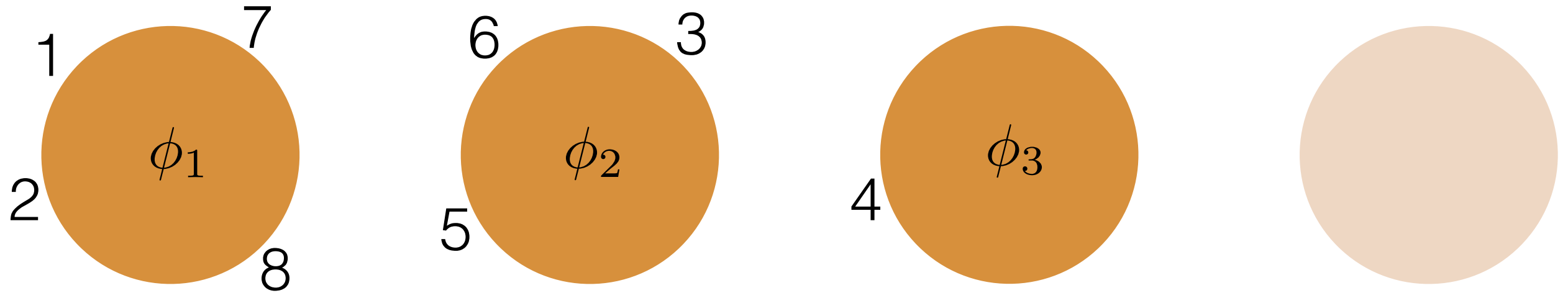


Nonparametric Bayesian Statistics: Part III

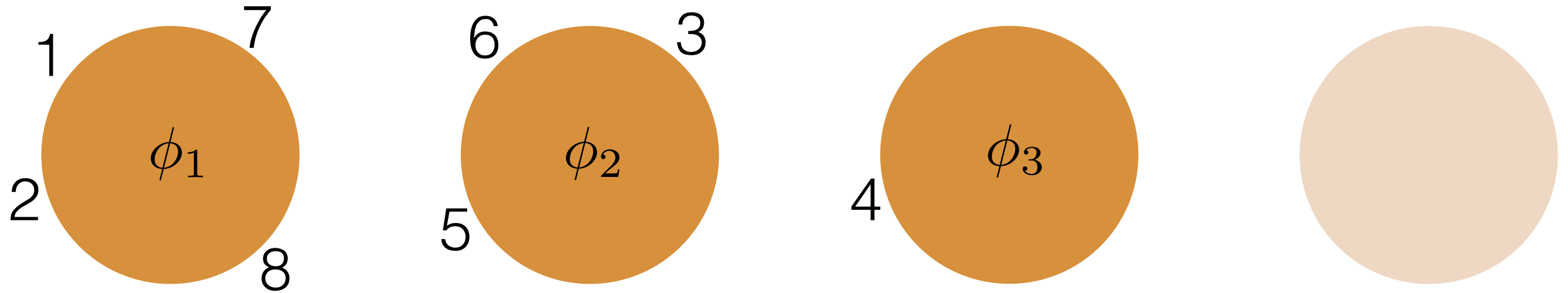
Tamara Broderick

ITT Career Development Assistant Professor
Electrical Engineering & Computer Science
MIT

Chinese restaurant process

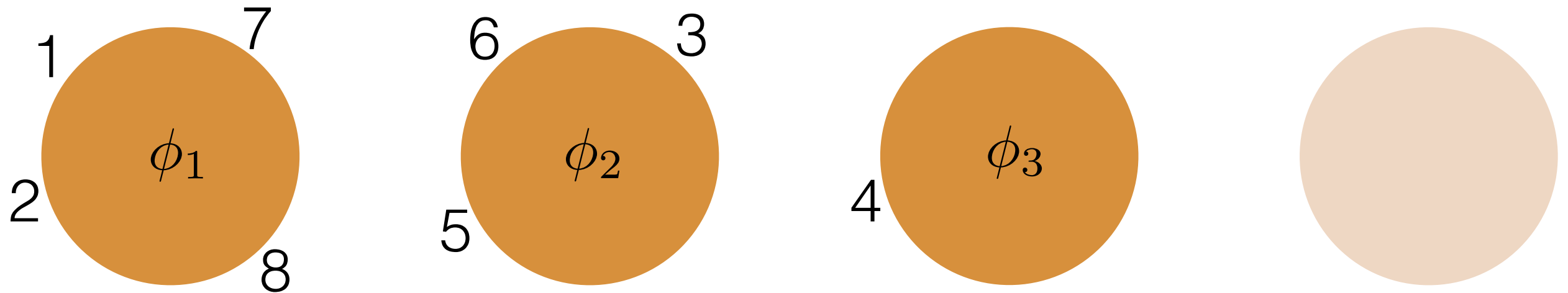


Chinese restaurant process



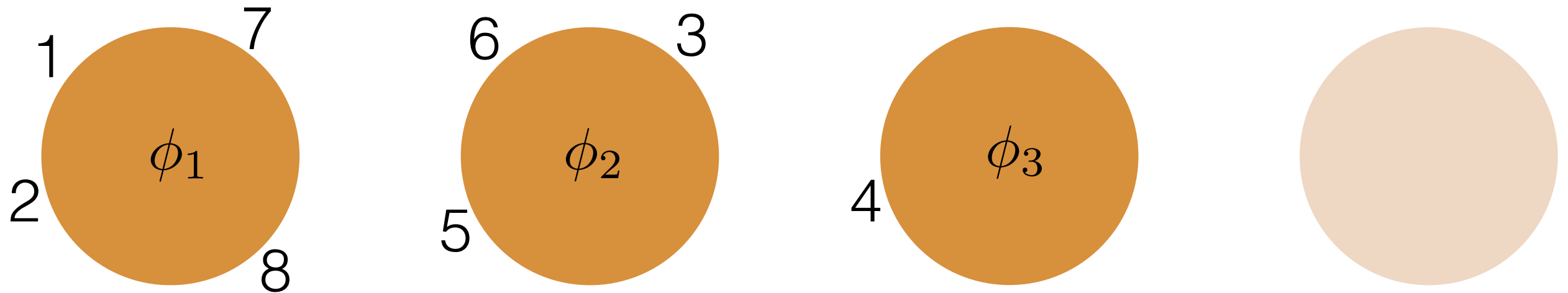
- Each customer walks into the restaurant

Chinese restaurant process



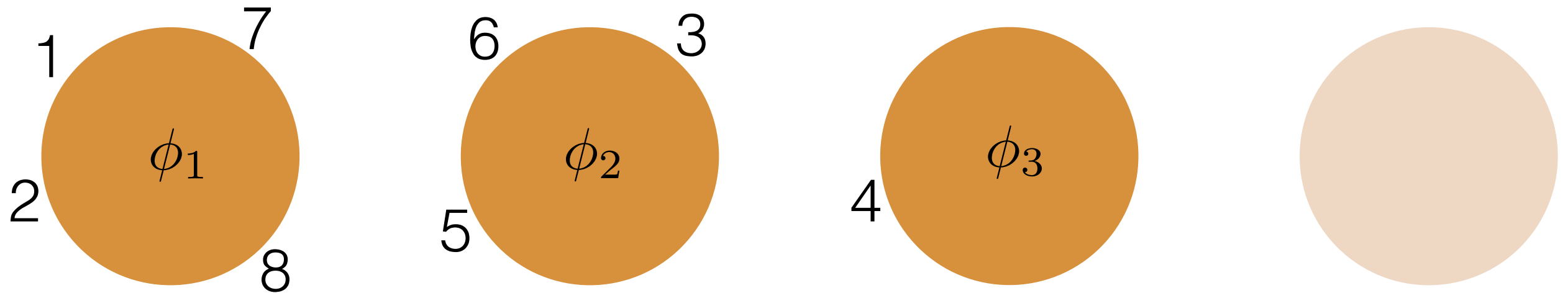
- Each customer walks into the restaurant
 - Sits at existing table with prob proportional to # people there

Chinese restaurant process



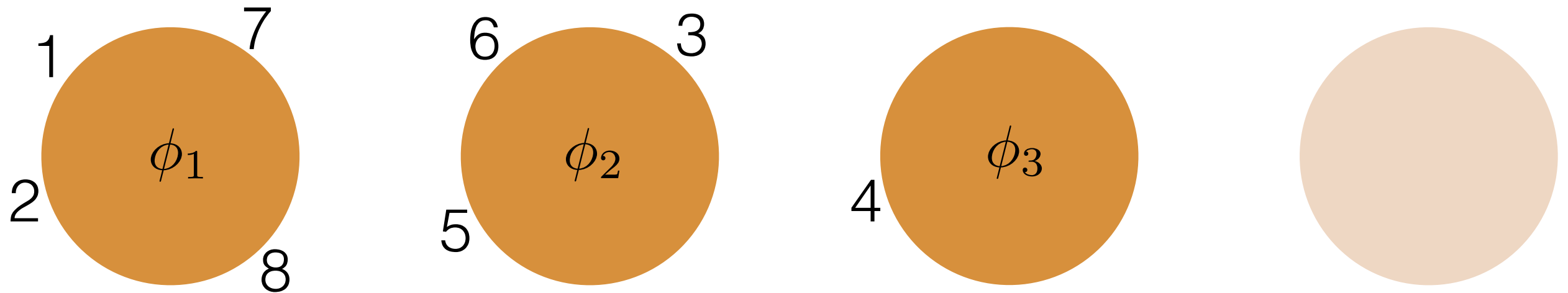
- Each customer walks into the restaurant
 - Sits at existing table with prob proportional to # people there
 - Forms new table with prob proportional to α

Chinese restaurant process



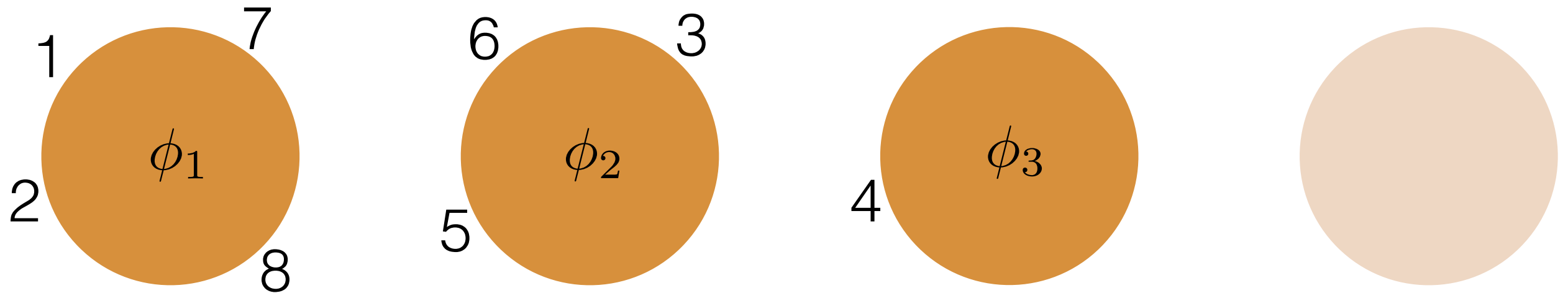
- Each customer walks into the restaurant
 - Sits at existing table with prob proportional to # people there
 - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior

Chinese restaurant process



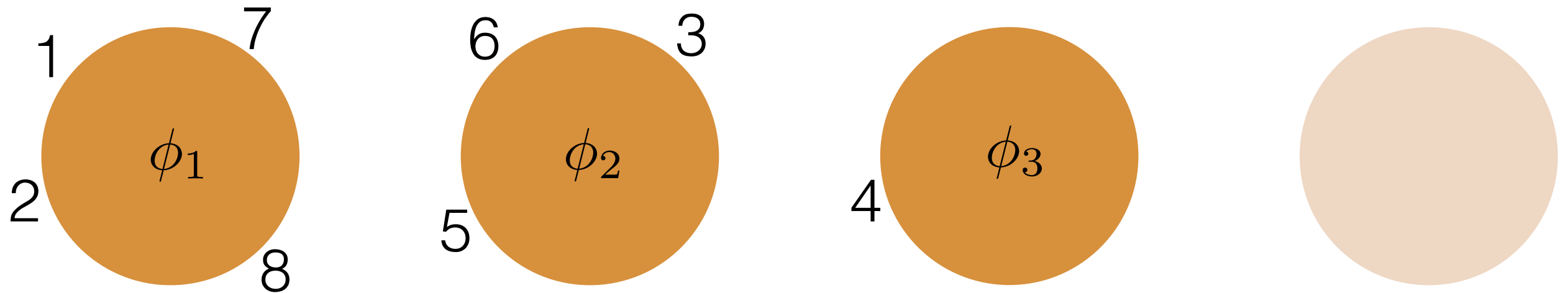
- Each customer walks into the restaurant
 - Sits at existing table with prob proportional to # people there
 - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior
 $z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$

Chinese restaurant process



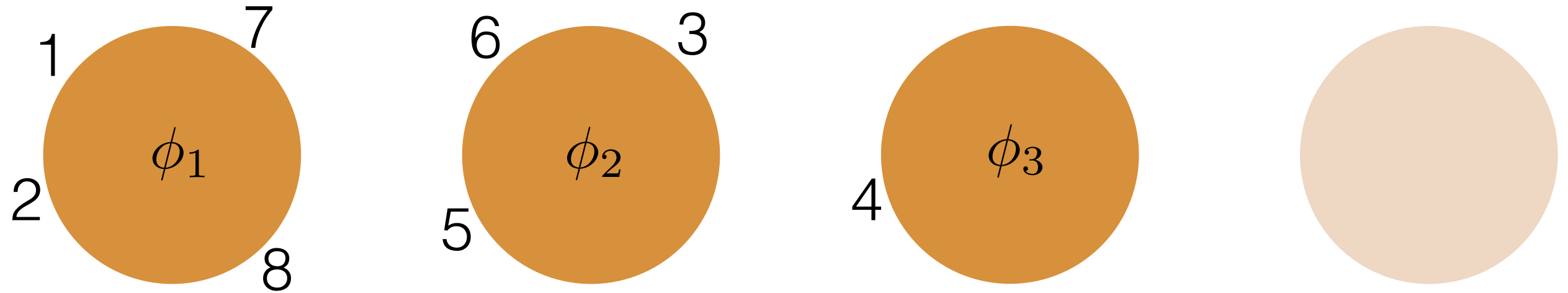
- Same thing we just did
- Each customer walks into the restaurant
 - Sits at existing table with prob proportional to # people there
 - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior
$$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$$
$$\Rightarrow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

Chinese restaurant process



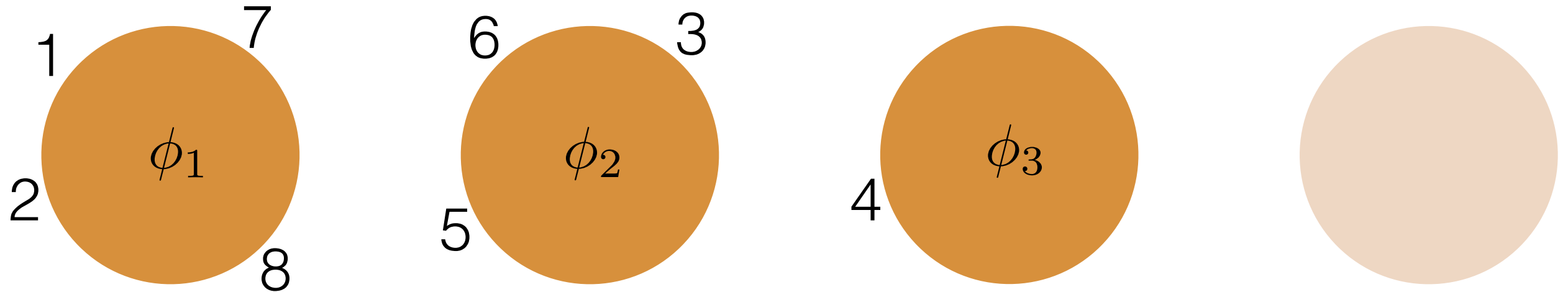
- Same thing we just did
- Each customer walks into the restaurant
 - Sits at existing table with prob proportional to # people there
 - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior
$$z_1 = z_2 = z_7 = z_8 = 1, z_3 = z_5 = z_6 = 2, z_4 = 3$$
$$\Rightarrow \Pi_8 = \{\{1, 2, 7, 8\}, \{3, 5, 6\}, \{4\}\}$$
- *Partition of [8]*: set of mutually exclusive & exhaustive sets of $[8] := \{1, \dots, 8\}$

Chinese restaurant process



- Probability of this seating:

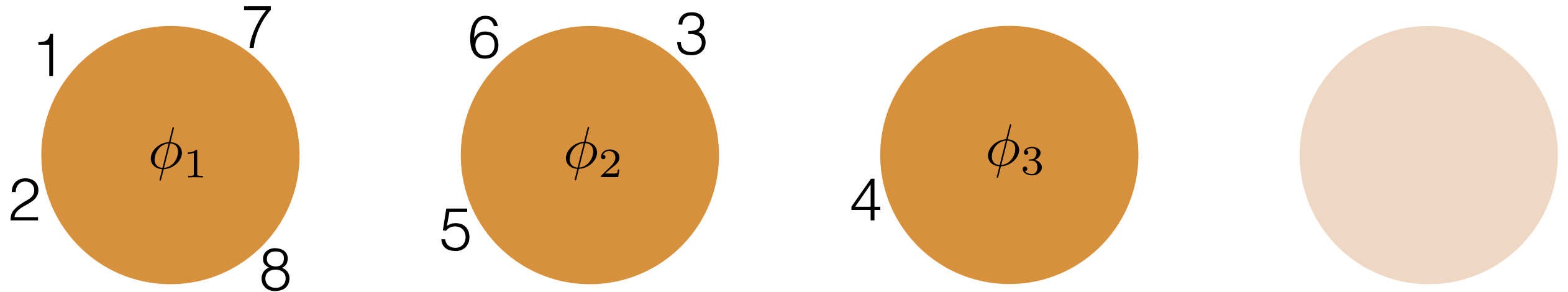
Chinese restaurant process



- Probability of this seating:

$$\frac{\alpha}{\alpha}$$

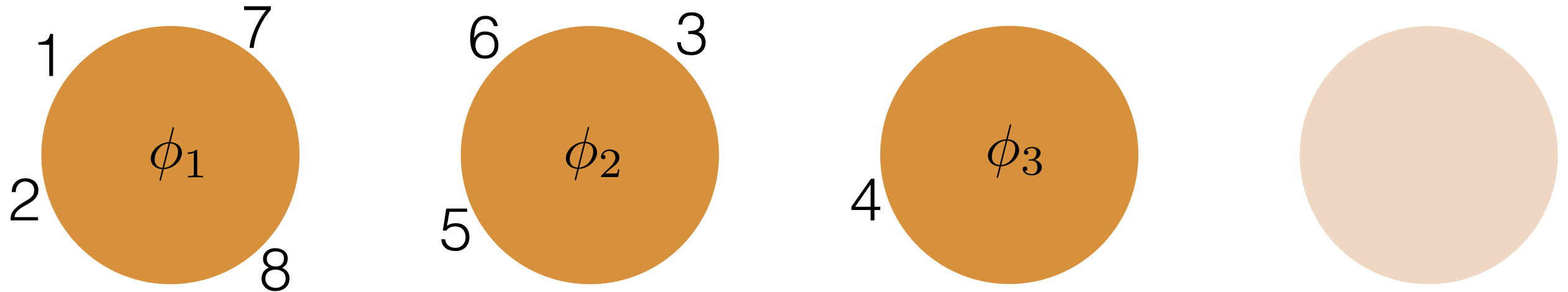
Chinese restaurant process



- Probability of this seating:

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

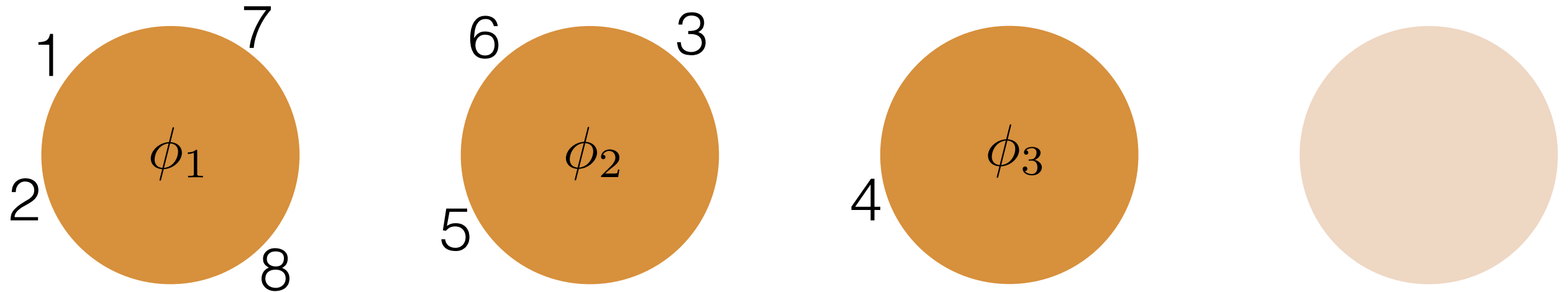
Chinese restaurant process



- Probability of this seating:

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

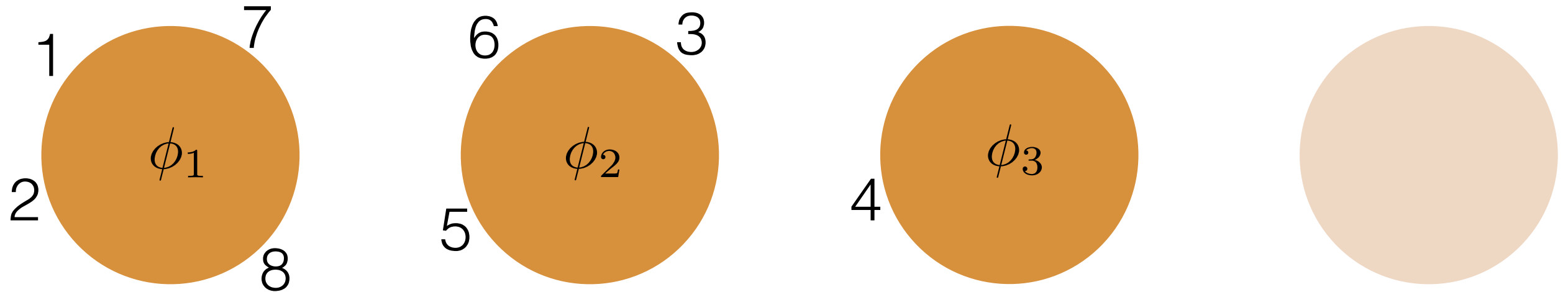
Chinese restaurant process



- Probability of this seating:

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

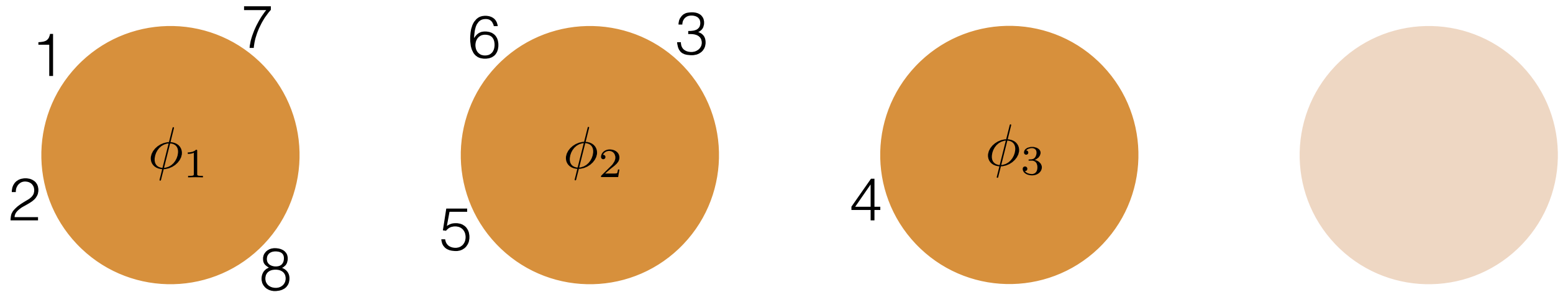
Chinese restaurant process



- Probability of this seating:

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

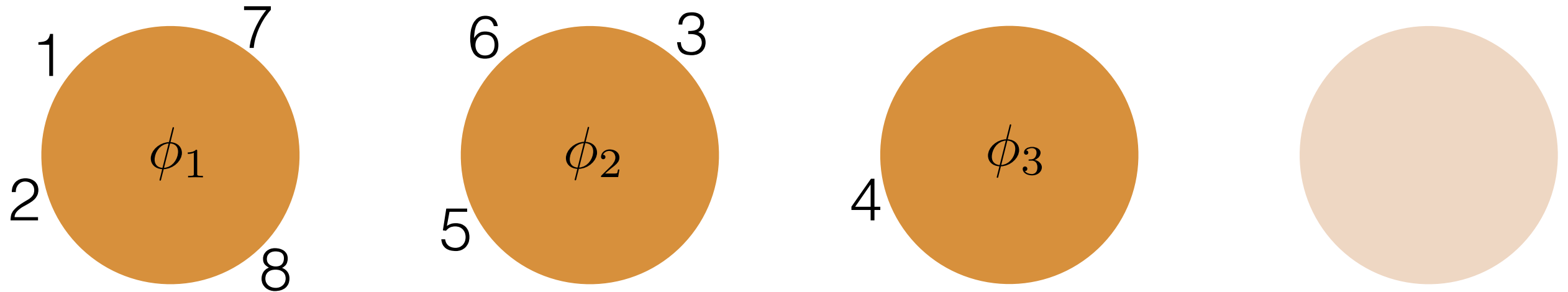
Chinese restaurant process



- Probability of this seating:

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

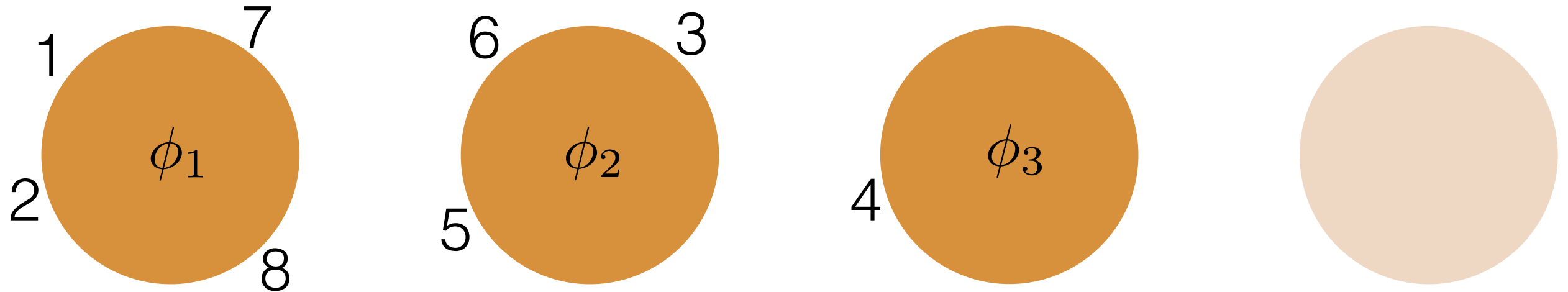
Chinese restaurant process



- Probability of this seating:

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

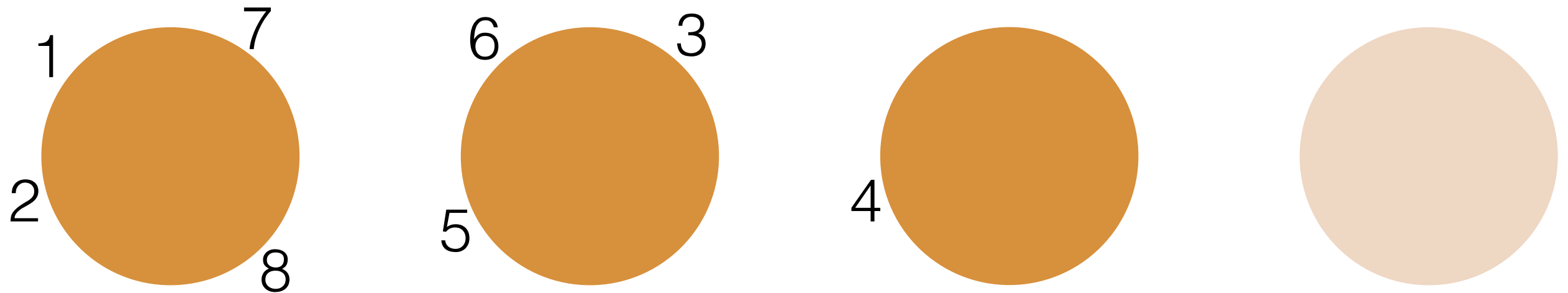
Chinese restaurant process



- 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

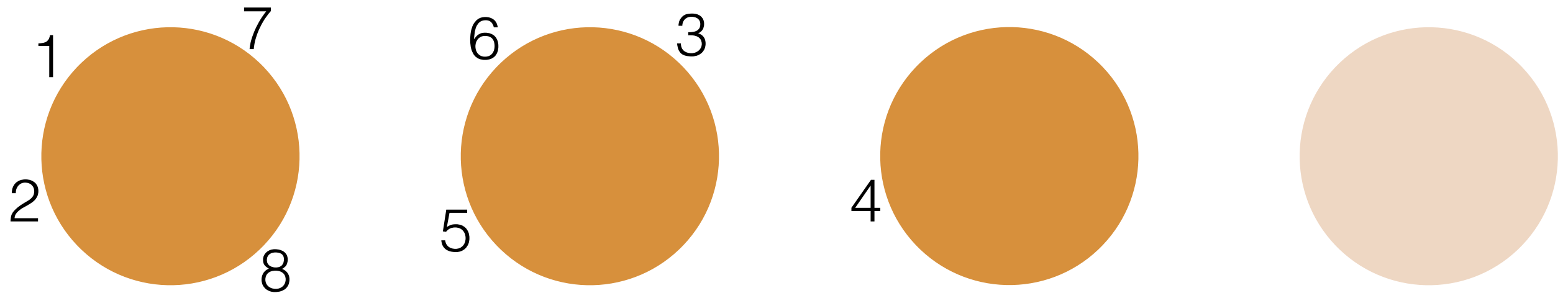


- Probability of this seating:

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

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

Chinese restaurant process

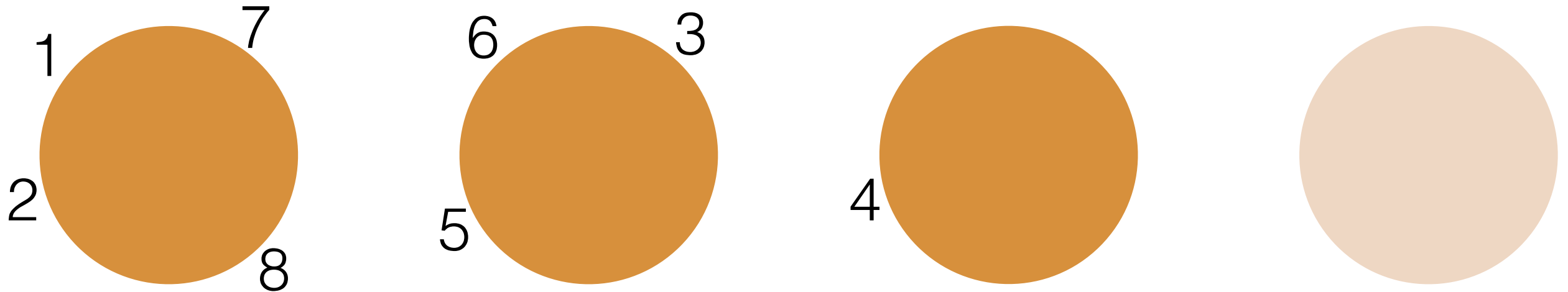


- Probability of this seating:

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

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

Chinese restaurant process



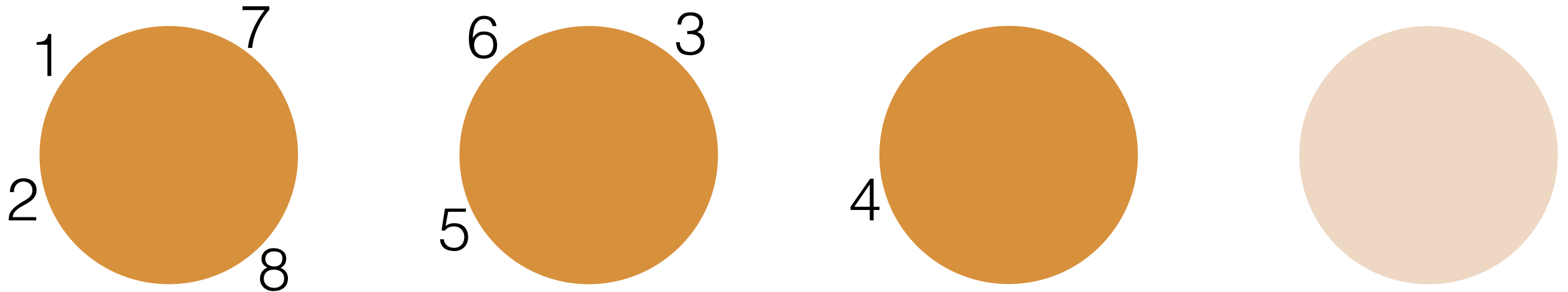
- Probability of this seating:

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

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

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

Chinese restaurant process



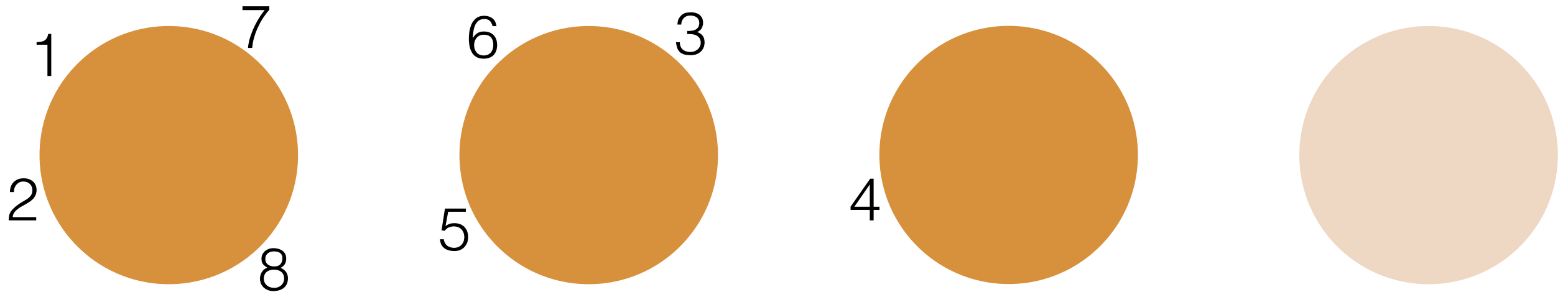
- Probability of this seating:

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

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

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

Chinese restaurant process



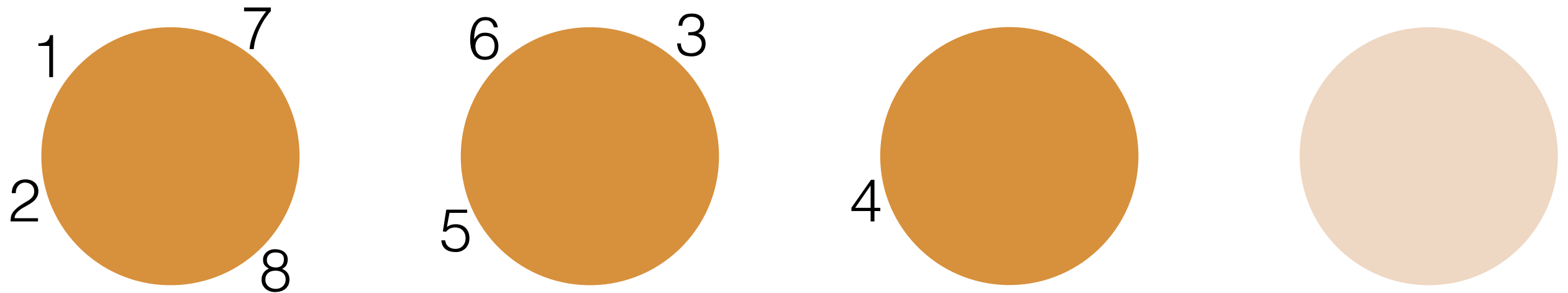
- Probability of this seating:

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

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

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

Chinese restaurant process



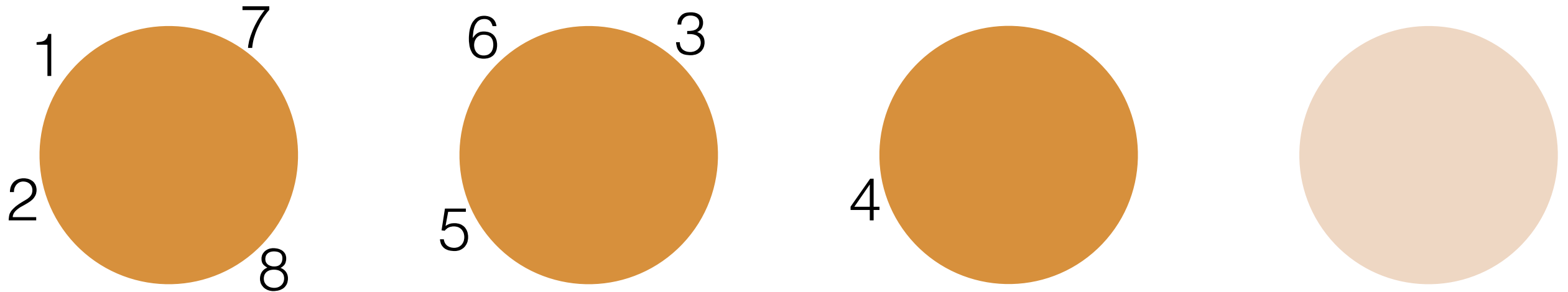
- Probability of this seating:

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

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

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

Chinese restaurant process



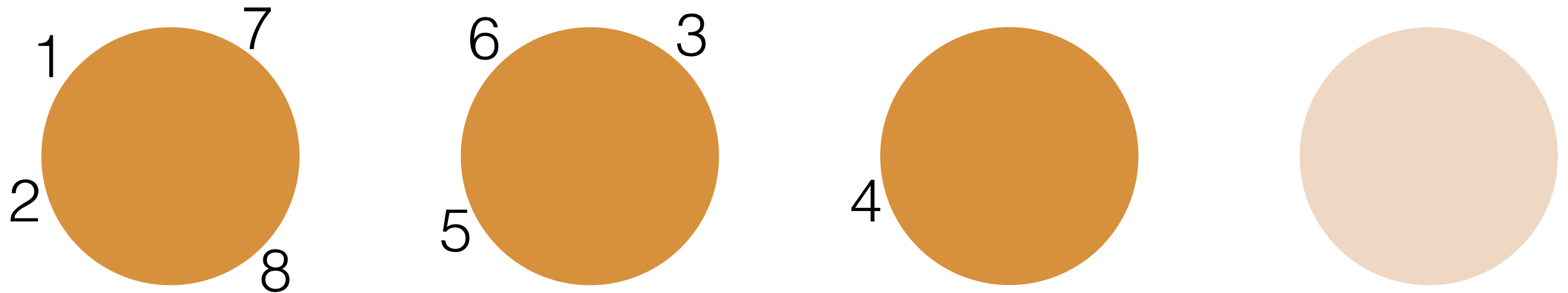
- Probability of this seating:

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

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

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

Chinese restaurant process



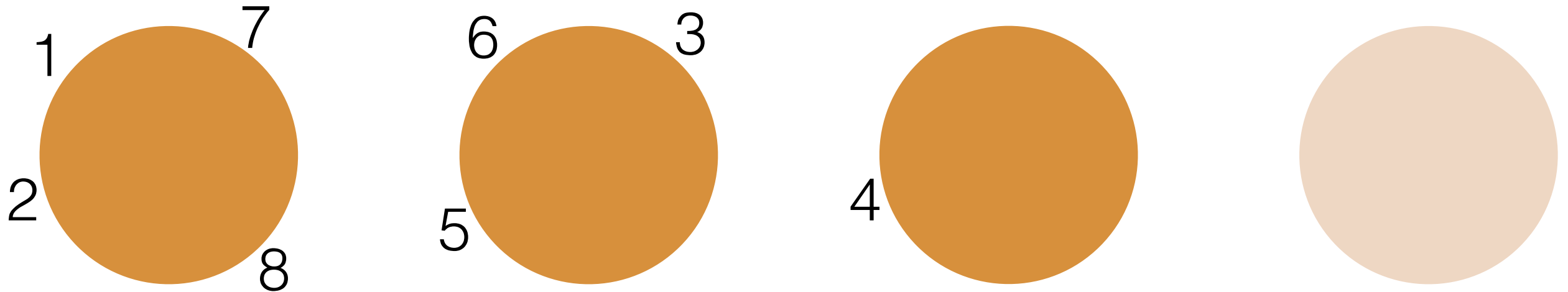
- Probability of this seating:

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

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

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

Chinese restaurant process



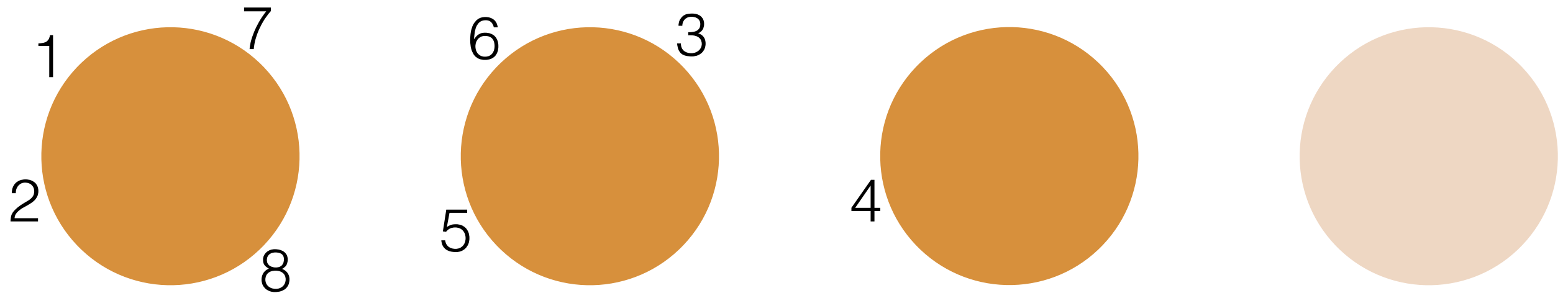
- Probability of this seating:

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

- 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



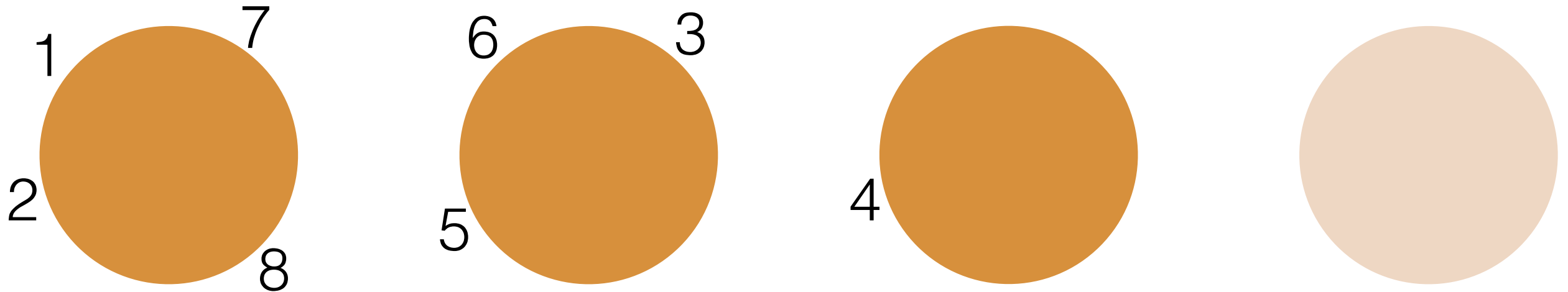
- Probability of this seating:

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

- 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



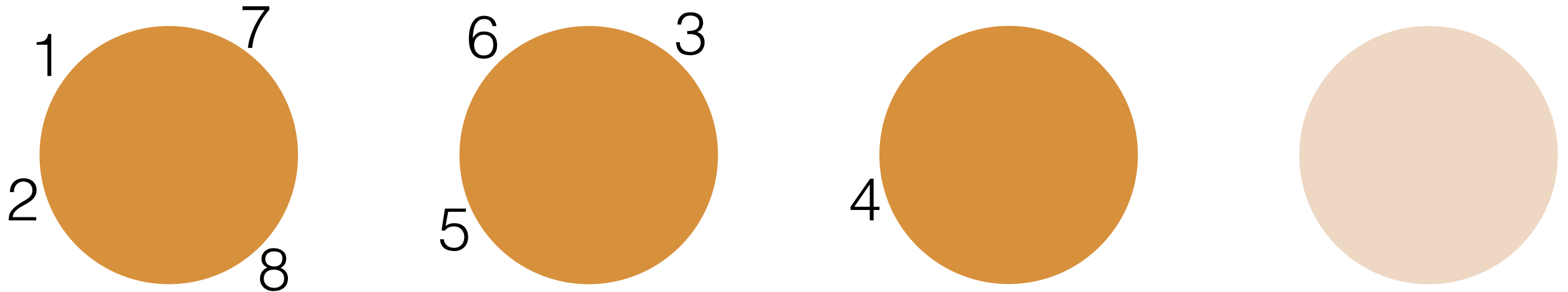
- Probability of this seating:

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

- 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:

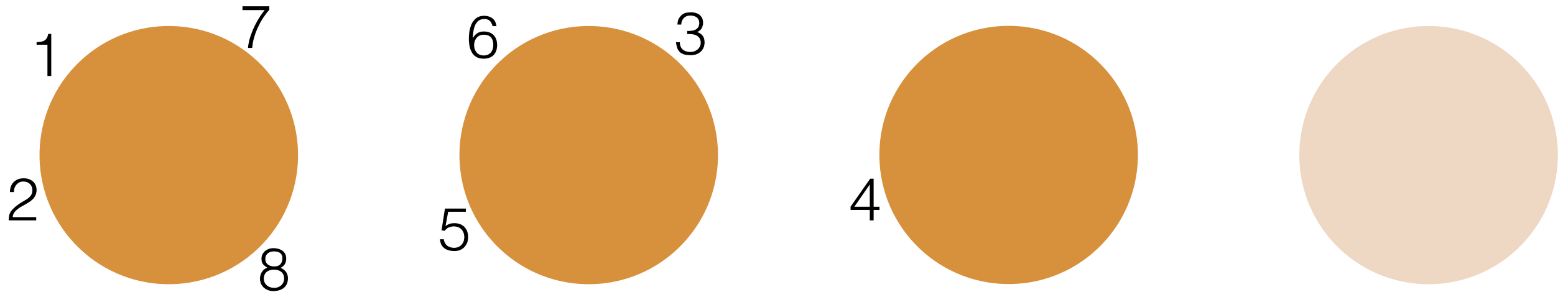
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}$$

- 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:

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

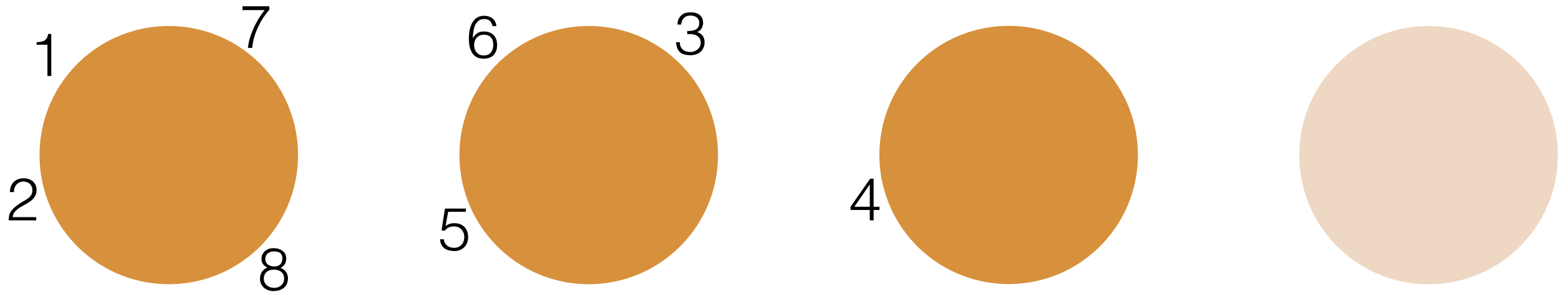
- Probability of N customers (K_N tables, $\#C$ at table C):

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

- 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:

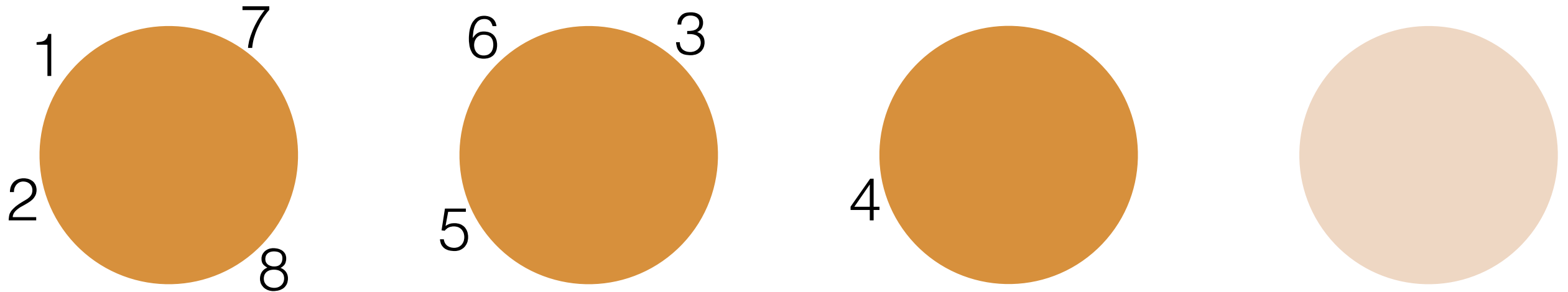
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha + 1} \cdot \frac{\alpha}{\alpha + 2} \cdot \frac{\alpha}{\alpha + 3} \cdot \frac{1}{\alpha + 4} \cdot \frac{2}{\alpha + 5} \cdot \frac{2}{\alpha + 6} \cdot \frac{3}{\alpha + 7}$$
- Probability of N customers (K_N tables, $\#C$ at table C):

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

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

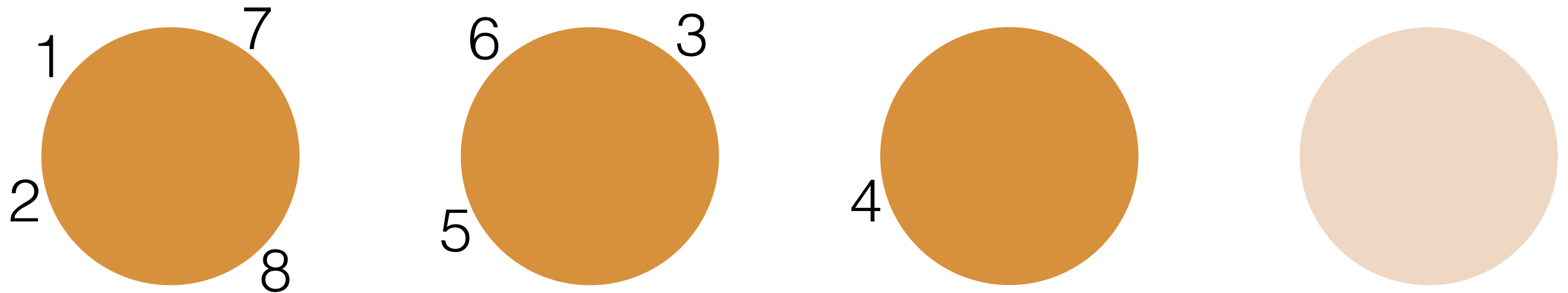
- Prob doesn't depend on customer order: *exchangeable*

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

- 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

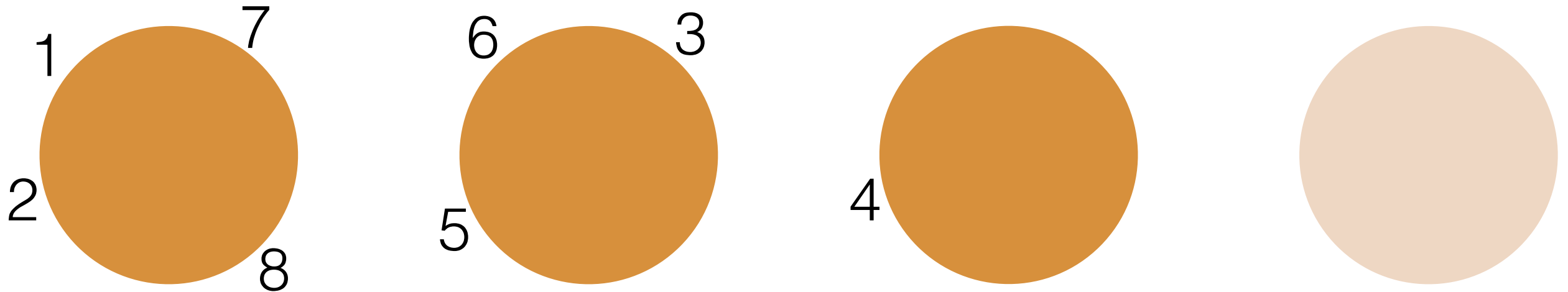


- Probability of N customers (K_N tables, $\#C$ at table C):

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

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

Chinese restaurant process

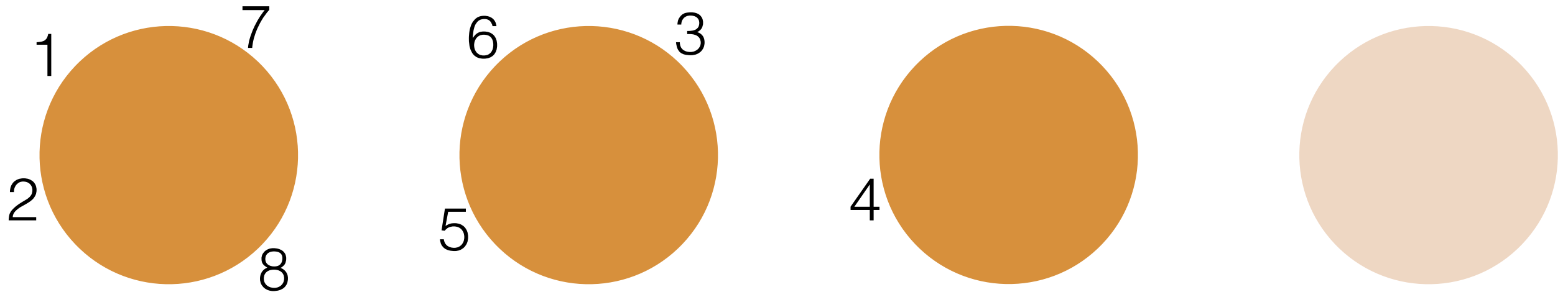


- Probability of N customers (K_N tables, $\#C$ at table C):

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

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

Chinese restaurant process

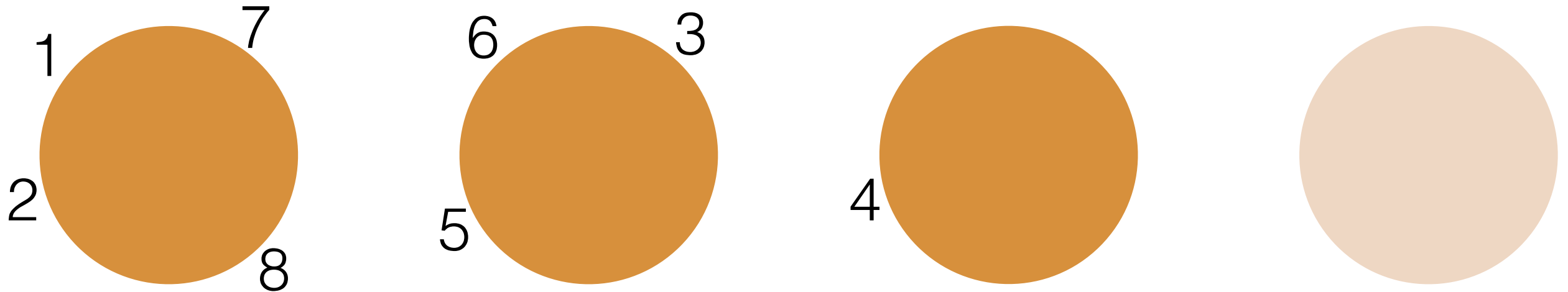


- Probability of N customers (K_N tables, $\#C$ at table C):

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

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

Chinese restaurant process

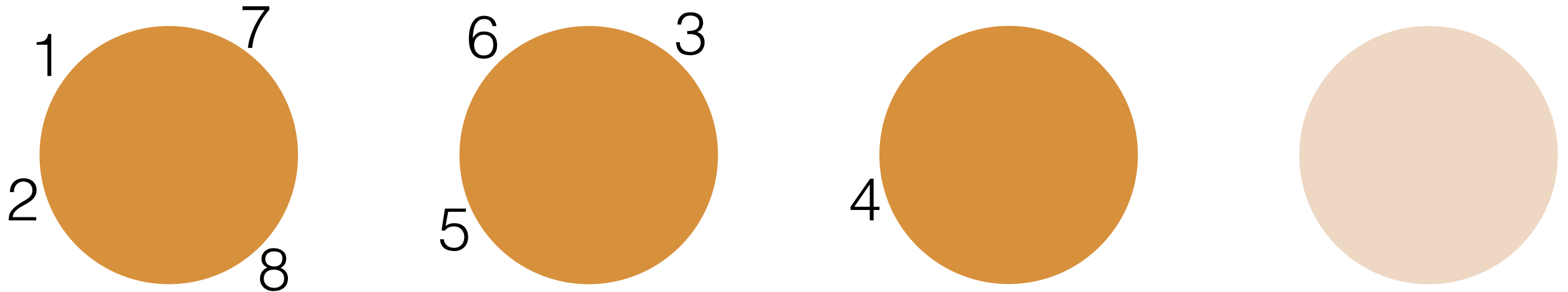


- Probability of N customers (K_N tables, $\#C$ at table C):

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

- 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

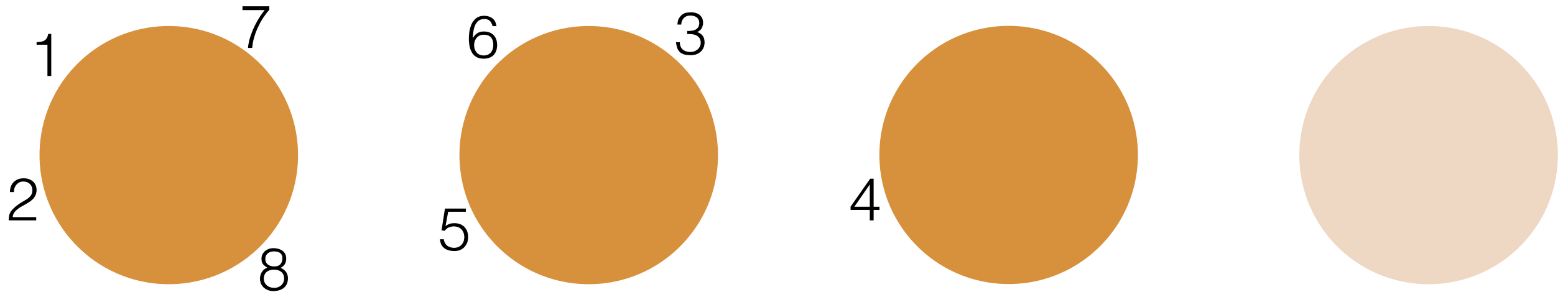


- Probability of N customers (K_N tables, $\#C$ at table C):

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

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

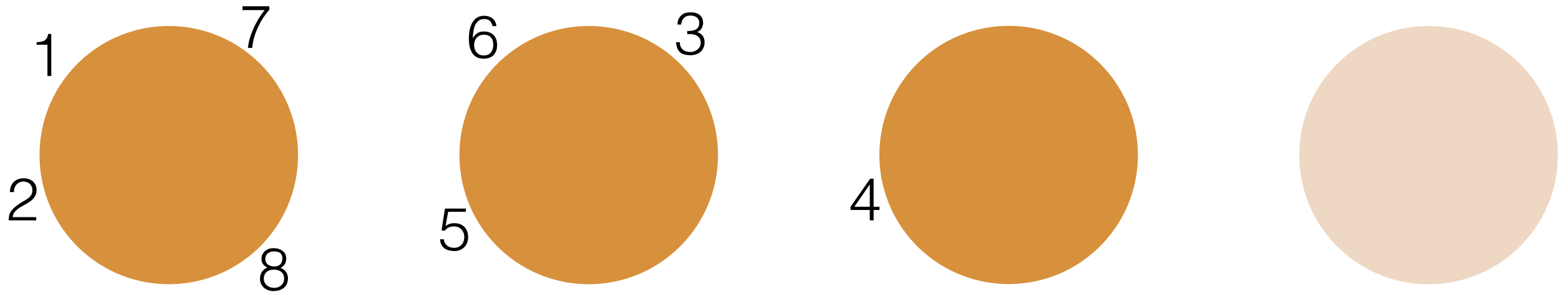
Chinese restaurant process



- 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:

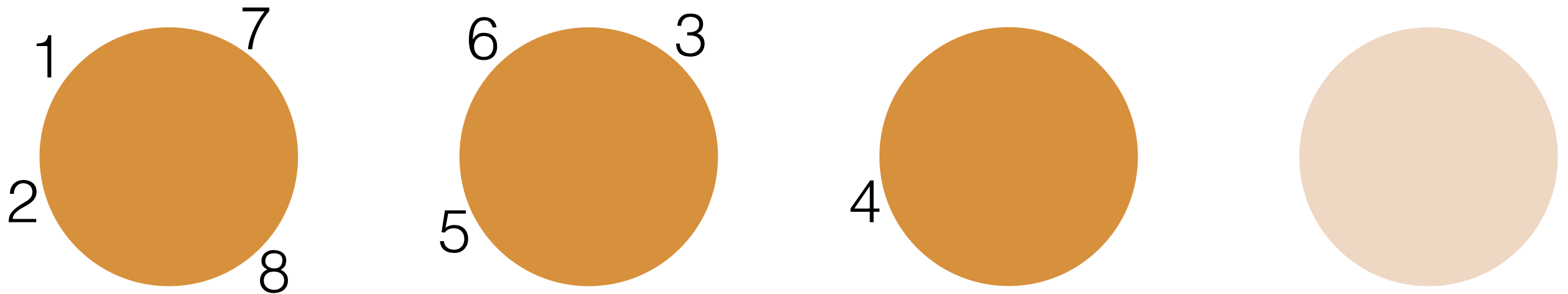
Chinese restaurant process



- 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 $p(v_1, v_2, v_3)$

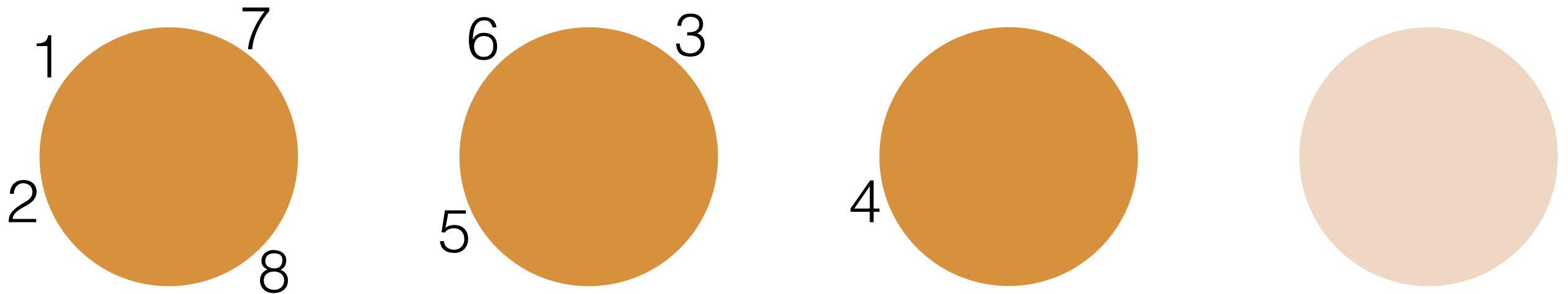
Chinese restaurant process



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

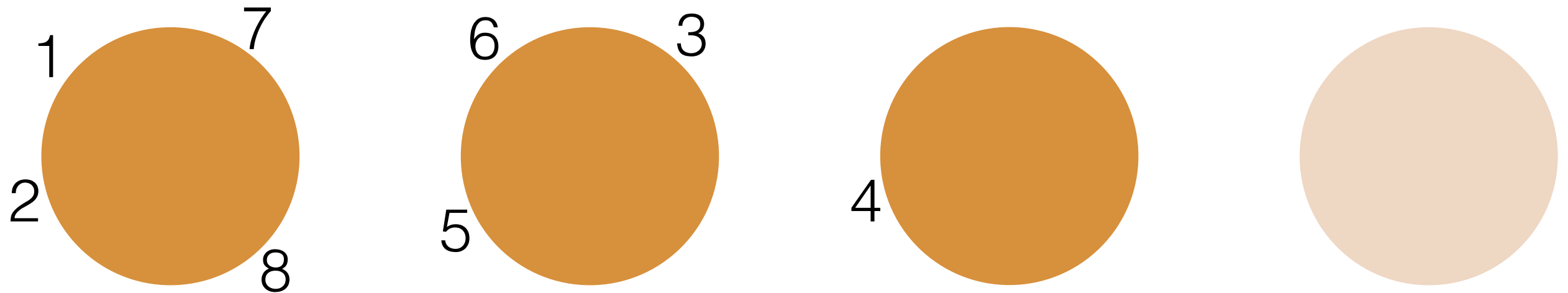
Chinese restaurant process



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

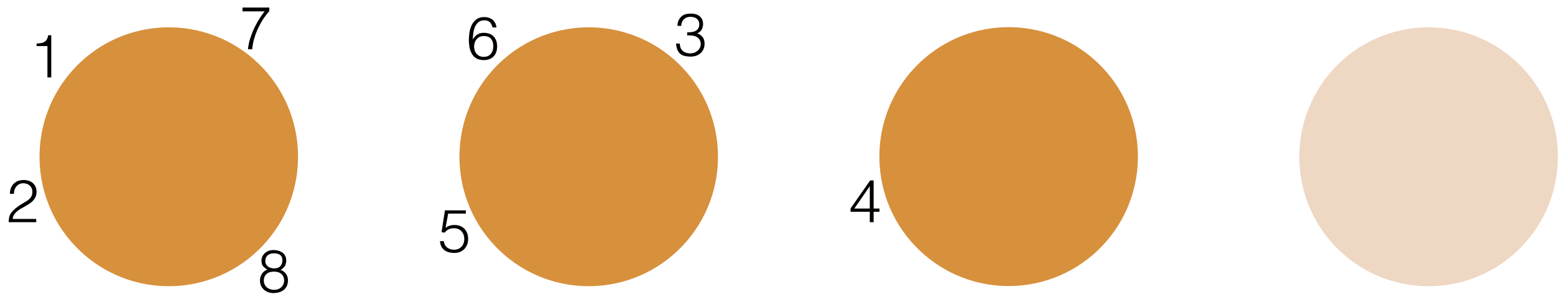
Chinese restaurant process



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

Chinese restaurant process



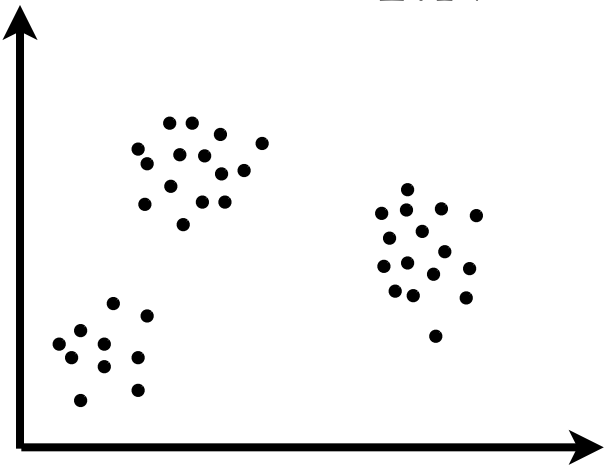
- Probability of N customers (K_N tables, $\#C$ at table C):

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

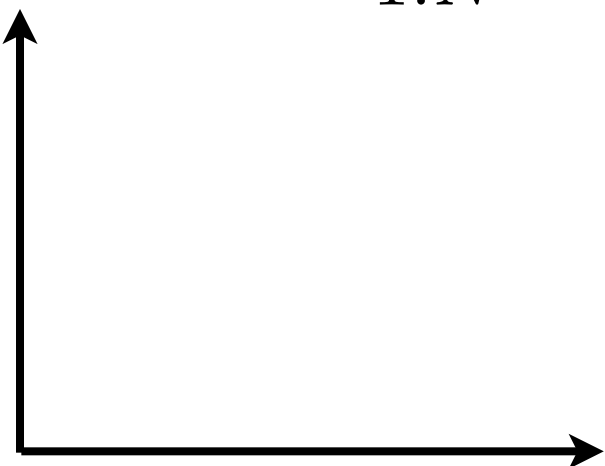
CRP mixture model: inference

- Data $x_{1:N}$



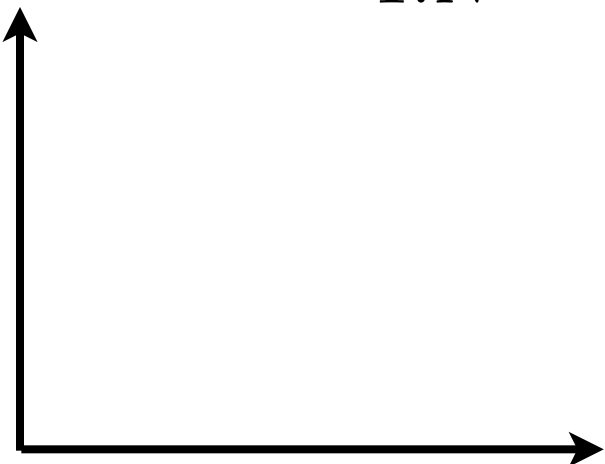
CRP mixture model: inference

- Data $x_{1:N}$



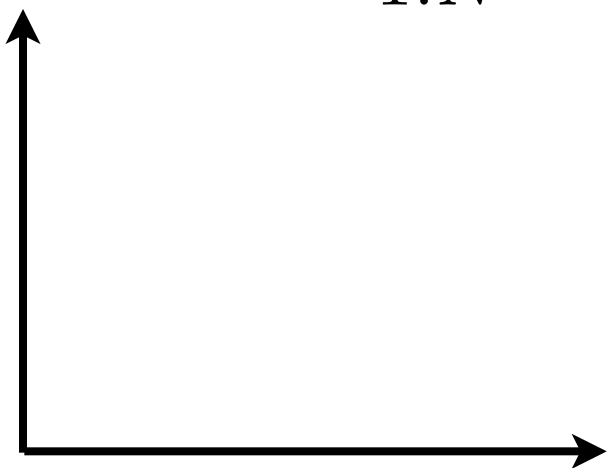
CRP mixture model: inference

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



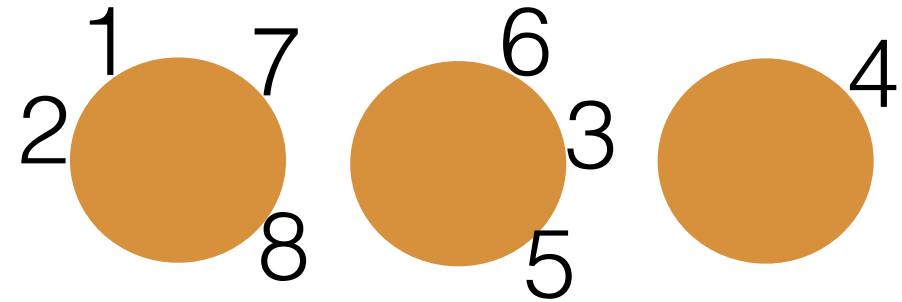
CRP mixture model: inference

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



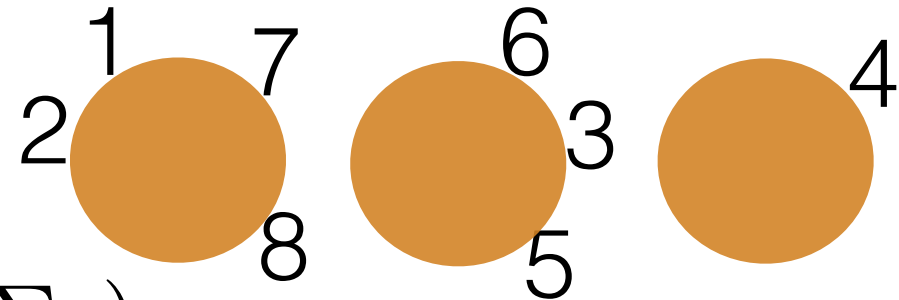
CRP mixture model: inference

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



CRP mixture model: inference

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



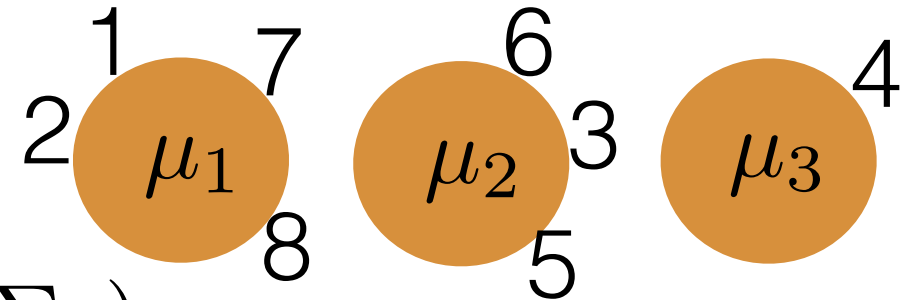
CRP mixture model: inference

- Data $x_{1:N}$

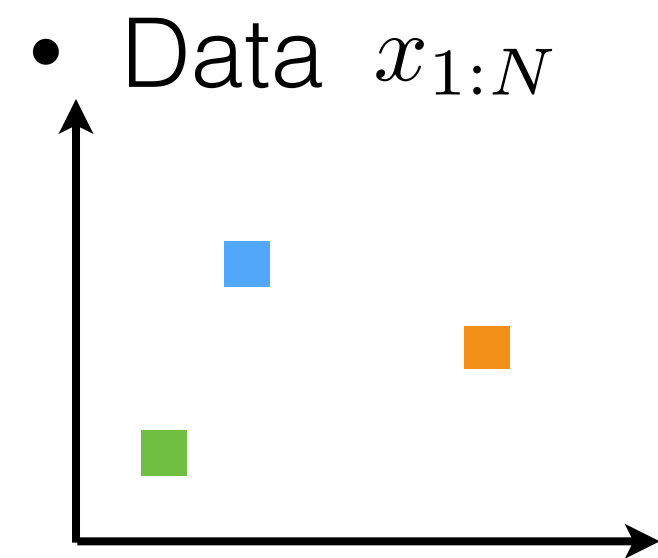
- Generative model

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

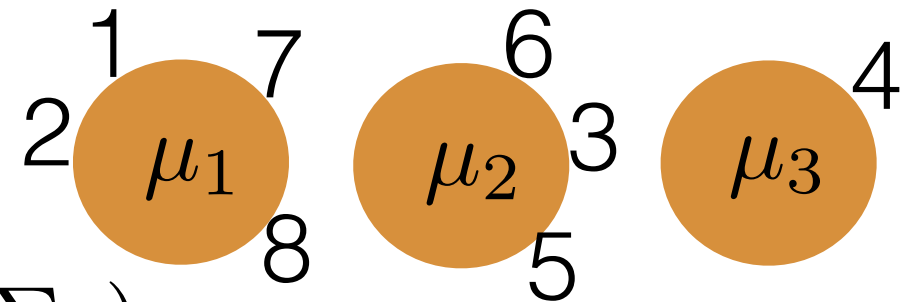
$$\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$



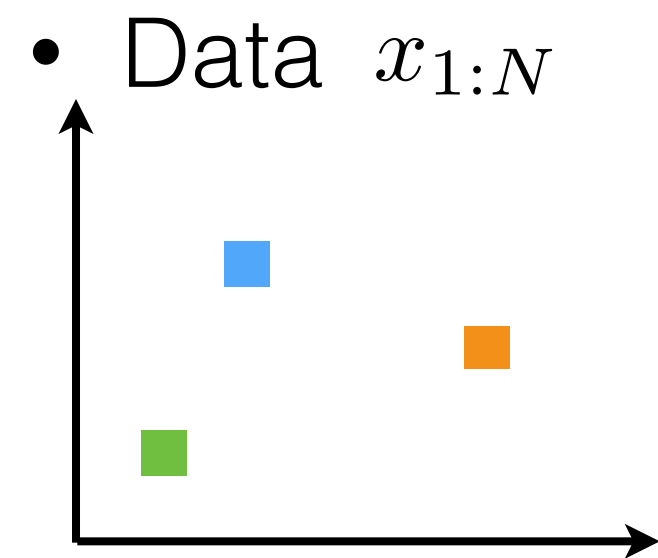
CRP mixture model: inference



- 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

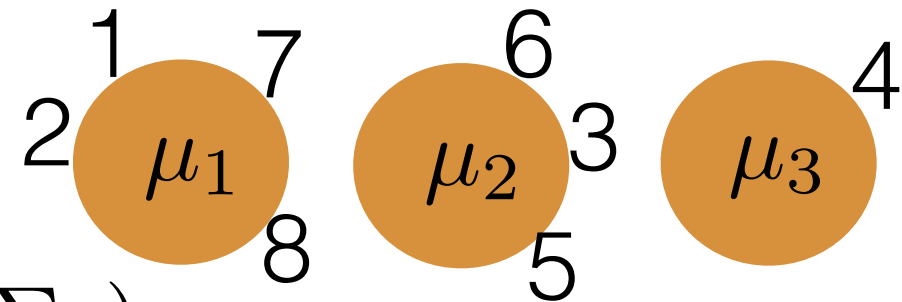


- Generative model

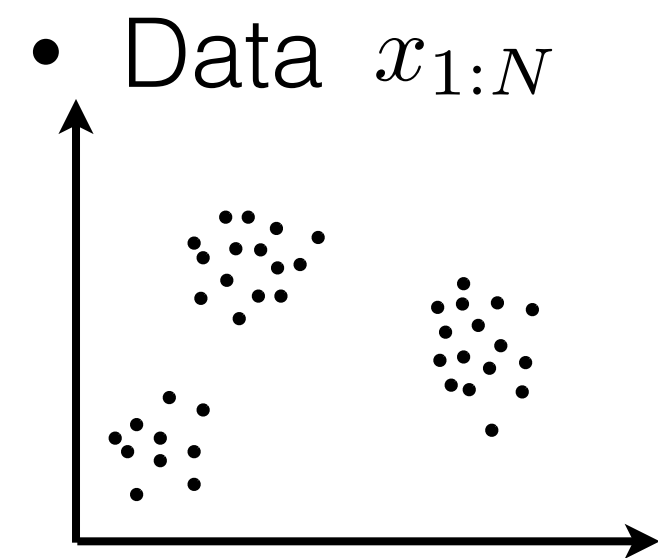
$$\Pi_N \sim \text{CRP}(N, \alpha)$$

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

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



CRP mixture model: inference

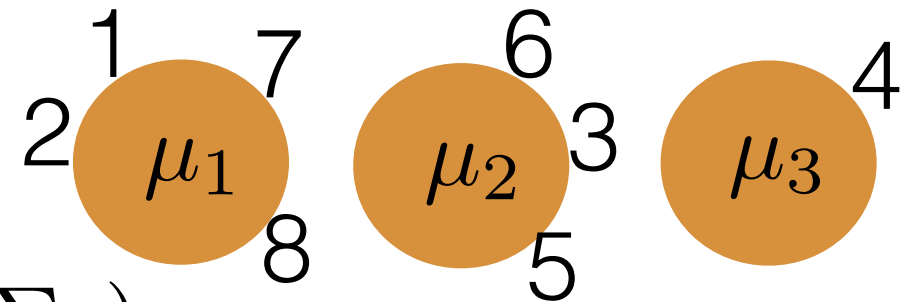


- Generative model

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

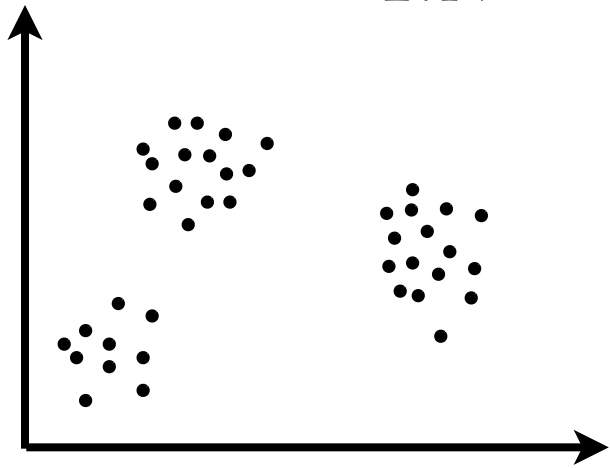
$$\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

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



CRP mixture model: inference

- Data $x_{1:N}$



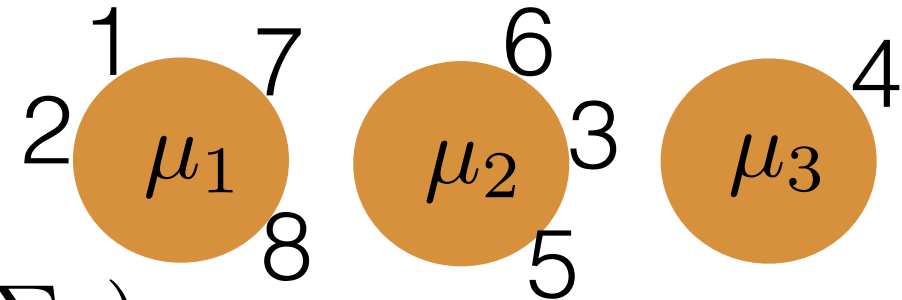
- Want: posterior

- Generative model

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

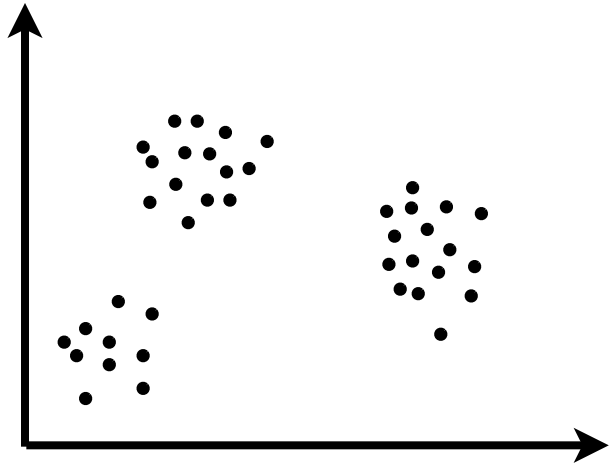
$$\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

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



CRP mixture model: inference

- Data $x_{1:N}$

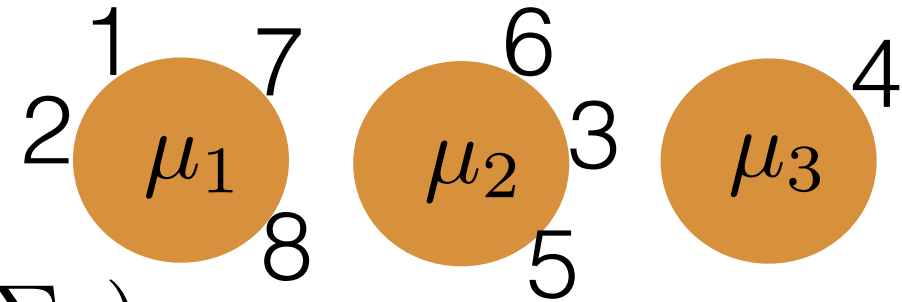


- Generative model

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

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

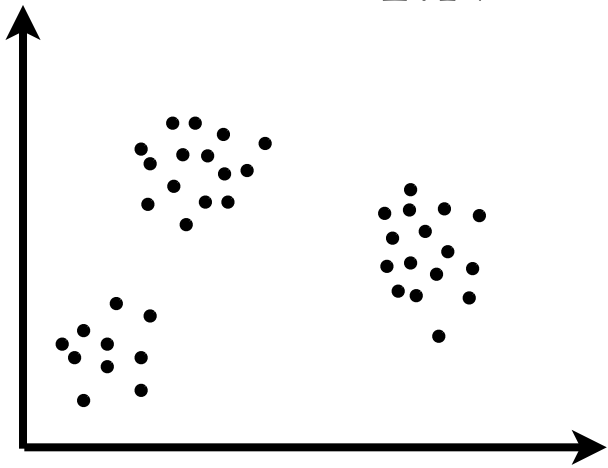
$$\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



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

CRP mixture model: inference

- Data $x_{1:N}$

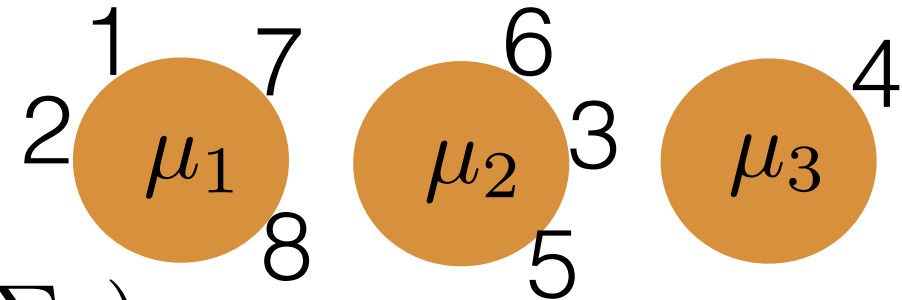


- Generative model

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

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

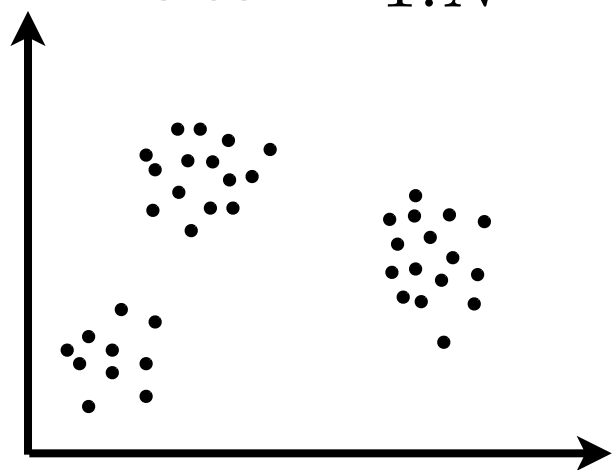
$$\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



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

CRP mixture model: inference

- Data $x_{1:N}$

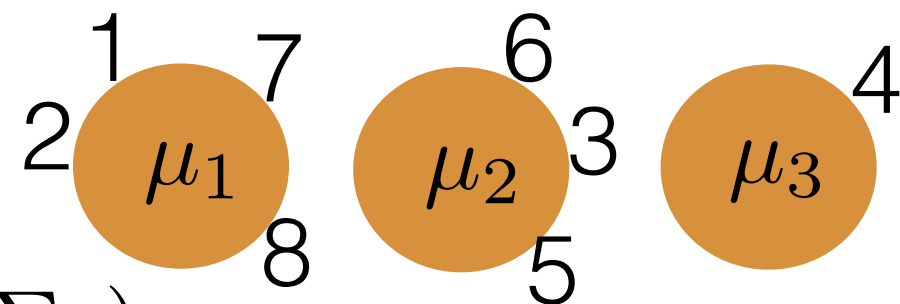


- Generative model

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

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

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



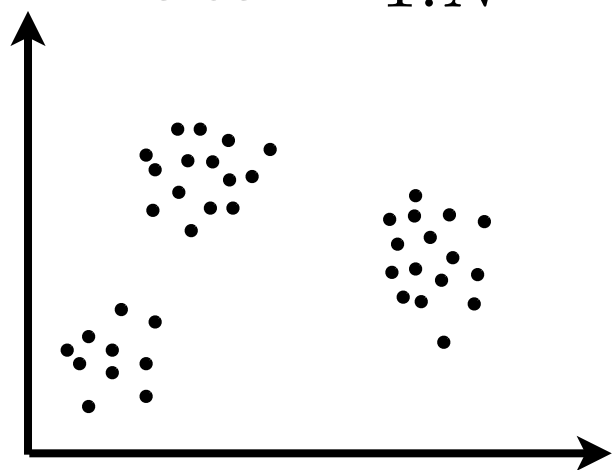
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

CRP mixture model: inference

- Data $x_{1:N}$

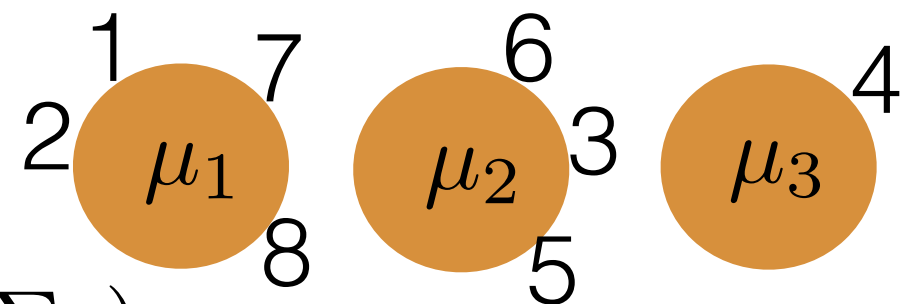


- Generative model

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

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

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



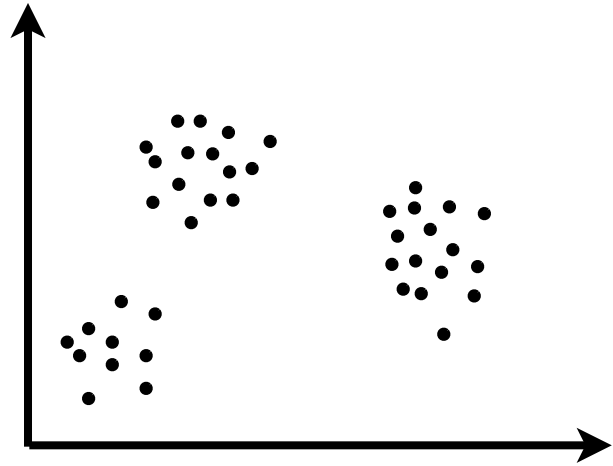
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

CRP mixture model: inference

- Data $x_{1:N}$

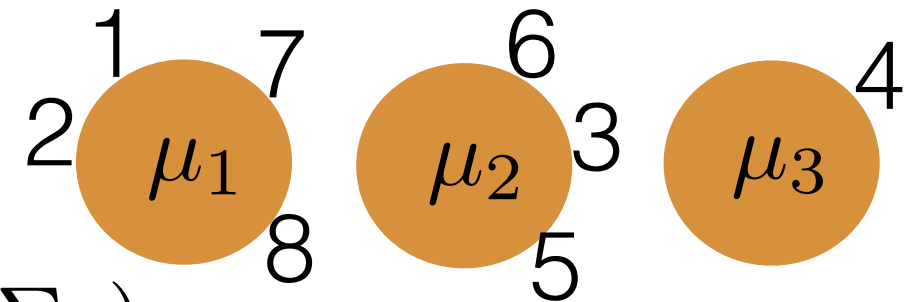


- Generative model

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

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

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



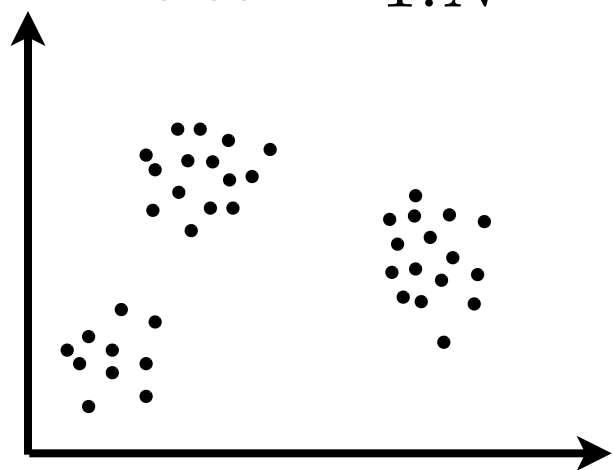
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

CRP mixture model: inference

- Data $x_{1:N}$

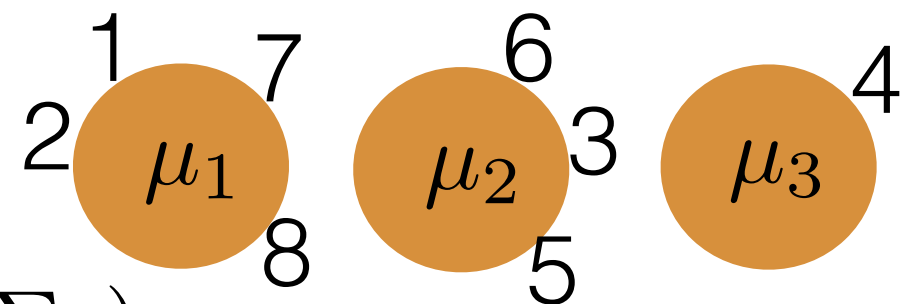


- Generative model

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

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

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



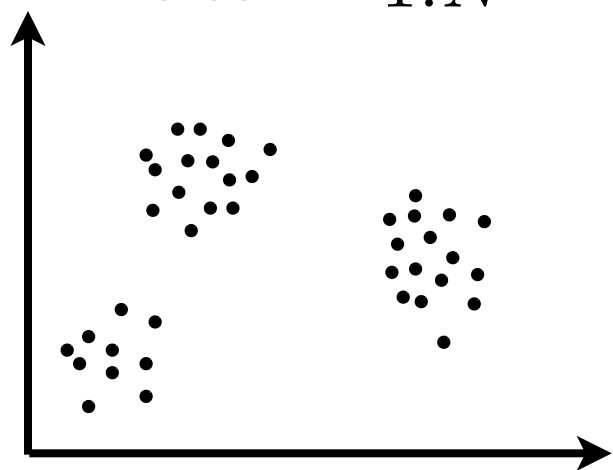
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

$$p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} \text{if } n \text{ joins cluster } C \\ \text{if } n \text{ starts a new cluster} \end{cases}$$

CRP mixture model: inference

- Data $x_{1:N}$

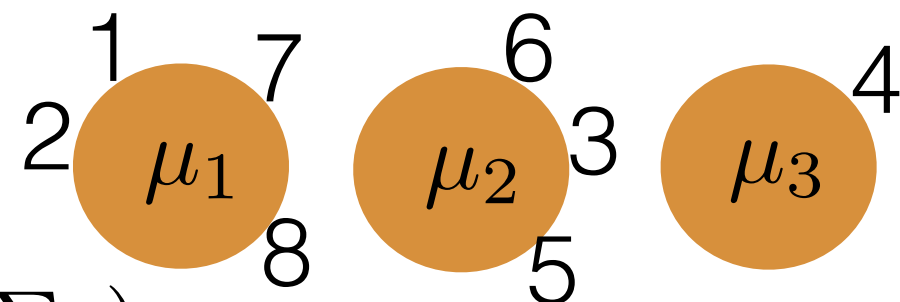


- Generative model

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

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

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



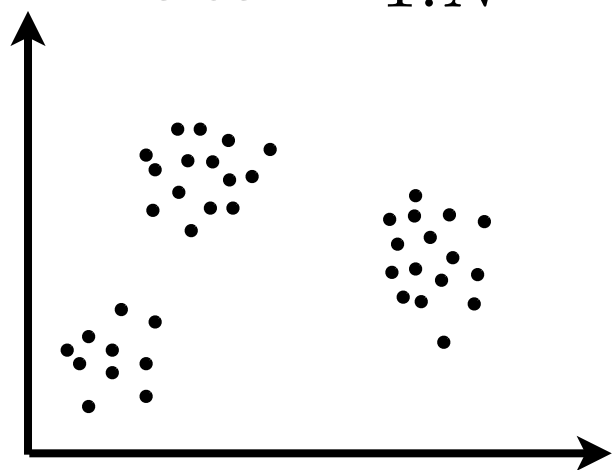
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

CRP mixture model: inference

- Data $x_{1:N}$

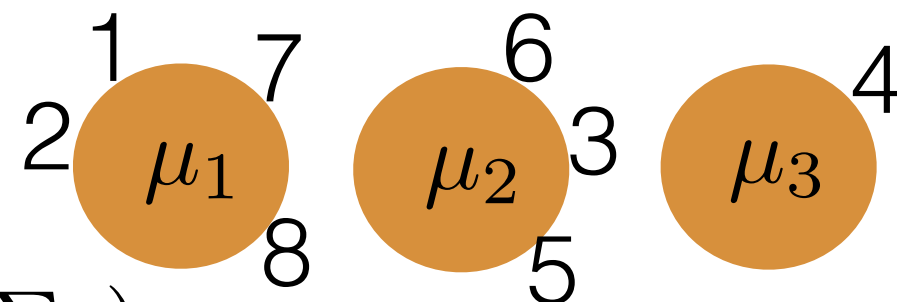


- Generative model

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

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

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



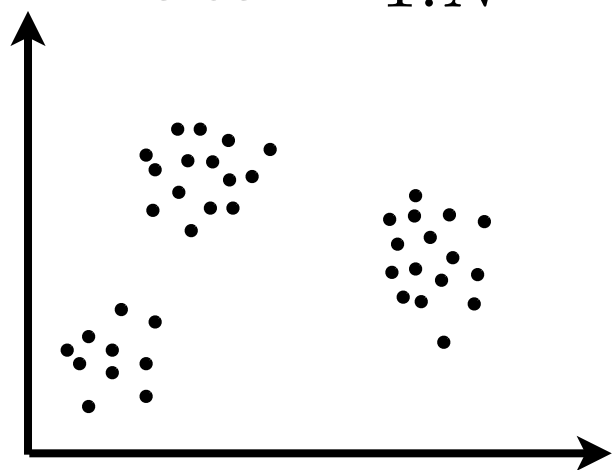
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

CRP mixture model: inference

- Data $x_{1:N}$

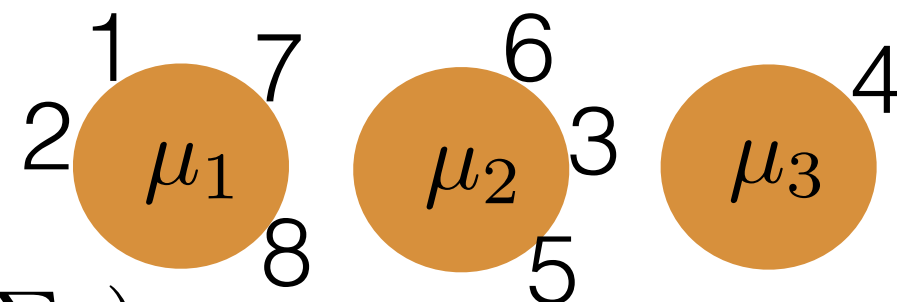


- Generative model

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

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

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



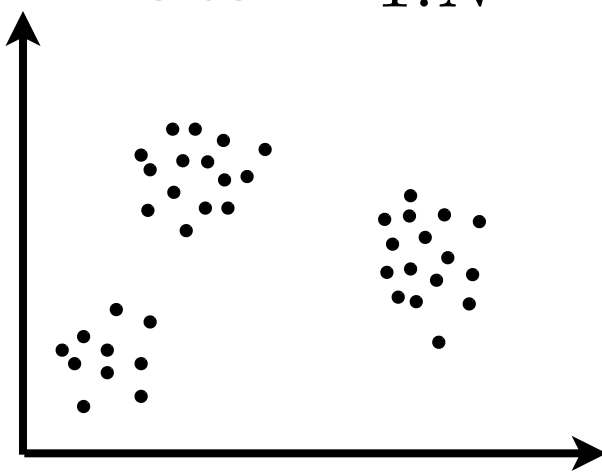
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

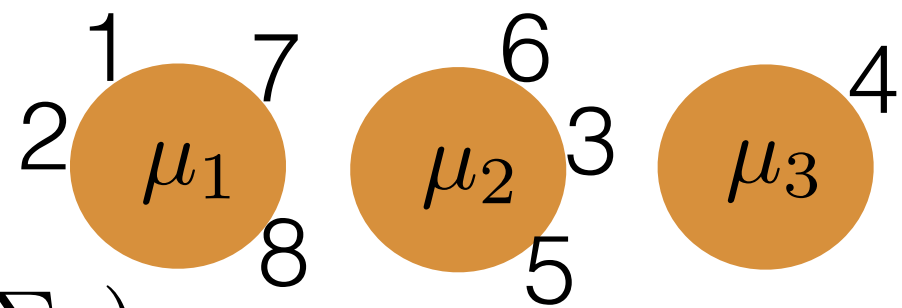
- For completeness: $p(x_{C \cup \{n\}} | x_C) =$

CRP mixture model: inference

- Data $x_{1:N}$

- Generative model

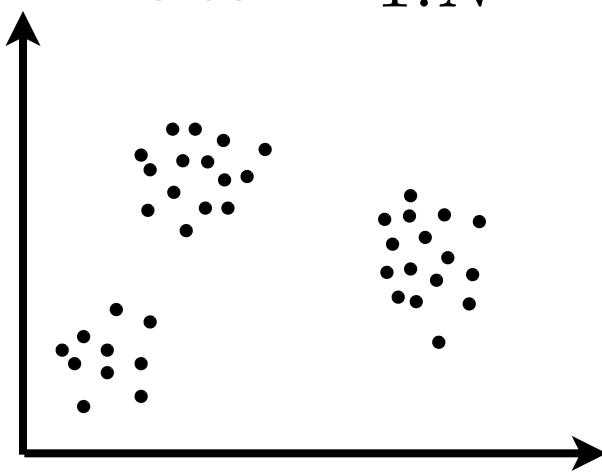
$$\Pi_N \sim \text{CRP}(N, \alpha)$$

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

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

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

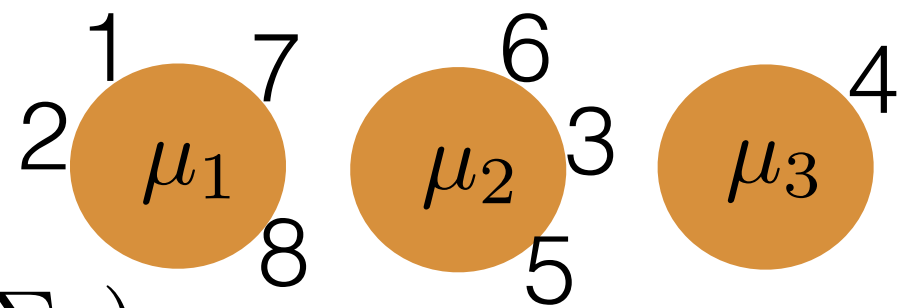
$$p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m},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

CRP mixture model: inference

- Data $x_{1:N}$

- Generative model

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

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

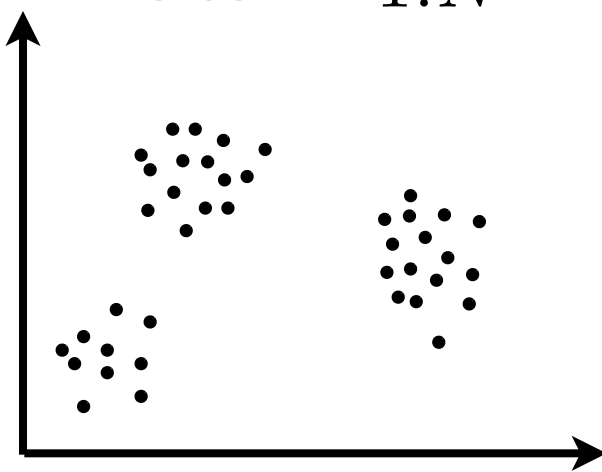
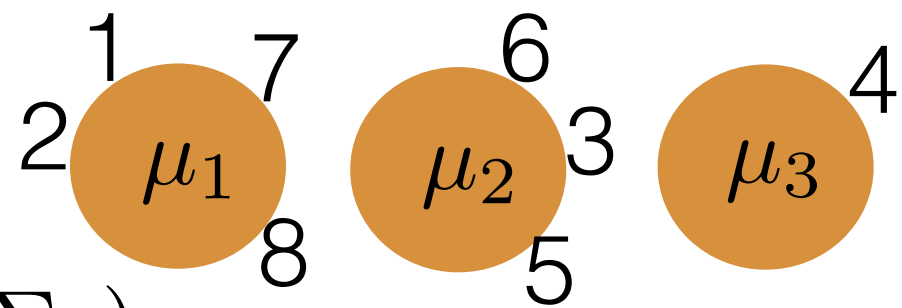
$$\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$

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

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

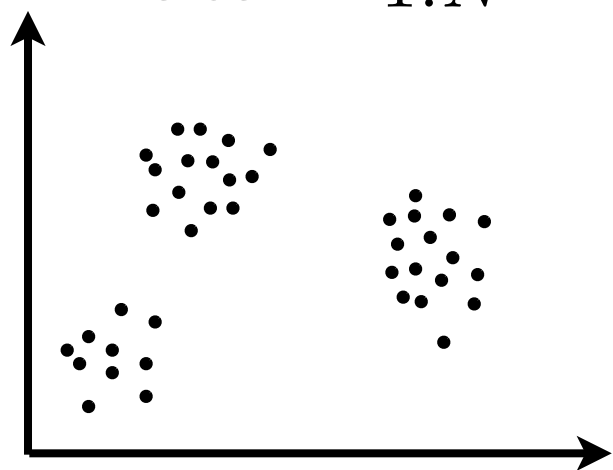
$$p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$
 - For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m},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

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

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

CRP mixture model: inference

- Data $x_{1:N}$

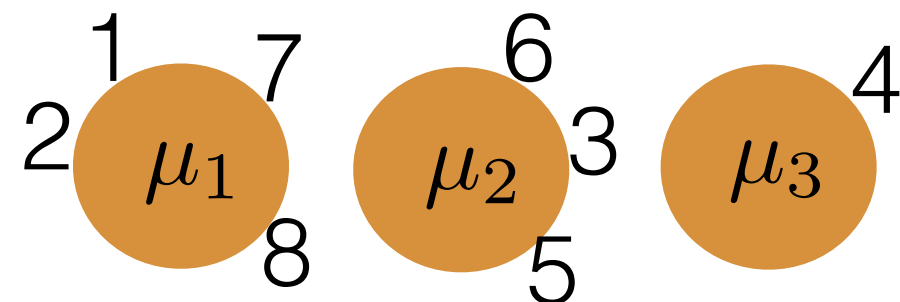


- Generative model

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

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

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



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

- Gibbs sampler:

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

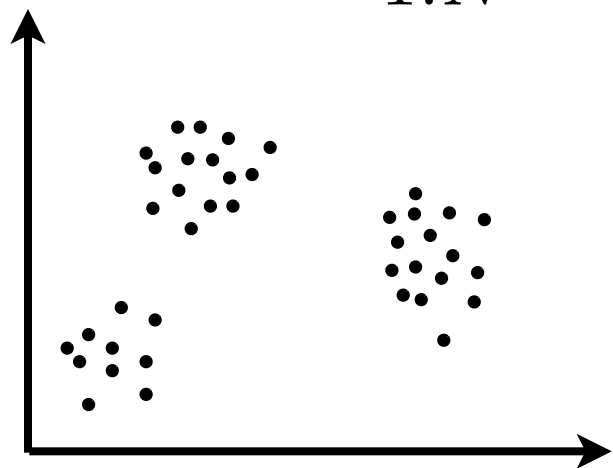
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

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

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

CRP mixture model: inference

- Data $x_{1:N}$

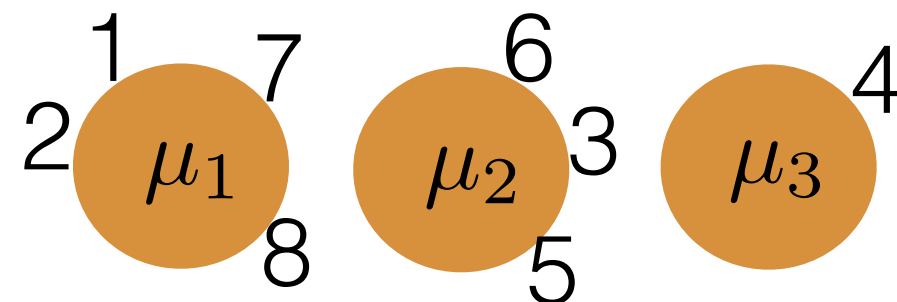


- Generative model

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

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

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



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

- Gibbs sampler:

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

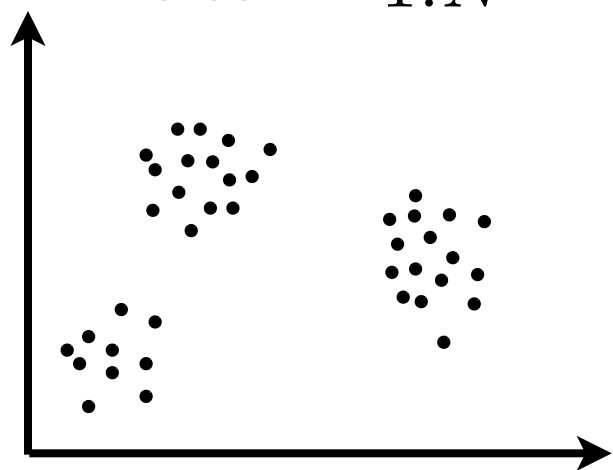
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

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

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

CRP mixture model: inference

- Data $x_{1:N}$

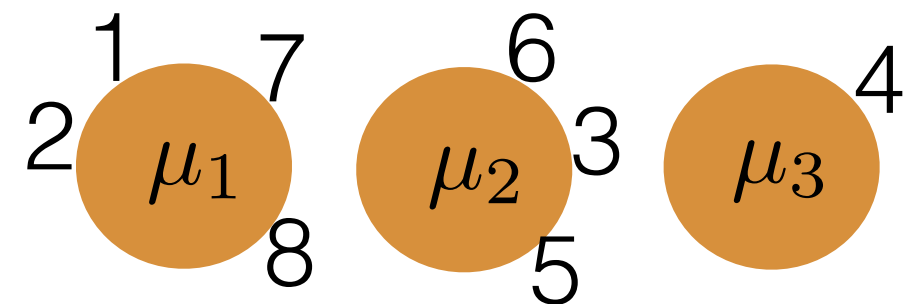


- Generative model

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

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

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

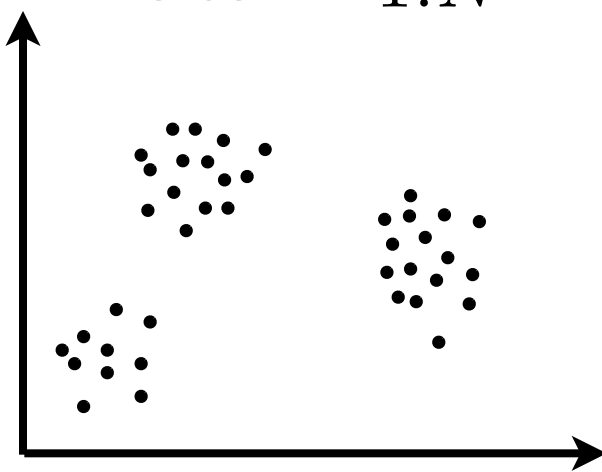
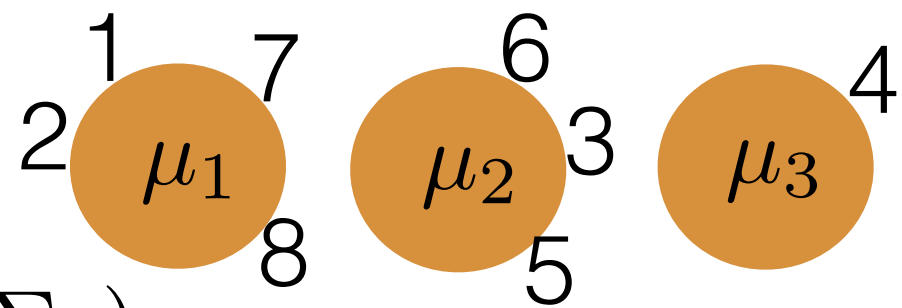


- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

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:

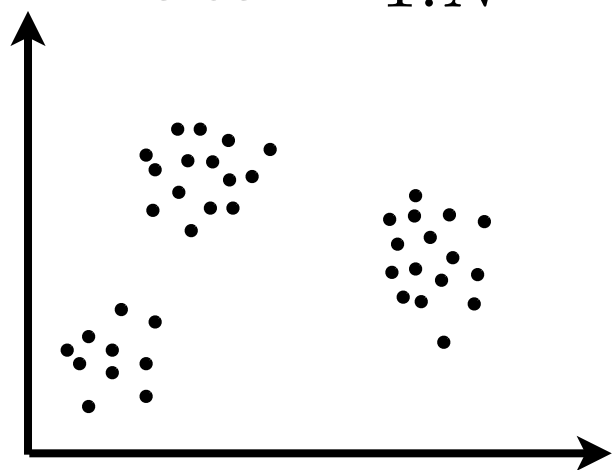
$$p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$
 - For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m},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

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

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

CRP mixture model: inference

- Data $x_{1:N}$

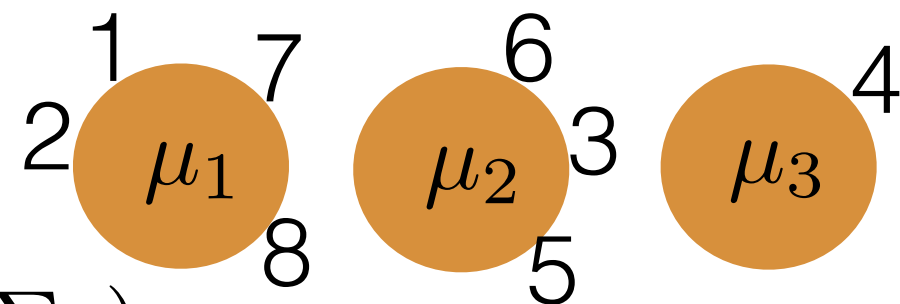


- Generative model

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

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

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



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

- Gibbs sampler:

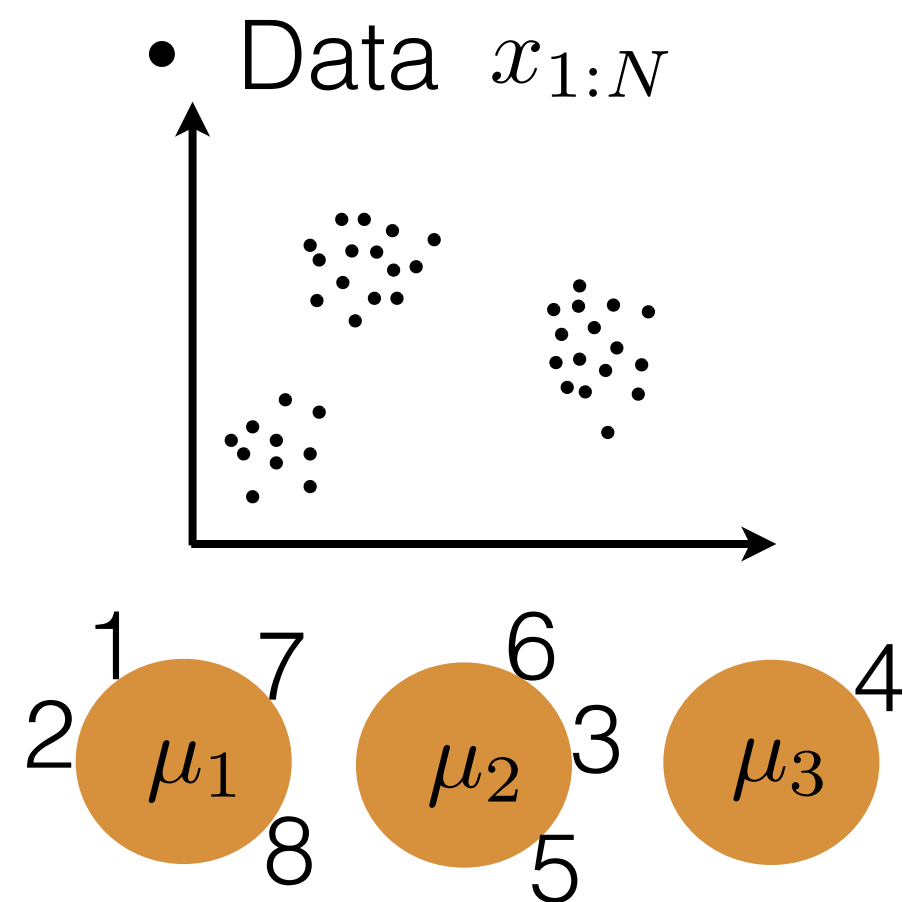
$$p(\Pi_N | \Pi_{N,-n}, x) = \begin{cases} \frac{\#C}{\alpha + N - 1} p(x_{C \cup \{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha + N - 1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

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

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

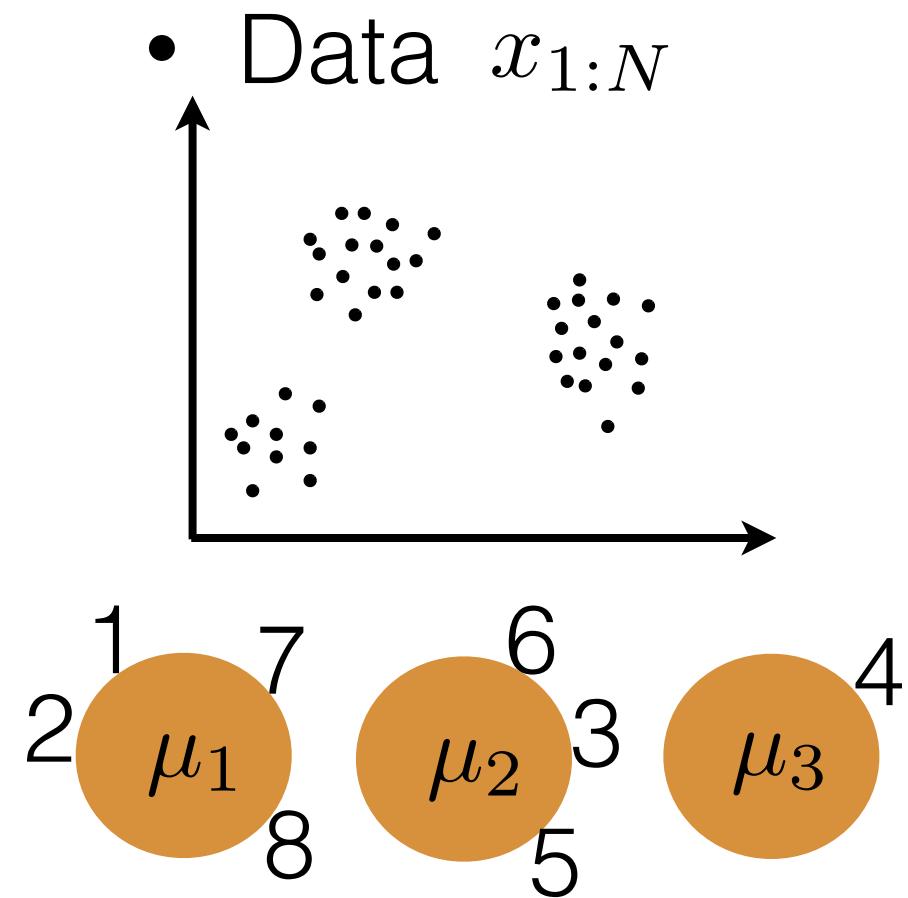
$$\tilde{m} := \tilde{\Sigma} \left(\Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1} \mu_0 \right) \quad [\text{demo}]$$

CRP mixture model exercises



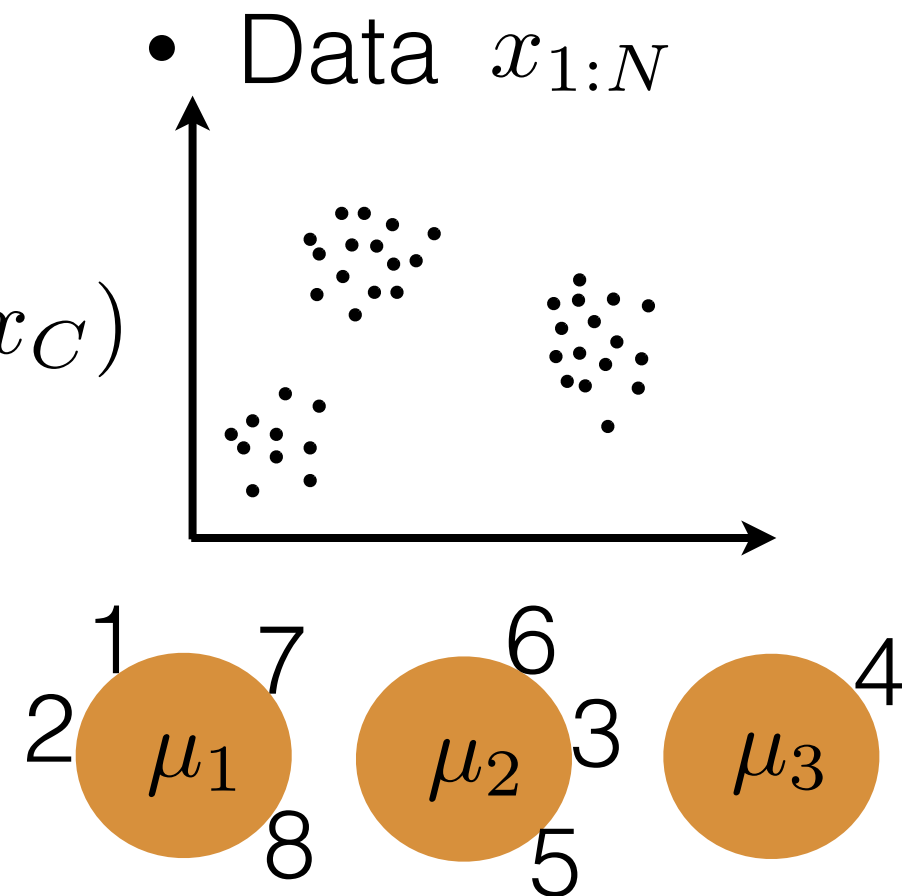
CRP mixture model exercises

- Code a CRP mixture model simulator



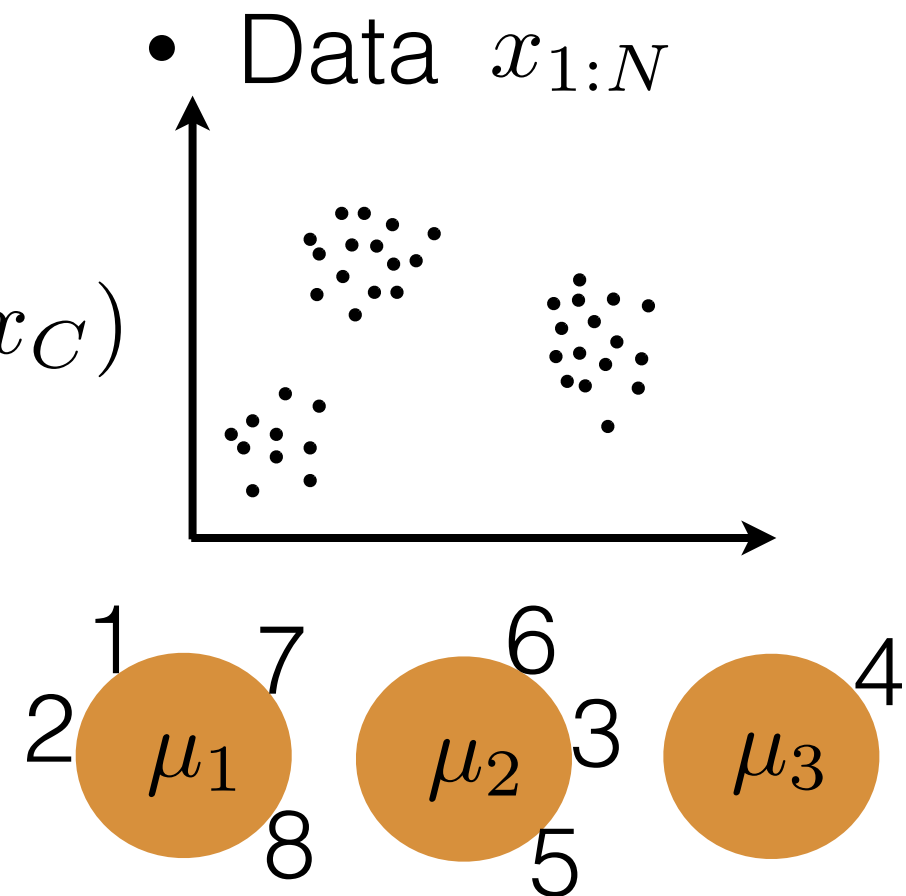
CRP mixture model exercises

- 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



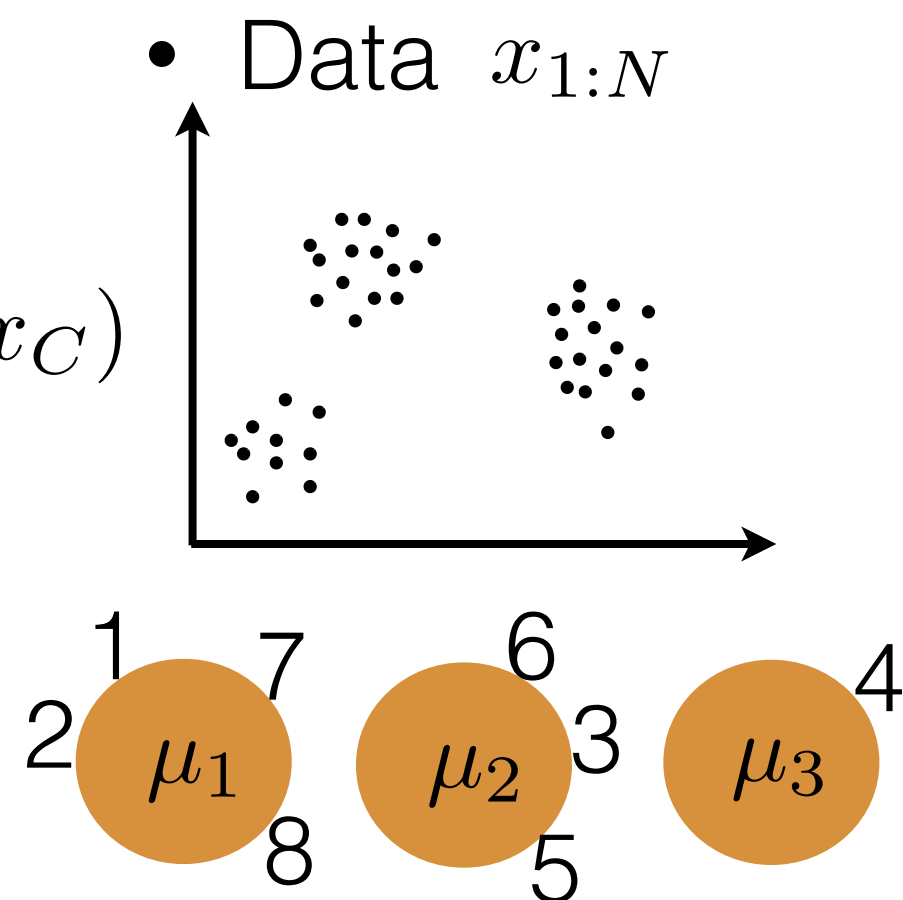
CRP mixture model exercises

- 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



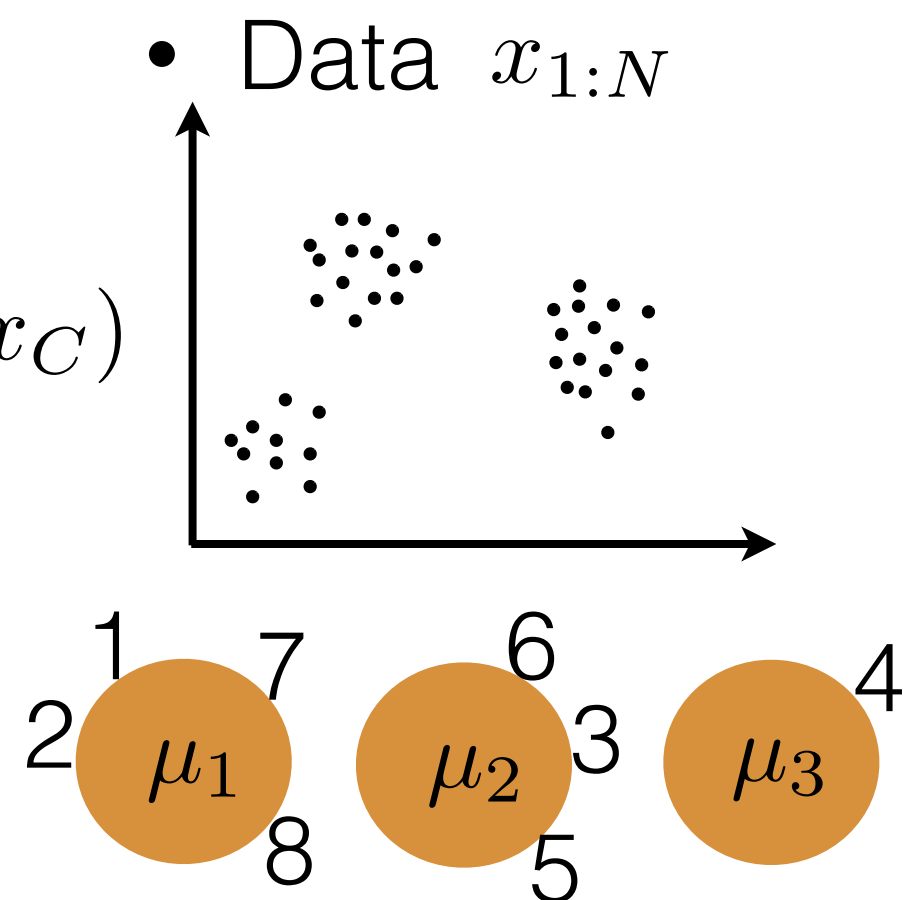
CRP mixture model exercises

- 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
- Read Neal 2000 and try out other samplers



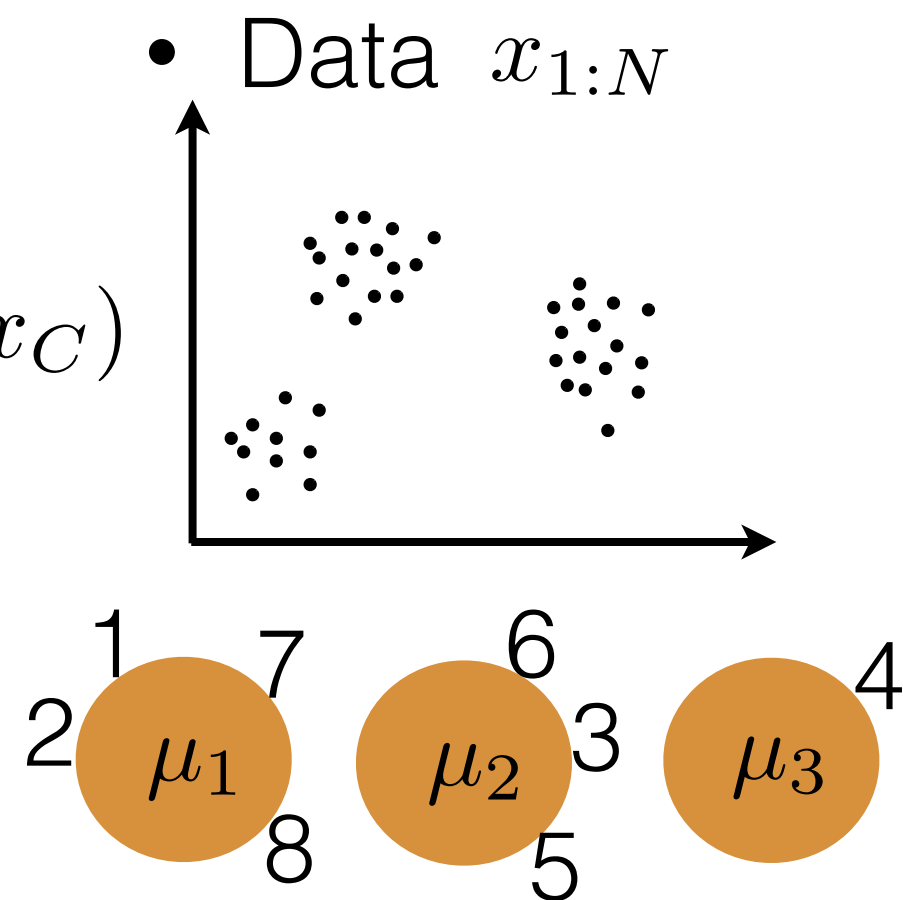
CRP mixture model exercises

- 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
- Read Neal 2000 and try out other samplers
- Further: Read Walker 2007; Kalli, Griffin, Walker 2011 and code a DPMM slice sampler; Read Wang, Blei 2012 and code a variational inference algorithm

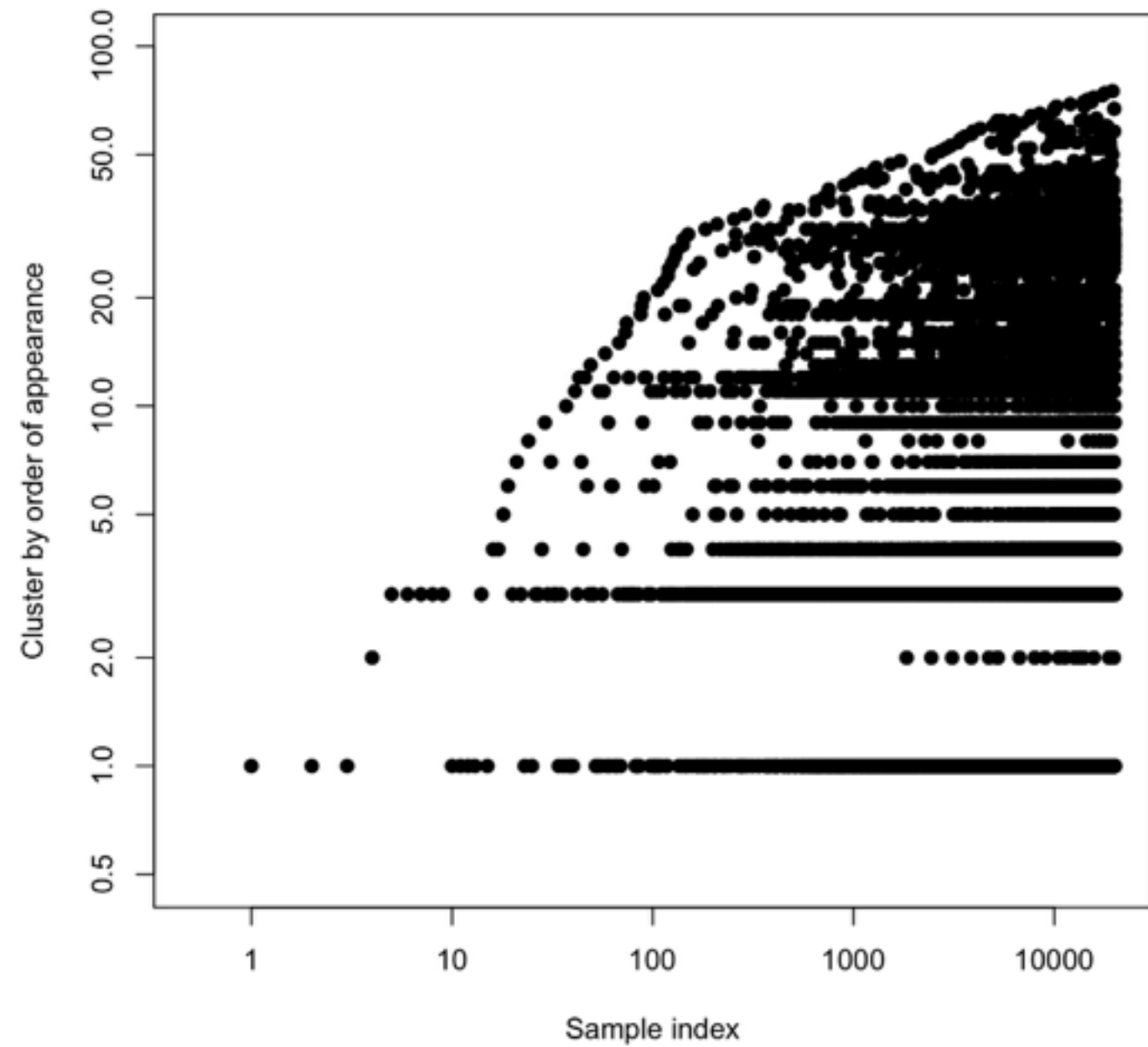


CRP mixture model exercises

- 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
- Read Neal 2000 and try out other samplers
- Further: Read Walker 2007; Kalli, Griffin, Walker 2011 and code a DPMM slice sampler; Read Wang, Blei 2012 and code a variational inference algorithm
- Read Broderick, Jordan, Pitman 2013 “Cluster and feature modeling [...]” for more background/extensions

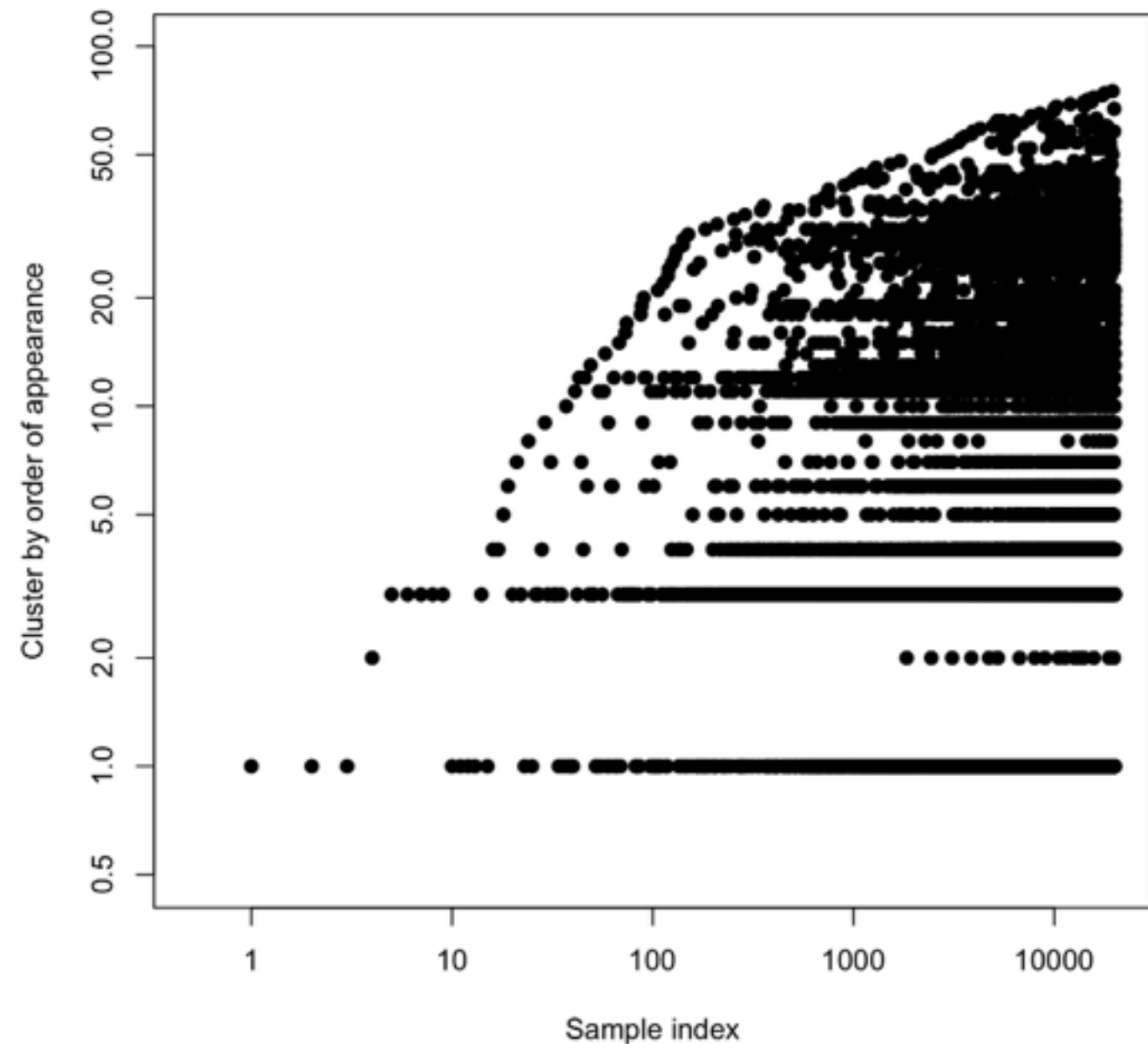


Power laws



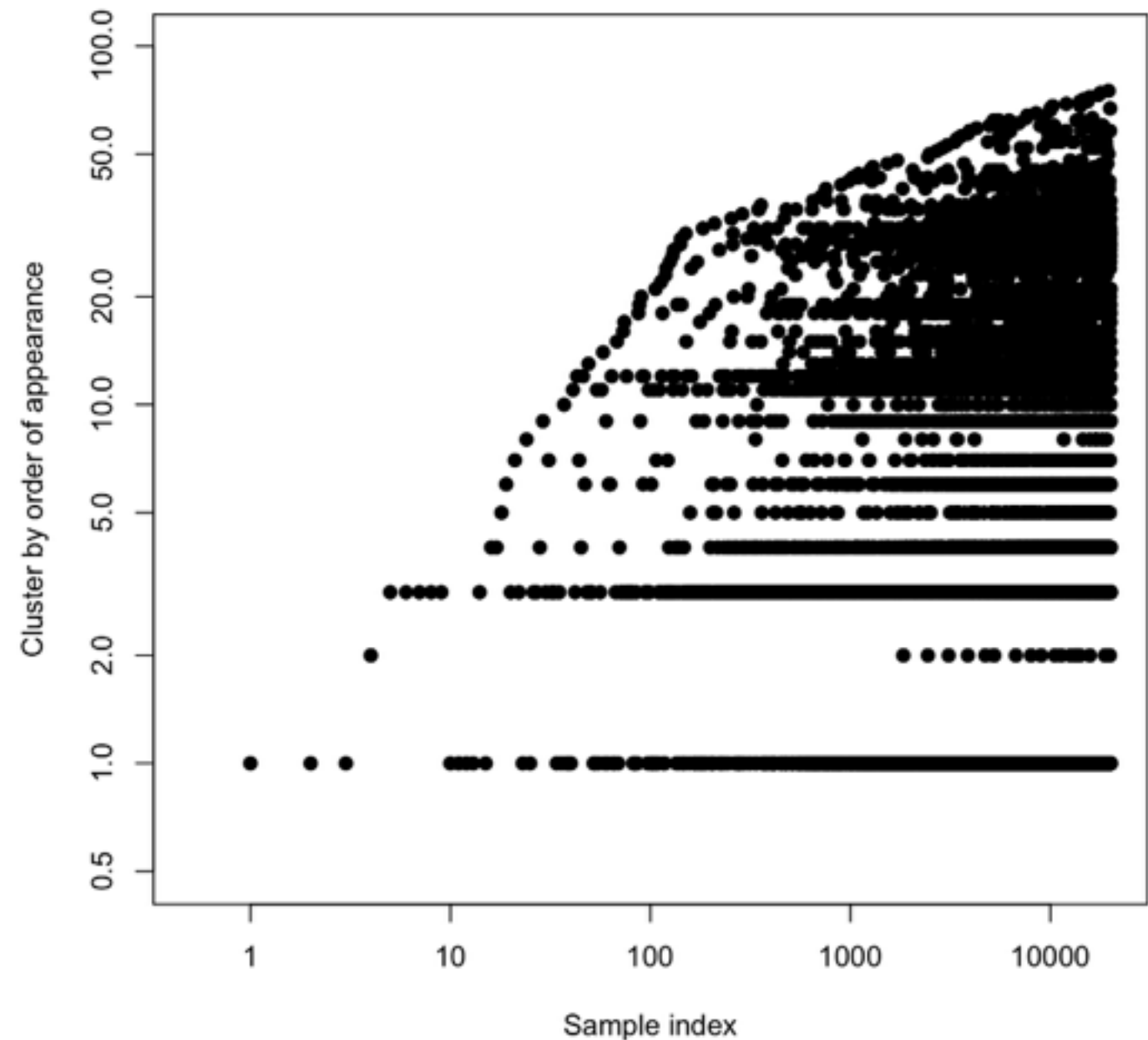
Power laws

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



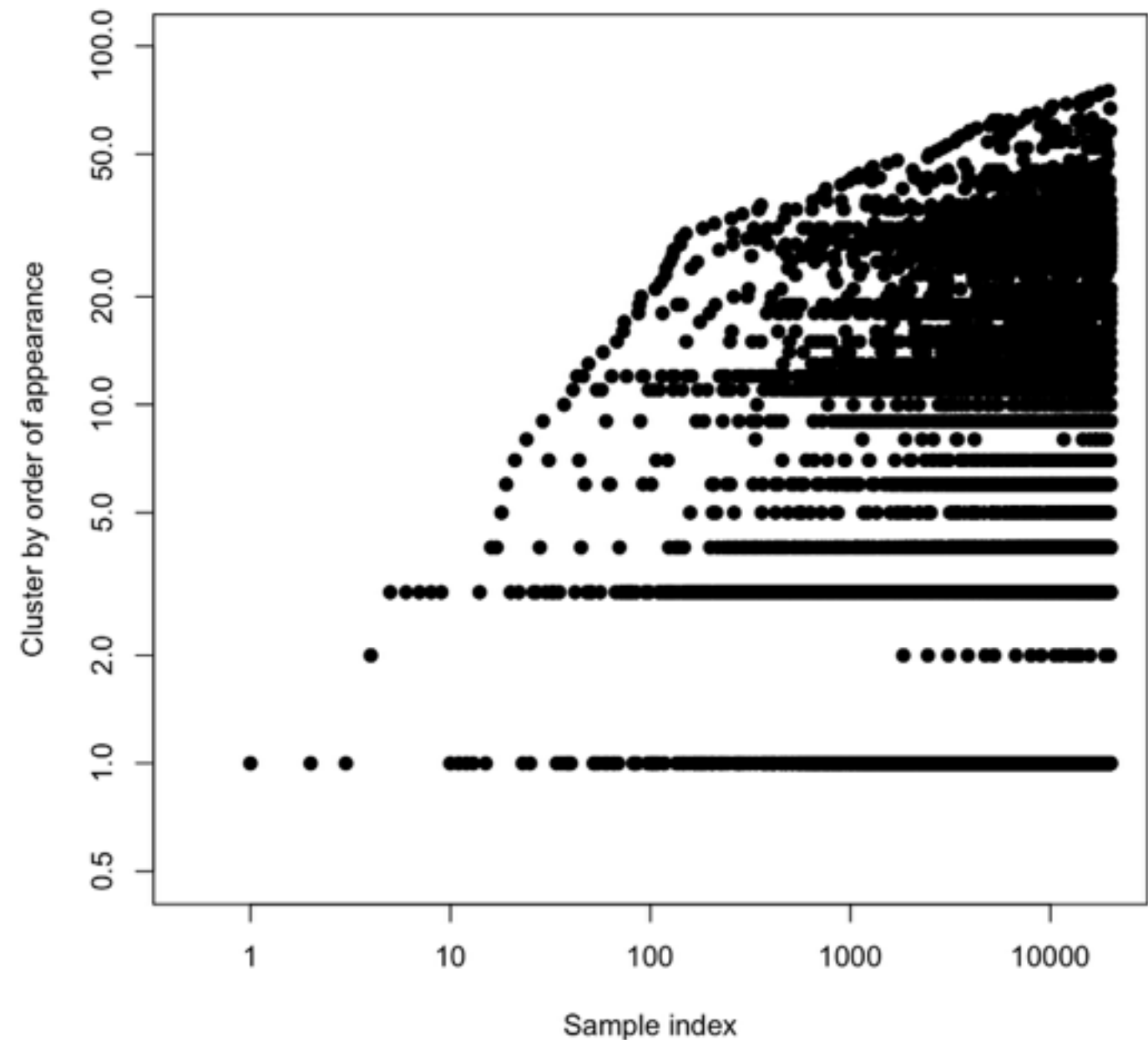
Power laws

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



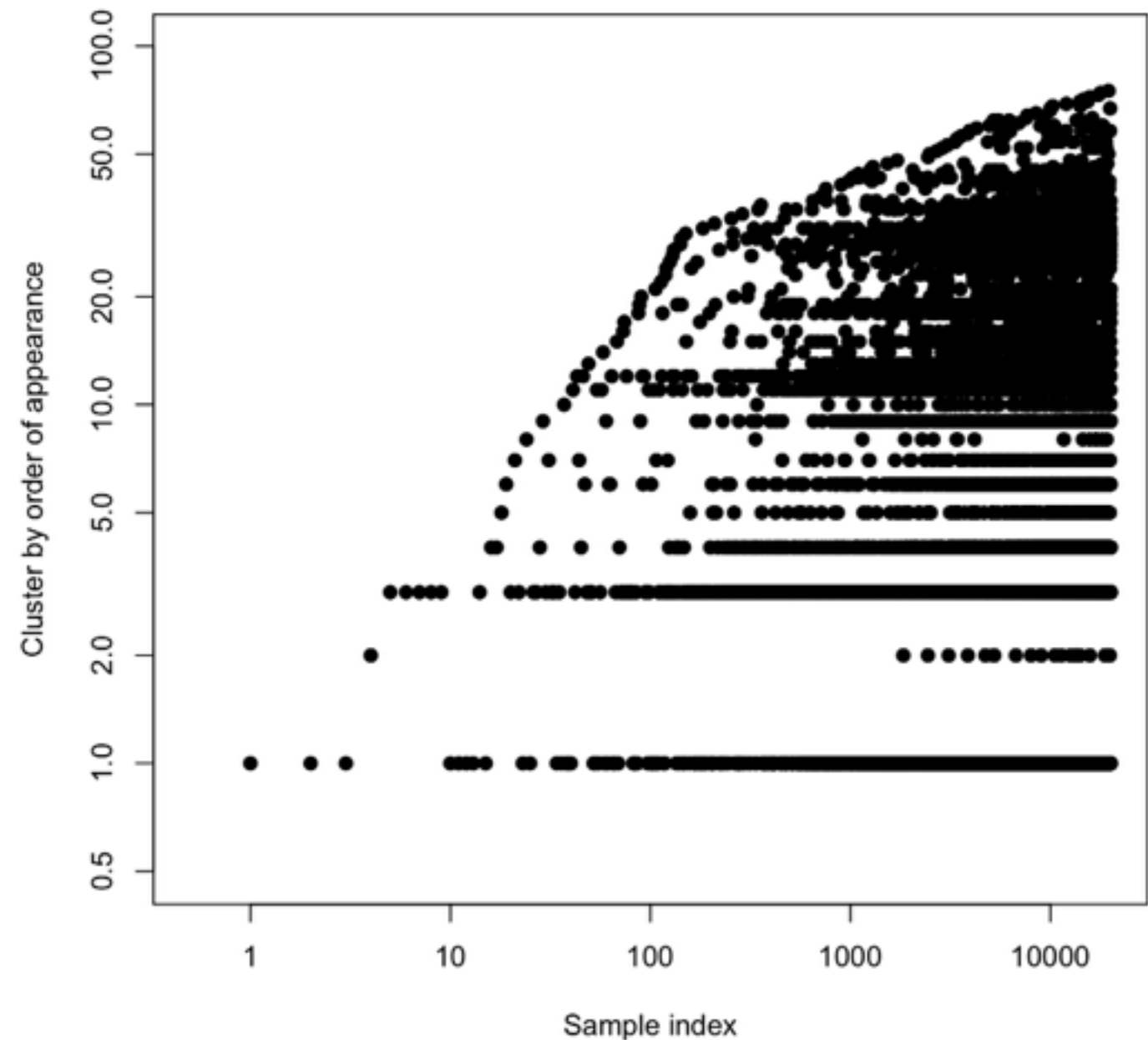
Power laws

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



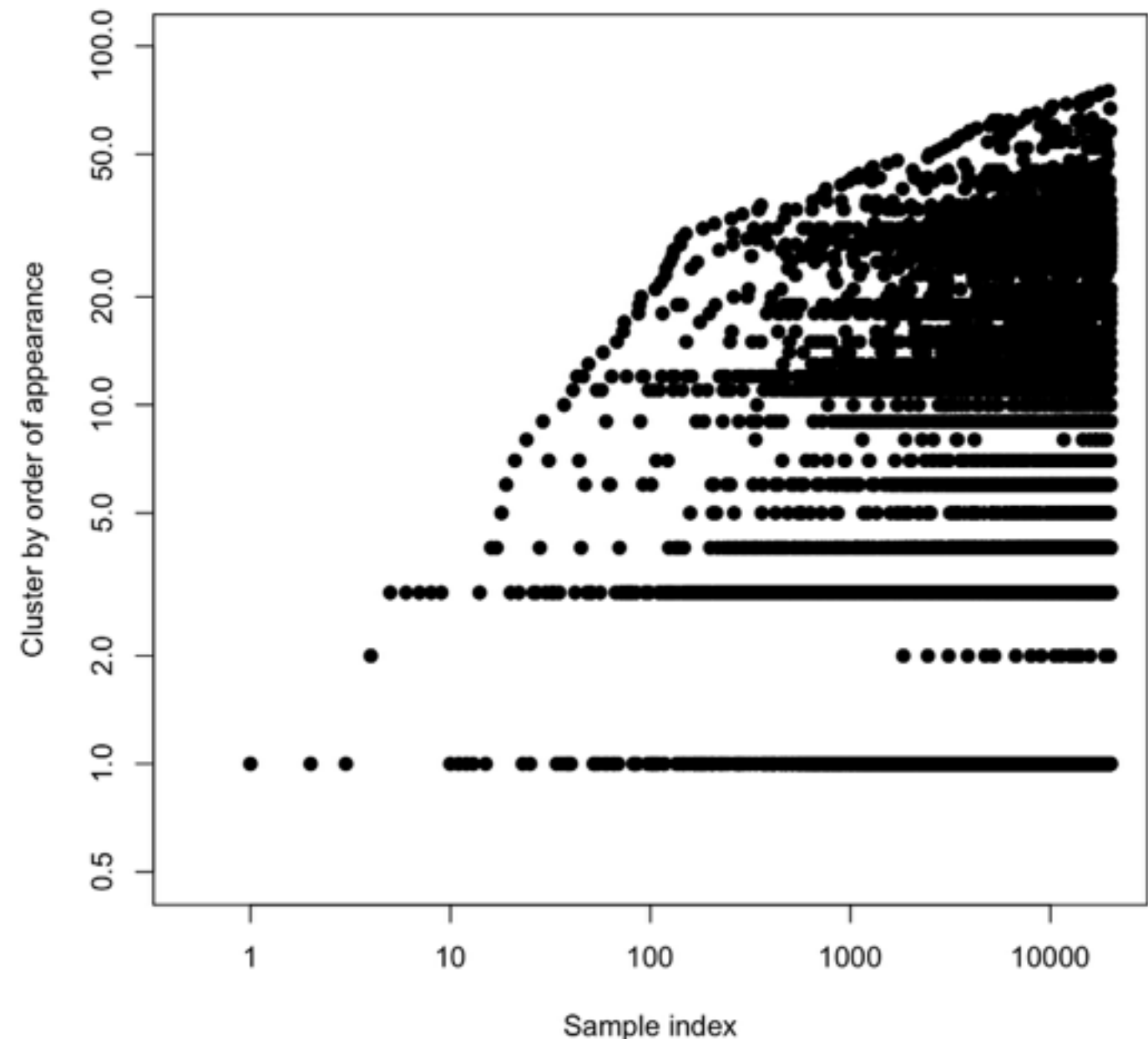
Power laws

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



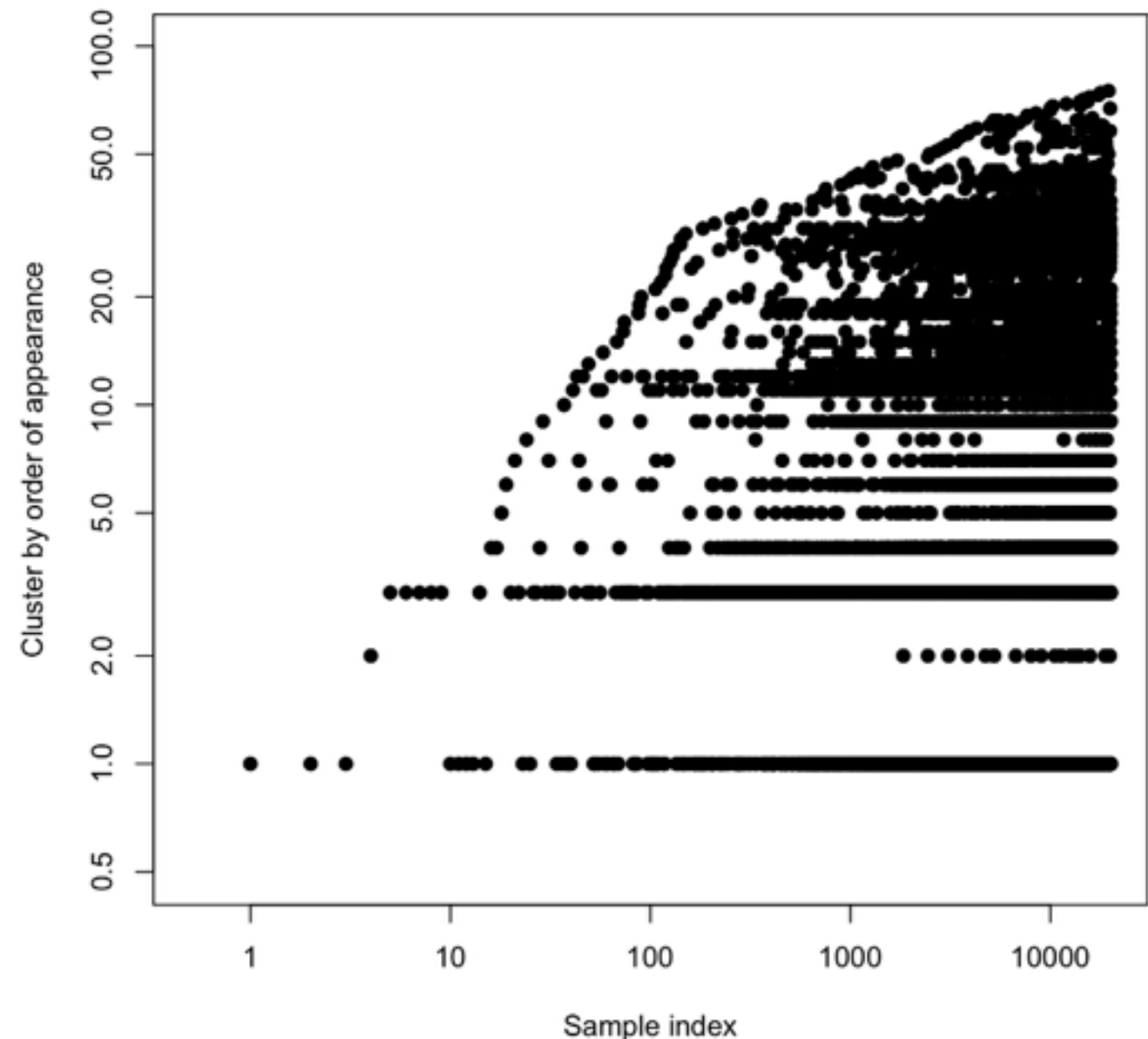
Power laws

- $K_N := \#$ 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:



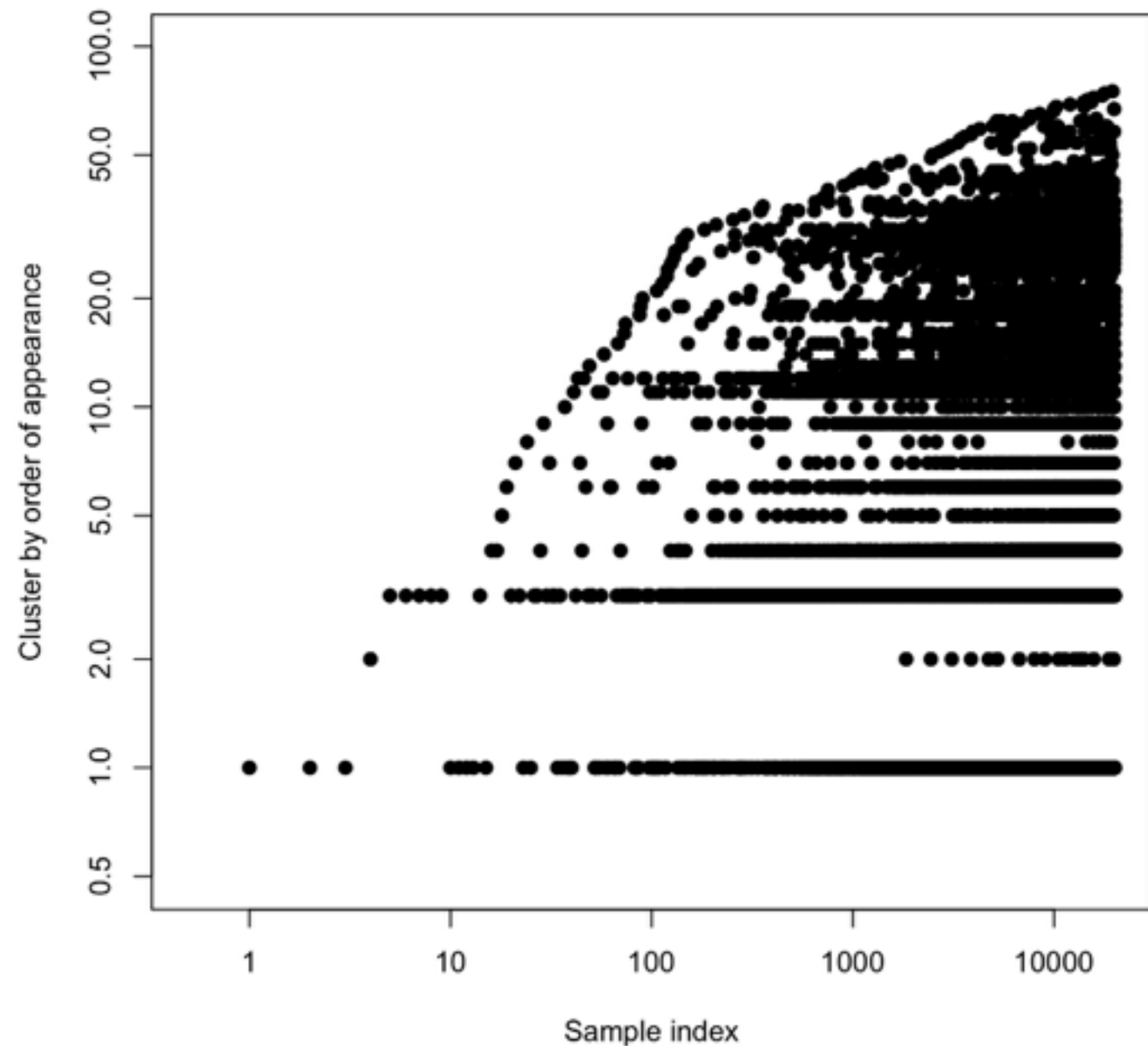
Power laws

- $K_N := \#$ 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:



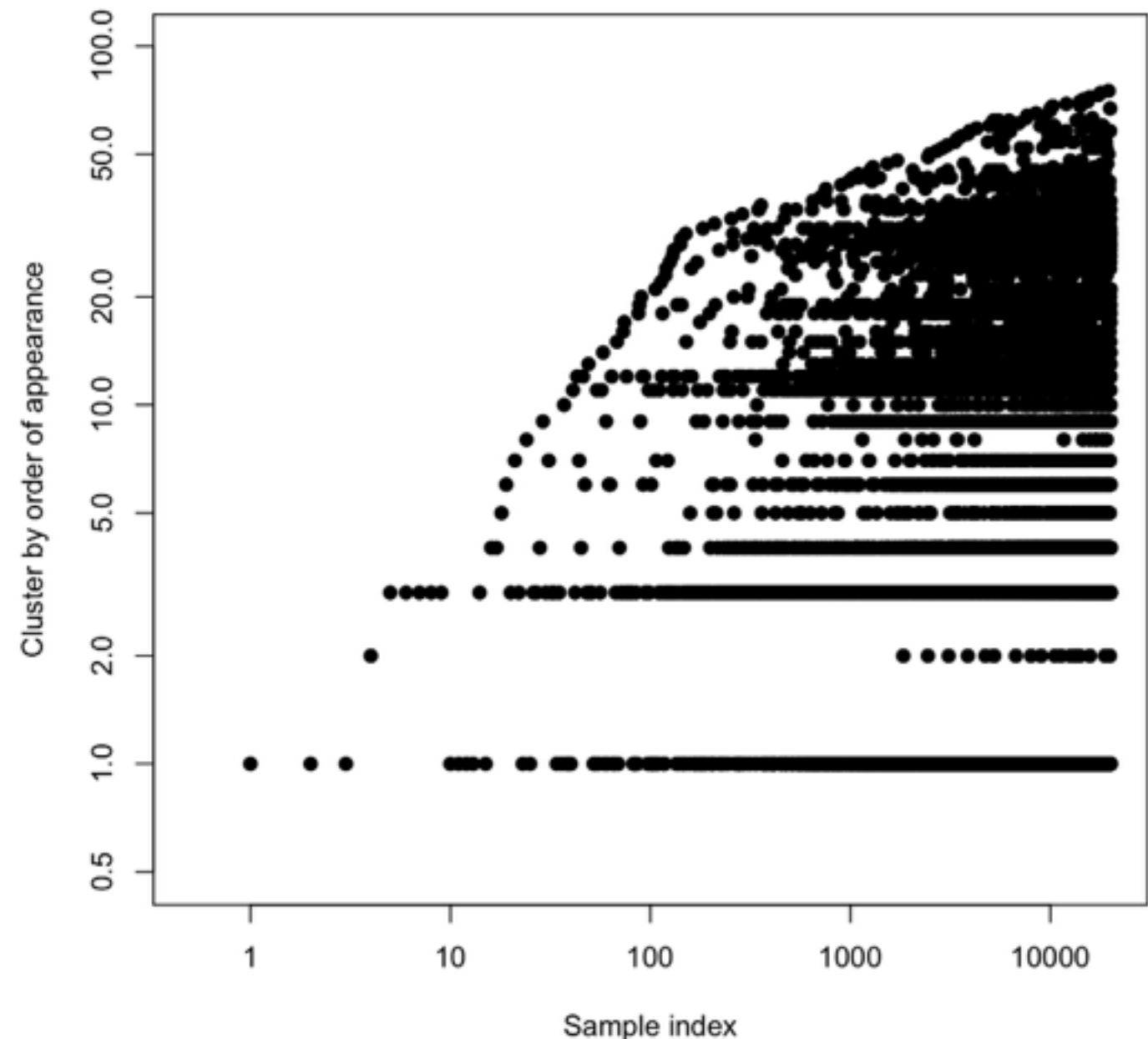
Power laws

- $K_N := \#$ 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



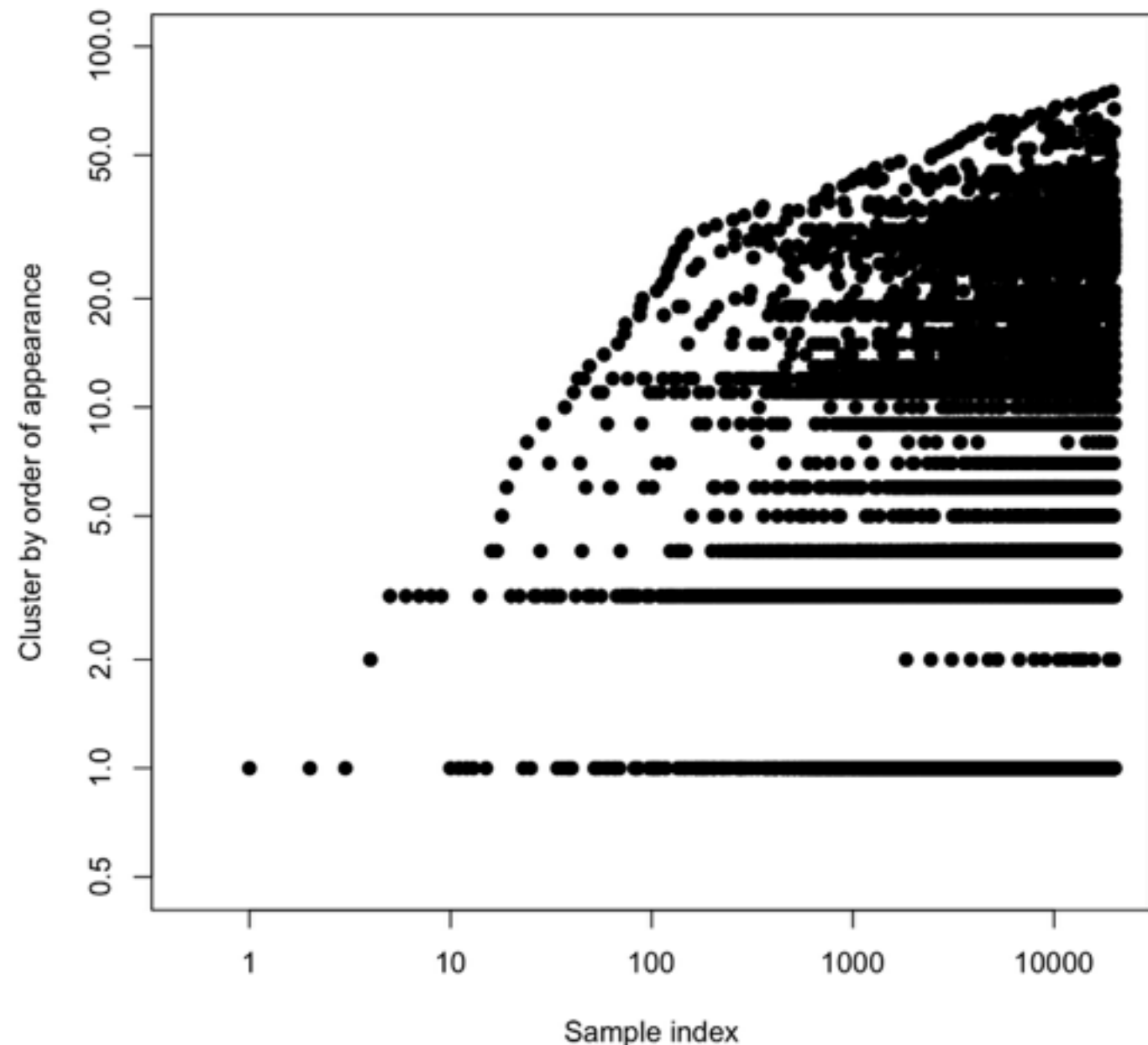
Power laws

- $K_N := \#$ 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



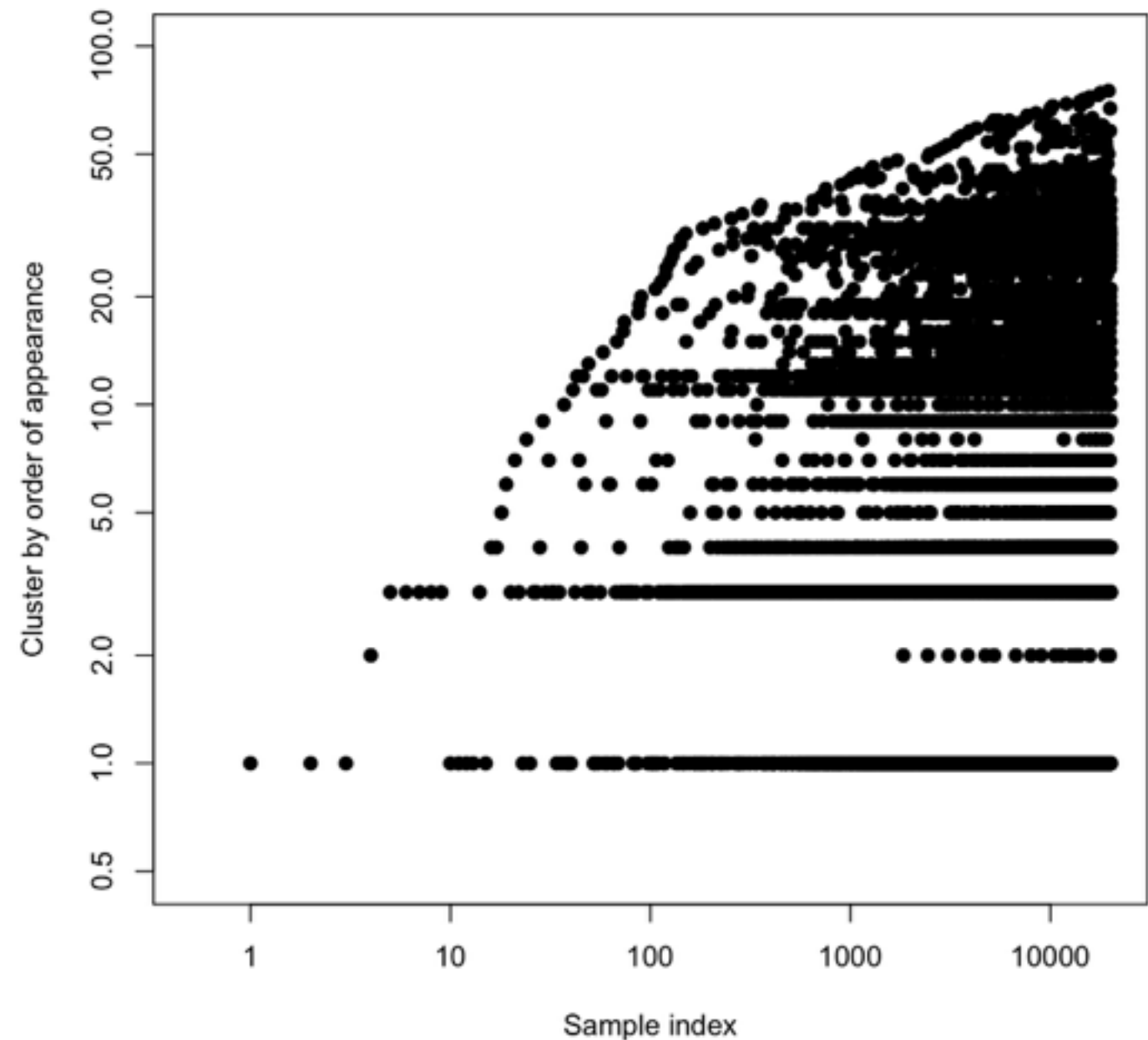
Power laws

- $K_N := \#$ 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)



Hierarchies

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

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

Hierarchies

- Hierarchical
Dirichlet process

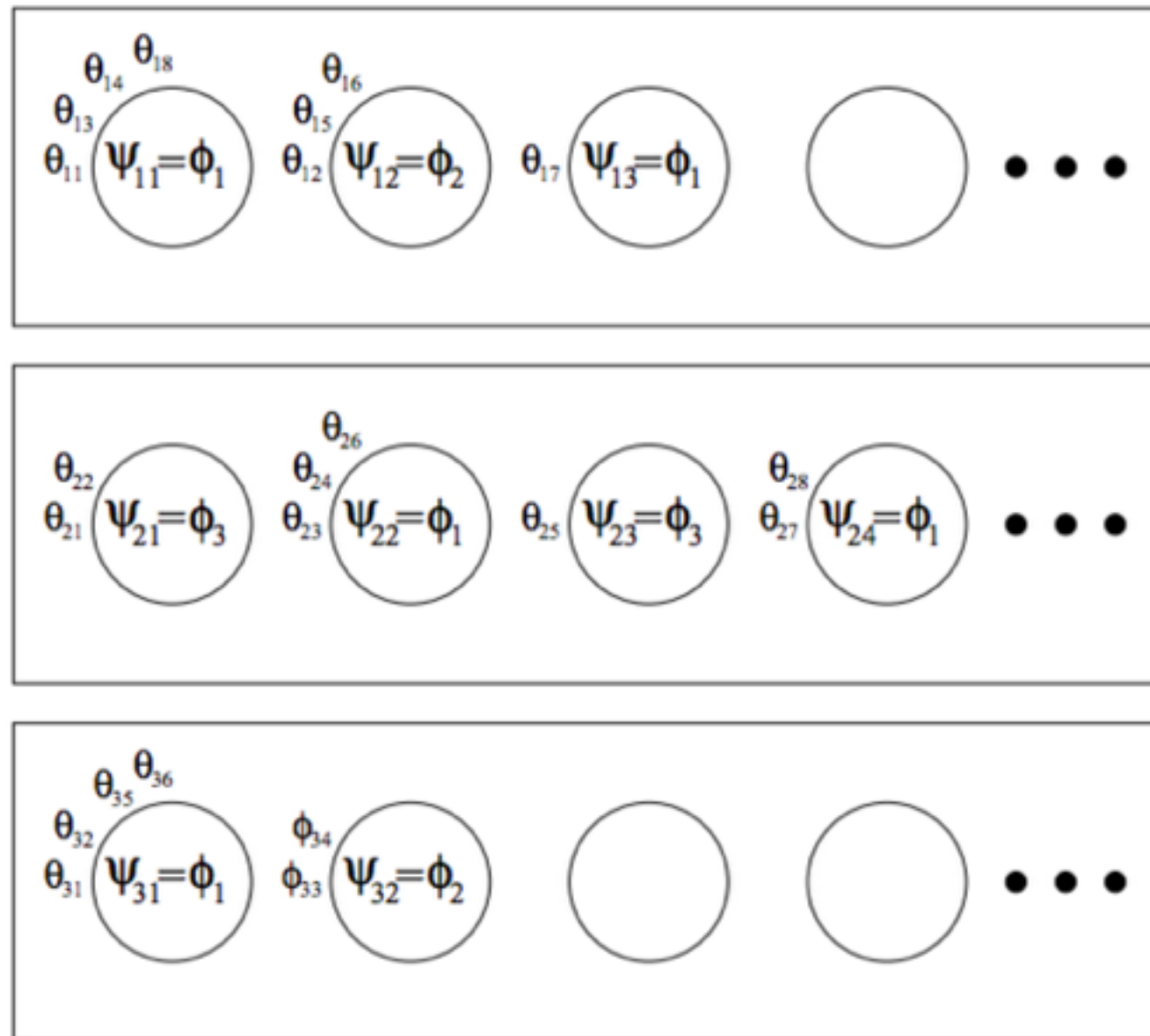
Hierarchies

- Hierarchical Dirichlet process

Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise

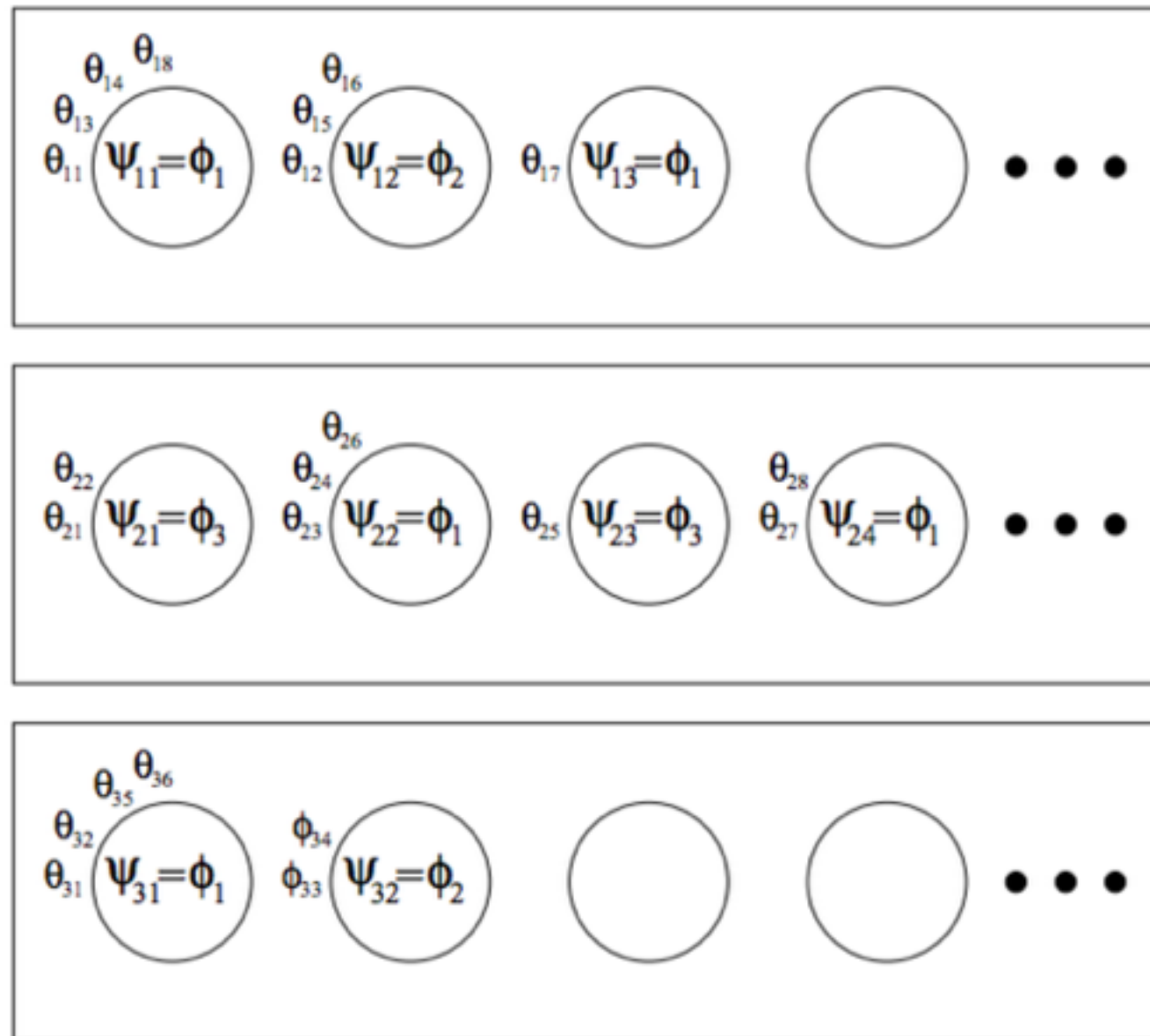
Hierarchies



- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006]

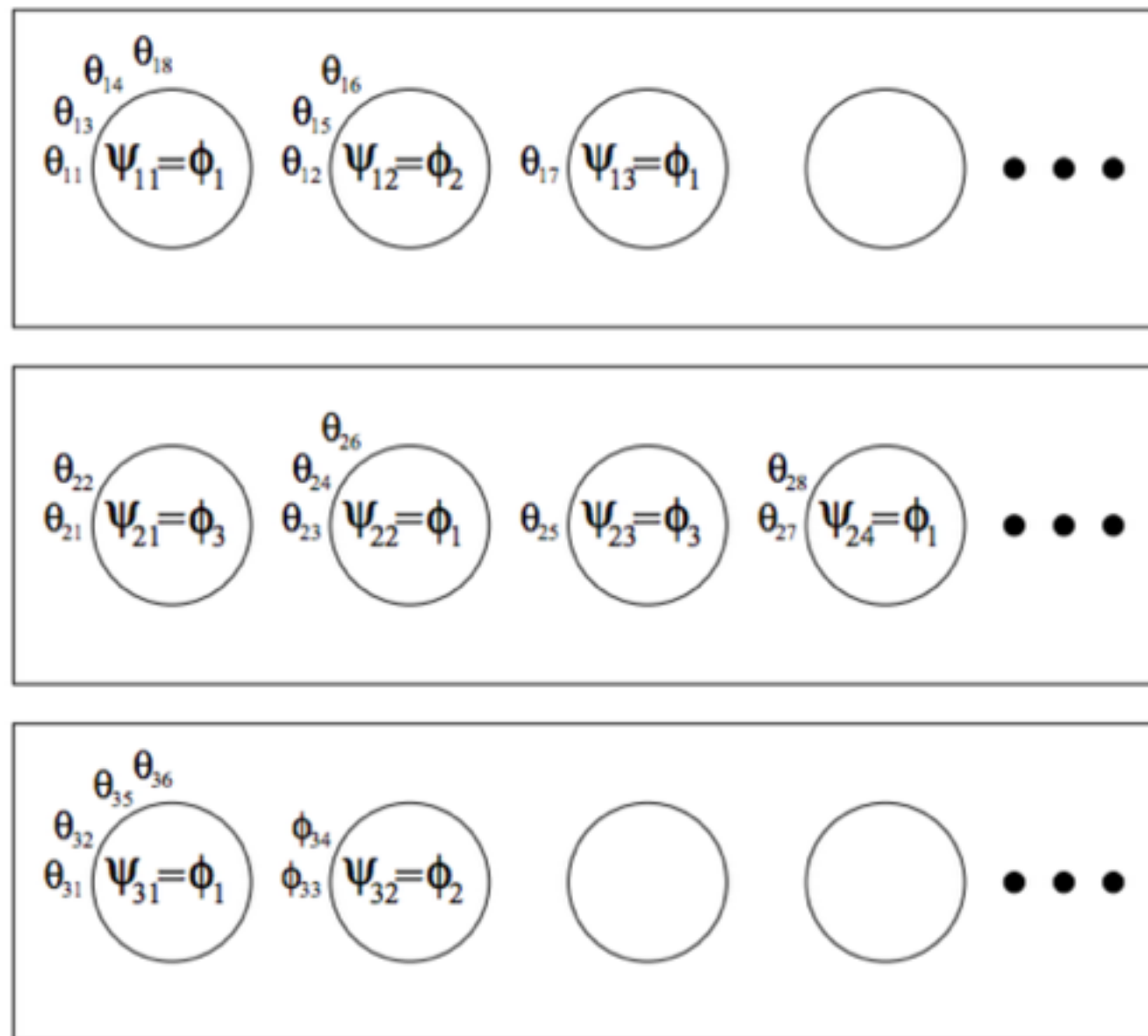
Hierarchies



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

[Teh et al 2006]

Hierarchies



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

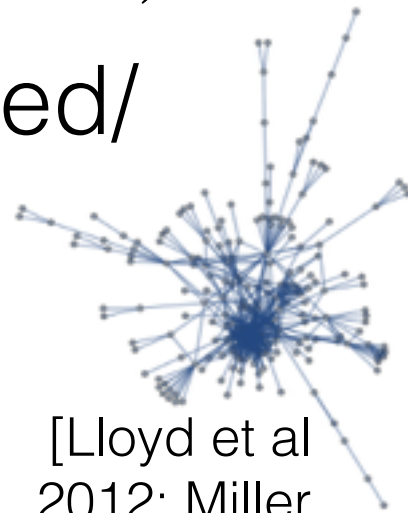
[Teh et al 2006]

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

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

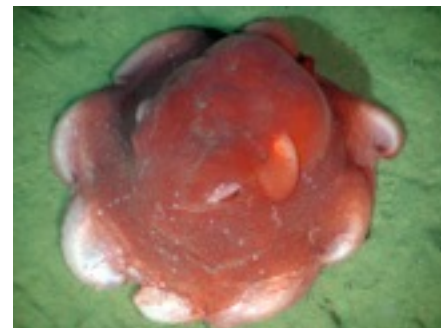
- Not parametric (i.e. not finite parameter, unbounded/ growing/infinite number of parameters)



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



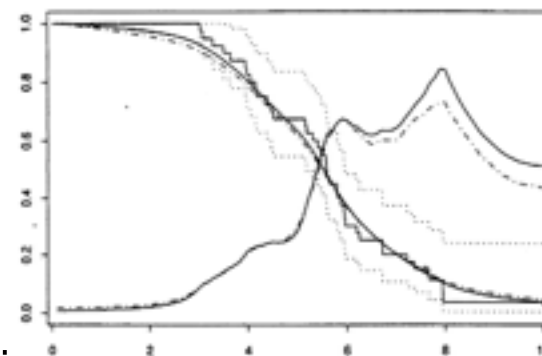
[wikipedia.org]



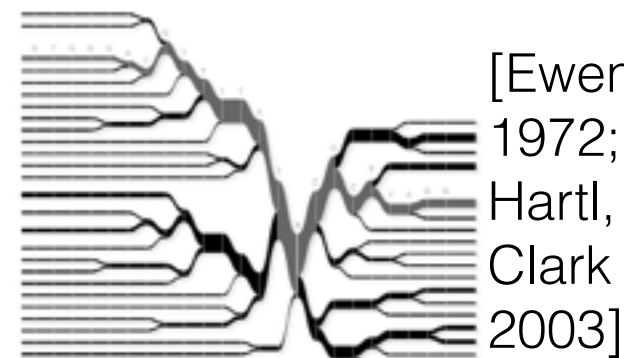
[Ed Bowlby, NOAA]



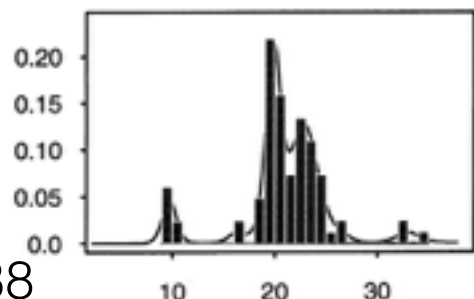
[Fox, et al 2014]



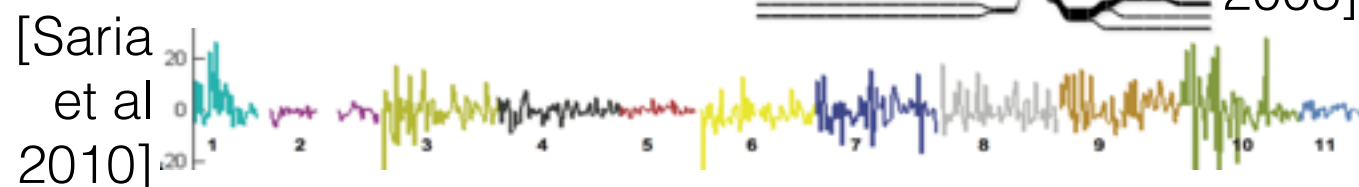
[Arjas, Gasbarra 1994]



[Ewens, 1972; Hartl, Clark 2003]



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



[Saria et al 2010]



[Sudderth, Jordan 2009]

References (page 1 of 4)

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.

E Bowlby. NOAA/Olympic Coast NMS; NOAA/OAR/Office of Ocean Exploration - NOAA Photo Library. Retrieved from: https://en.wikipedia.org/wiki/Opisthoteuthis_californiana#/media/File:Opisthoteuthis_californiana.jpg

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. arXiv preprint arXiv:1410.6843, 2014

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.

References (page 2 of 4)

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.

EB Fox, personal website. Retrieved 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.

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.

References (page 3 of 4)

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

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.

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.

References (page 4 of 4)

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.