD Escobar and M West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 1995.

W Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 1972.

W Ewens. Population genetics theory -- the past and the future. *Mathematical and Statistical Developments of Evolutionary Theory*, 1987.

TS Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1973.

TS Ferguson. Bayesian density estimation by mixtures of normal distributions. *Recent Advances in Statistics*, 1983.



# References (3/6)

EB Fox, personal website. Retrieved in 2016 from: <http://www.stat.washington.edu/~ebfox/research.html> --- Associated paper: EB Fox, MC Hughes, EB Sudderth, and MI Jordan. *The Annals of Applied Statistics*, 2014.

S Ghosal, JK Ghosh, and RV Ramamoorthi. Posterior consistency of Dirichlet mixtures in density estimation. *The Annals of Statistics*, 1999.

S Goldwater, TL Griffiths, and M Johnson. Interpolating between types and tokens by estimating power-law generators. *NIPS*, 2005.

A Gnedin, B Hansen, and J Pitman. Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probability Surveys*, 2007.

TL Griffiths and Z Ghahramani. Infinite latent feature models and the Indian buffet process. *NIPS*, 2005.

K Harris, TL Parsons, UZ Ijaz, L Lahti, I Holmes, and C Quince. Linking statistical and ecological theory: Hubbell's unified neutral theory of biodiversity as a hierarchical Dirichlet process. *Proceedings of the IEEE*, 2017.

DL Hartl and AG Clark. *Principles of Population Genetics, Fourth Edition*. 2003.

E Hewitt and LJ Savage. Symmetric measures on Cartesian products. *Transactions of the American Mathematical Society*, 1955.

NL Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *The Annals of Statistics*, 1990.

DN Hoover. Relations on probability spaces and arrays of random variables, *Preprint, Institute for Advanced Study*, 1979.

FM Hoppe. Pólya-like urns and the Ewens' sampling formula. *Journal of Mathematical Biology*, 1984.

H Ishwaran and LF James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 2001.

L James. Poisson latent feature calculus for generalized Indian buffet processes. arXiv preprint arXiv:1411.2936, 2014.

Y Kim. Nonparametric Bayesian estimators for counting processes. *The Annals of Statistics*, 1999.

# References (4/6)

JFC Kingman. The representation of partition structures. *Journal of the London Mathematical Society*, 1978.

JFC Kingman. On the genealogy of large populations. *Journal of Applied Probability*, 1982.

JFC Kingman. *Poisson processes*, 1992.

AS Lan, D Vats, AE Waters, and RG Baraniuk. Mathematical language processing: Automatic grading and feedback for open response mathematical questions. In *Proceedings of the Second (2015) ACM Conference on Learning@Scale*, 2015.

JR Lloyd, P Orbanz, Z Ghahramani, and DM Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. *NIPS*, 2012.

SN MacEachern and P Müller. Estimating mixtures of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 1998.

JW McCloskey. A model for the distribution of individuals by species in an environment. *Ph.D. thesis, Michigan State University*, 1965.

K Miller, MI Jordan, and TL Griffiths. Nonparametric latent feature models for link prediction. *NIPS*, 2009.

RM Neal. Density modeling and clustering using Dirichlet diffusion trees. *Bayesian Statistics*, 2003.

P Orbanz. Construction of nonparametric Bayesian models from parametric Bayes equations. *NIPS*, 2009.

P Orbanz. Conjugate Projective Limits. arXiv preprint arXiv:1012.0363, 2010.

P Orbanz, DM Roy. Bayesian models of graphs, arrays and other exchangeable random structures. *IEEE TPAMI*, 2015.

GP Patil and C Taillie. Diversity as a concept and its implications for random communities. *Bulletin of the International Statistical Institute*, 1977.

J Pitman and M Yor. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 1997.

# References (5/6)

- S Prabhakaran, E Azizi, A Carr, and D Pe'er. Dirichlet process mixture model for correcting technical variation in single-cell gene expression data. *ICML*, 2016.
- A Rodríguez, DB Dunson & AE Gelfand. The nested Dirichlet process. *Journal of the American Statistical Association*, 2008.
- S Saria, D Koller, and A Penn. Learning individual and population traits from clinical temporal data. *NIPS*, 2010.
- J Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 1994.
- EB Sudderth and MI Jordan. Shared segmentation of natural scenes using dependent Pitman-Yor processes. *NIPS*, 2009.
- YW Teh. A hierarchical Bayesian language model based on Pitman-Yor processes. *ACL*, 2006.
- YW Teh, C Blundell, and L Elliott. Modelling genetic variations using fragmentation-coagulation processes. *NIPS*, 2011.
- YW Teh, MI Jordan, MJ Beal, and DM Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 2006.
- R Thibaux and MI Jordan. Hierarchical beta processes and the Indian buffet process. *ICML*, 2007.
- J Wakeley. *Coalescent Theory: An Introduction*, Chapter 3, 2008.
- M West, P Müller, and MD Escobar. Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation. *Aspects of Uncertainty*, 1994.
- Y Xu, P Müller, Y Yuan, K Gulukota, and Y Ji. MAD Bayes for tumor heterogeneity—feature allocation with exponential family sampling. *Journal of the American Statistical Association*, 2015.

# Image References (6/6)

E Bowlby. NOAA/Olympic Coast NMS; NOAA/OAR/Office of Ocean Exploration - NOAA Photo Library. Retrieved in 2016 from: [https://en.wikipedia.org/wiki/Opisthoteuthis\\_californiana#/media/File:Opisthoteuthis\\_californiana.jpg](https://en.wikipedia.org/wiki/Opisthoteuthis_californiana#/media/File:Opisthoteuthis_californiana.jpg)

JW Cassidy, C Caldas, and A Bruna. Maintaining tumor heterogeneity in patient-derived tumor xenografts. *Cancer research*, 2015.

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](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)

MIT xPro, 2017. Obtained in 2018 from: [https://mitxpro.mit.edu/courses/course-v1:MITProfessionalX+DSx+2017\\_T2/about](https://mitxpro.mit.edu/courses/course-v1:MITProfessionalX+DSx+2017_T2/about)