Nonparametric Bayesian Statistics: Part III

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

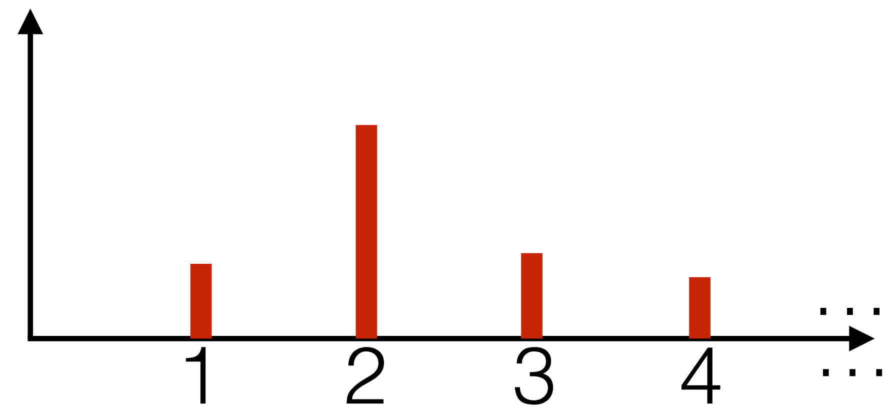
Recall: Part I

Recall: Part I

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

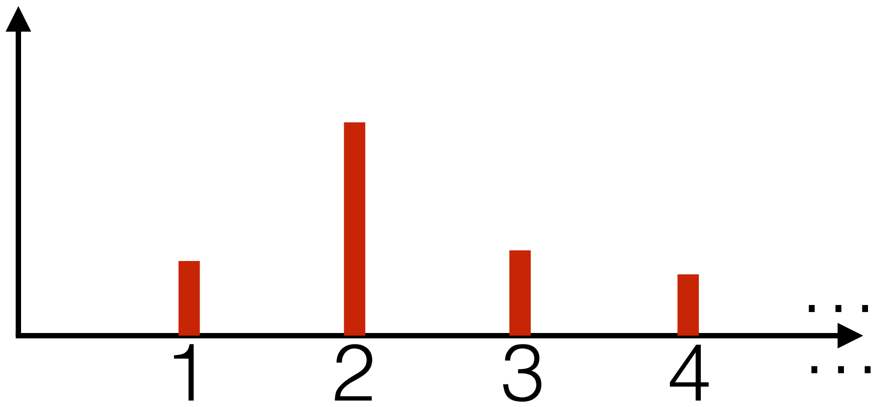
Recall: Part I

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$



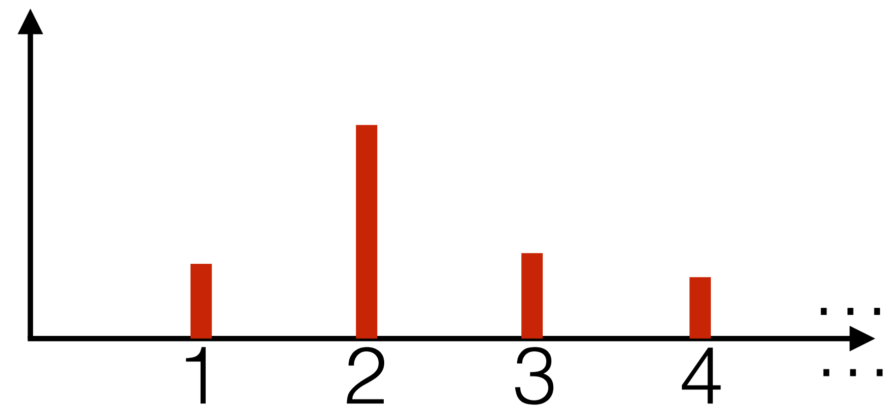
Recall: Part I

$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$



$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho)$$

Recall: Part I and Part II

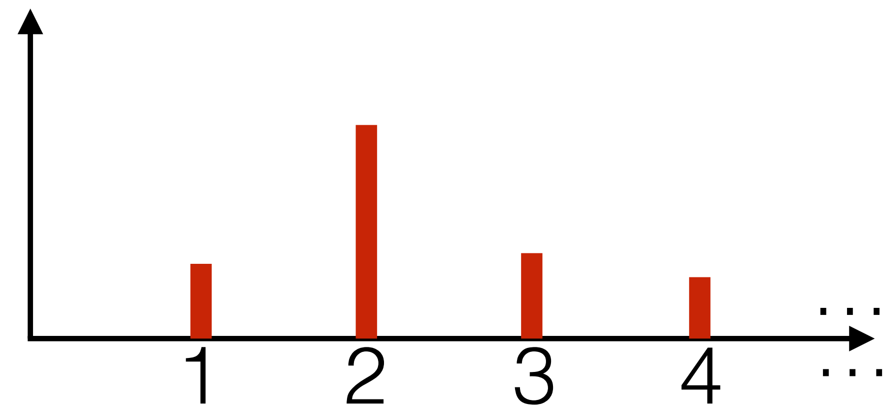


$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$

$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

Recall: Part I and Part II

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

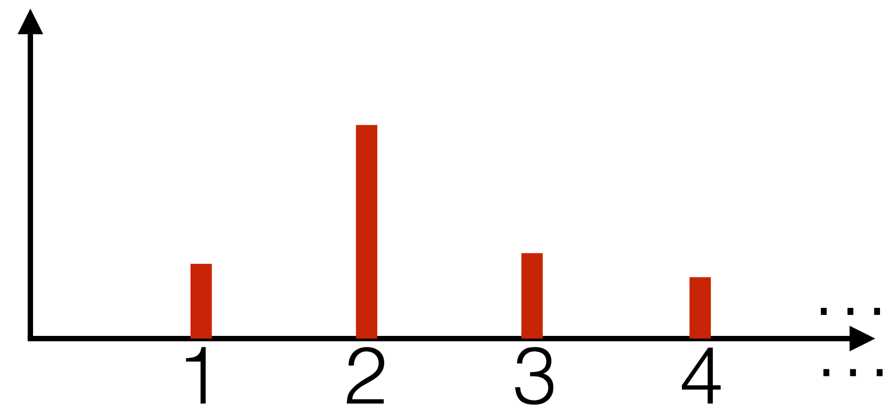


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

- Part of Dirichlet Process mixture model

Recall: Part I and Part II

$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$



$$z_n \stackrel{iid}{\sim} \mathrm{Categorical}(\rho)$$

- Part of Dirichlet Process mixture model

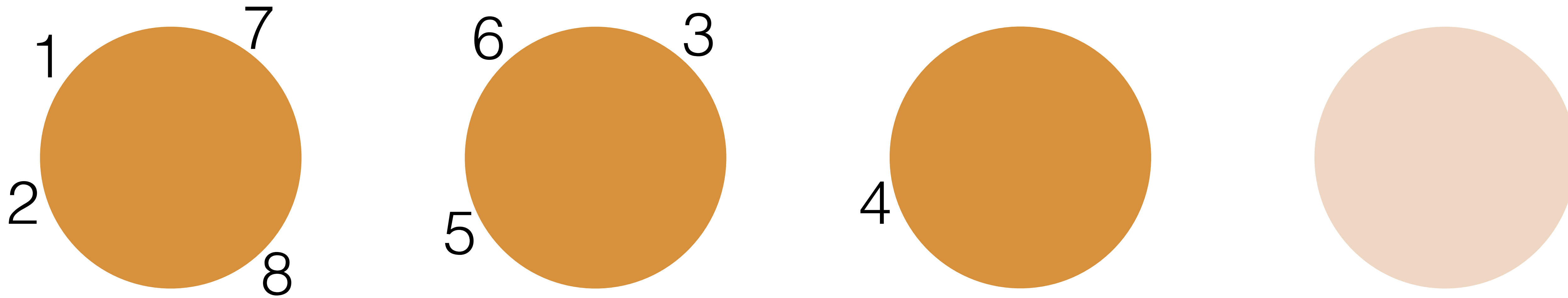
- Finite representation for inference?

Recall: Part II

Recall: Part II

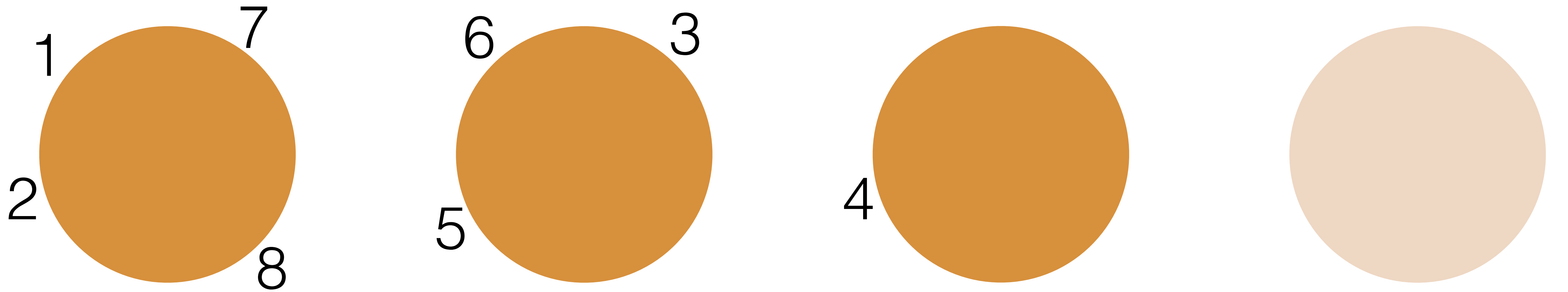
- Chinese restaurant process

Recall: Part II



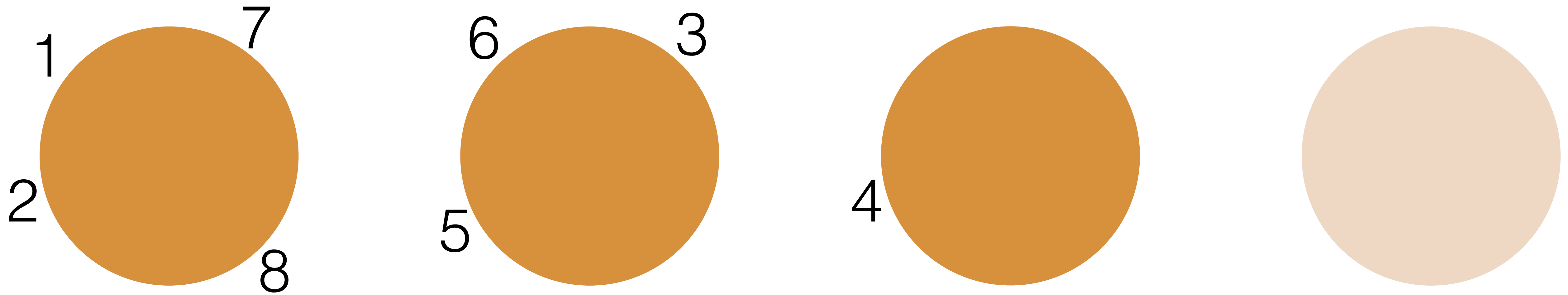
- Chinese restaurant process

Recall: Part II



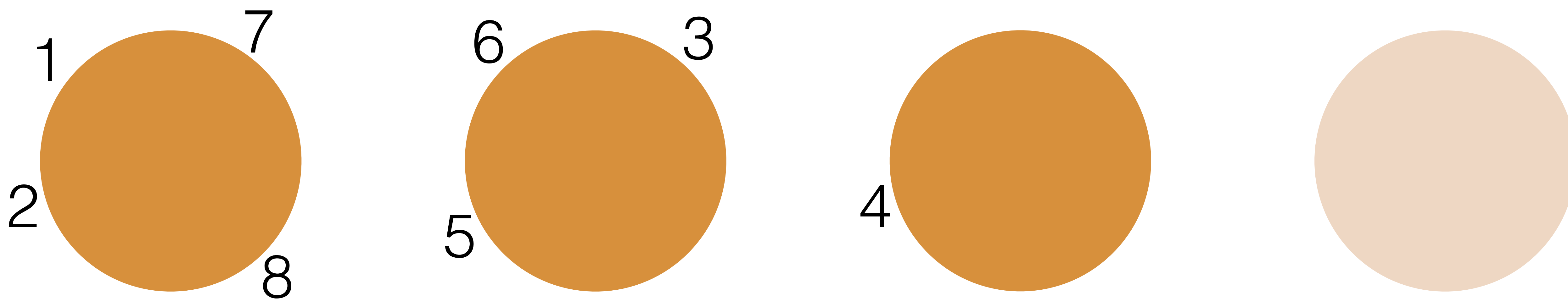
- Chinese restaurant process
- Each customer walks into the restaurant

Recall: Part II



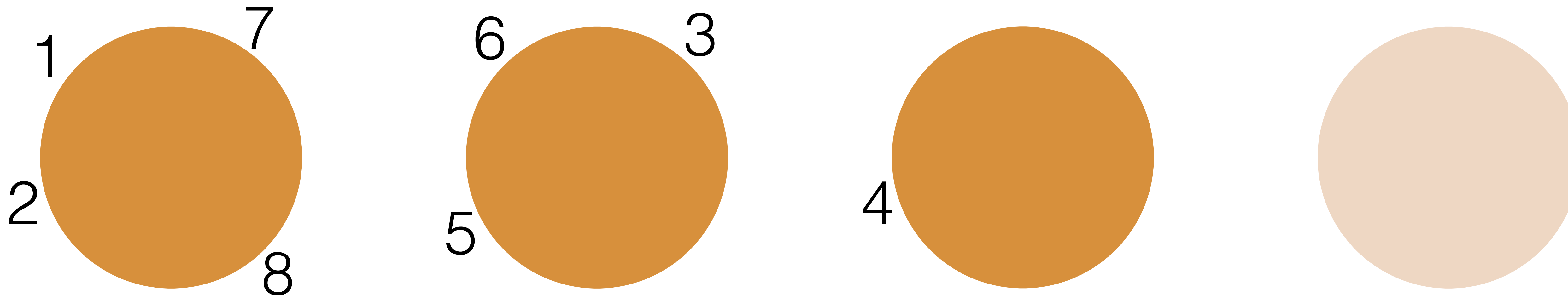
- Chinese restaurant process
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there

Recall: Part II



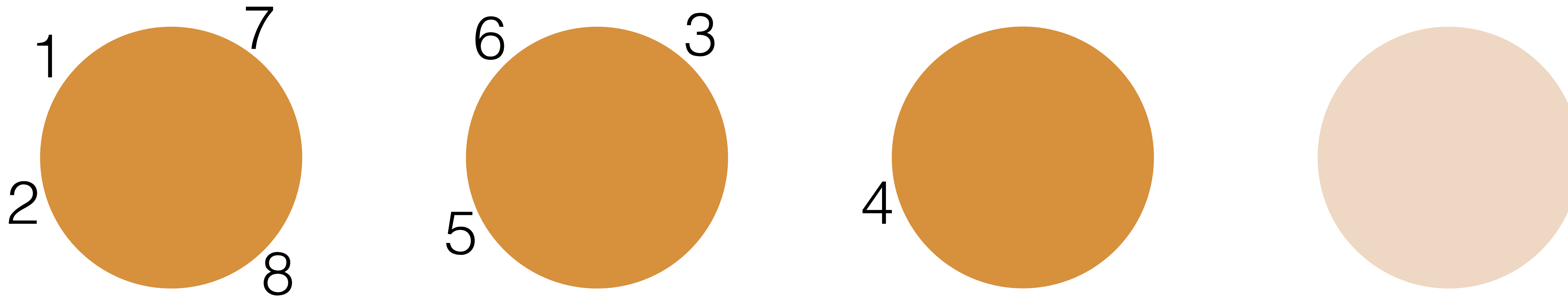
- Chinese restaurant process
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Recall: Part II



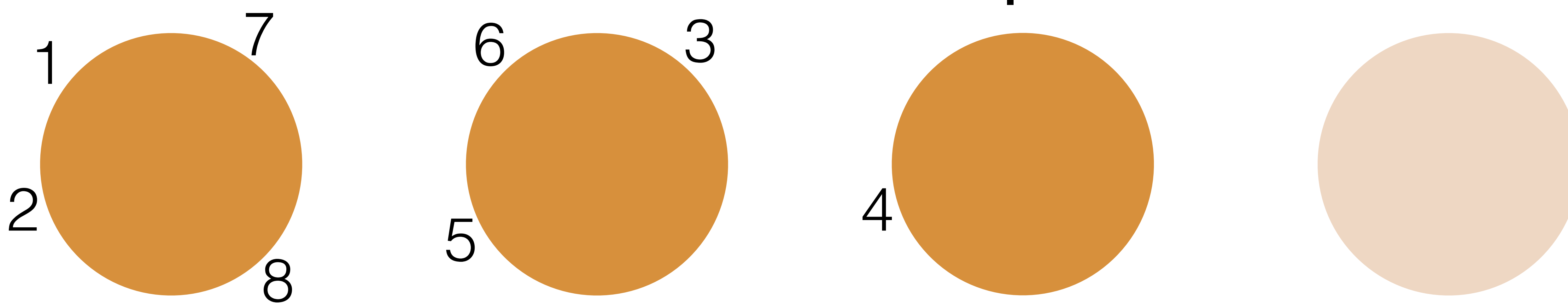
- Chinese restaurant process
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α
- "Partition" $\Pi_8 = \{\{1,2,7,8\},\{3,5,6\},\{4\}\}$

Recall: Part II



- 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
- "Partition" $\Pi_8 = \{\{1,2,7,8\},\{3,5,6\},\{4\}\}$
- Partition from a GEM(α) with categorical draws = same distribution as partition from a CRP(α)

Chinese restaurant process

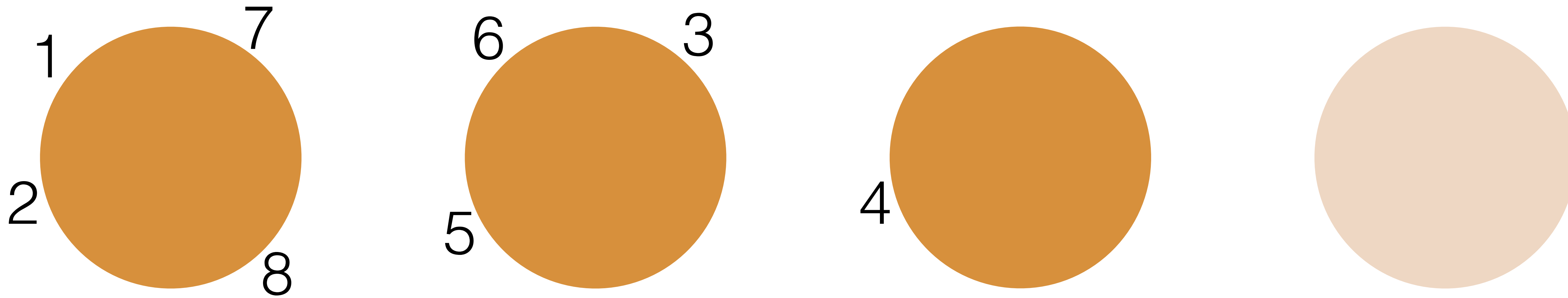


- Probability of this seating:

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

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

Chinese restaurant process

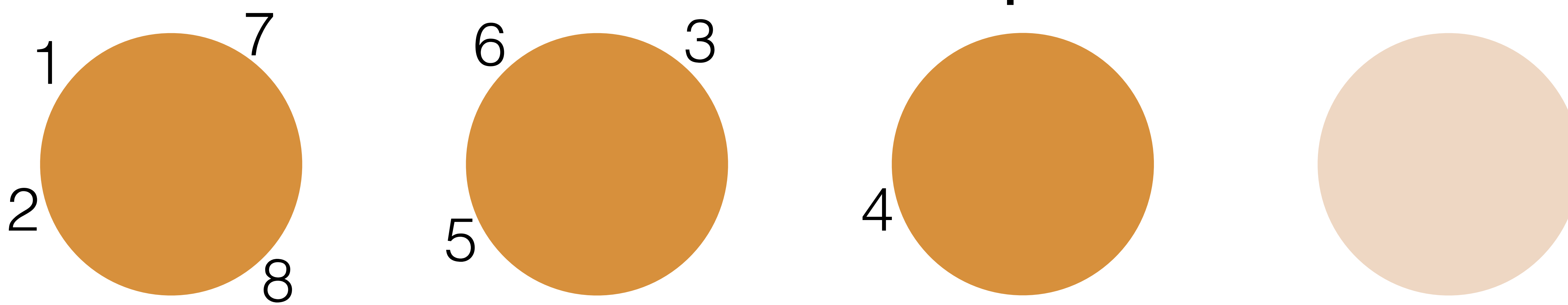


- Probability of this seating:

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

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

Chinese restaurant process



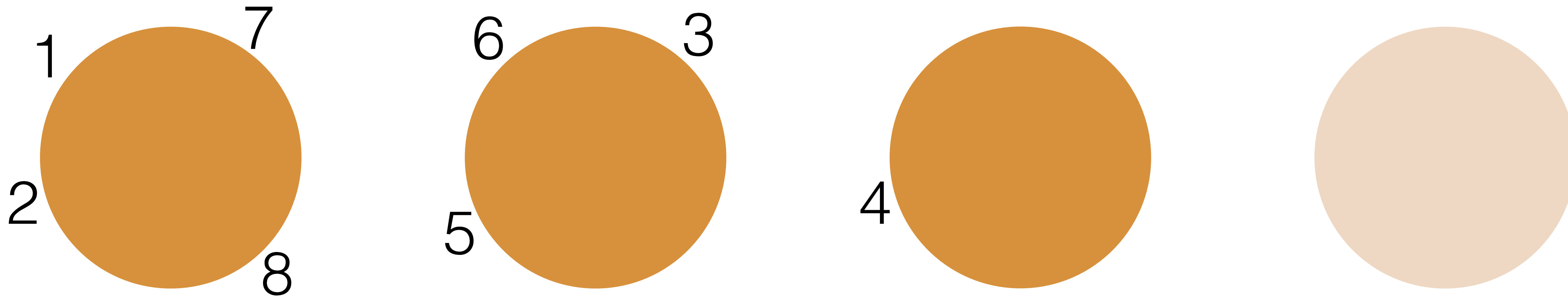
- Probability of this seating:

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

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

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

Chinese restaurant process



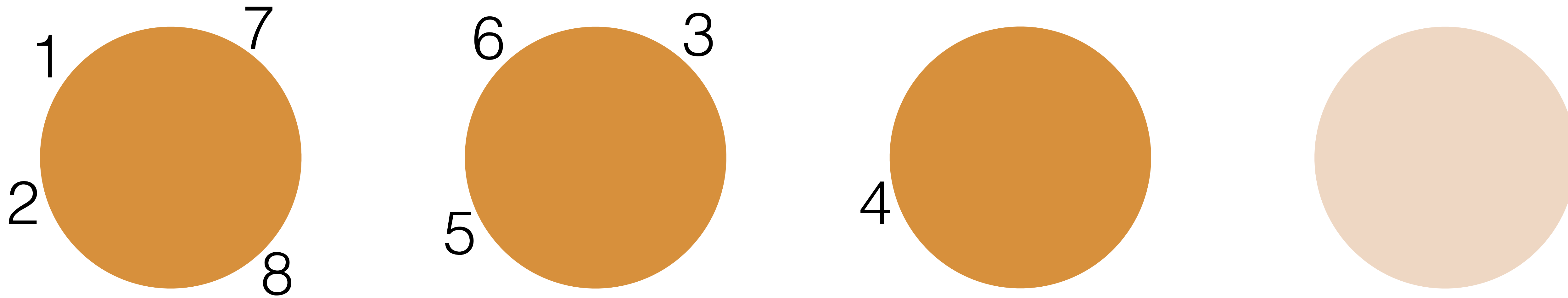
- Probability of this seating:

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

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

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

Chinese restaurant process



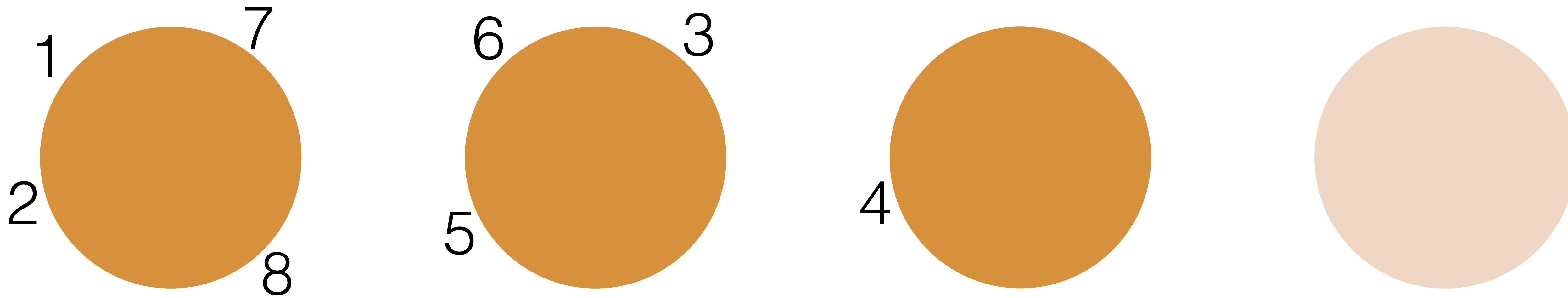
- Probability of this seating:

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

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

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

Chinese restaurant process



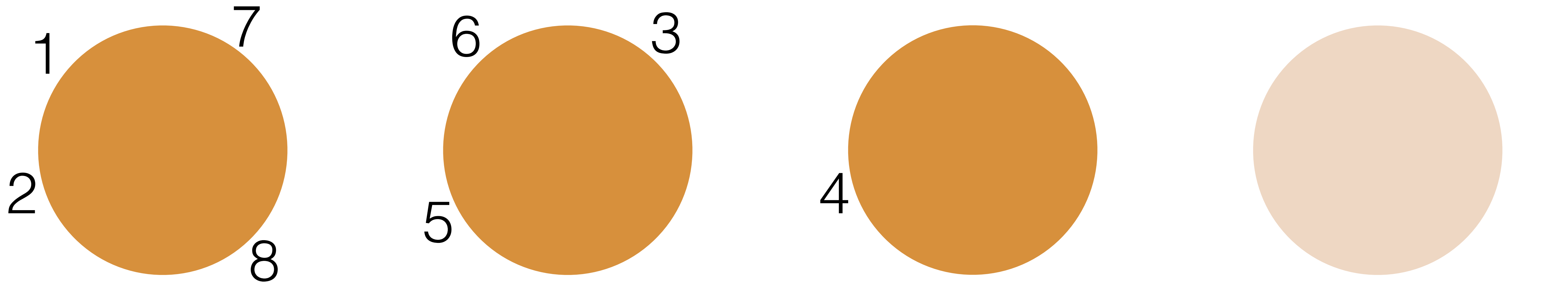
- Probability of this seating:

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

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

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

Chinese restaurant process



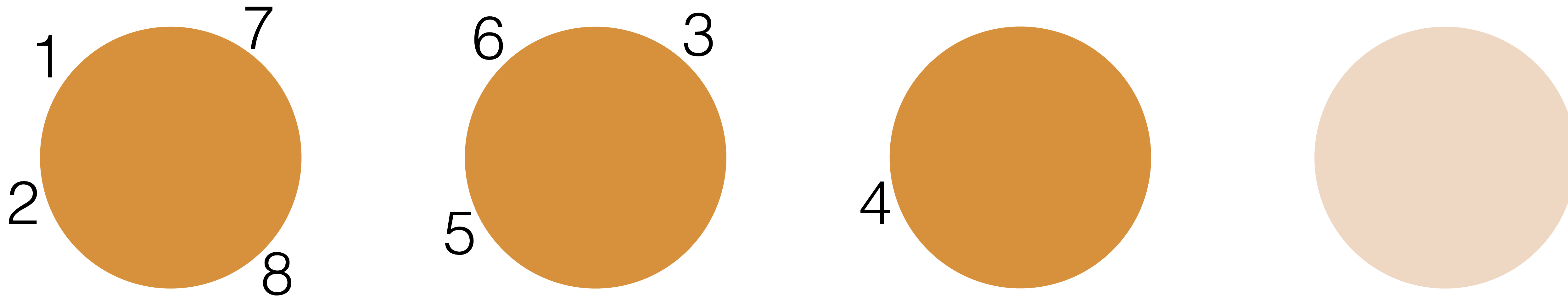
- Probability of this seating:

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

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

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

Chinese restaurant process



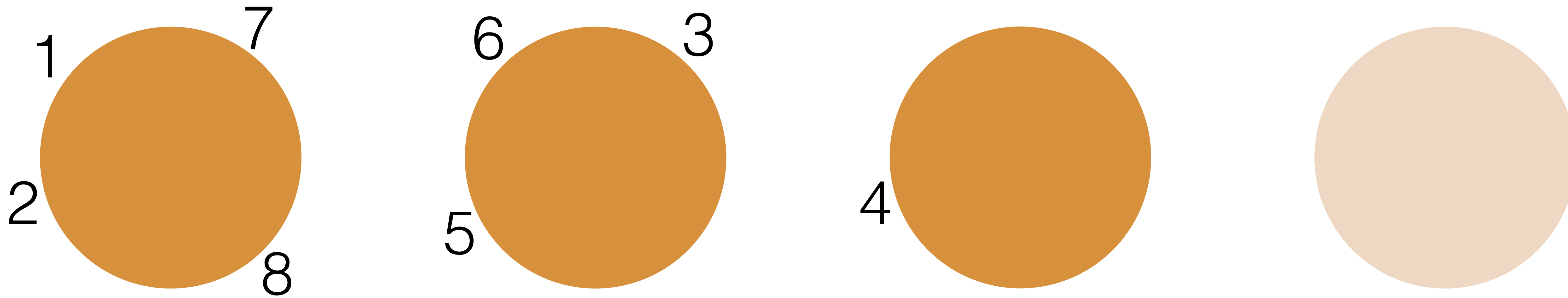
- Probability of this seating:

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

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

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

Chinese restaurant process



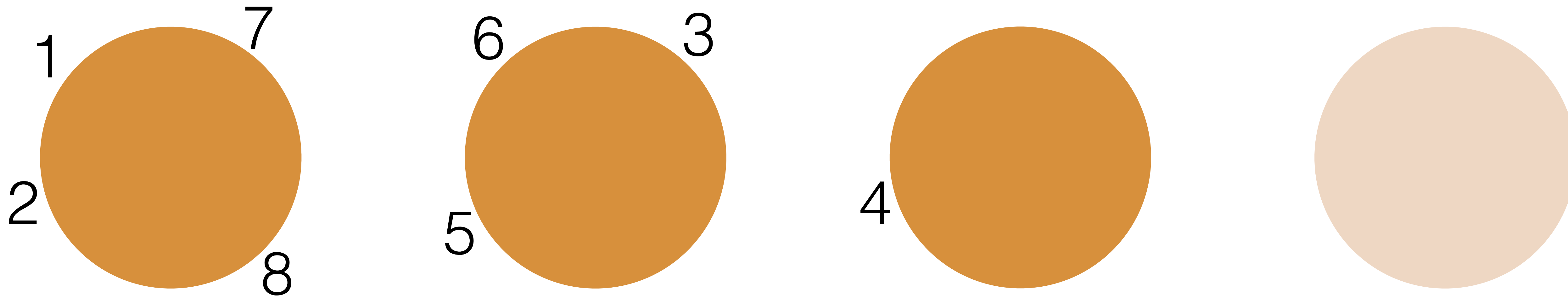
- Probability of this seating:

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

- 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



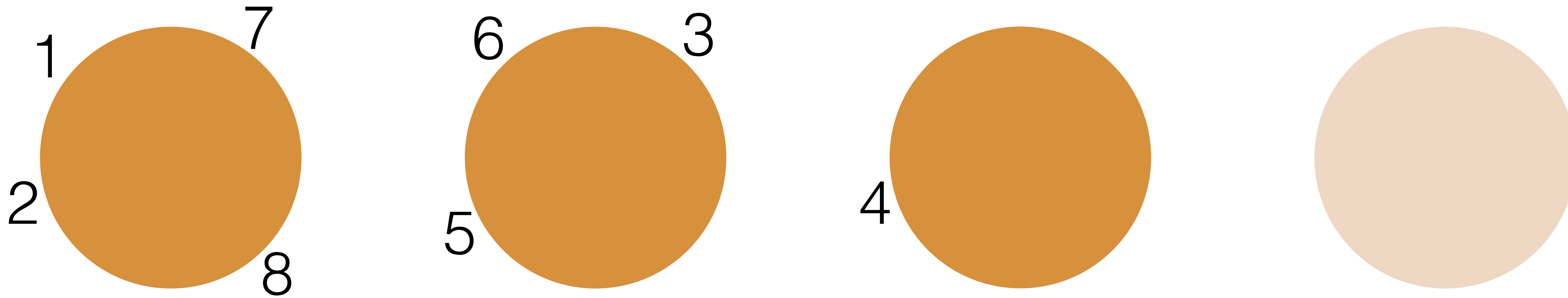
- Probability of this seating:

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

- 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



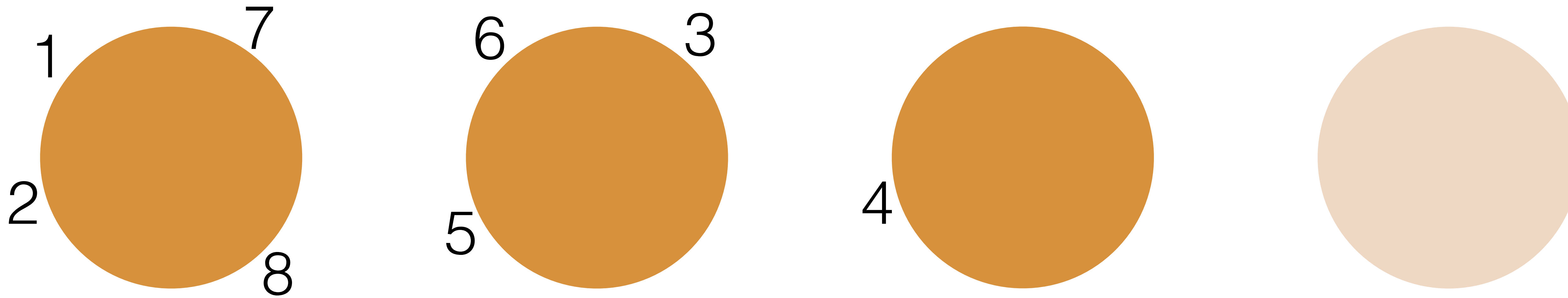
- Probability of this seating:

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

- 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:

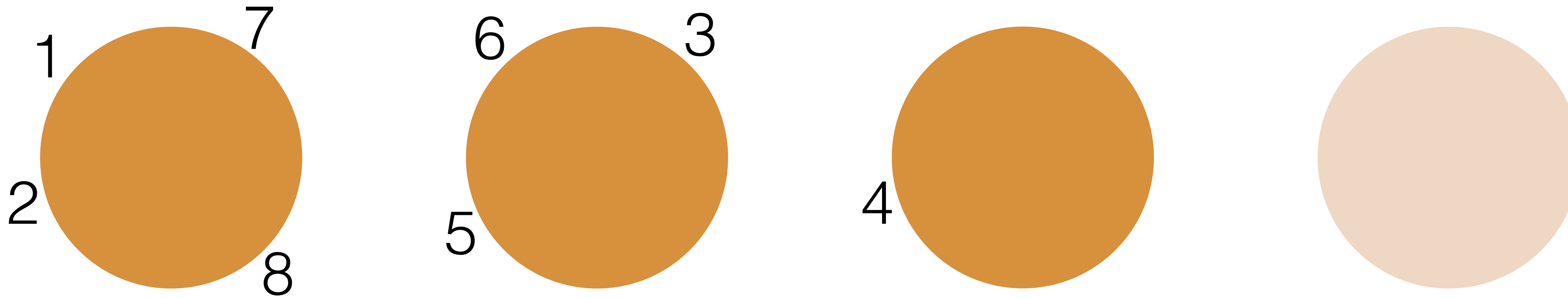
$$\frac{\alpha}{\alpha} \cdot \frac{1}{\alpha+1} \cdot \frac{\alpha}{\alpha+2} \cdot \frac{\alpha}{\alpha+3} \cdot \frac{1}{\alpha+4} \cdot \frac{2}{\alpha+5} \cdot \frac{2}{\alpha+6} \cdot \frac{3}{\alpha+7}$$

- 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:

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

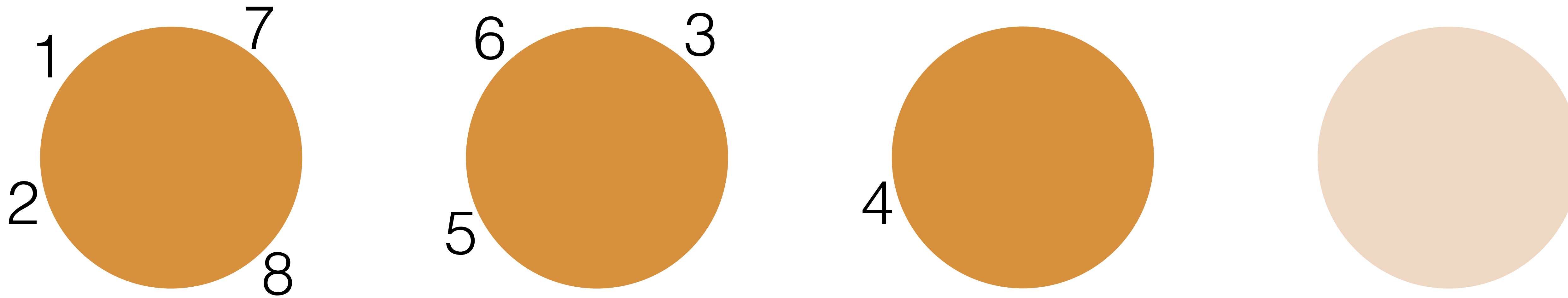
- Probability of *N* customers (K_N tables, #*C* at table *C*):

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

- 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*):

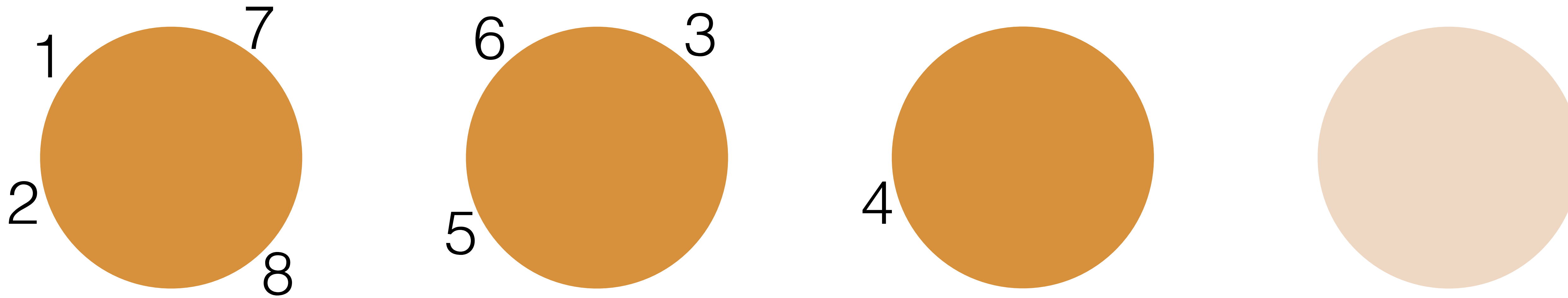
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

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

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*):

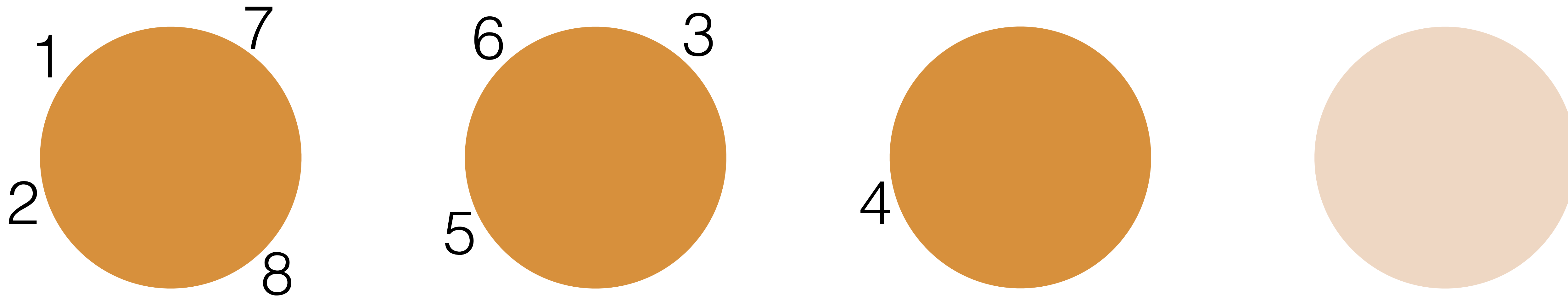
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- 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 $p(\Pi_N | \Pi_{N,\text{-}n})$
    - e.g. $\Pi_{8,\text{-}5} = \{\{1,2,7,8\},\{3,6\},\{4\}\}$

Chinese restaurant process



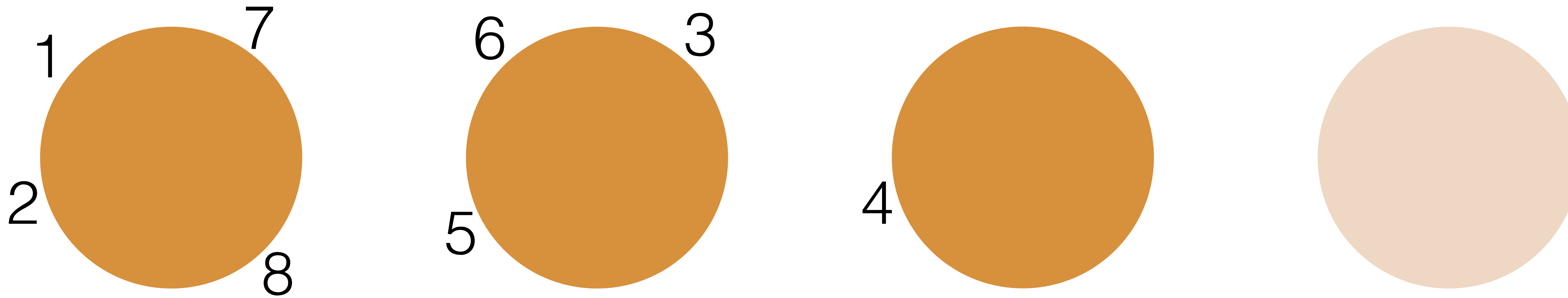
- Probability of N customers (K_N tables, $\#C$ at table C):

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

- So:

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

Chinese restaurant process



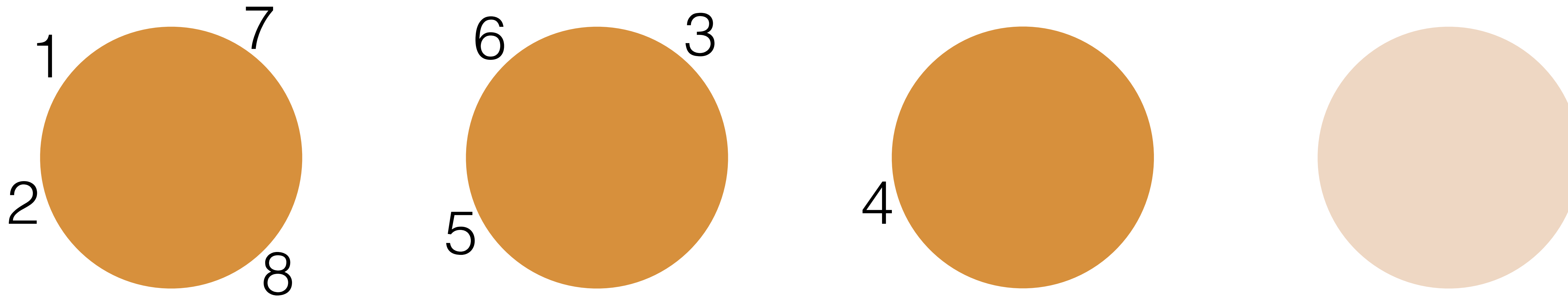
- Probability of N customers (K_N tables, $\#C$ at table C):

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

- So:

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

Chinese restaurant process

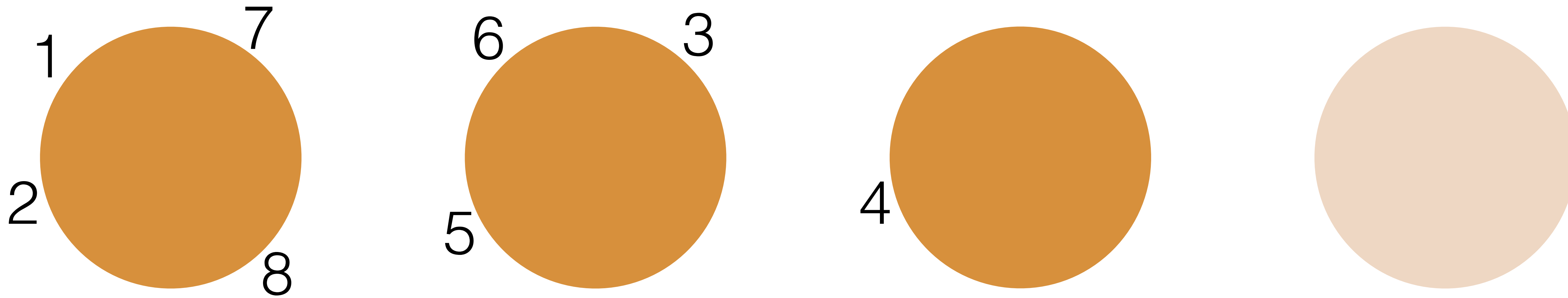


- Probability of N customers (K_N tables, $\#C$ at table C):

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

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

Chinese restaurant process



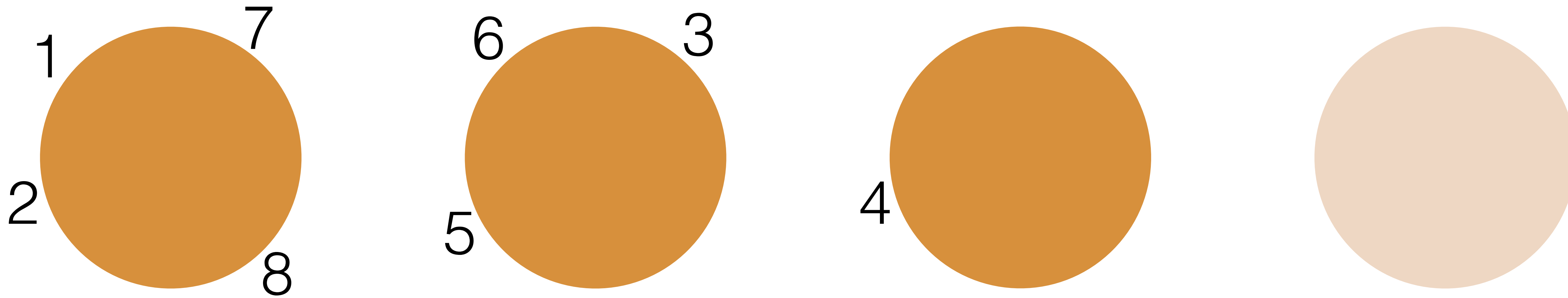
- Probability of N customers (K_N tables, #C at table C):

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

- So:

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

Chinese restaurant process



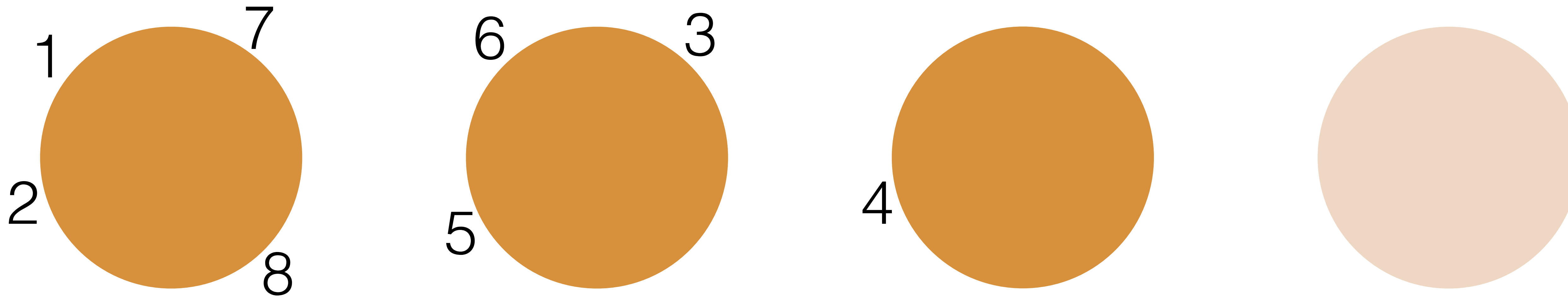
- Probability of *N* customers (K_N tables, #*C* at table *C*):

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

- So:

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

Chinese restaurant process



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

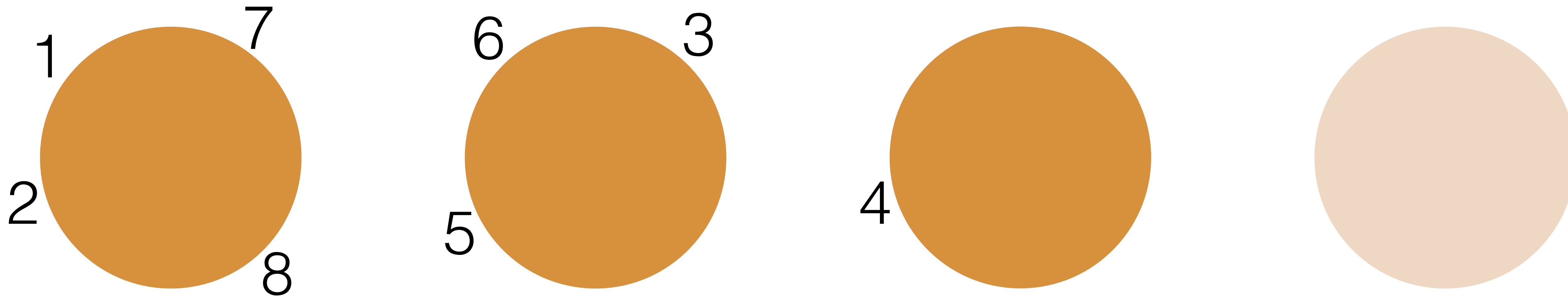
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- 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):

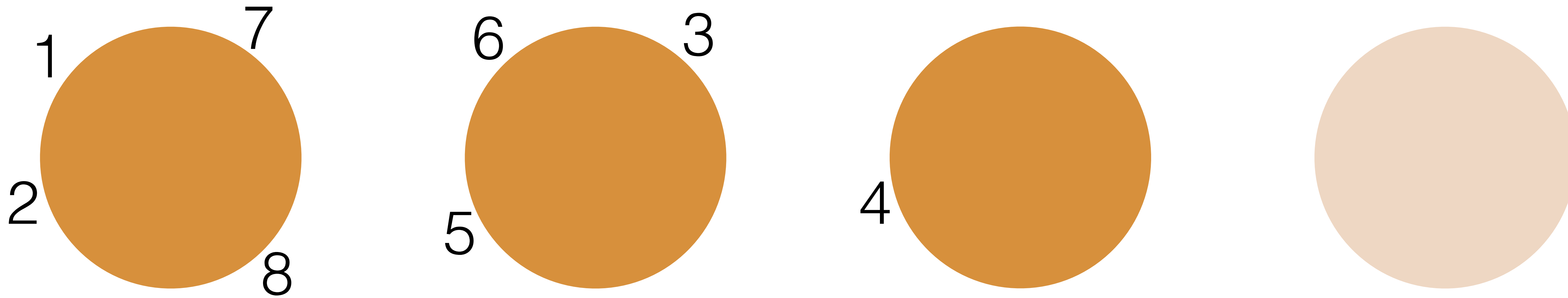
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

- 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



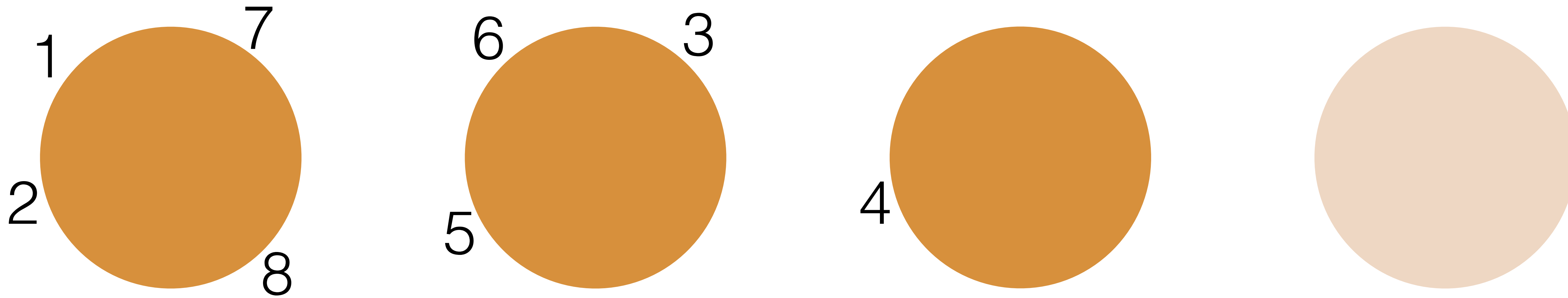
- Probability of *N* customers (K_N tables, #*C* at table *C*):

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

- 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



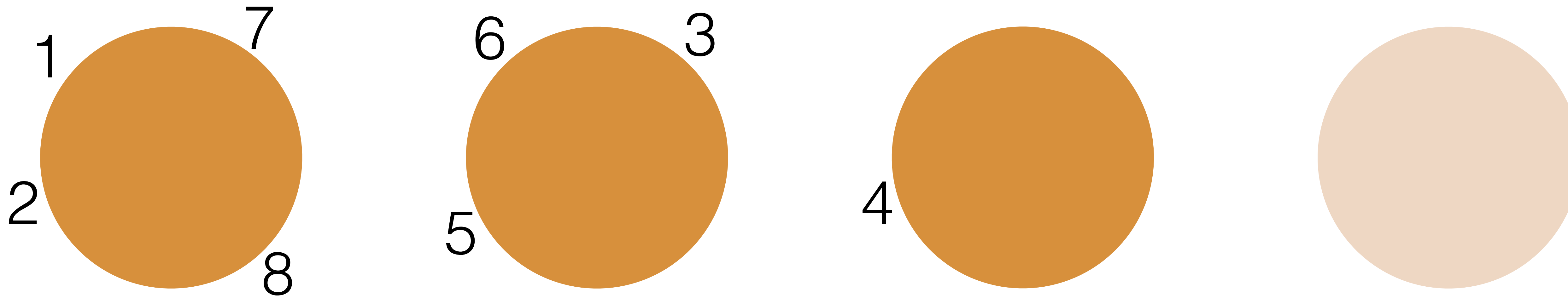
- Probability of N customers (K_N tables, $\#C$ at table C):

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

- 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):

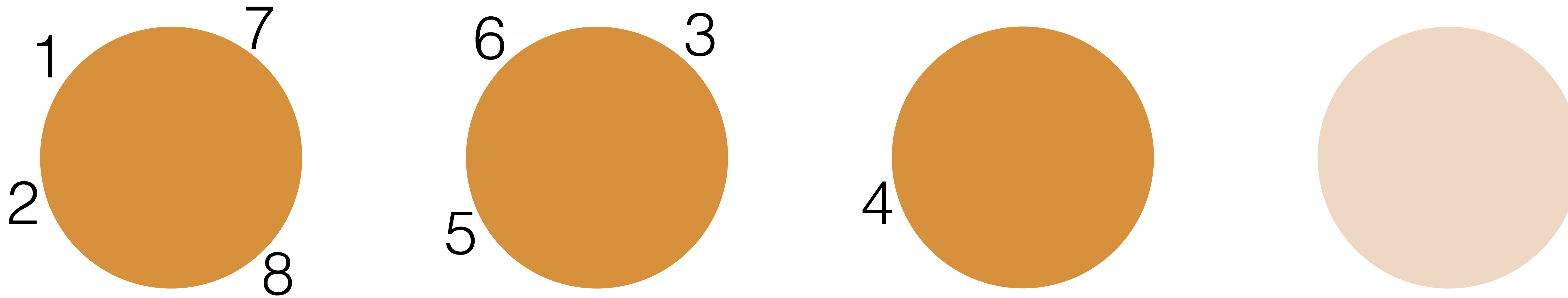
$$\frac{\alpha^{K_N} \prod_{C \in \Pi_N} (\#C - 1)!}{\alpha \cdots (\alpha + N - 1)} = \mathbb{P}(\Pi_N = \pi_N)$$

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

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

Chinese restaurant process



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

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

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

- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - *t* th step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

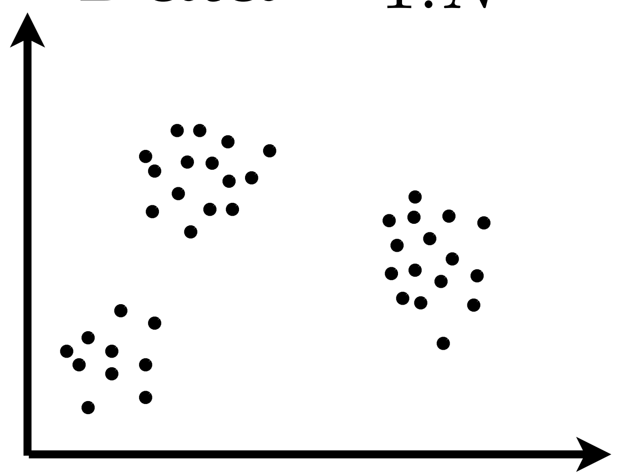
    $v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$

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

CRP mixture model: inference

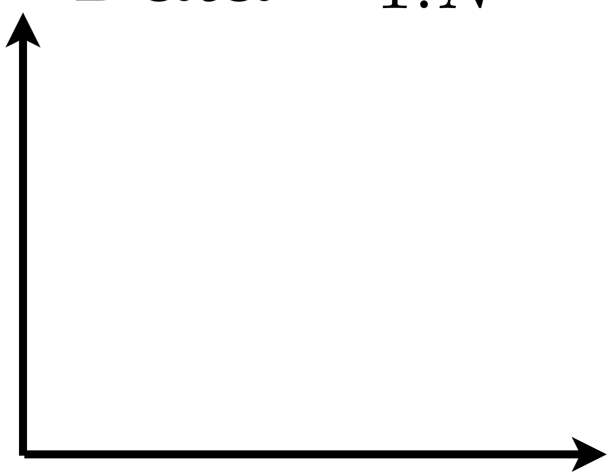
CRP mixture model: inference

- Data $x_{1:N}$



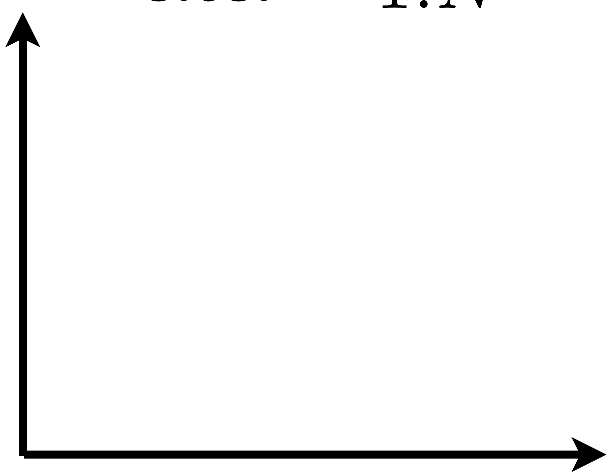
CRP mixture model: inference

- Data $x_{1:N}$



CRP mixture model: inference

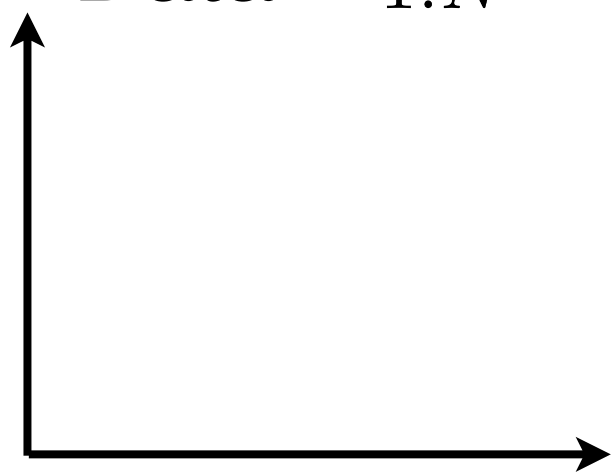
- Data $x_{1:N}$



- Generative model

CRP mixture model: inference

- Data $x_{1:N}$

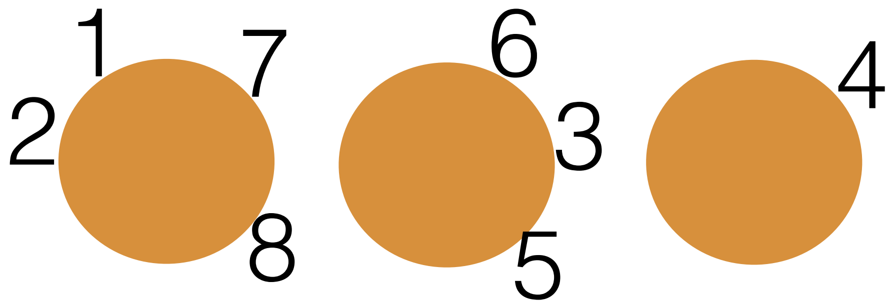


- Generative model
  $\Pi_N \sim \mathrm{CRP}(N, \alpha)$

CRP mixture model: inference

- Data $x_{1:N}$

- Generative model
  $\Pi_N \sim \mathrm{CRP}(N, \alpha)$



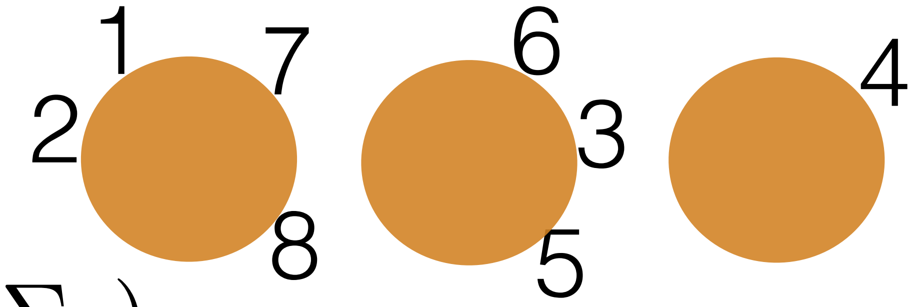
CRP mixture model: inference

- Data $x_{1:N}$

- Generative model

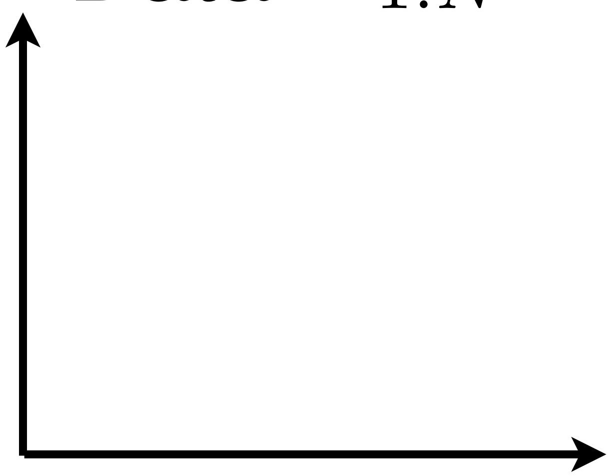
$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

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



CRP mixture model: inference

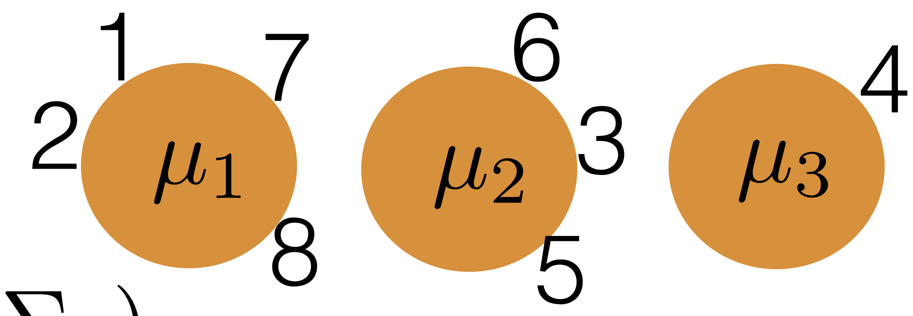
- Data $x_{1:N}$



- Generative model

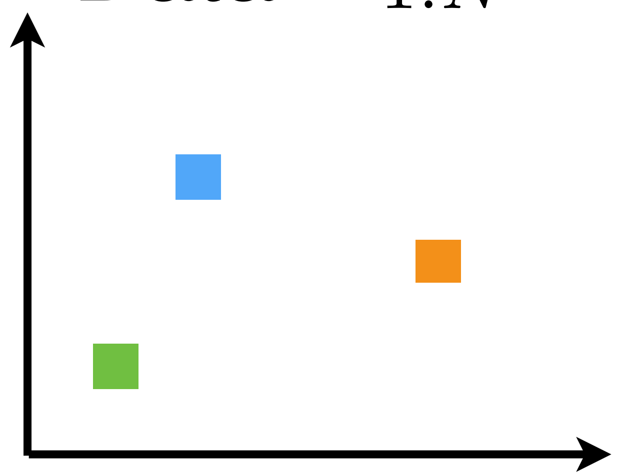
$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

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



CRP mixture model: inference

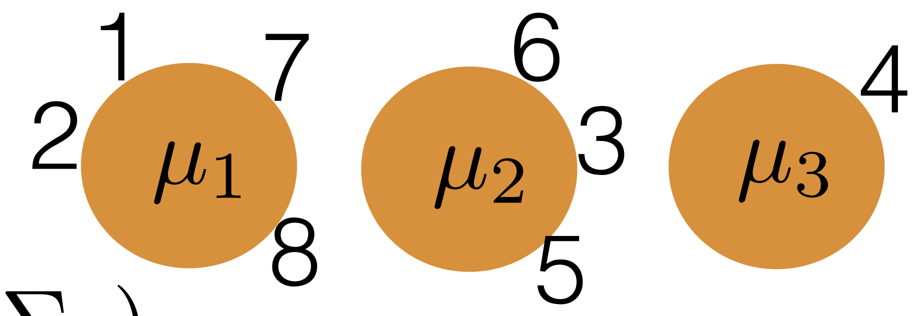
- Data $x_{1:N}$



- Generative model

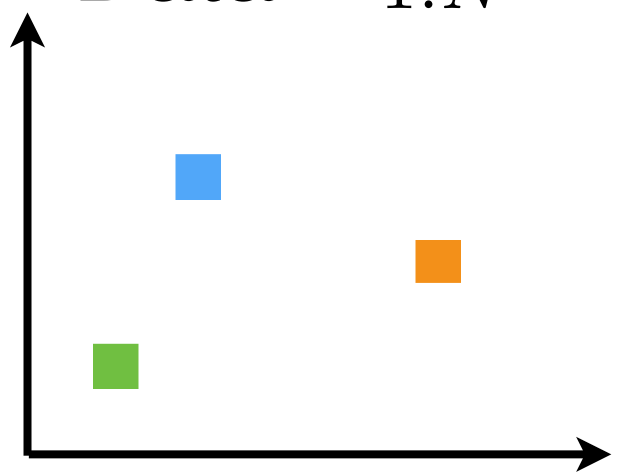
$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$

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



CRP mixture model: inference

- Data $x_{1:N}$

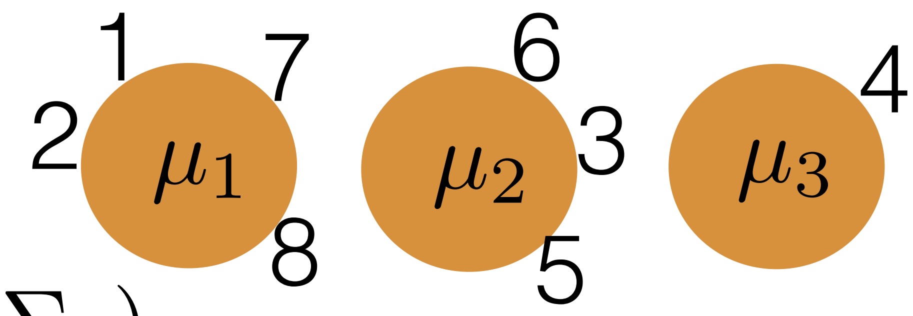


- Generative model

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

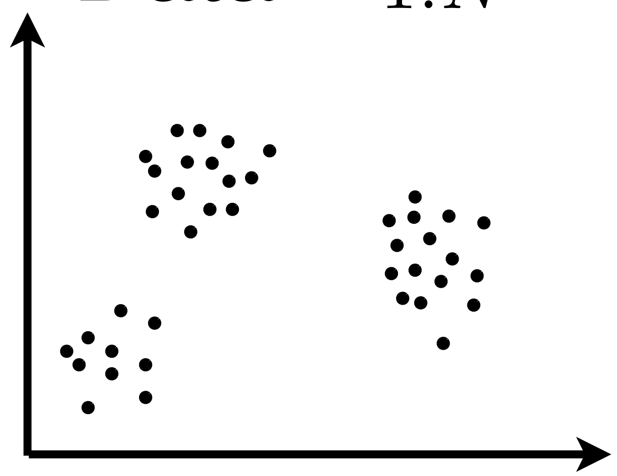
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

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



CRP mixture model: inference

- Data $x_{1:N}$

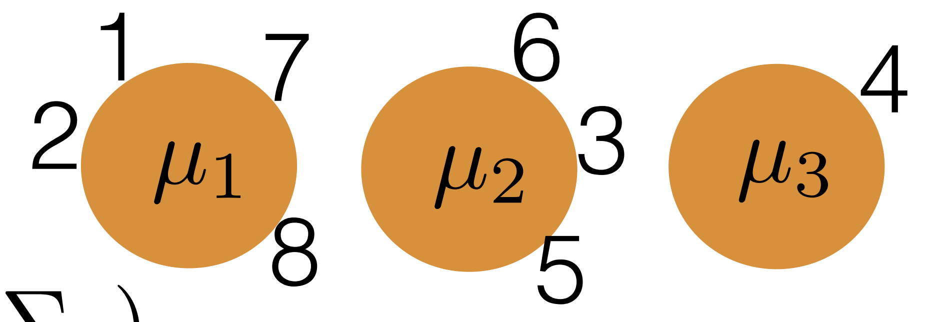


- Generative model

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

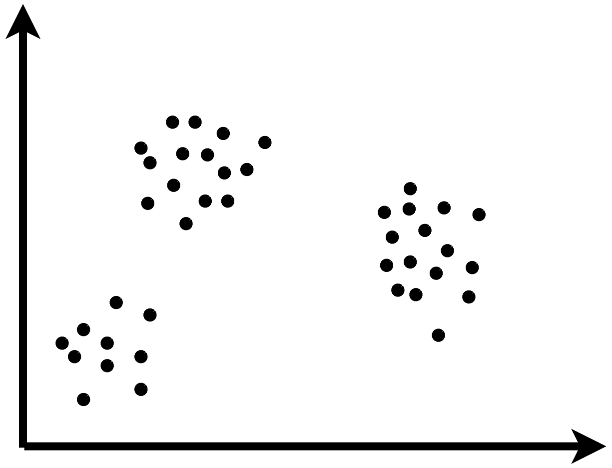
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

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



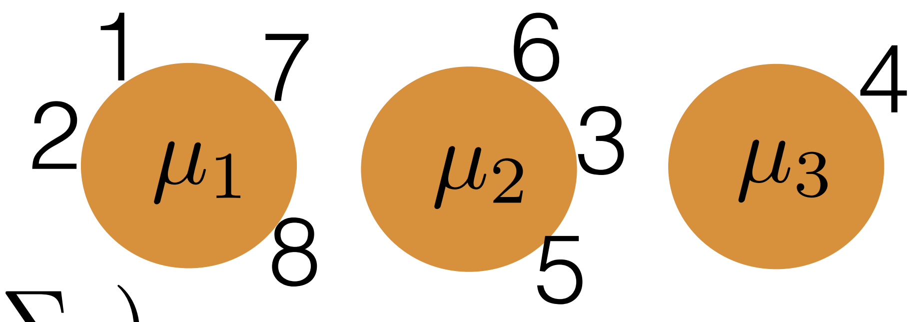
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

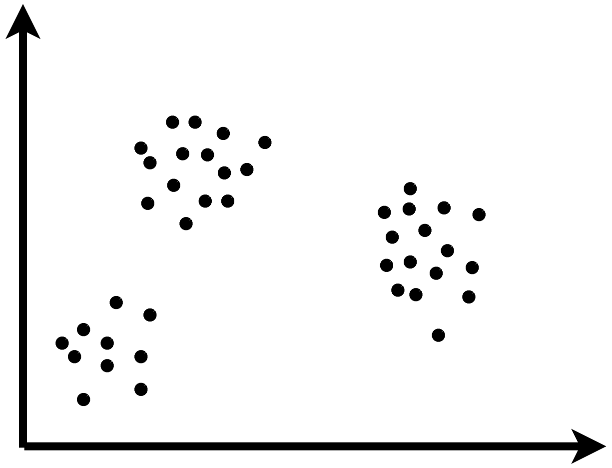
$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



- Want: posterior

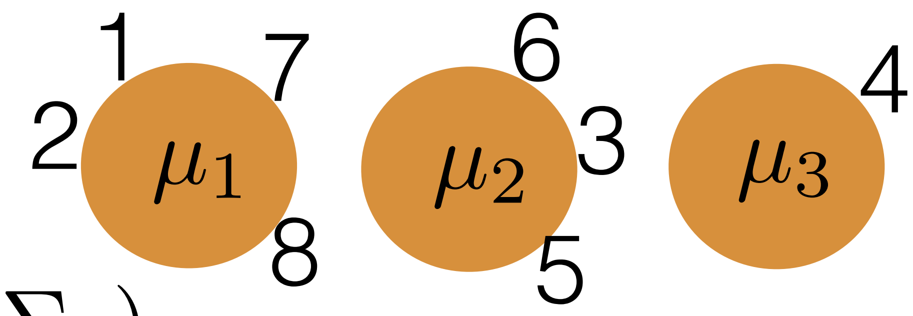
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

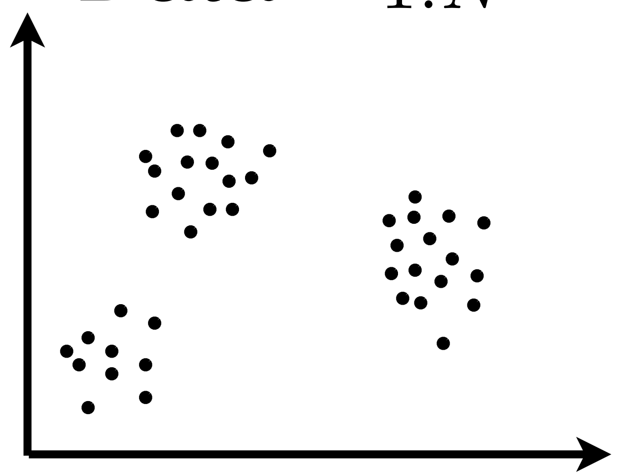
$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \mu_C \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



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

CRP mixture model: inference

- Data $x_{1:N}$

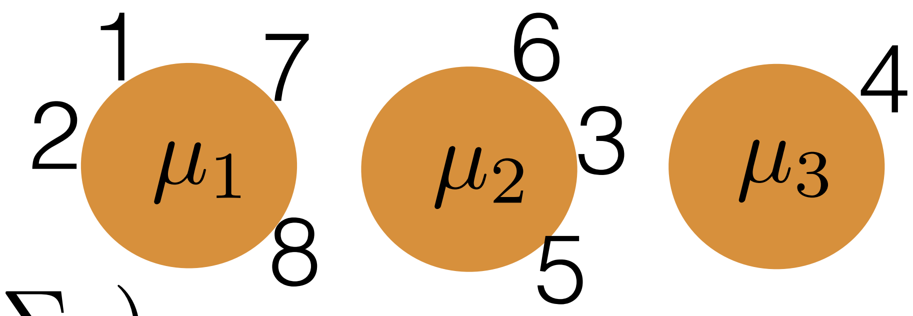


- Generative model

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

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

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

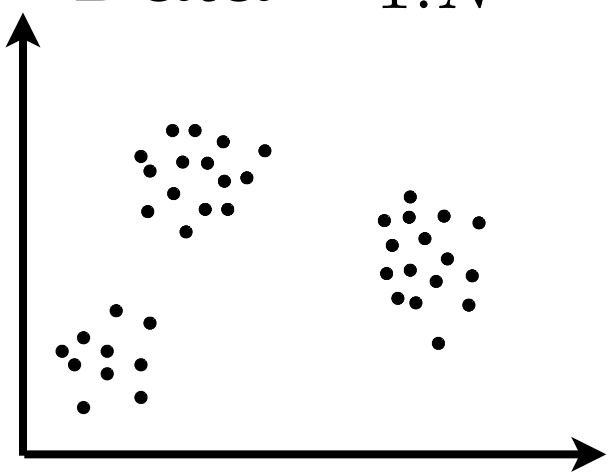


- Want: posterior $p(\Pi_N | x_{1:N})$

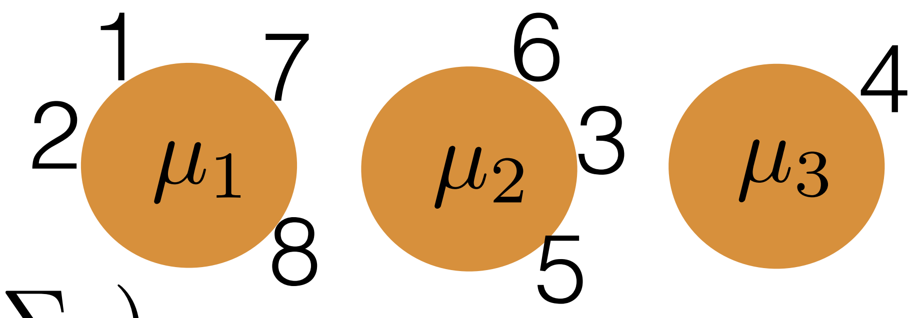
- Gibbs sampler:

CRP mixture model: inference

- Data $x_{1:N}$



- Generative model



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

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

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

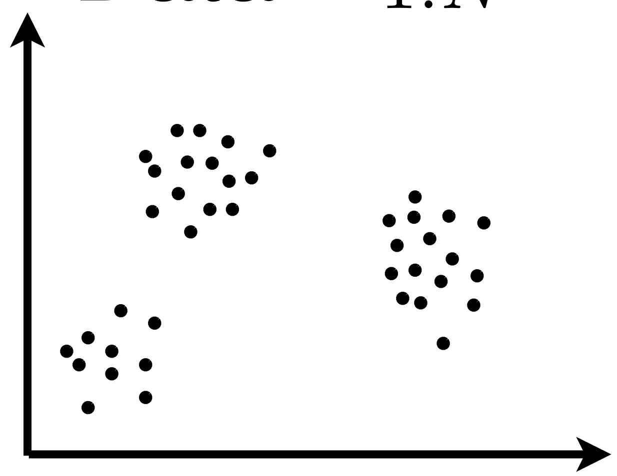
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

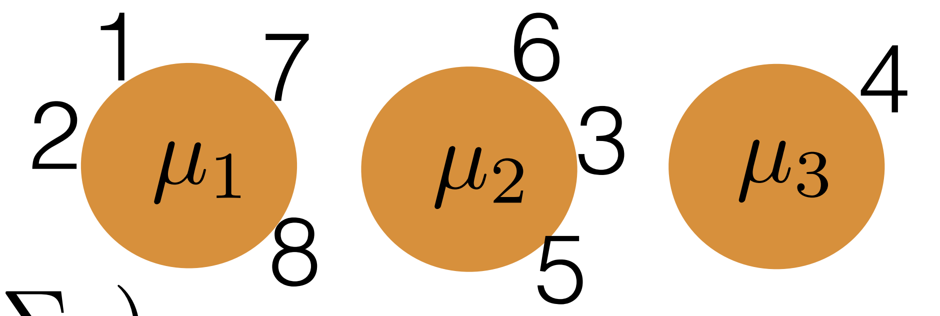
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



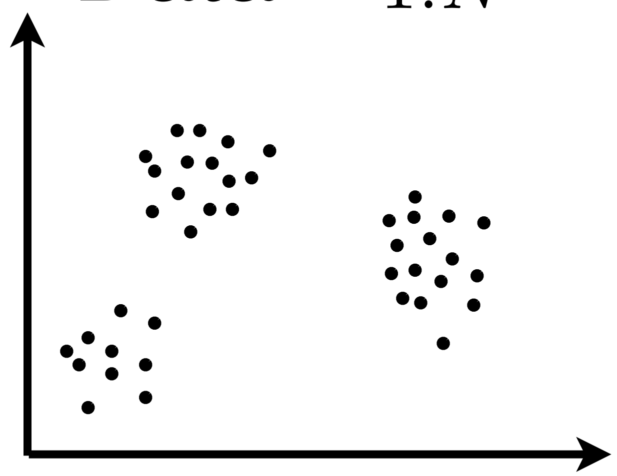
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

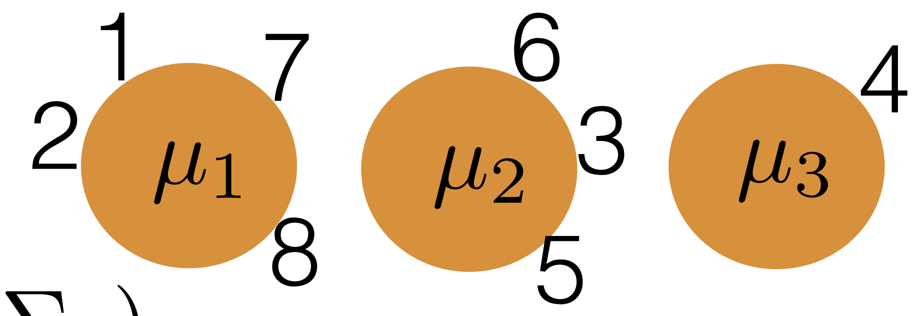
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



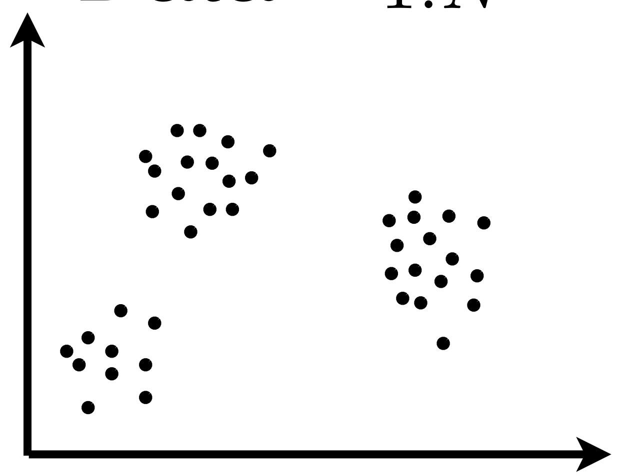
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

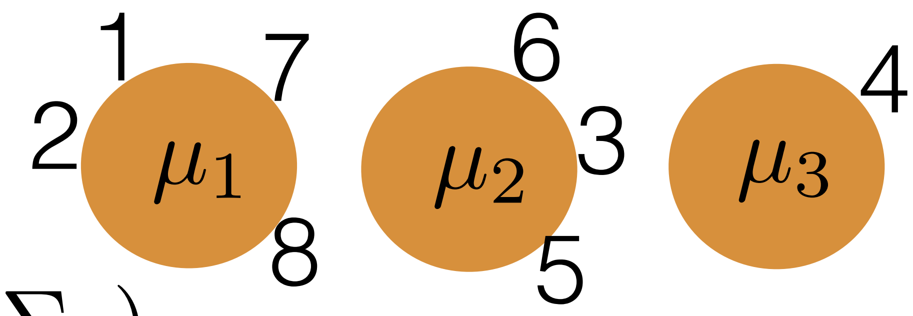
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



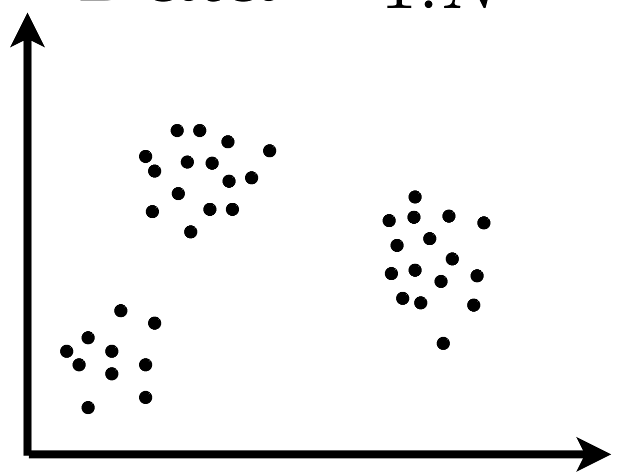
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

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

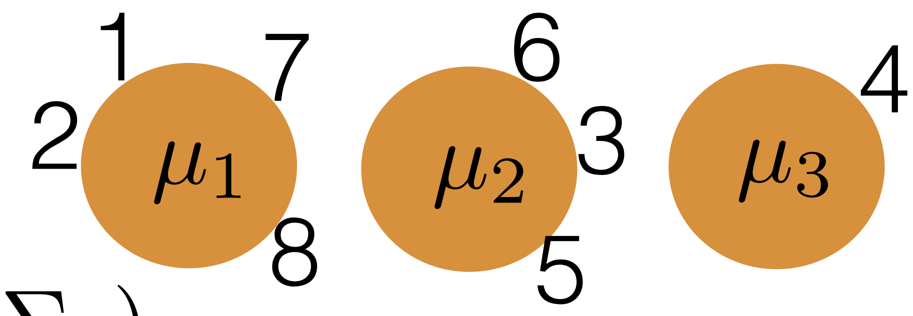
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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

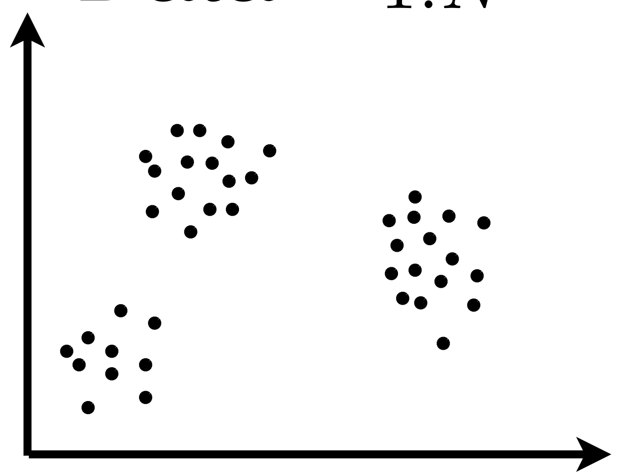


- Want: posterior $p(\Pi_N | x_{1:N})$
- Gibbs sampler:

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

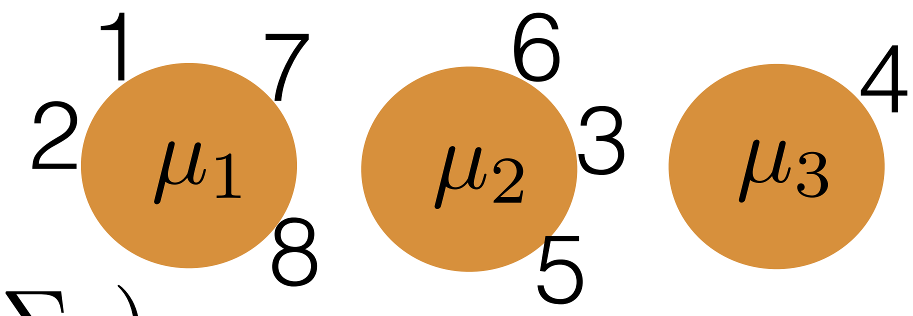
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



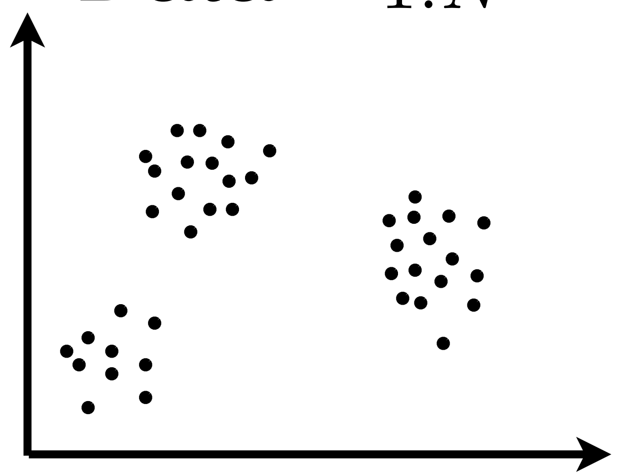
- Want: posterior $p(\Pi_N | x_{1:N})$

- Gibbs sampler:

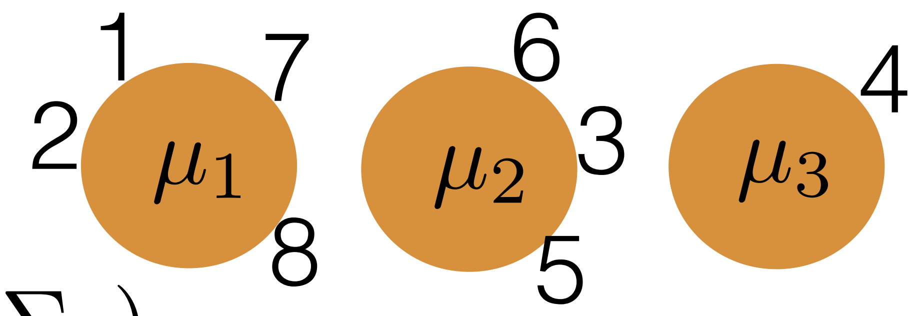
$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

CRP mixture model: inference

- Data $x_{1:N}$



- Generative model



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

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

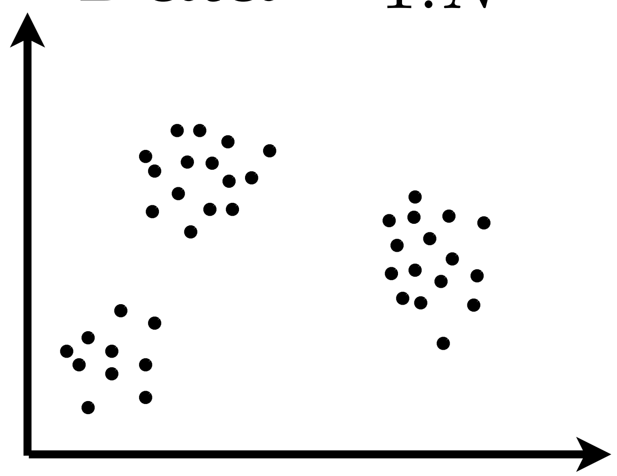
- Gibbs sampler:

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

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

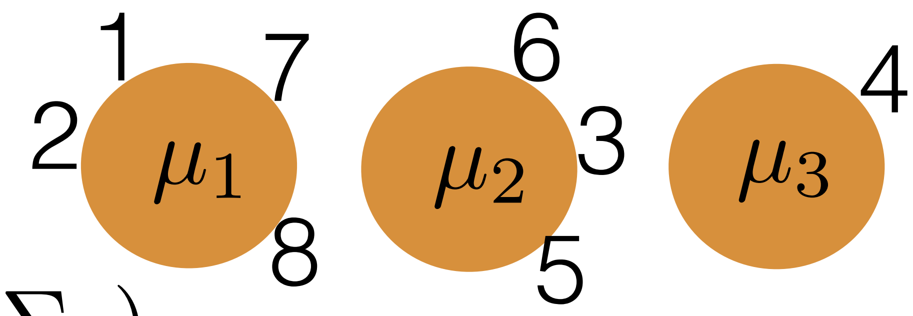
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



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

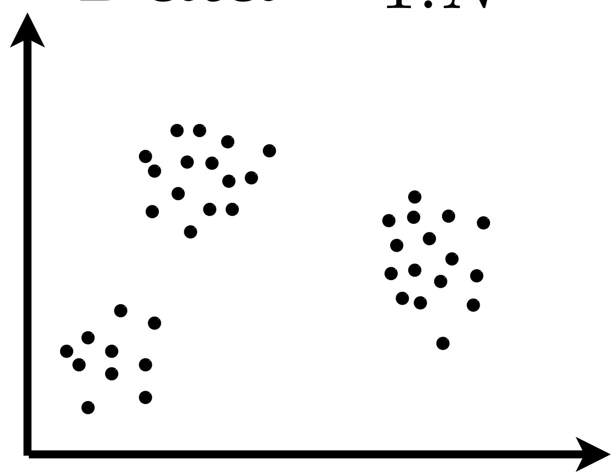
- Gibbs sampler:

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

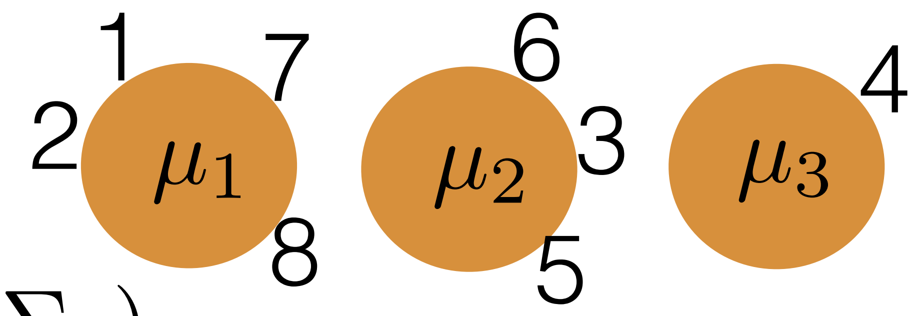
- For completeness: $p(x_{C\cup\{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

CRP mixture model: inference

- Data $x_{1:N}$



- Generative model



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

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

- Gibbs sampler:

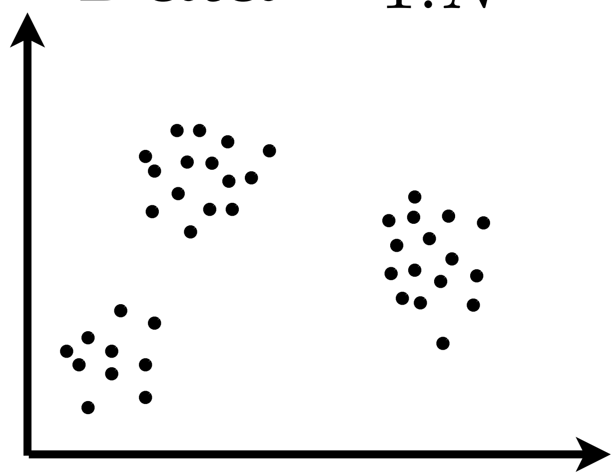
$$p(\Pi_N | \Pi_{N,\text{-}n}, x) = \begin{cases} \frac{\#C}{\alpha+N-1} p(x_{C\cup\{n\}} | x_C) & \text{if } n \text{ joins cluster } C \\ \frac{\alpha}{\alpha+N-1} p(x_{\{n\}}) & \text{if } n \text{ starts a new cluster} \end{cases}$$

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

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

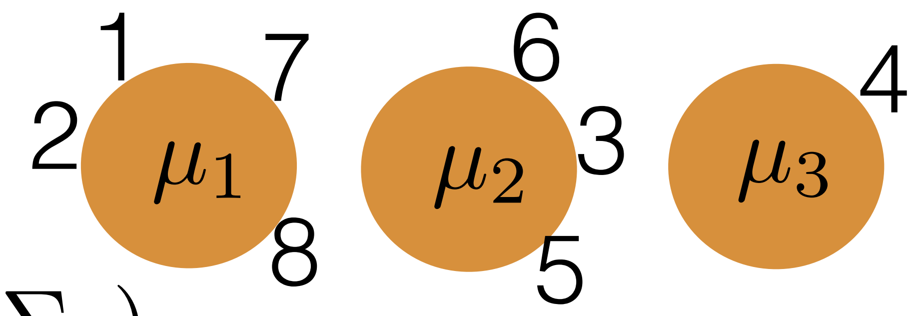
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



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

- Gibbs sampler:

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

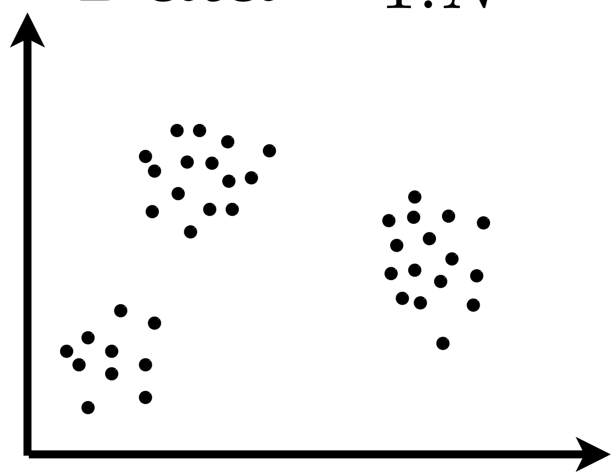
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

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

[MacEachern 1994; Neal 1992; Neal 2000]

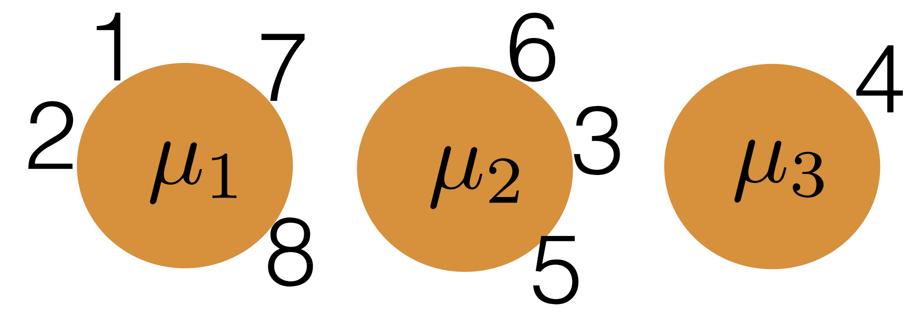
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

$$\Pi_N \sim \mathrm{CRP}(N, \alpha)$$
$$\forall C \in \Pi_N, \phi_C \overset{iid}{\sim} G_0$$
$$\forall C \in \Pi_N, \forall n \in C, x_n \overset{indep}{\sim} \mathcal{N}(\mu_C, \Sigma)$$



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

- Gibbs sampler:

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

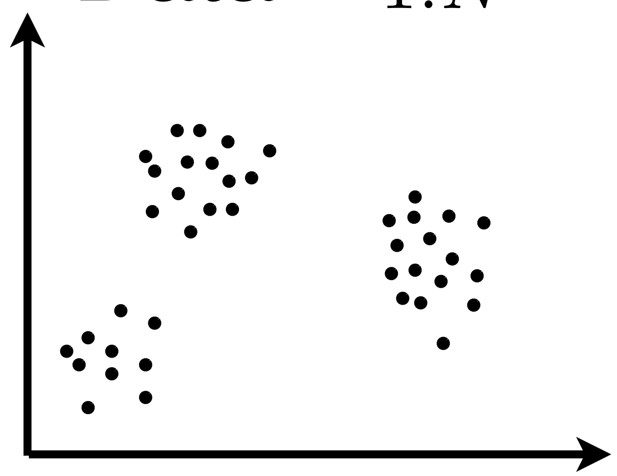
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

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

[MacEachern 1994; Neal 1992; Neal 2000]

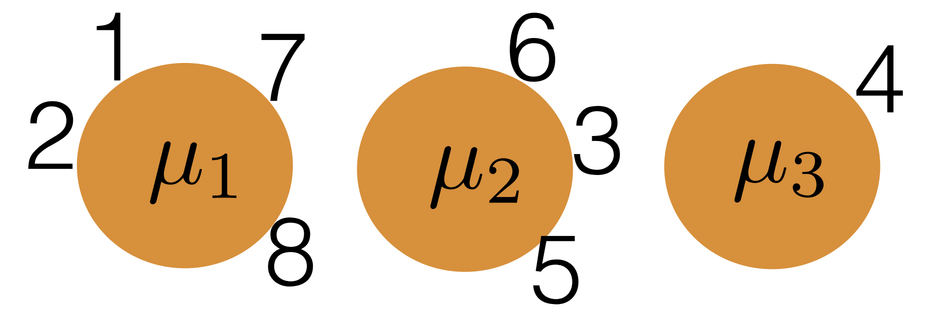
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



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

- Gibbs sampler:

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

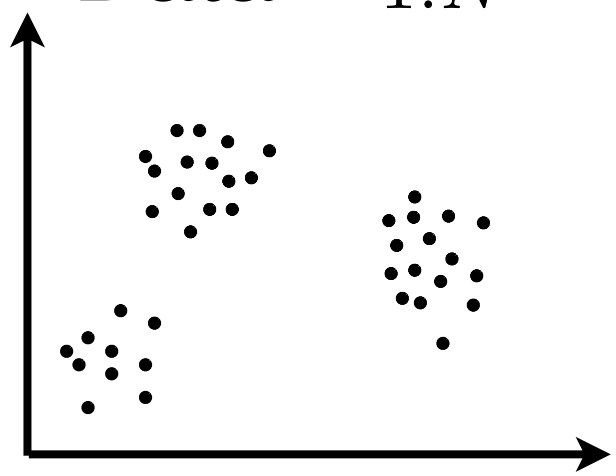
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

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

[MacEachern 1994; Neal 1992; Neal 2000]

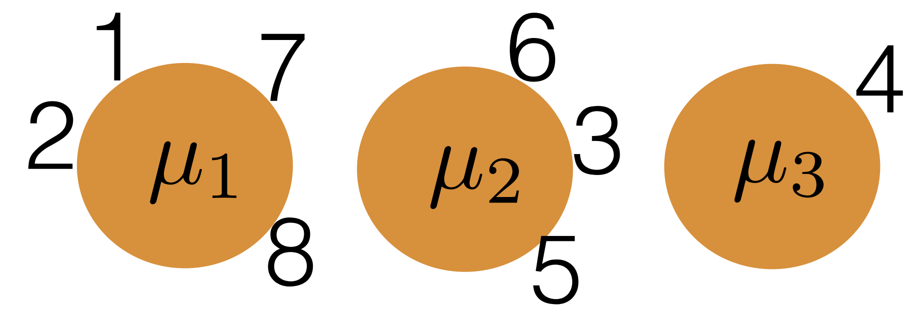
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



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

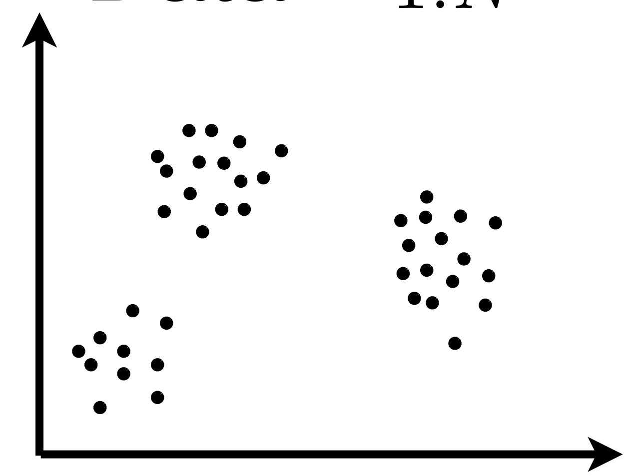
- Gibbs sampler:

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

[MacEachern 1994; Neal 1992; Neal 2000]

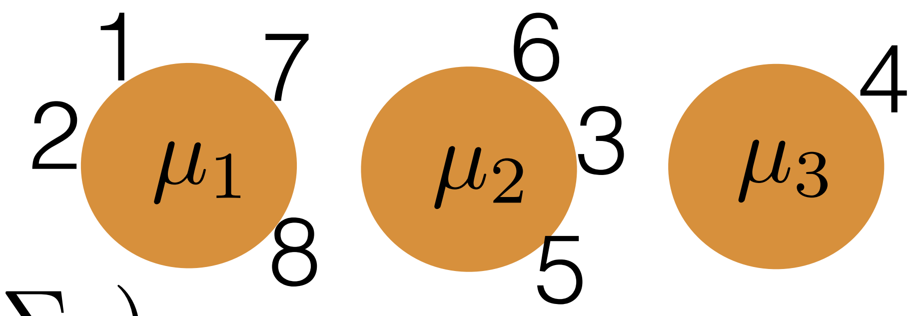
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



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

- Gibbs sampler:

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

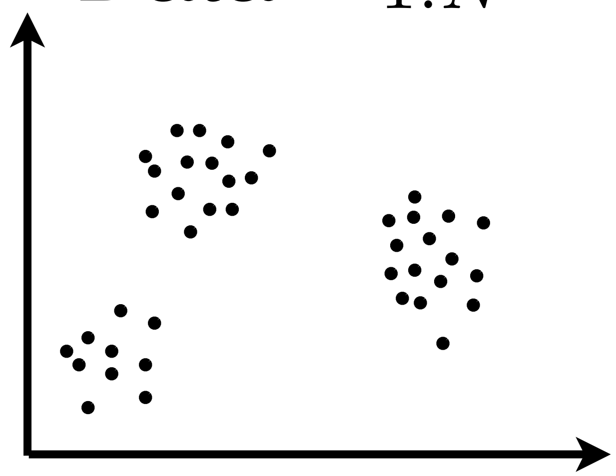
- For completeness: $p(x_{C \cup \{n\}} | x_C) = \mathcal{N}(\tilde{m}, \tilde{\Sigma} + \Sigma)$

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

[MacEachern 1994; Neal 1992; Neal 2000]

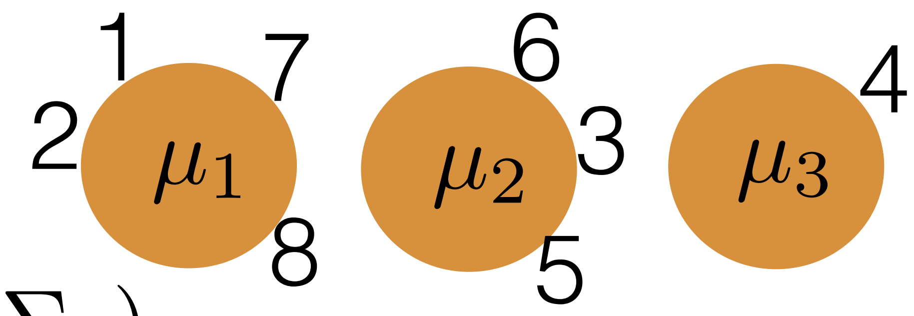
CRP mixture model: inference

- Data $x_{1:N}$



- Generative model

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



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

- Gibbs sampler:

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

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

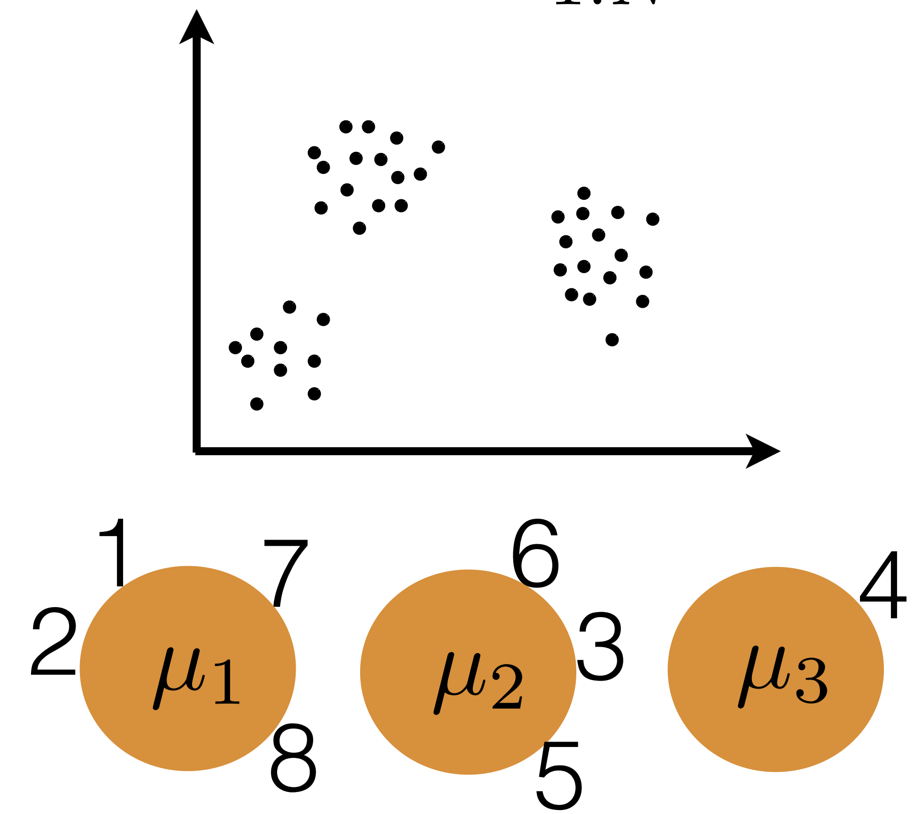
$$\tilde{\Sigma}^{-1} := \Sigma_0^{-1} + (\#C)\Sigma^{-1}$$
$$\tilde{m} := \tilde{\Sigma}\left(\Sigma^{-1} \sum_{m \in C} x_m + \Sigma_0^{-1}\mu_0\right)$$

[demo]

[MacEachern 1994; Neal 1992; Neal 2000]

CRP mixture model exercises

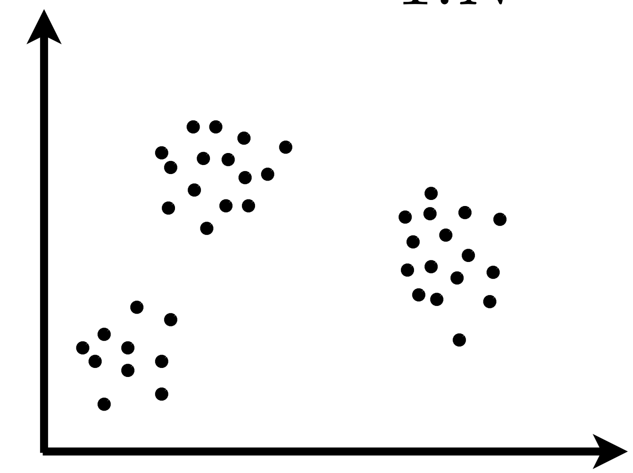
- Data $x_{1:N}$

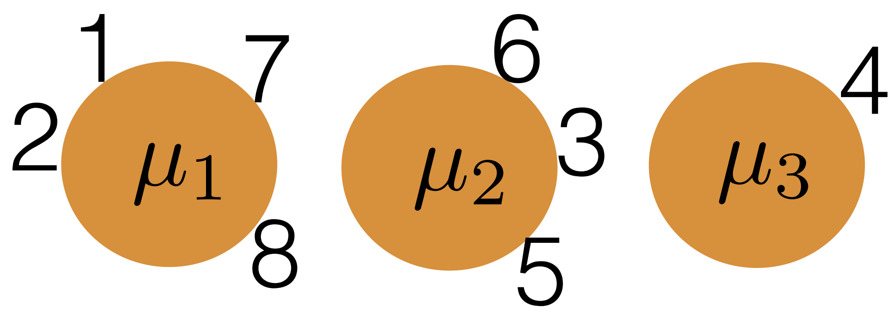


CRP mixture model exercises

- Code a CRP mixture model simulator

- Data $x_{1:N}$

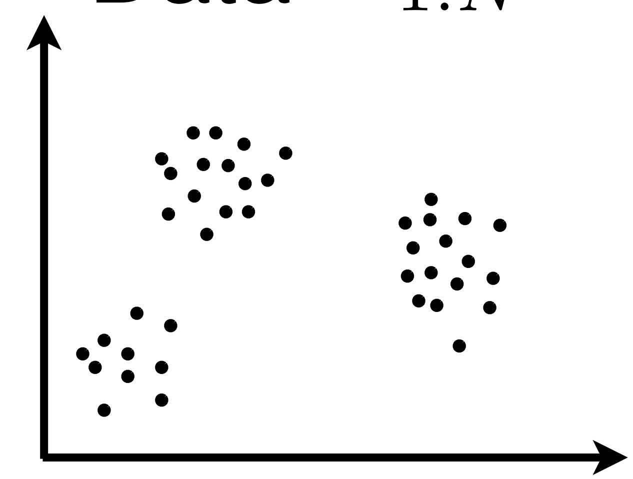


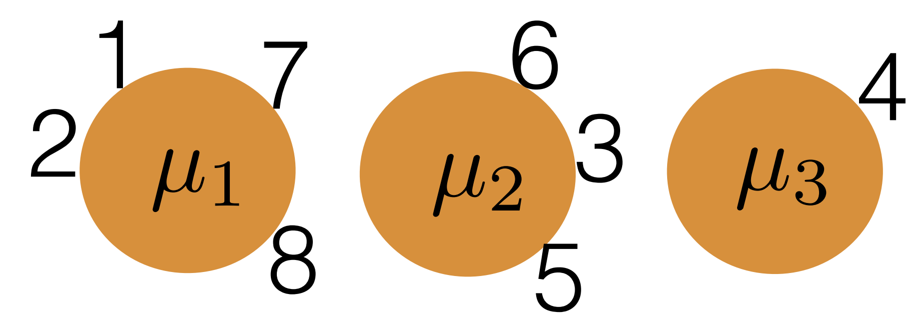


CRP mixture model exercises

- 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

- Data $x_{1:N}$

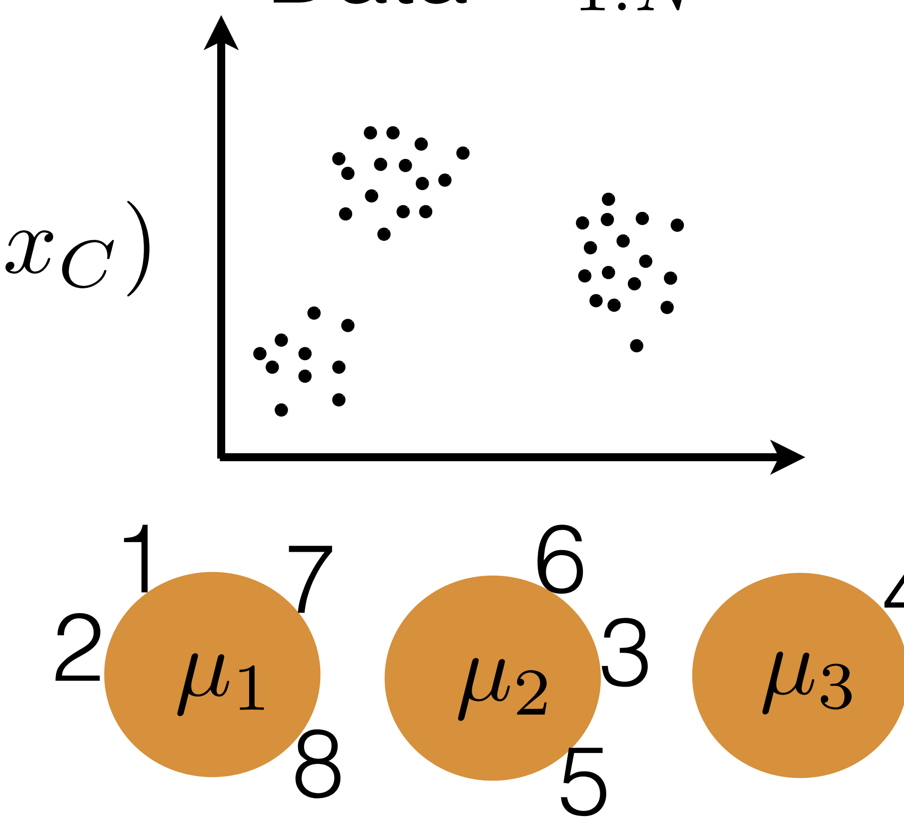




CRP mixture model exercises

- 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

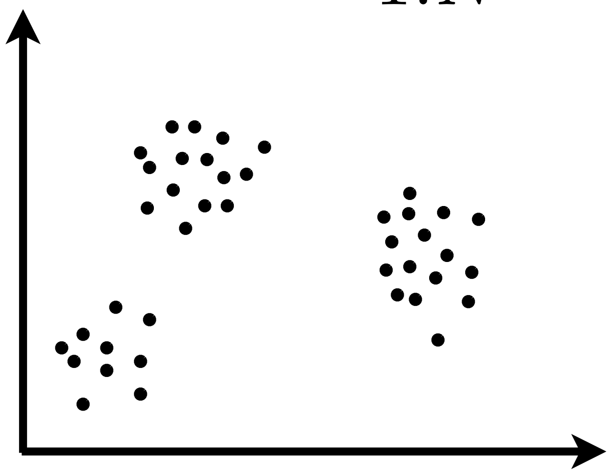
- Data $x_{1:N}$

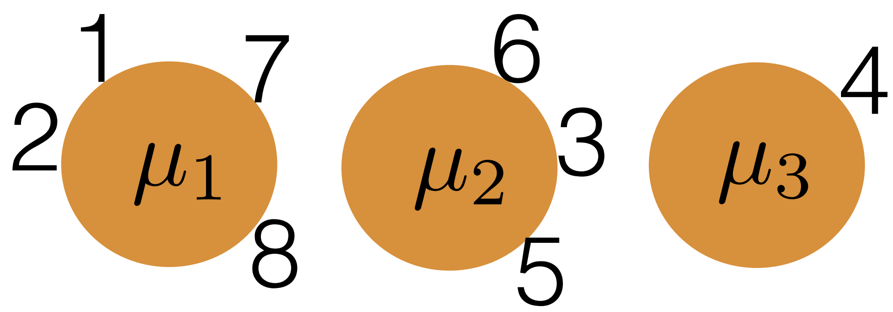


CRP mixture model exercises

- 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

- Data $x_{1:N}$

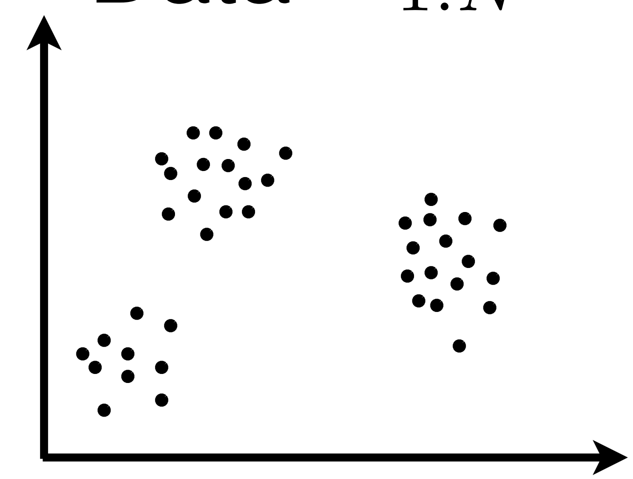


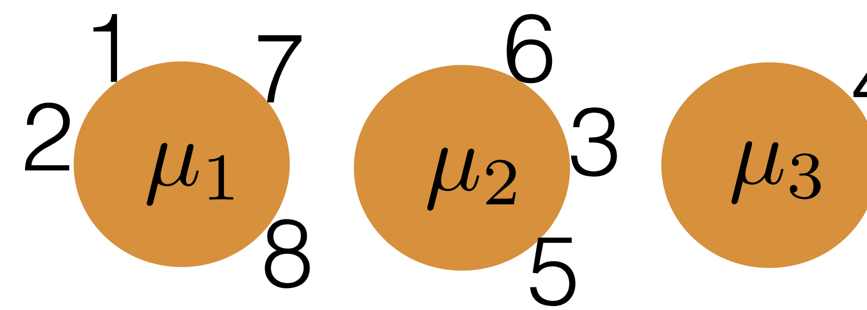


CRP mixture model exercises

- 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

- Data $x_{1:N}$

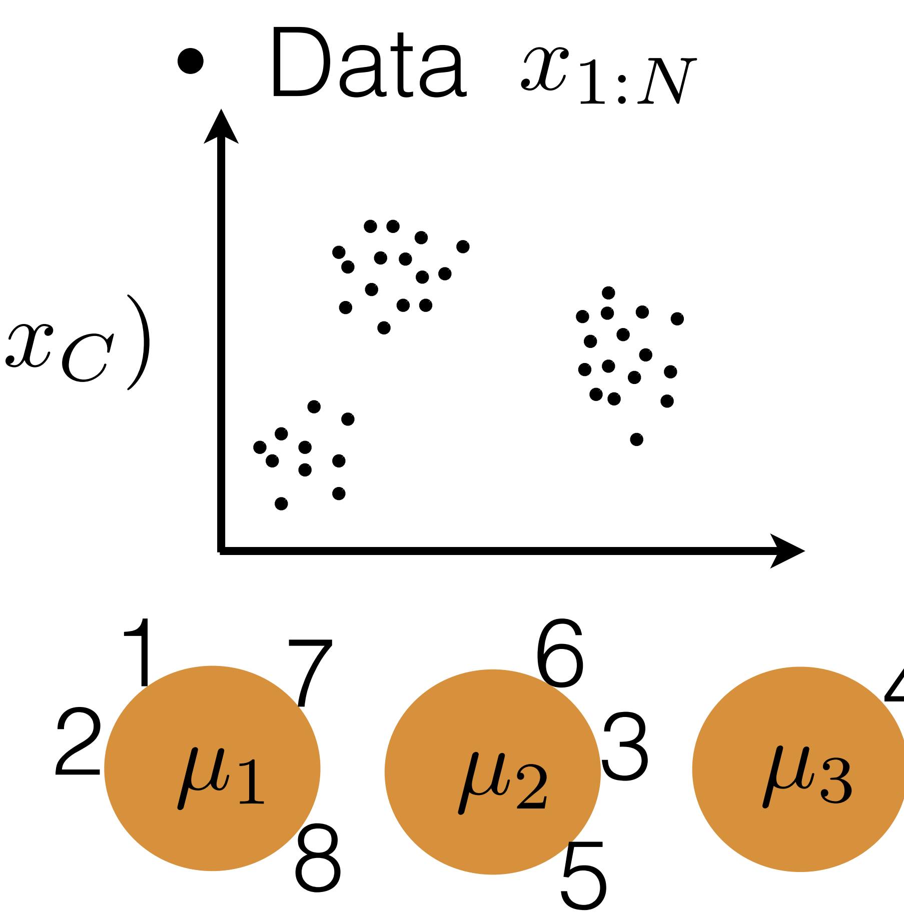




- Read Neal 2000 and try out other samplers
- Further: Read Walker 2007; Kalli, Griffin, Walker 2011 and code a DPMM slice sampler; Read Wang, Blei 2012 and code a variational inference algorithm

CRP mixture model exercises

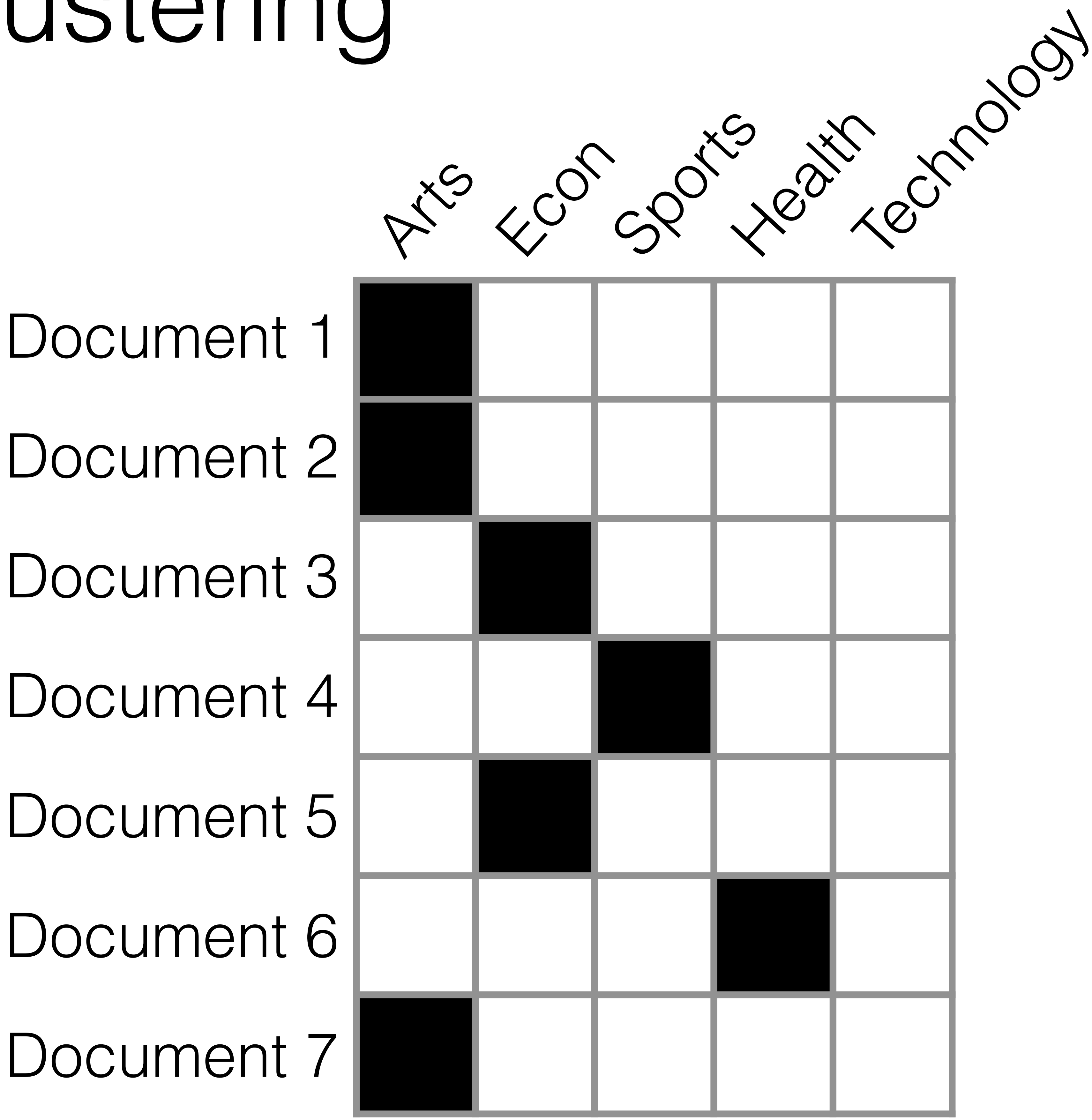
- Code a CRP mixture model simulator
- Derive the CRP mixture model Gibbs sampler in the slides; derive $p(x_{C \cup \{n\}} | x_C)$ explicitly for a Gaussian mixture
- Extend the CRP mixture model Gibbs sampler in the slides to sample the cluster-specific parameters as well



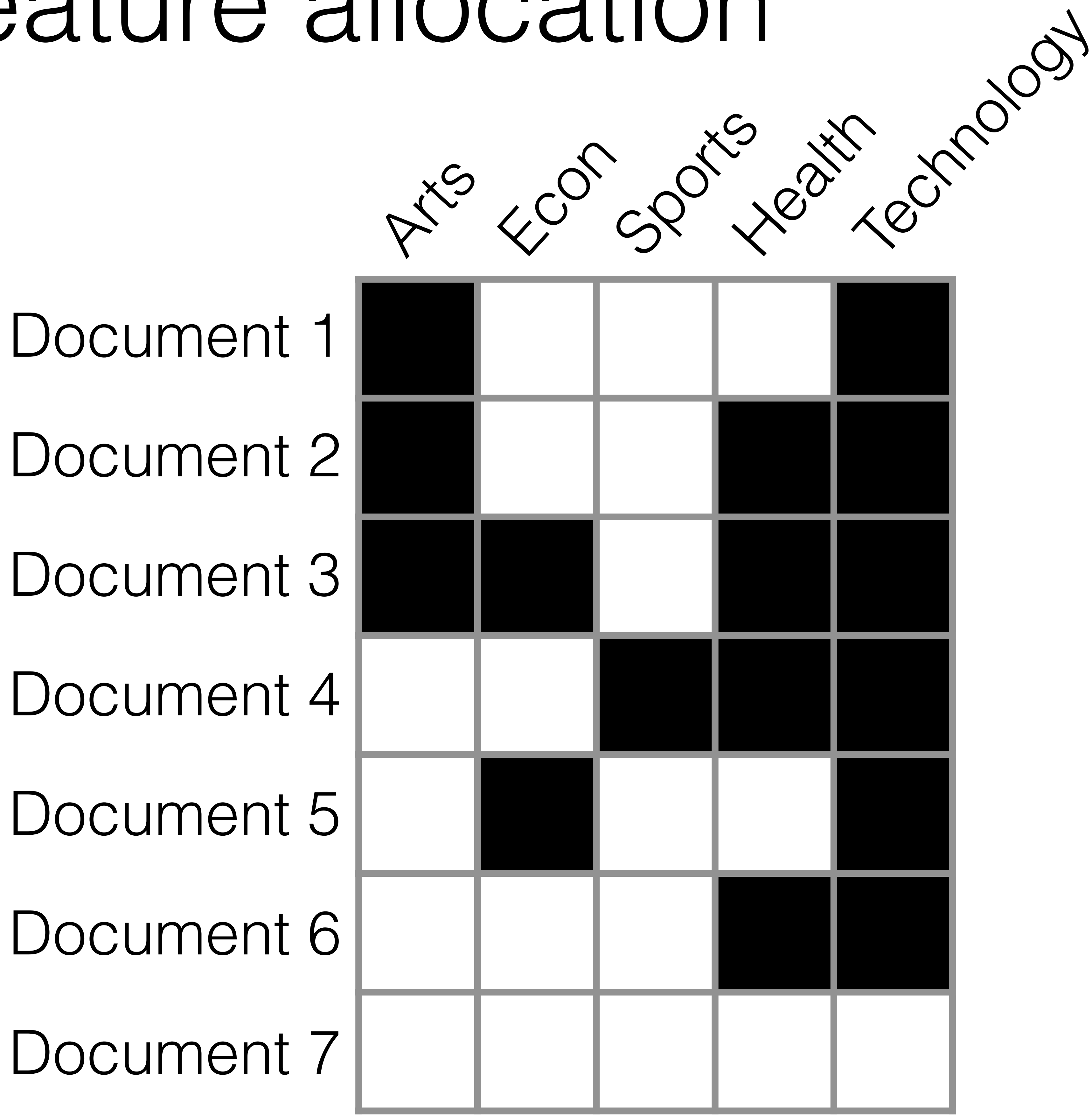
- Read Neal 2000 and try out other samplers
- Further: Read Walker 2007; Kalli, Griffin, Walker 2011 and code a DPMM slice sampler; Read Wang, Blei 2012 and code a variational inference algorithm
- Read Broderick, Jordan, Pitman 2013 "Cluster and feature modeling […]" for more background/extensions

[Tiia Monto, https://commons.wikimedia.org/wiki/File:Trail_and_mountain.jpg]

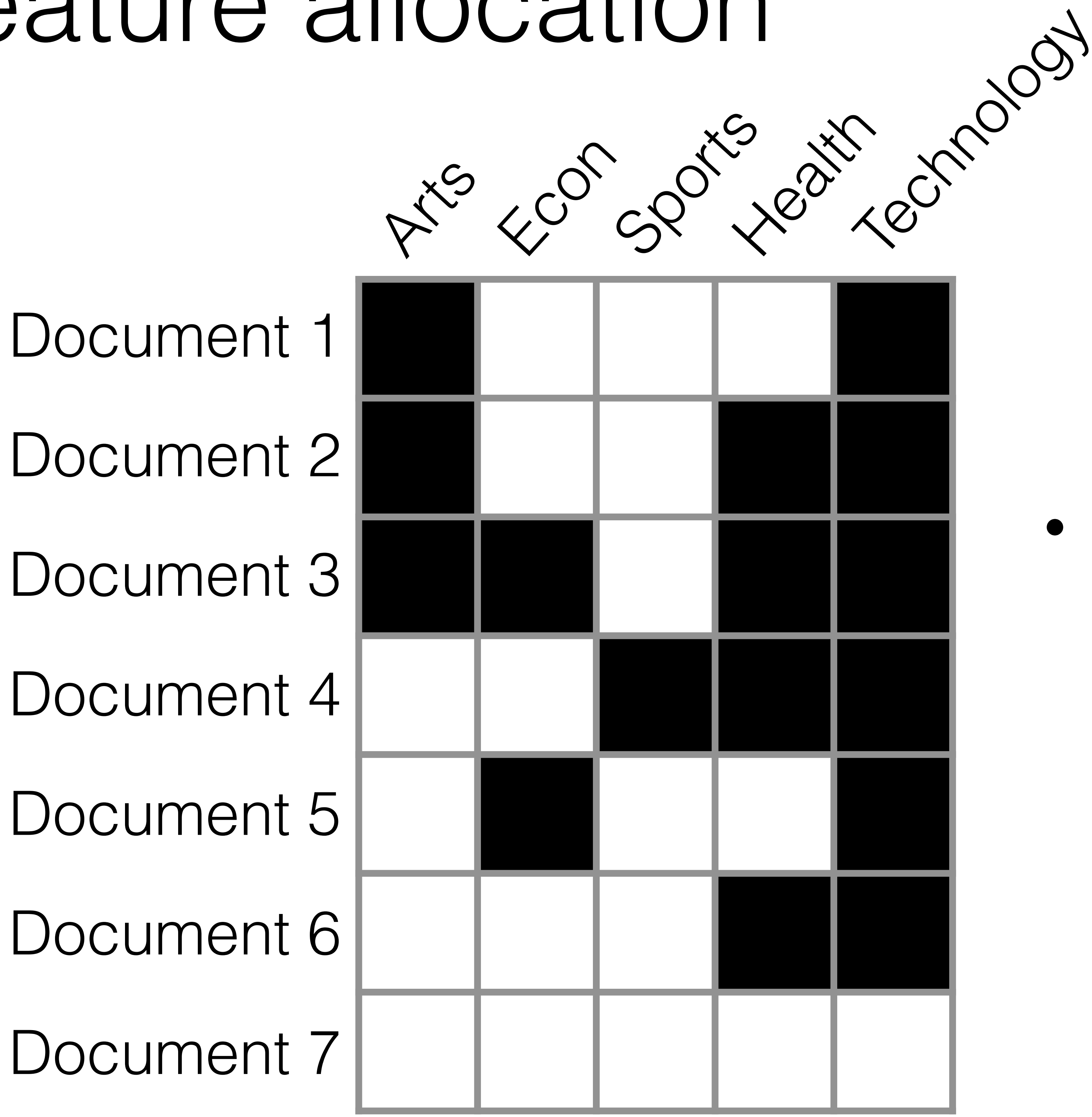
Clustering


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

Feature allocation

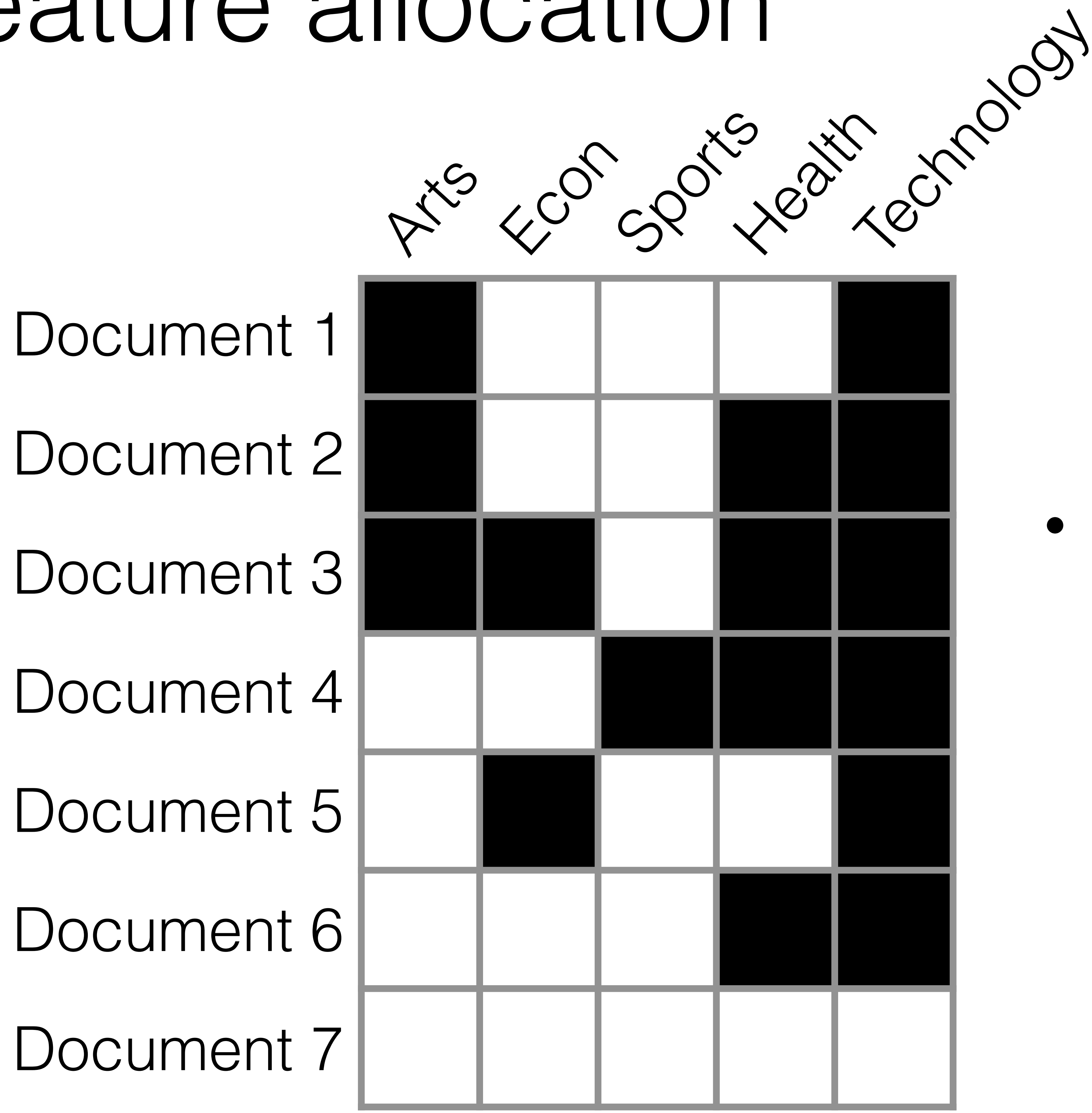


Feature allocation



- Indian buffet process

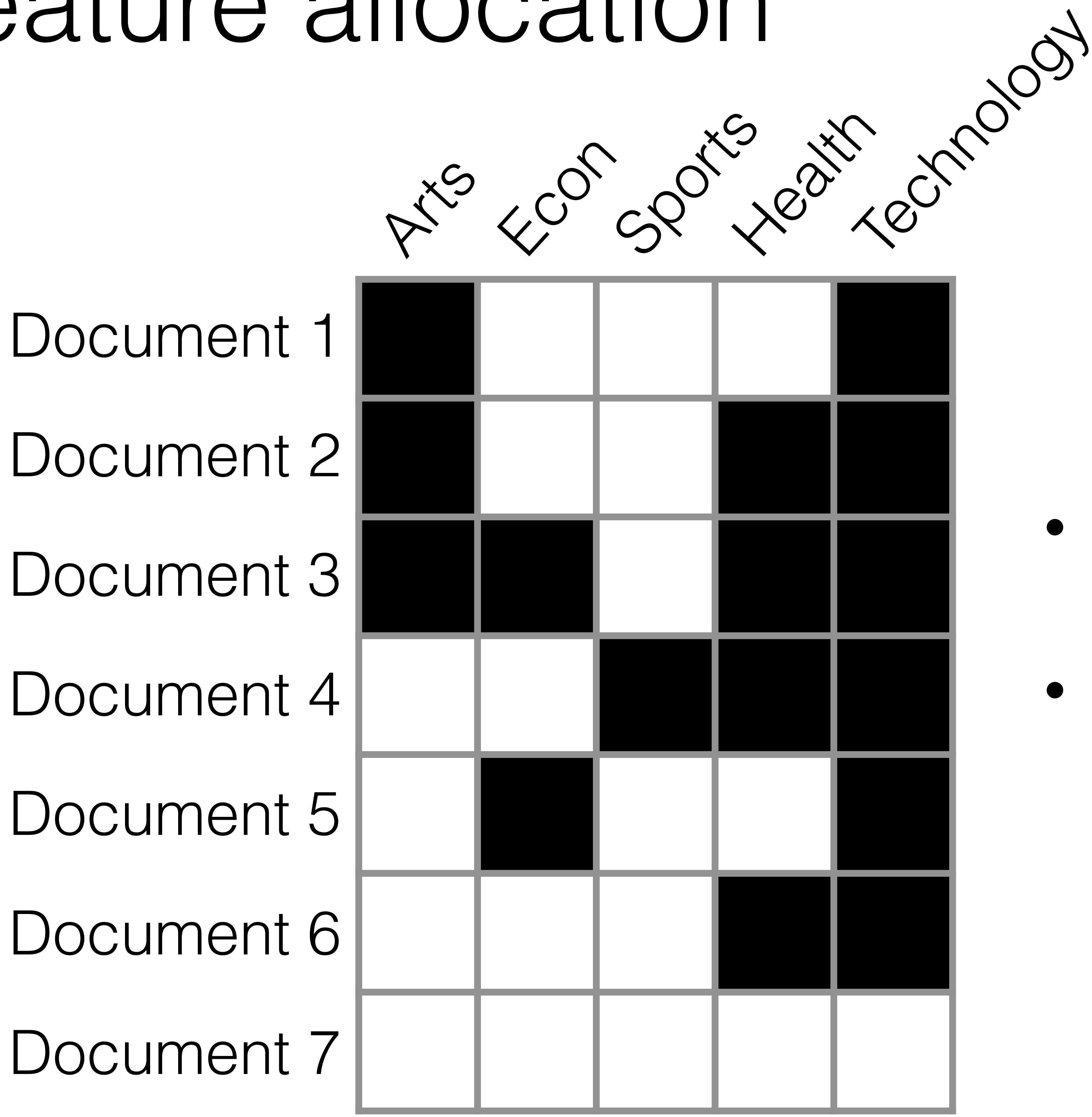
Feature allocation



- Indian buffet process

[Griffiths, Ghahramani 2005

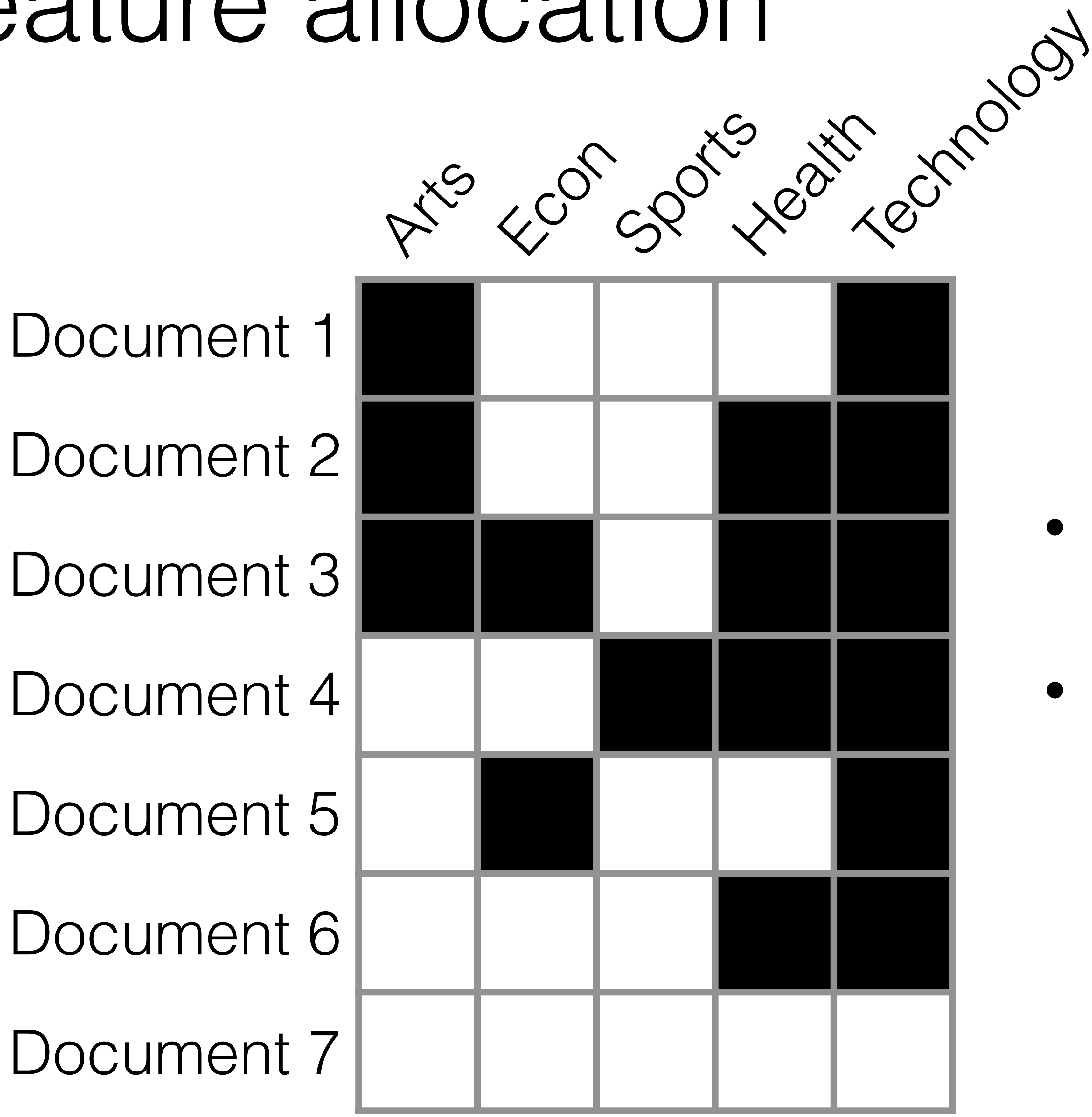
Feature allocation



- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005

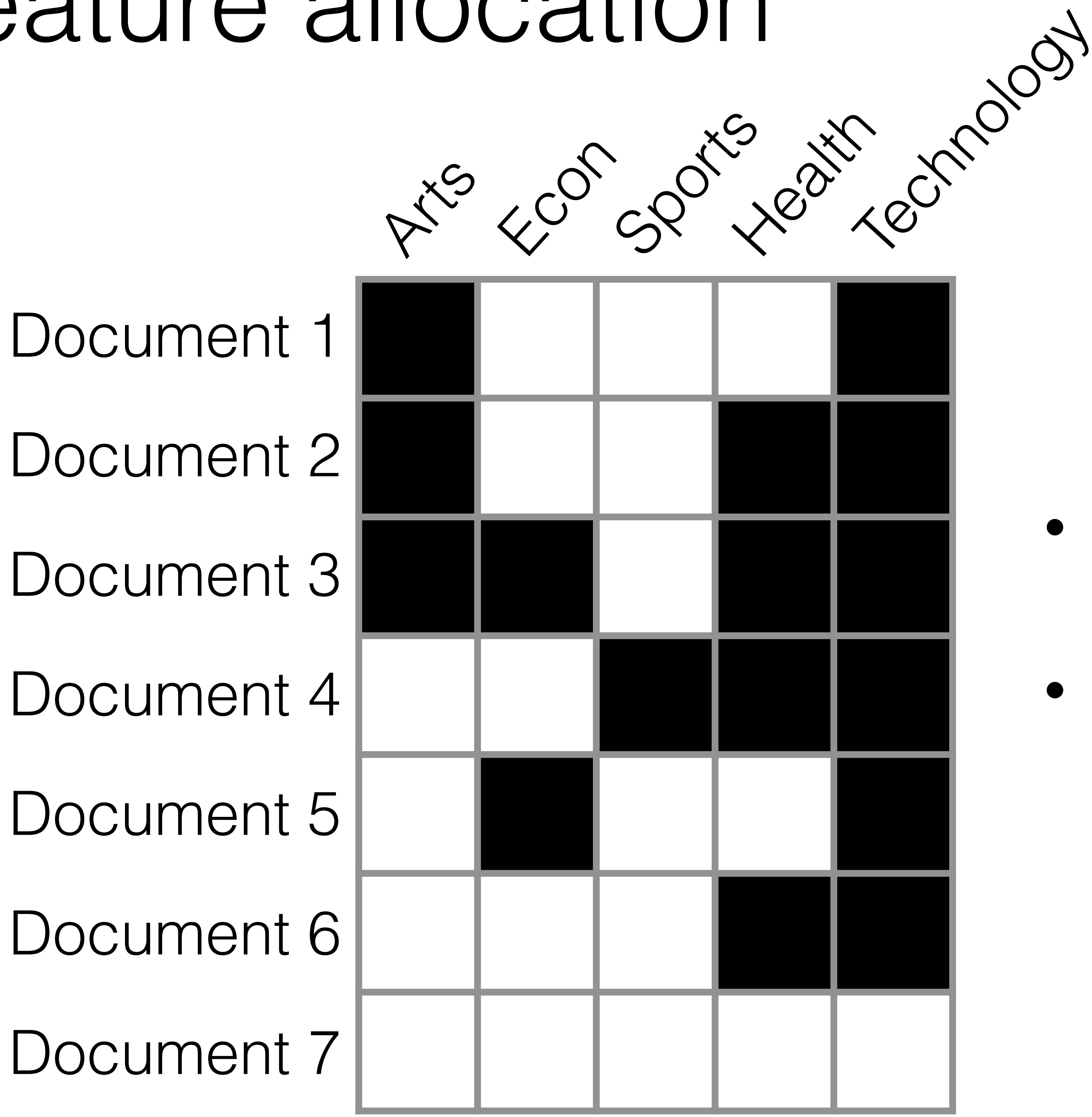
Feature allocation



- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005, Hjort 1990

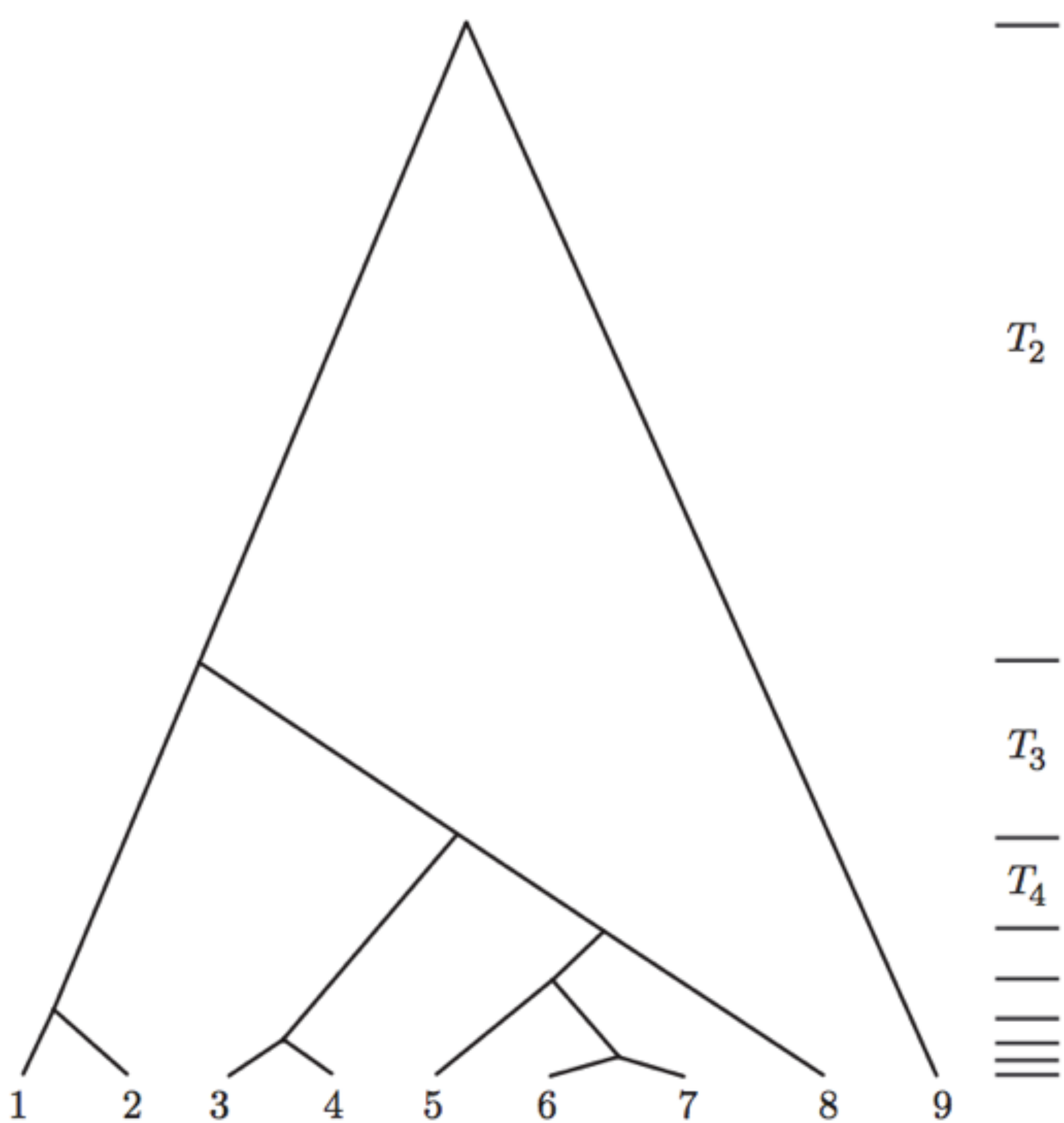
Feature allocation



- Indian buffet process
- Beta process

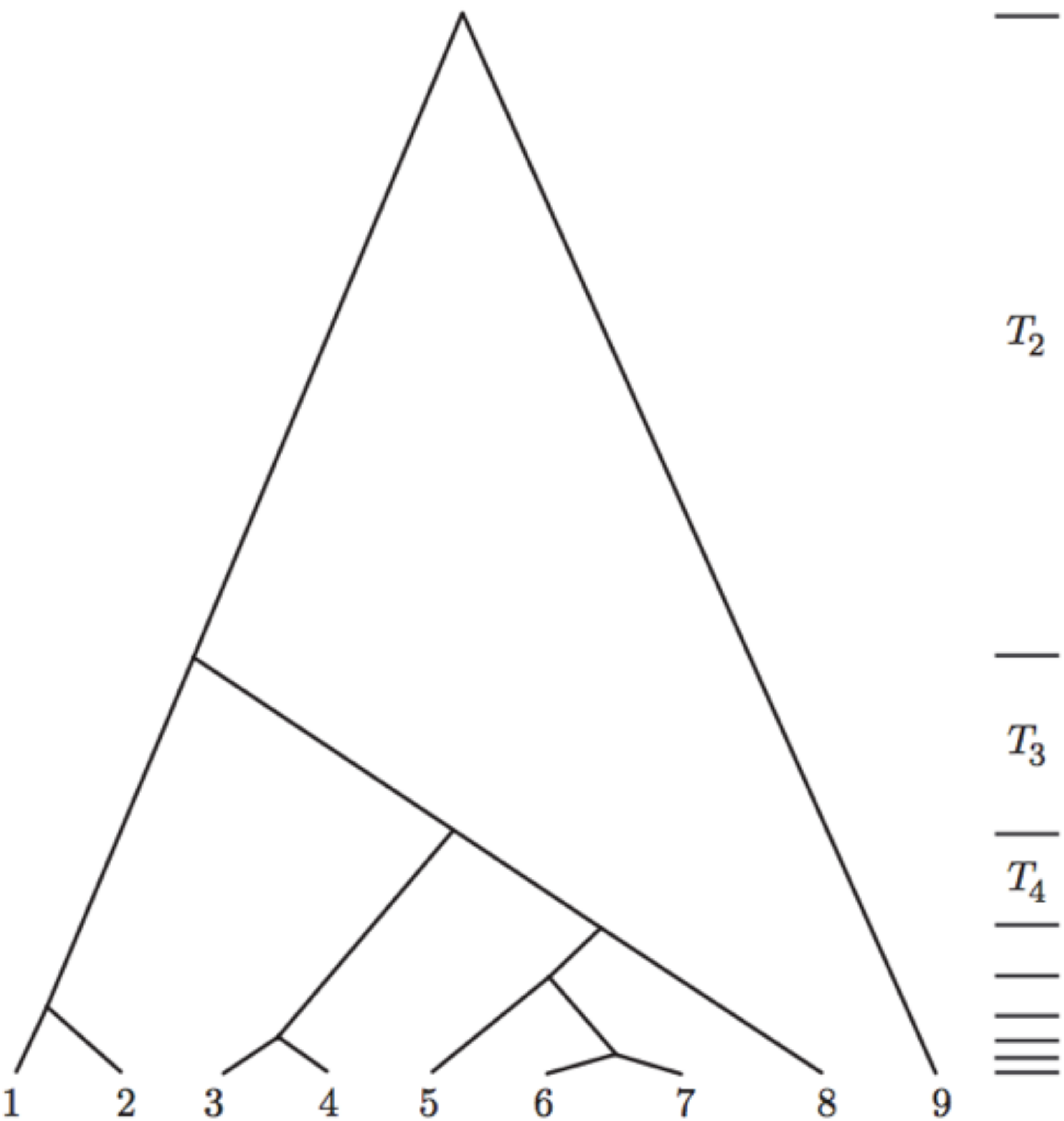
[Griffiths, Ghahramani 2005, Hjort 1990, Kim 1999, Thibaux, Jordan 2007, Broderick, Jordan, Pitman 2013]

Genealogy, trees, beyond trees



[Wakeley 2008]

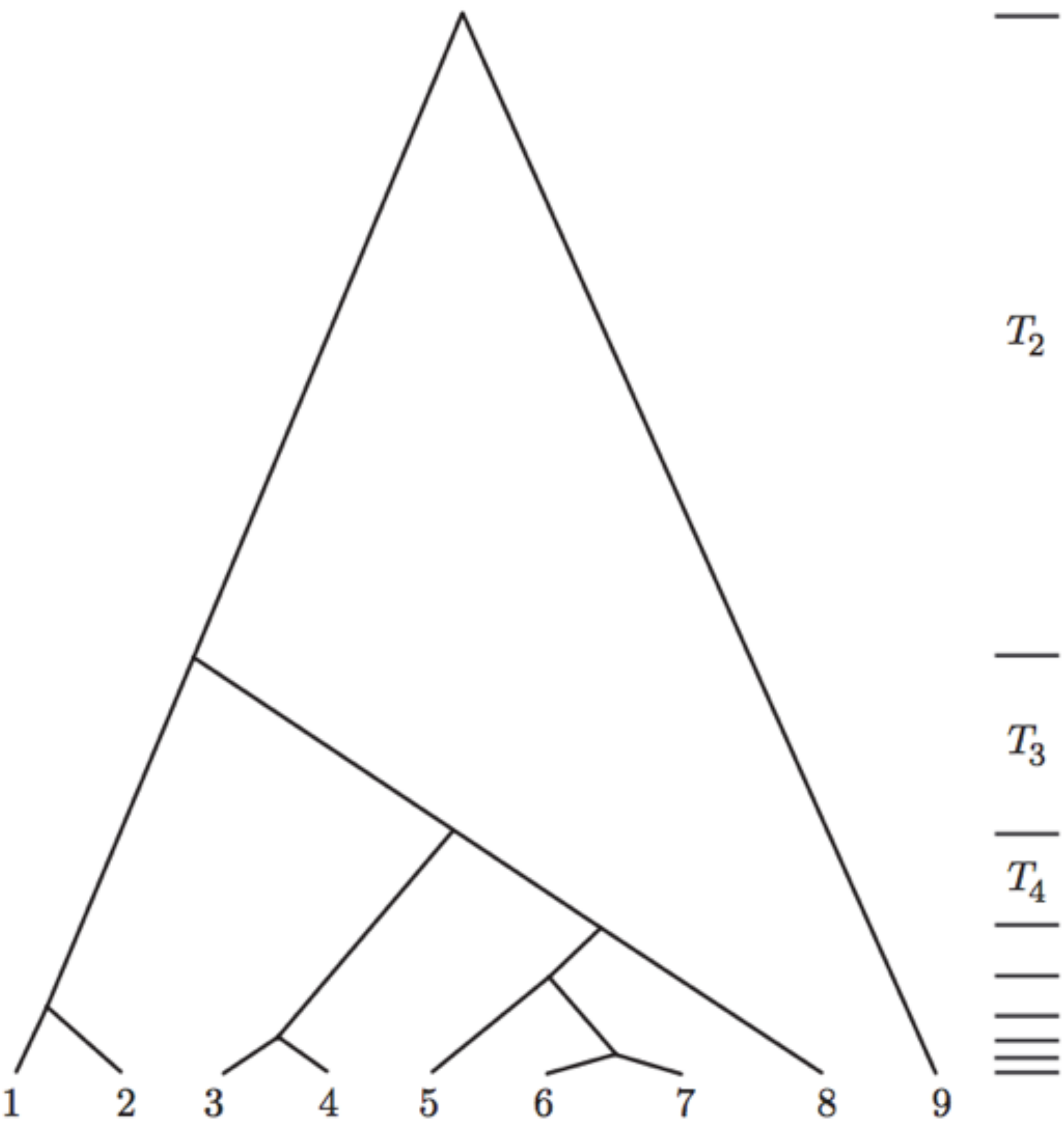
Genealogy, trees, beyond trees



[Wakeley 2008]

- Kingman coalescent

Genealogy, trees, beyond trees

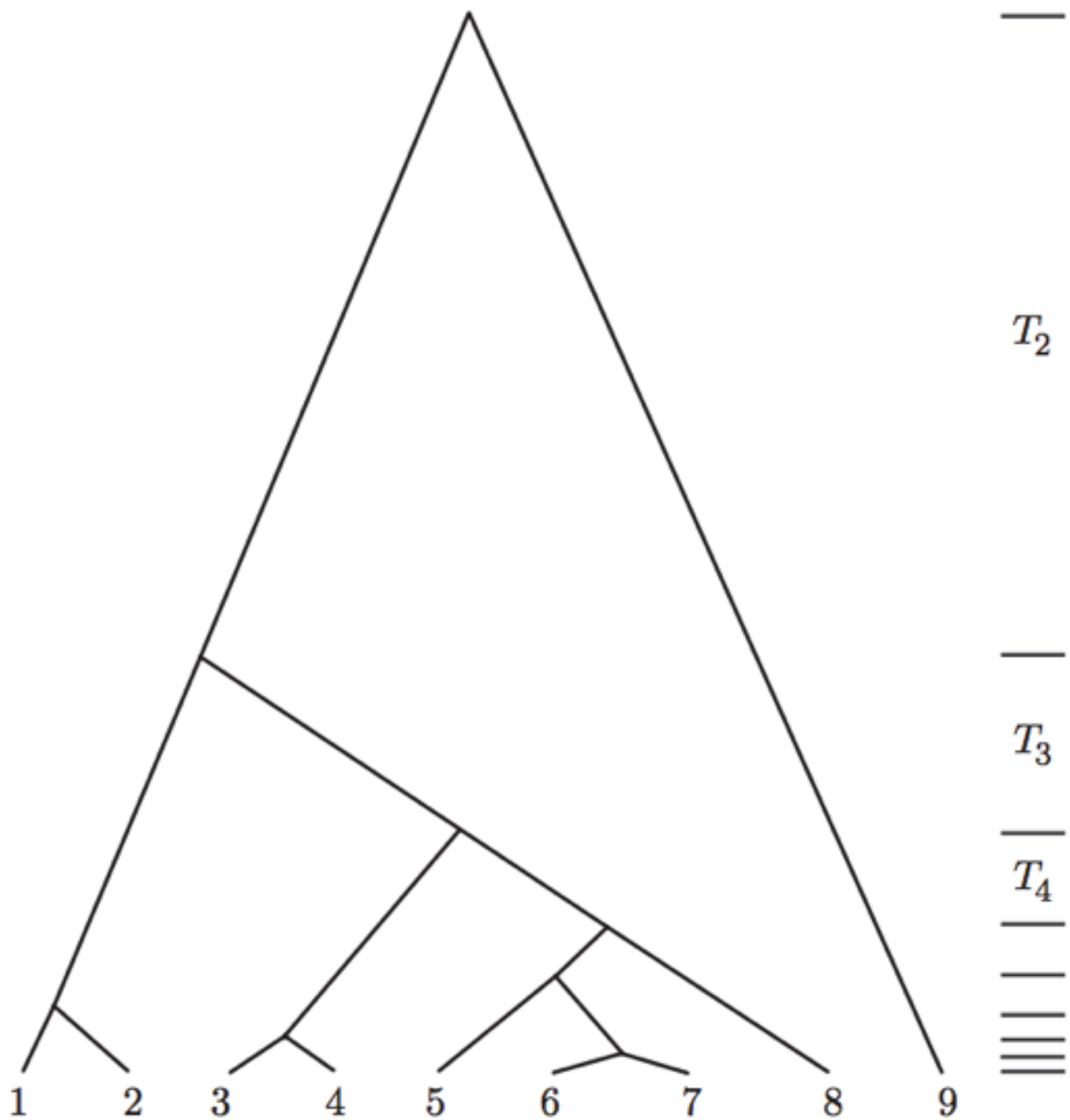


[Wakeley 2008]

- Kingman coalescent

[Kingman 1982

Genealogy, trees, beyond trees

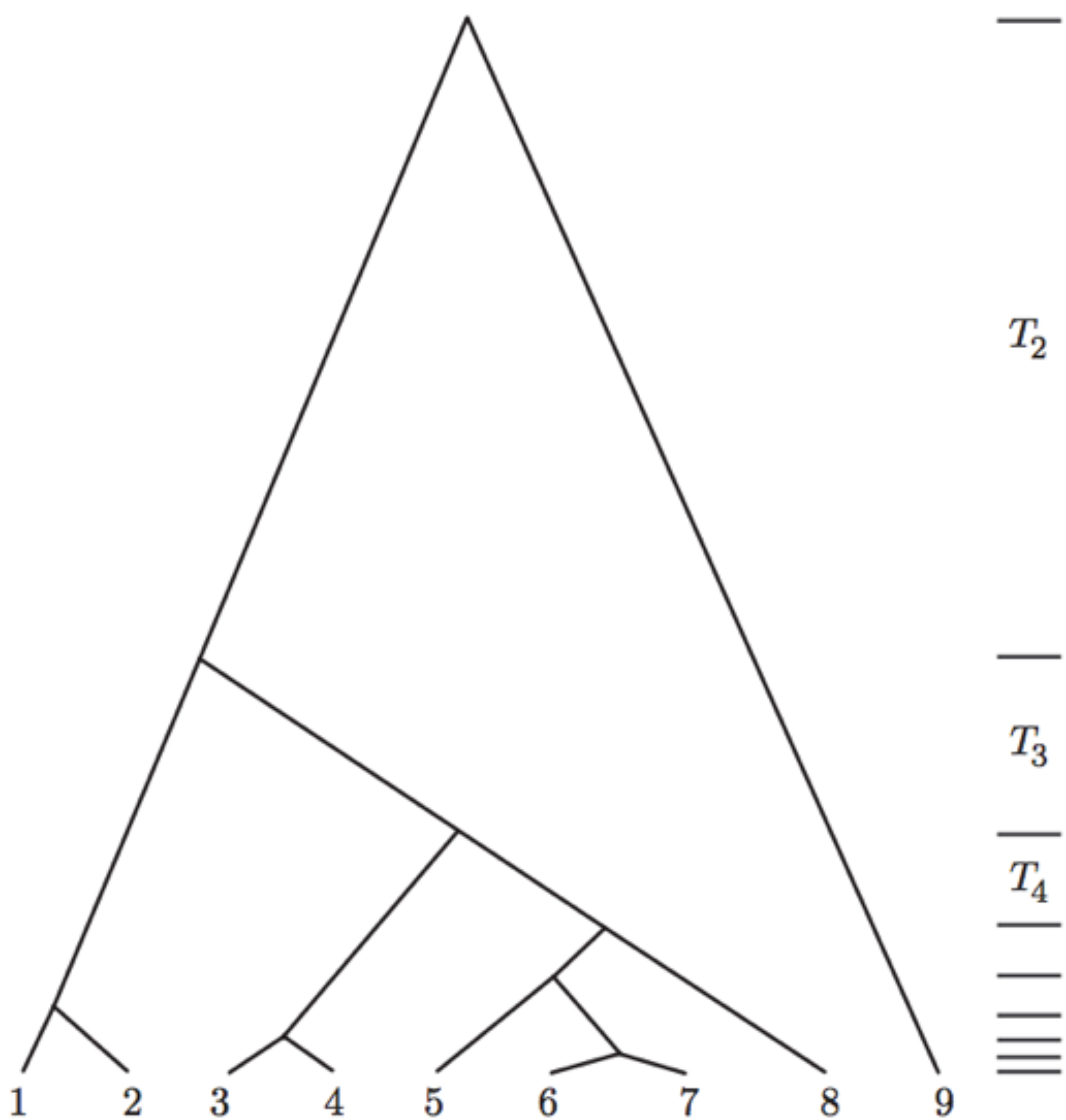


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation

[Kingman 1982

Genealogy, trees, beyond trees

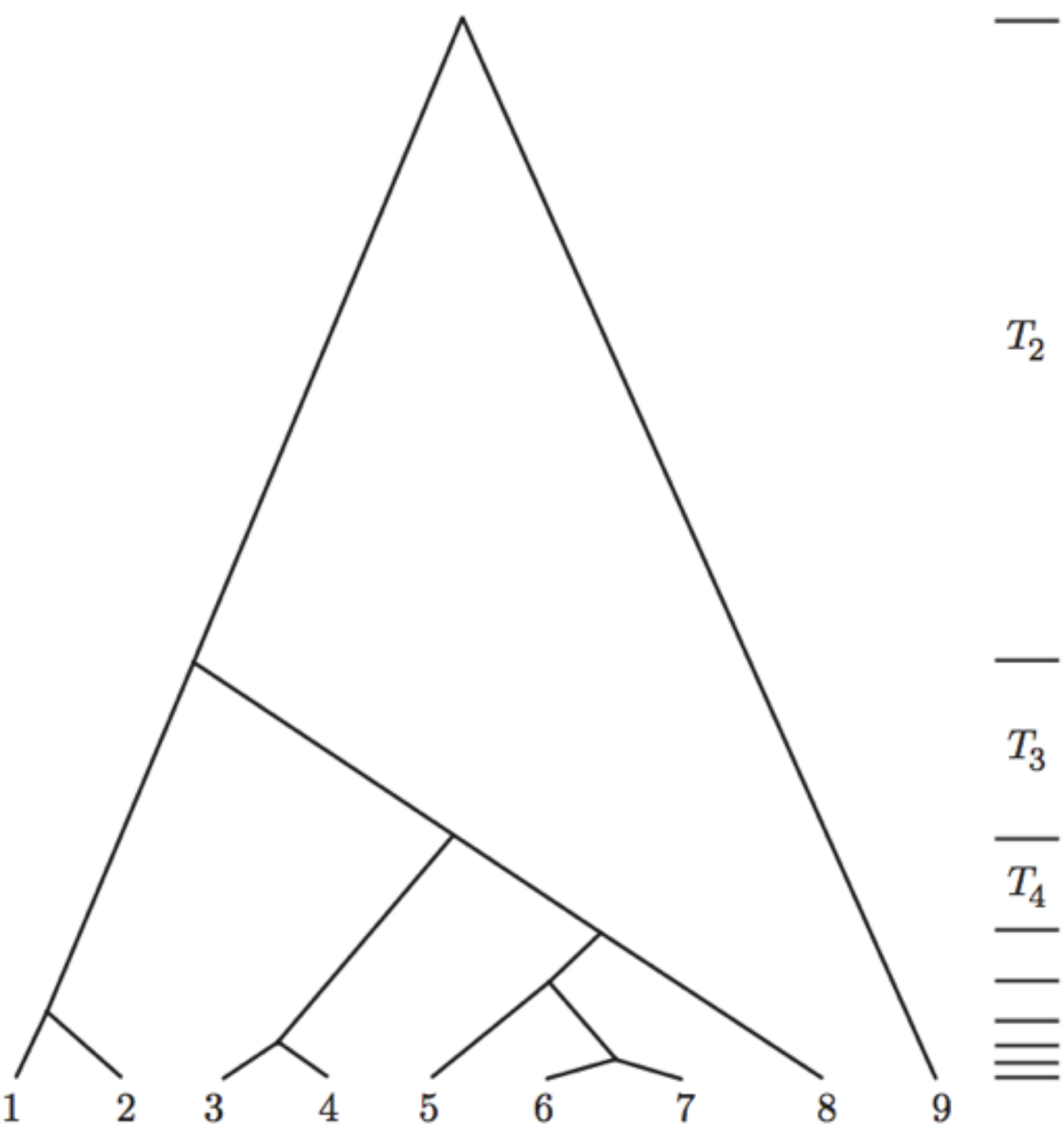


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation

[Kingman 1982, Bertoin 2006, Teh et al 2011

Genealogy, trees, beyond trees

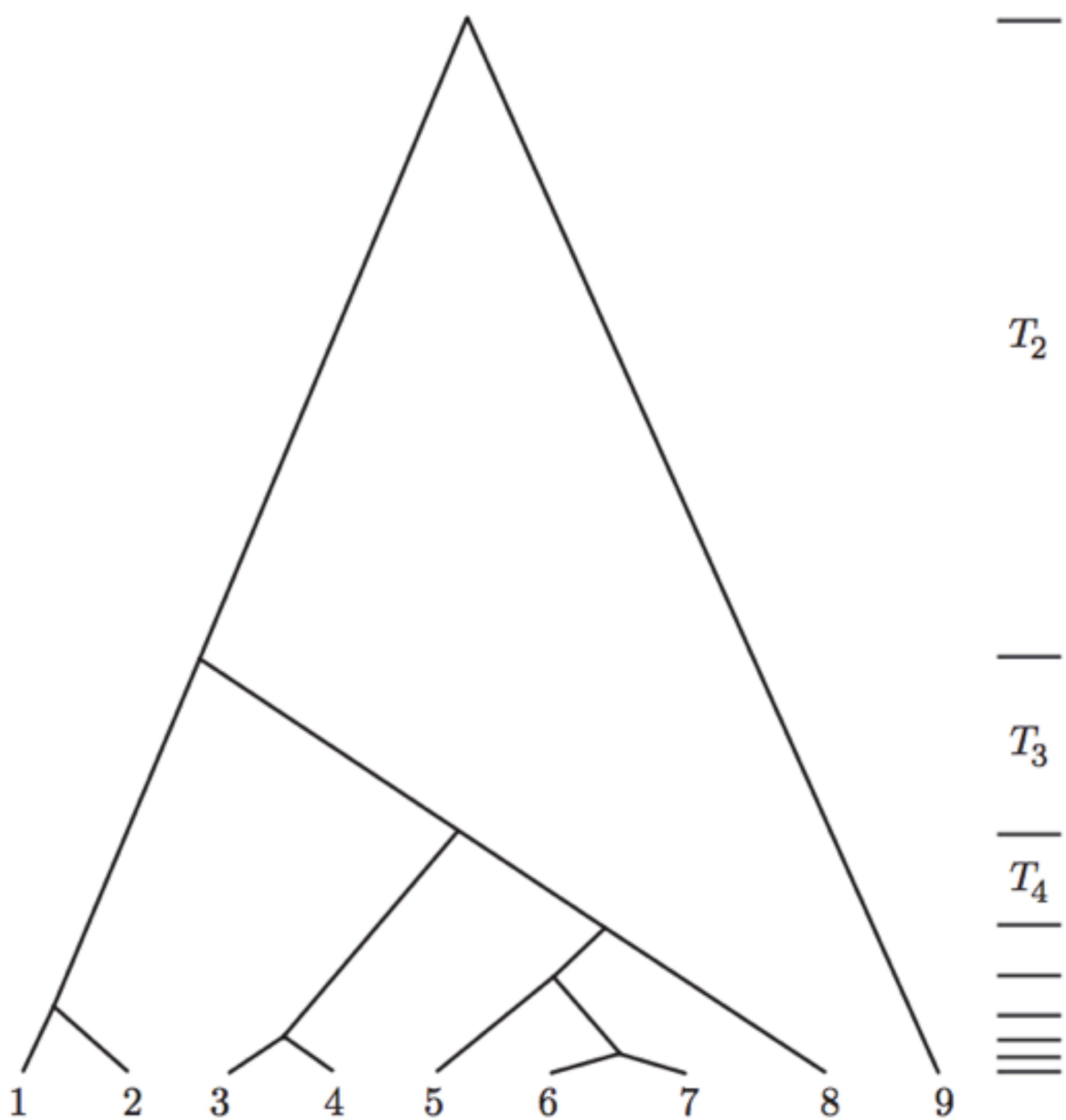


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

[Kingman 1982, Bertoin 2006, Teh et al 2011

Genealogy, trees, beyond trees

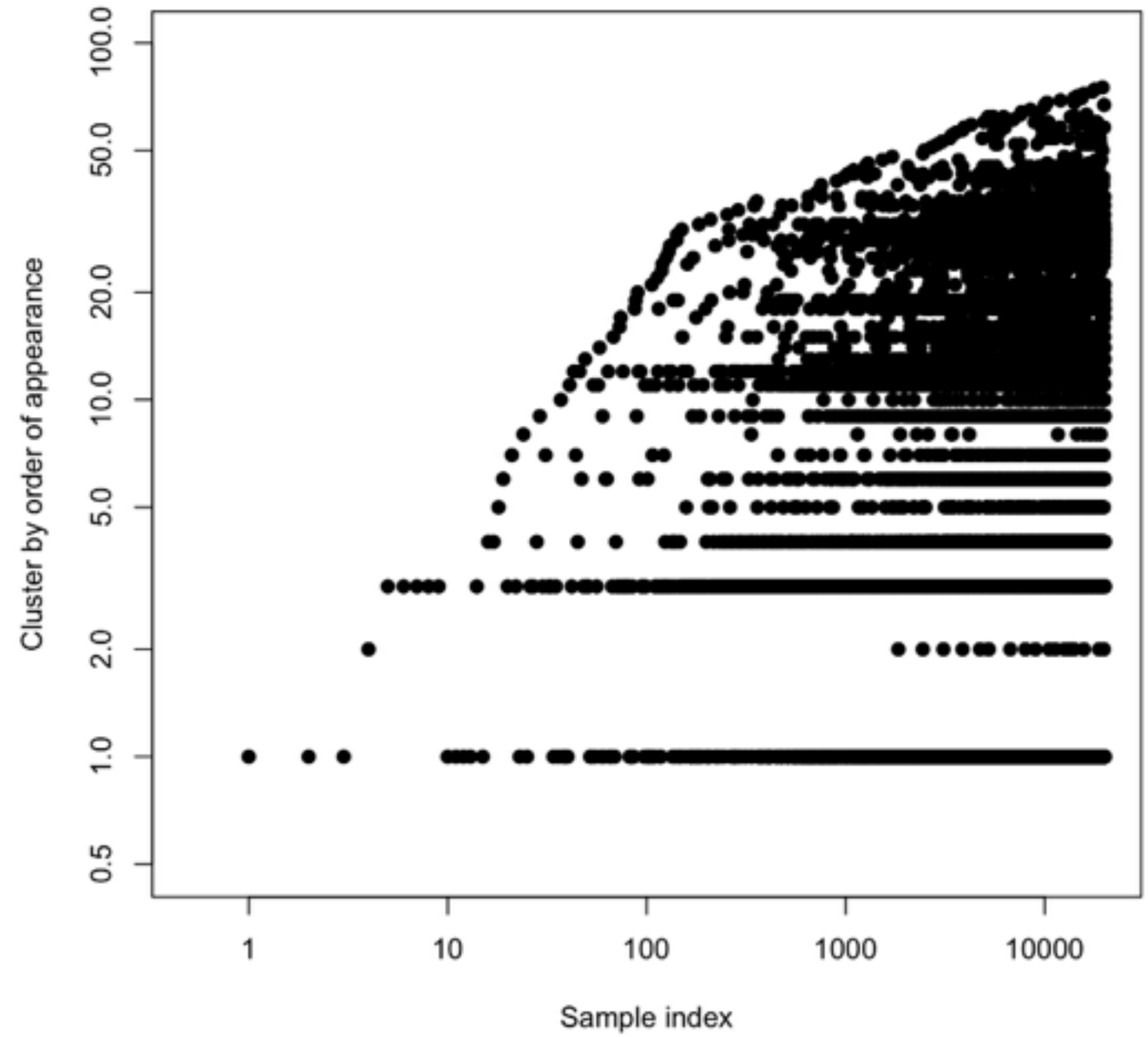


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

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

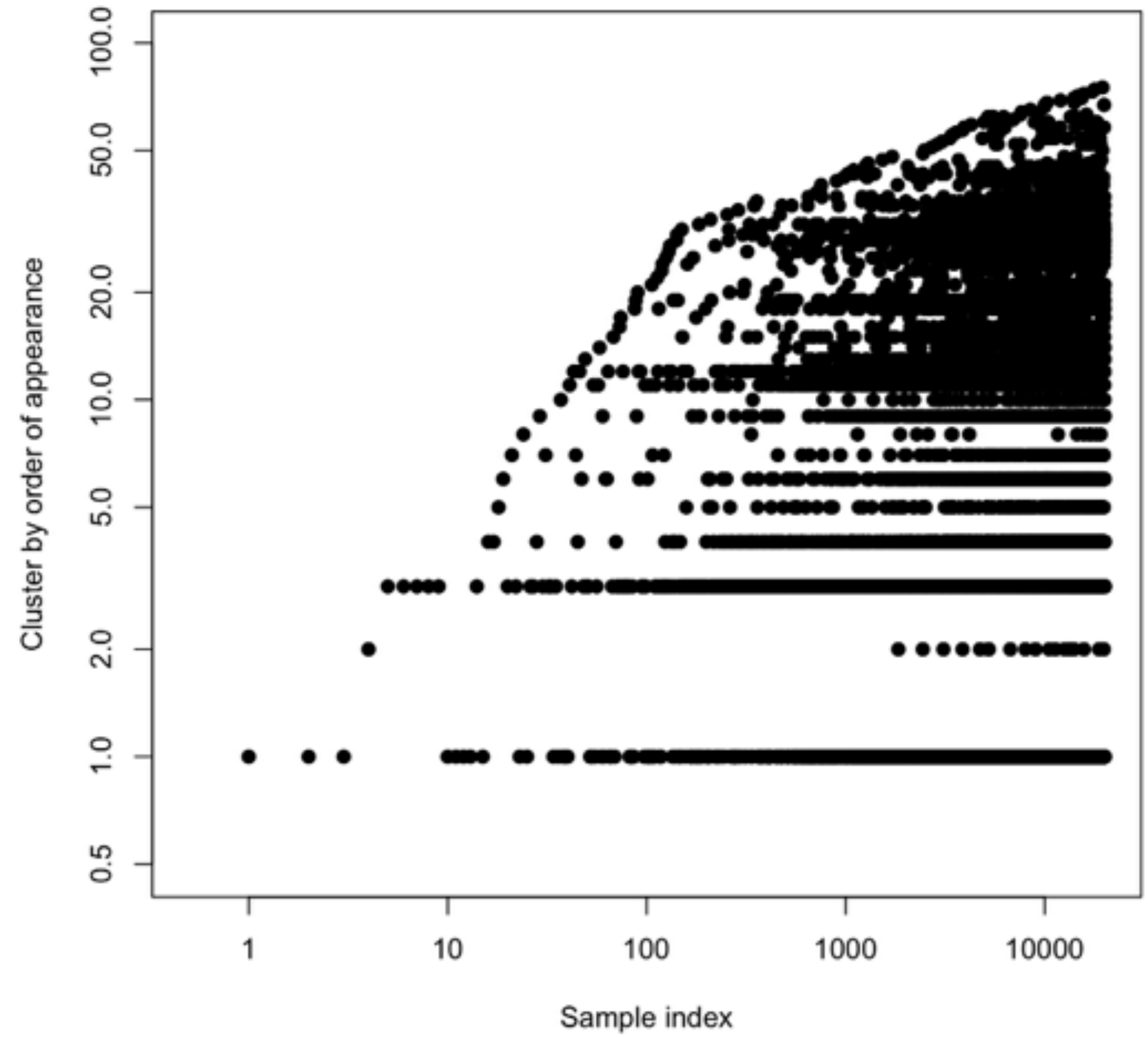
Power laws



Cluster by order of appearance
100.0
50.0
20.0
10.0
5.0
2.0
1.0
0.5
1
10
100
1000
10000
Sample index

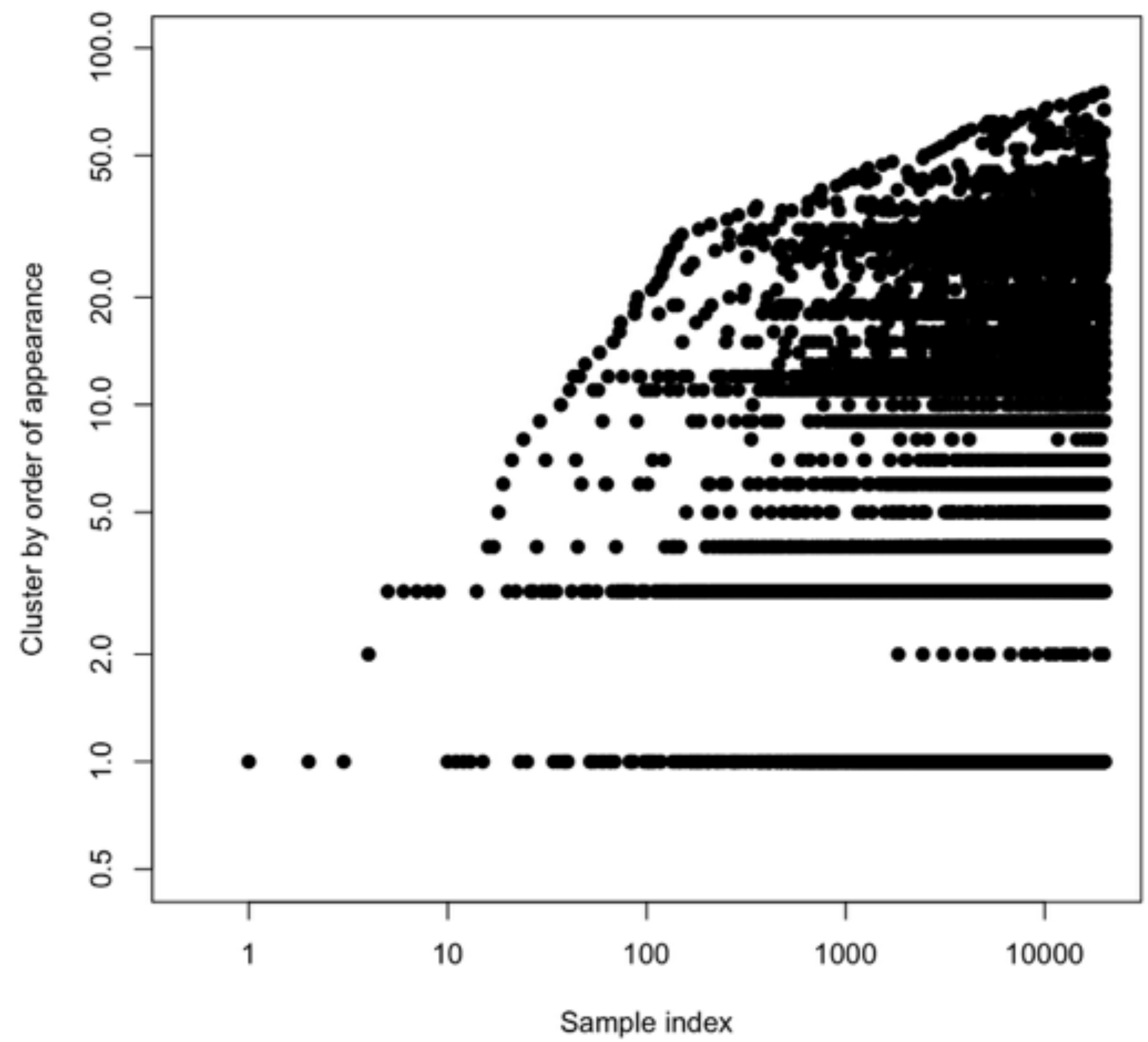
Power laws

- K_N := # clusters occupied by N data points



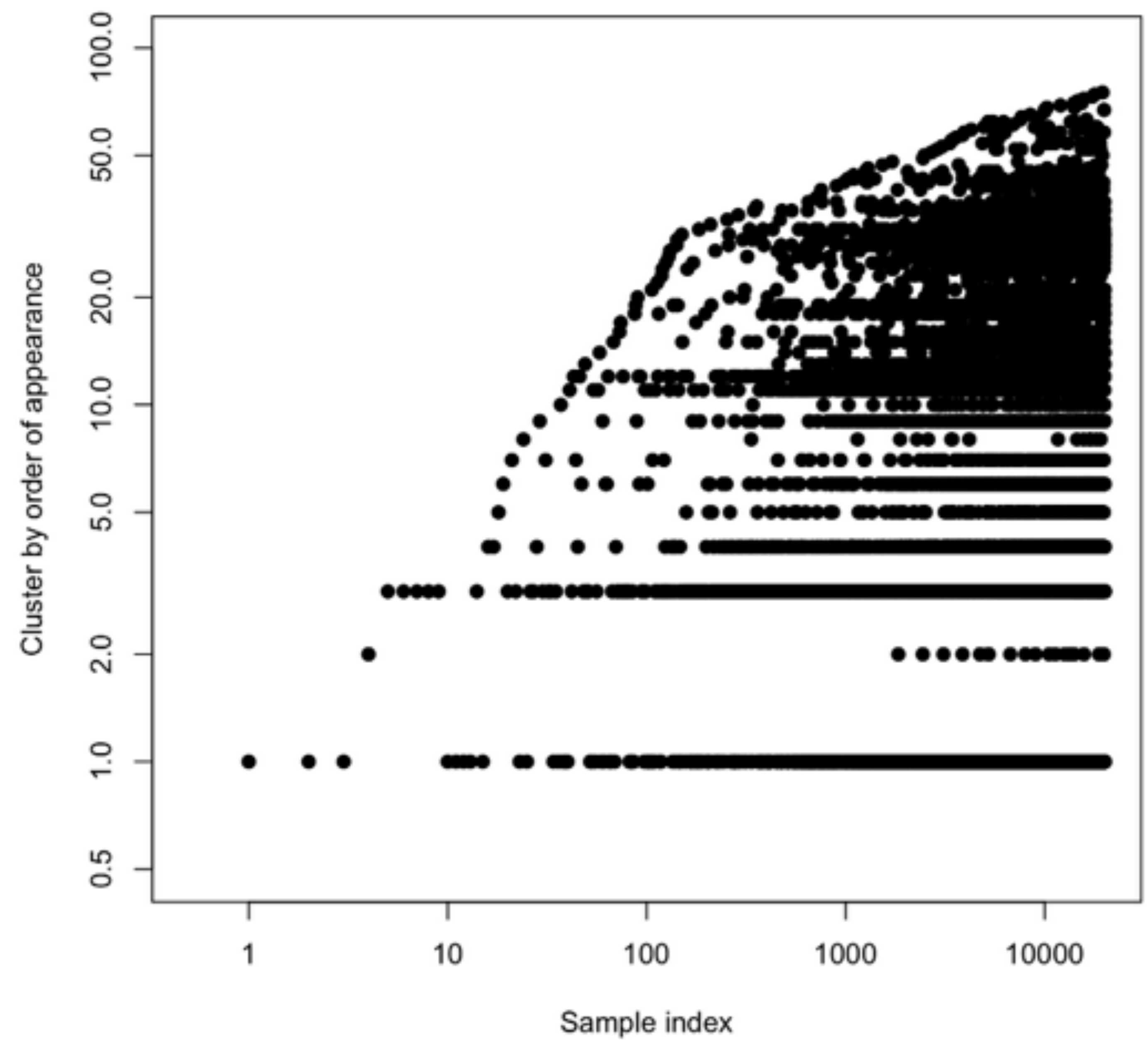
Power laws

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



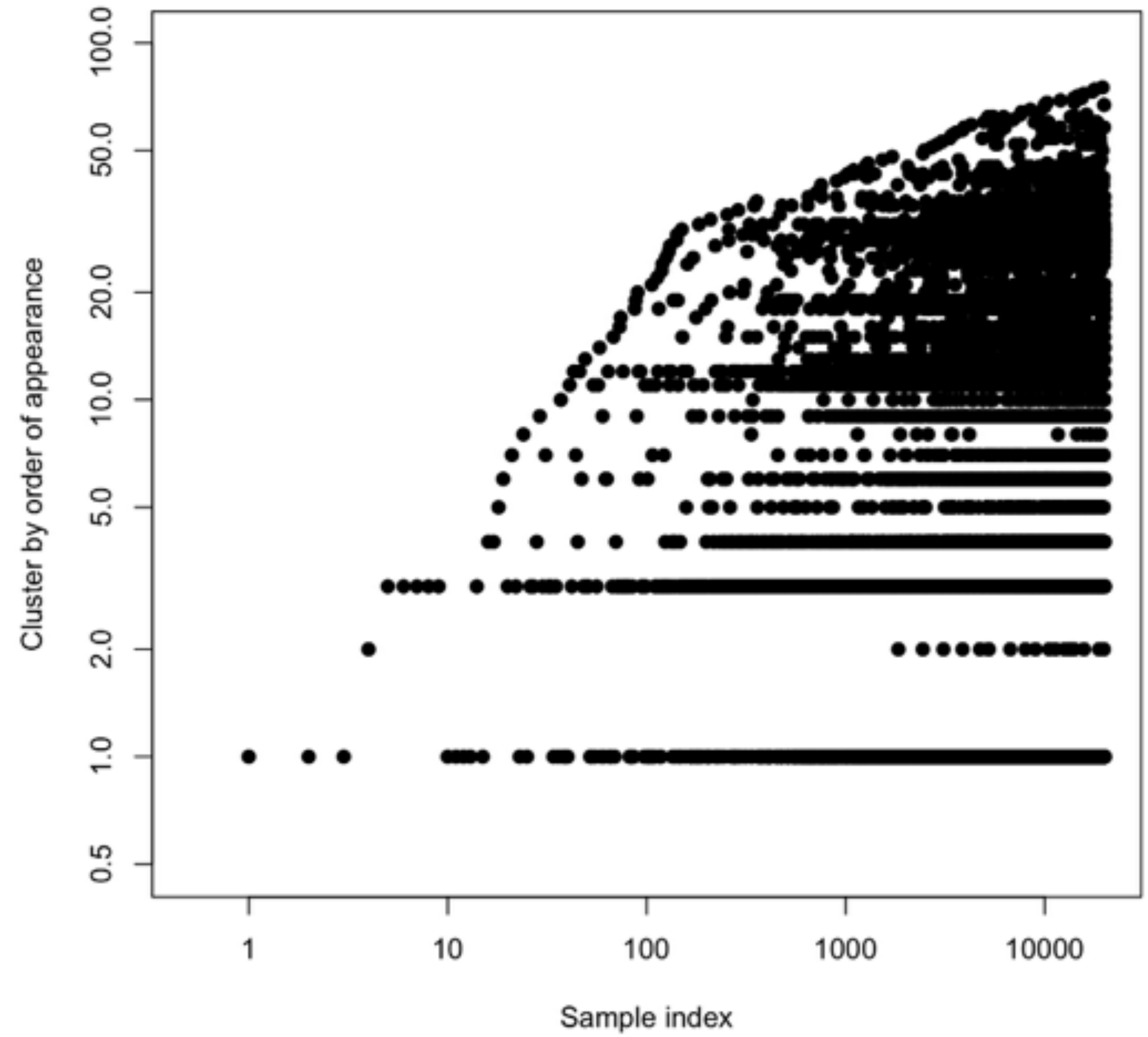
Power laws

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



Power laws

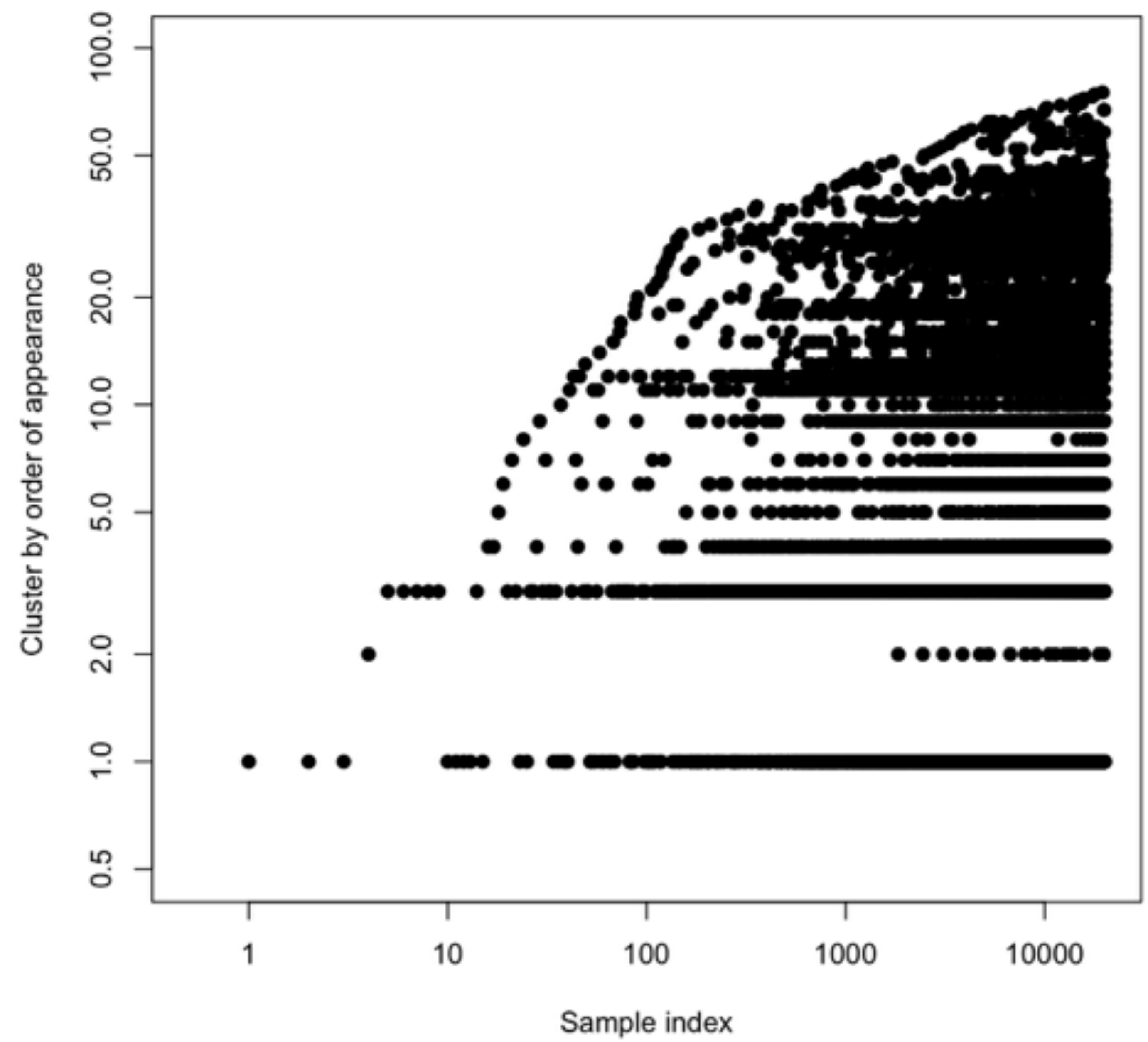
- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
    - vs. Heaps' law, Herdan's law, etc



[Gnedin, et al 2007

Power laws

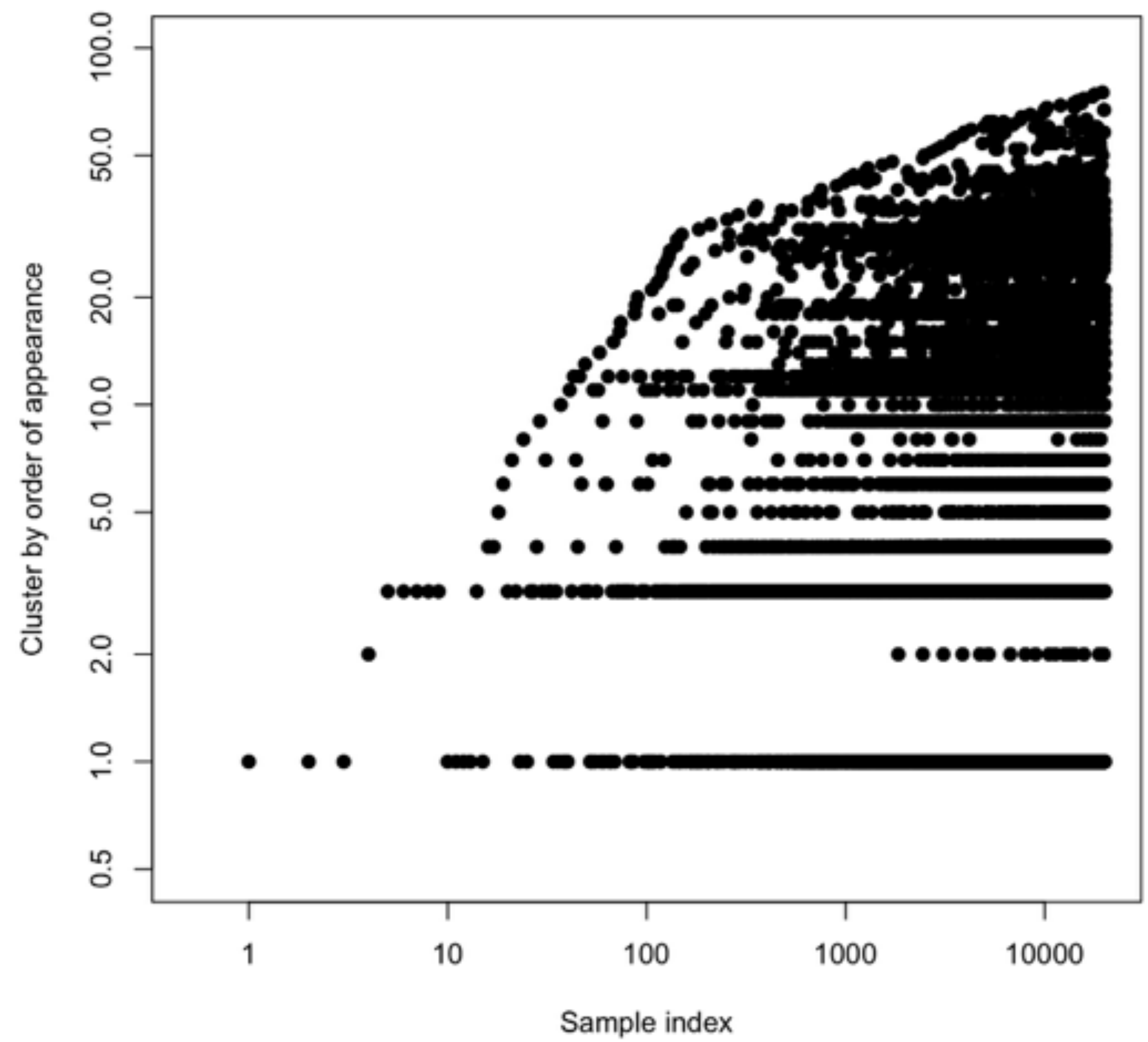
- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:



[Gnedin, et al 2007

Power laws

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

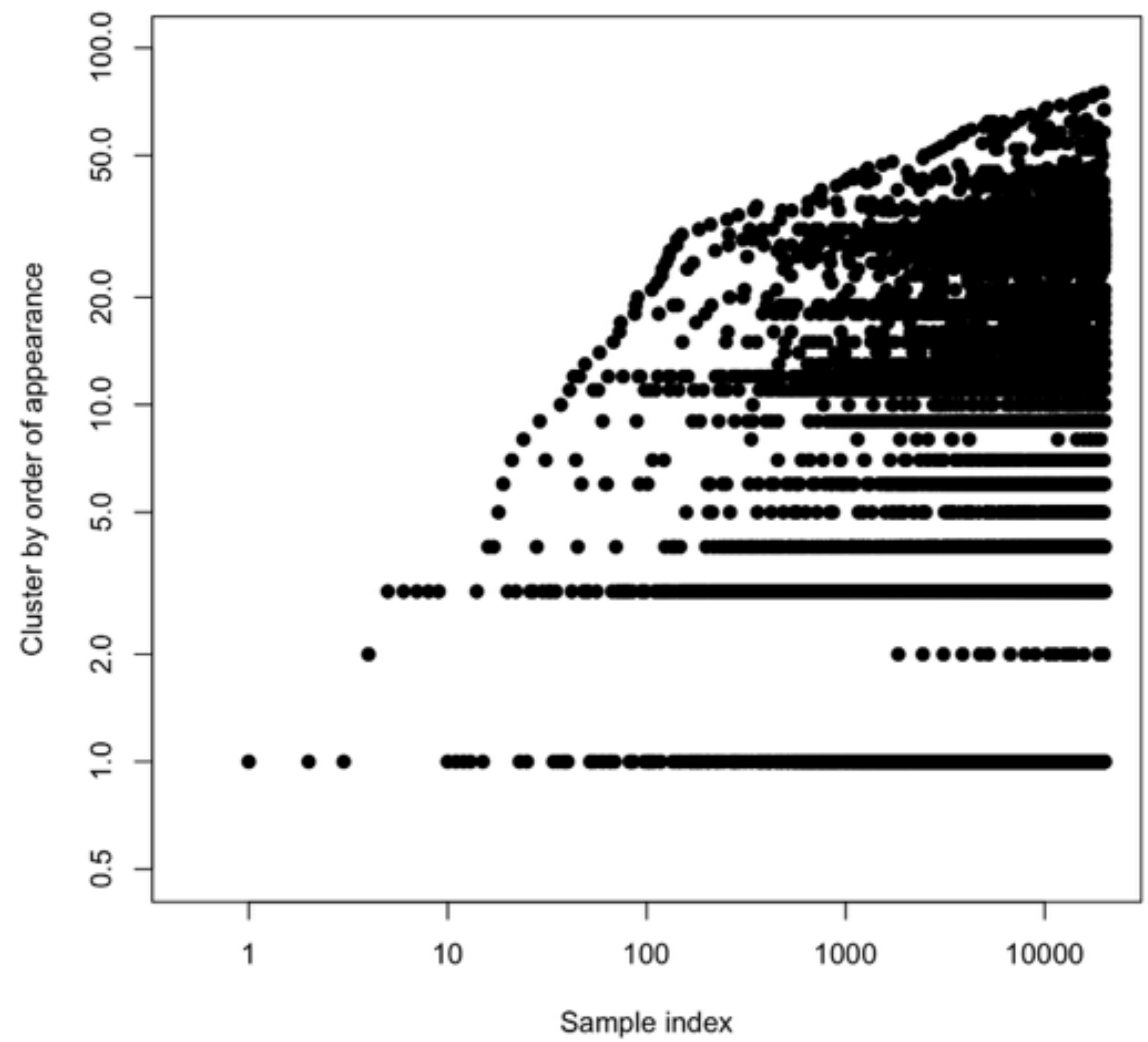


[Gnedin, et al 2007, Pitman, Yor 1997

Power laws

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

$$K_N \sim \alpha N^{\sigma} \text{ w.p. } 1$$



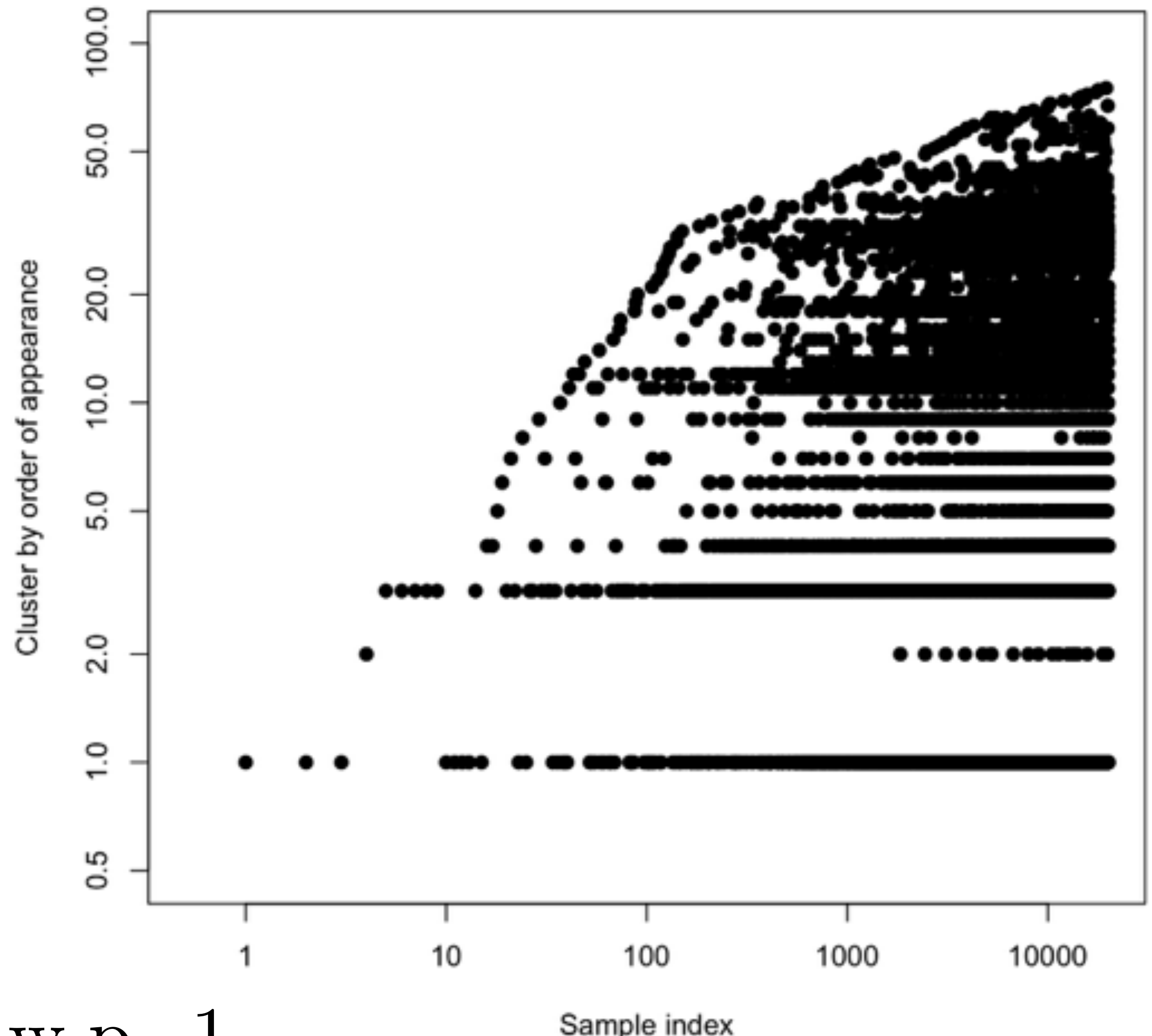
[Gnedin, et al 2007, Pitman, Yor 1997

Power laws

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

$$K_N \sim \alpha N^{\sigma} \text{ w.p. } 1$$

$$\Leftrightarrow \rho_j^{\downarrow} \sim C(\sigma) j^{-\sigma}, j \to \infty, \text{ w.p. } 1$$



[Gnedin, et al 2007, Pitman, Yor 1997

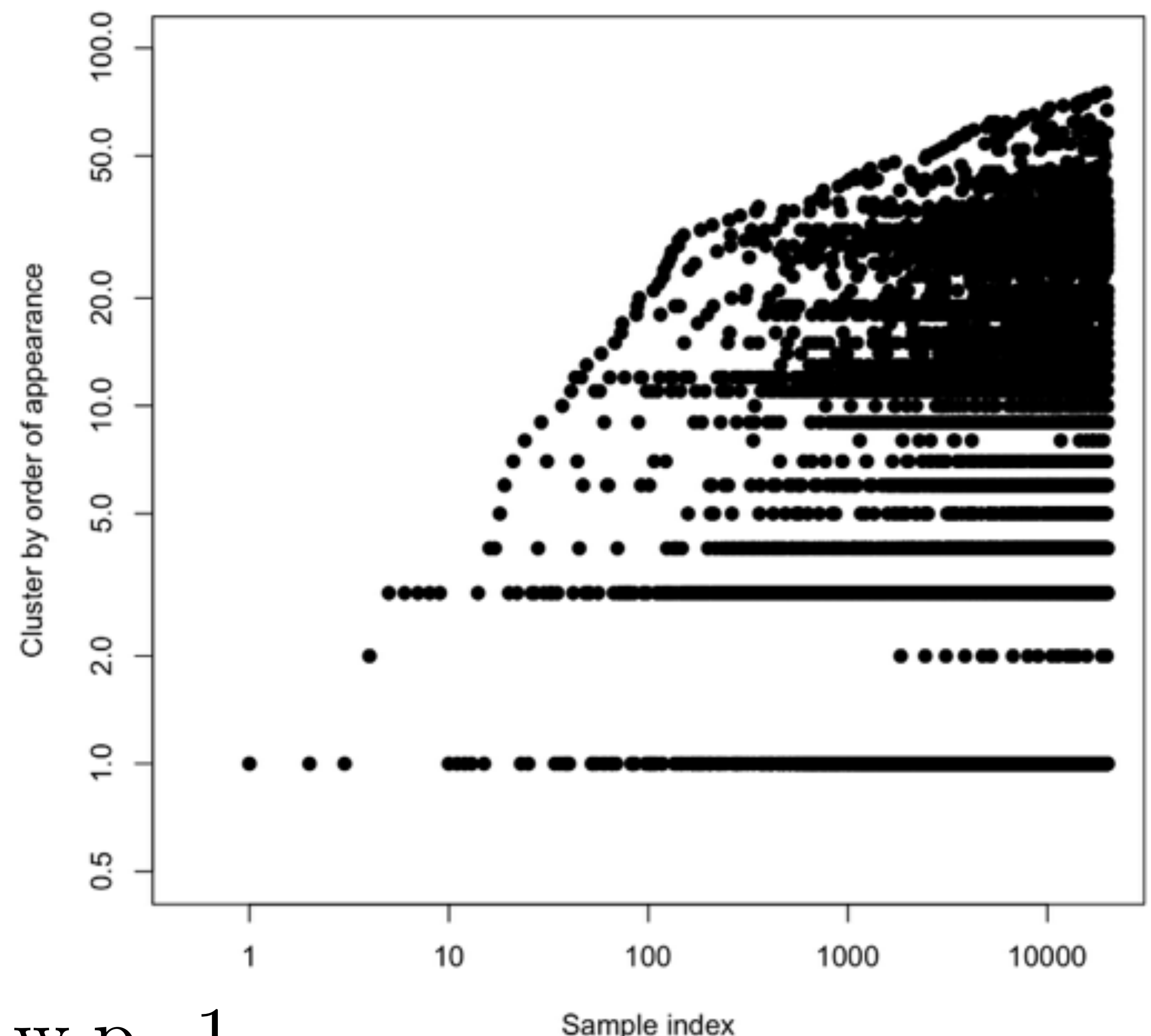
Power laws

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

$$K_N \sim \alpha N^\sigma \text{ w.p. } 1$$

$$\Leftrightarrow \rho_j^{\downarrow} \sim C(\sigma) j^{-\sigma}, j \to \infty, \text{ w.p. } 1$$

  - Zipf's law



[Gnedin, et al 2007, Pitman, Yor 1997

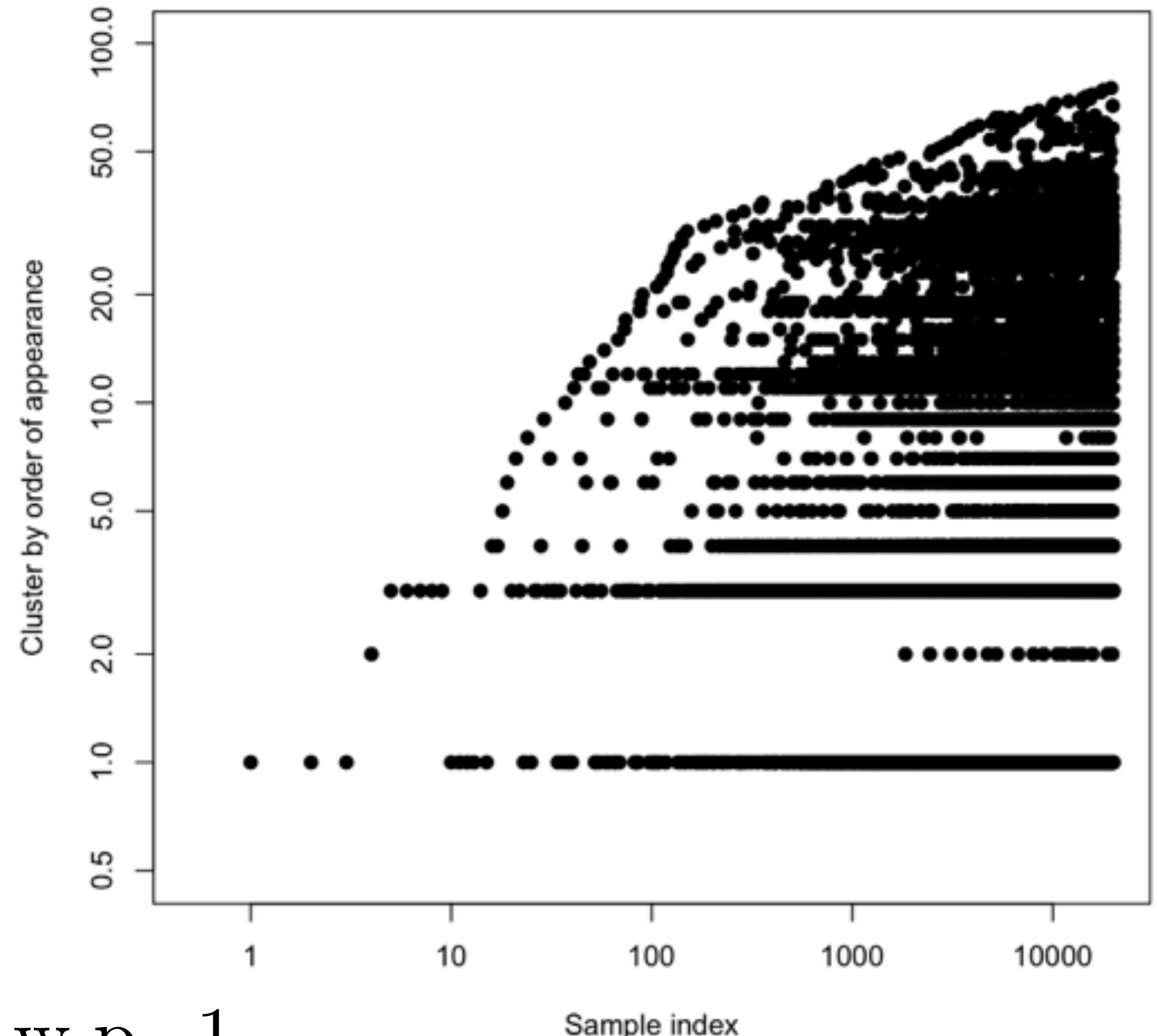
Power laws

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

$$K_N \sim \alpha N^{\sigma} \text{ w.p. } 1$$

$$\Leftrightarrow \rho_j^{\downarrow} \sim C(\sigma) j^{-\sigma}, j \to \infty, \text{ w.p. } 1$$

  - Zipf's law



[Gnedin, et al 2007, Pitman, Yor 1997, Goldwater et al 2005, Teh 2006

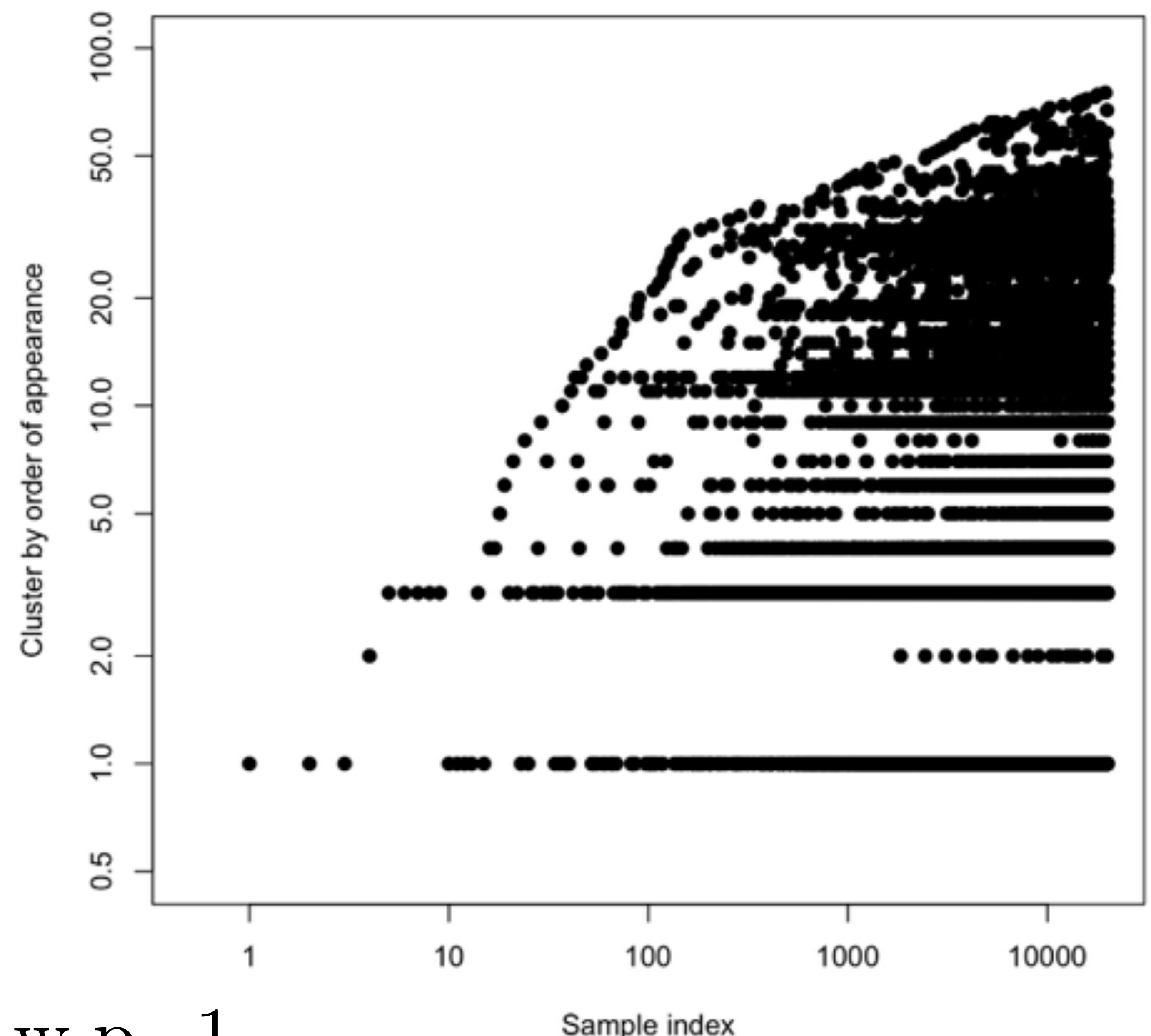
Power laws

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

$$K_N \sim \alpha N^{\sigma} \text{ w.p. } 1$$

$$\Leftrightarrow \rho_j^{\downarrow} \sim C(\sigma) j^{-\sigma}, j \to \infty, \text{ w.p. } 1$$

    - Zipf's law



[Gnedin, et al 2007, Pitman, Yor 1997, Goldwater et al 2005, Teh 2006, Broderick et al 2012]

Hierarchies

Hierarchies

- Hierarchical Dirichlet process

Hierarchies

- Hierarchical Dirichlet process

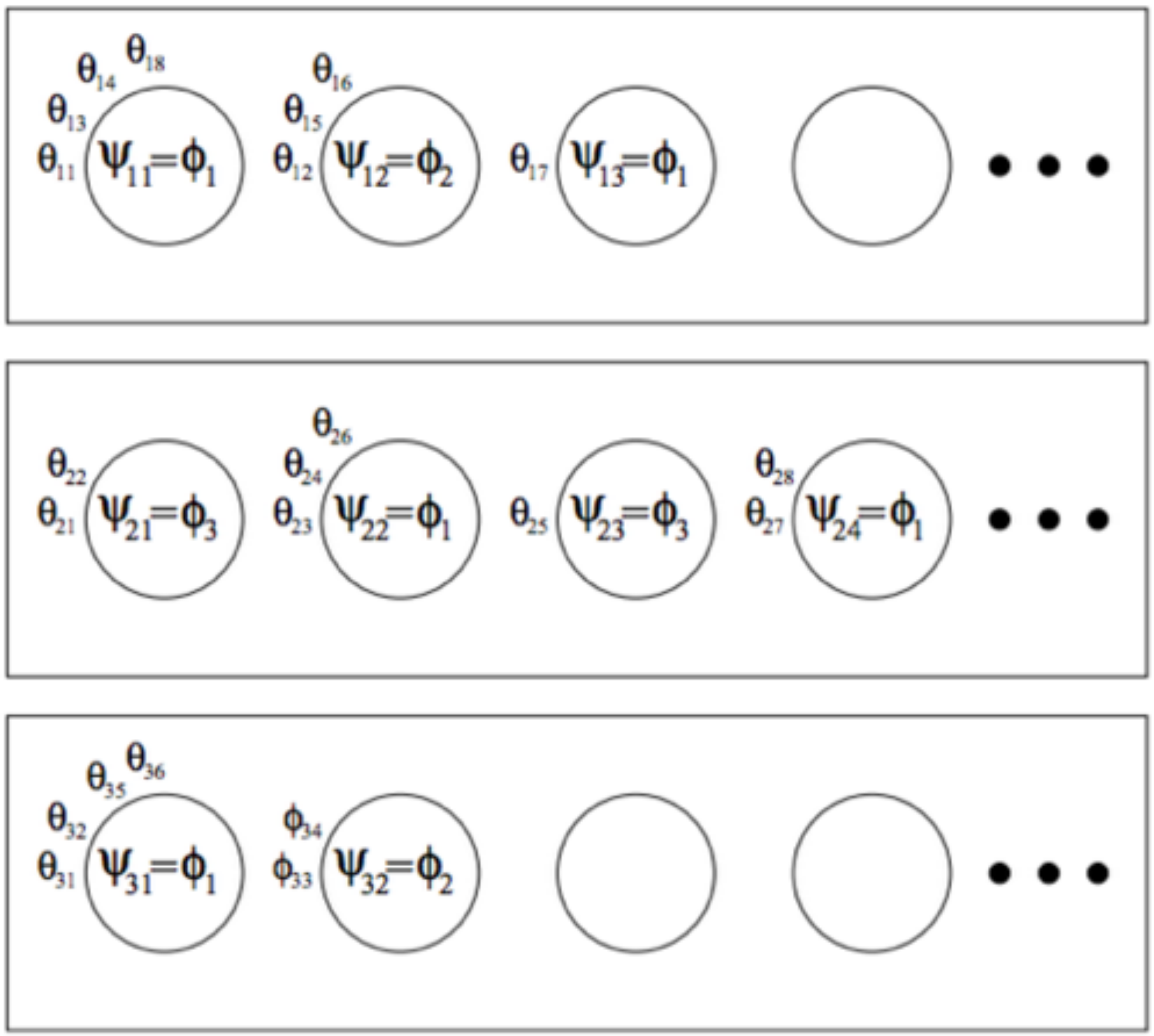
[Teh et al 2006, Rodríguez et al 2008

Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008

Hierarchies

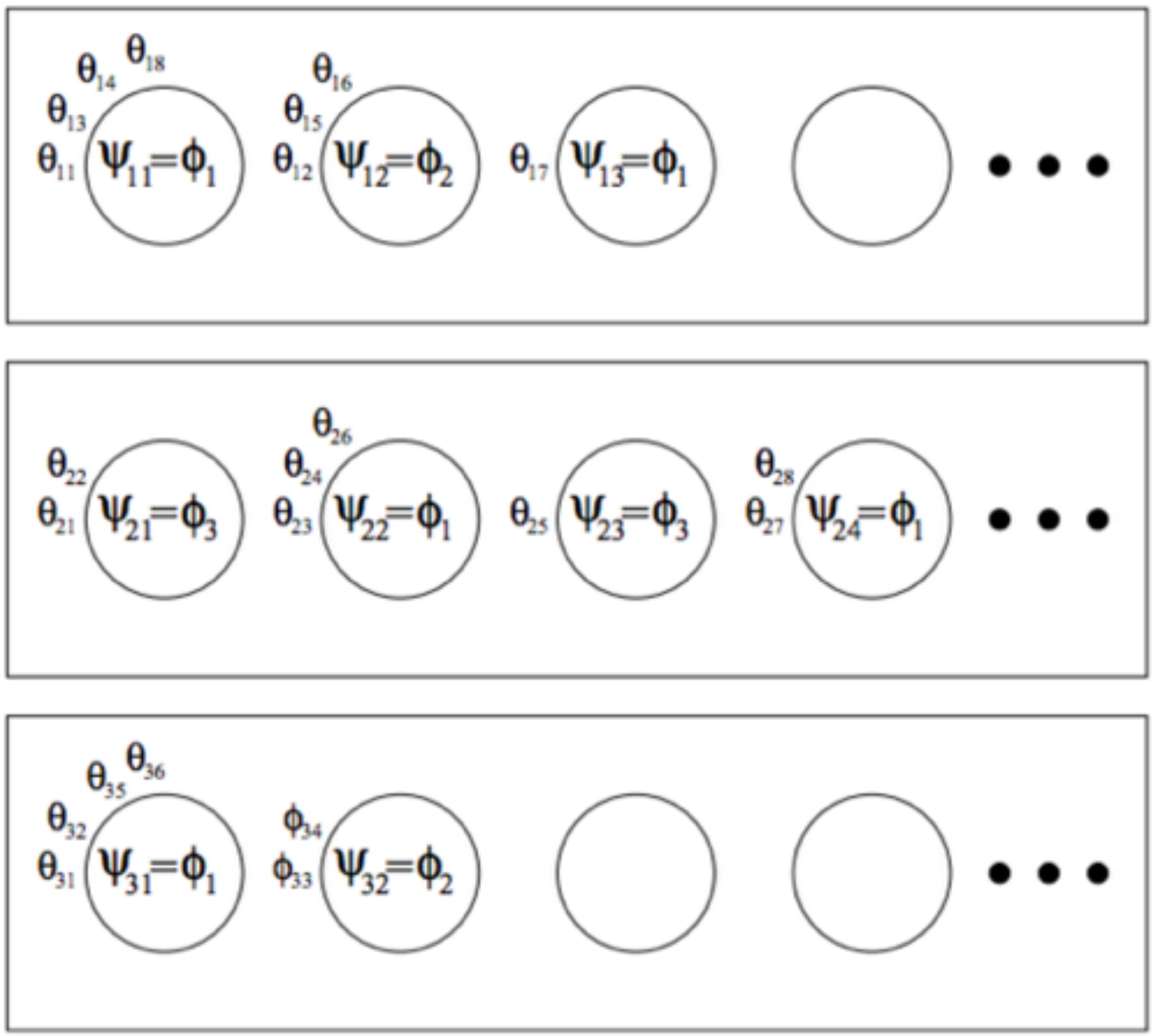


[Teh et al 2006]

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008

Hierarchies

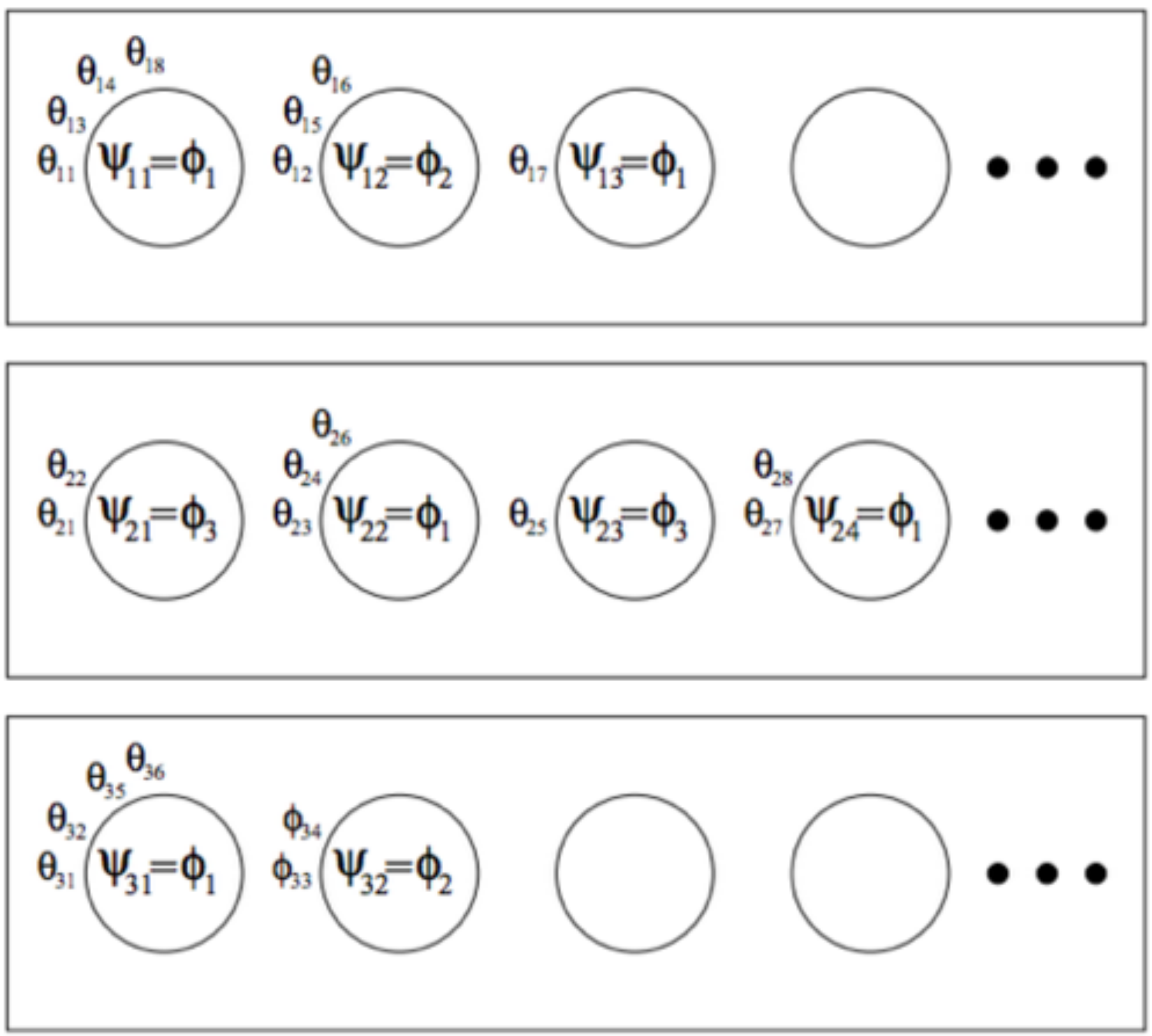


[Teh et al 2006]

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

[Teh et al 2006, Rodríguez et al 2008

Hierarchies



[Teh et al 2006]

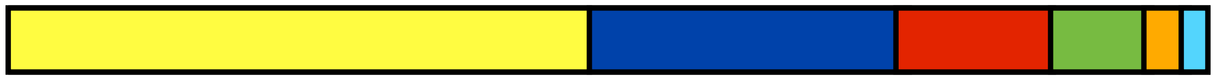
- Hierarchical Dirichlet process
- Chinese restaurant franchise
- Hierarchical beta process

[Teh et al 2006, Rodríguez et al 2008, Thibaux, Jordan 2007]

De Finetti mixing measures

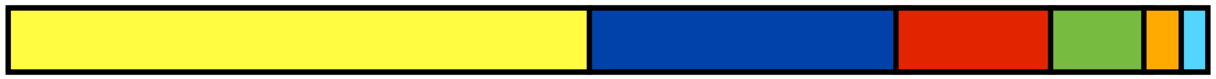
De Finetti mixing measures

- Clustering: Kingman paintbox



De Finetti mixing measures

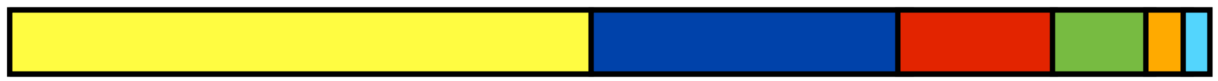
- Clustering: Kingman paintbox



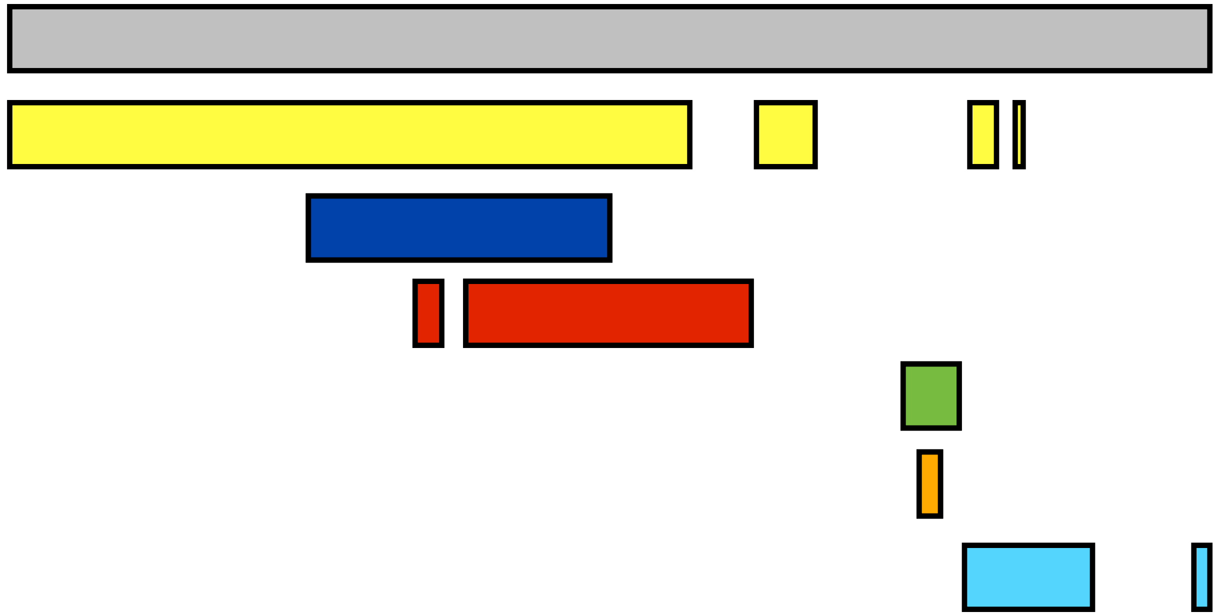
[Kingman 1978

De Finetti mixing measures

- Clustering: Kingman paintbox



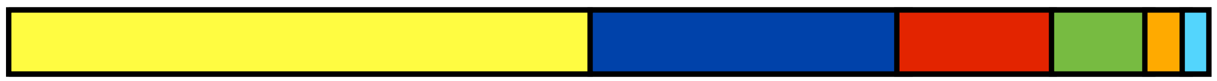
- Feature allocation: Feature paintbox



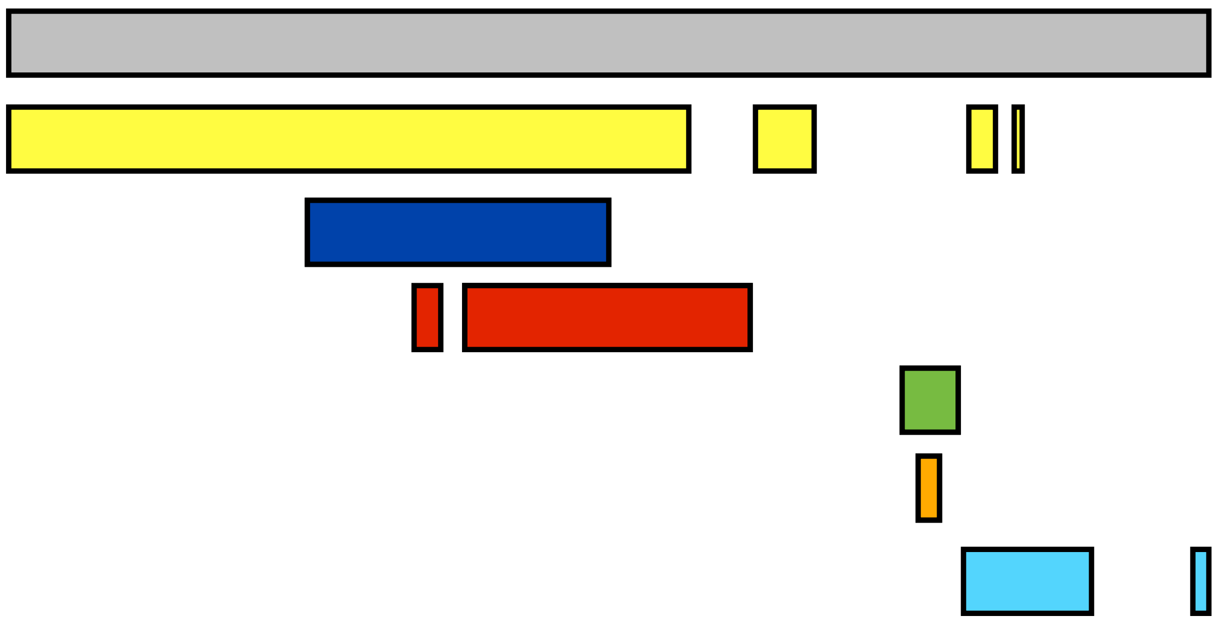
[Kingman 1978

De Finetti mixing measures

- Clustering: Kingman paintbox



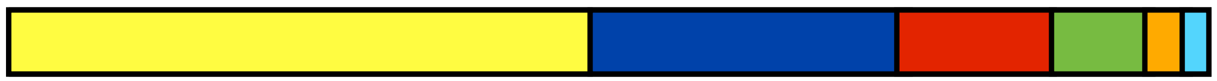
- Feature allocation: Feature paintbox



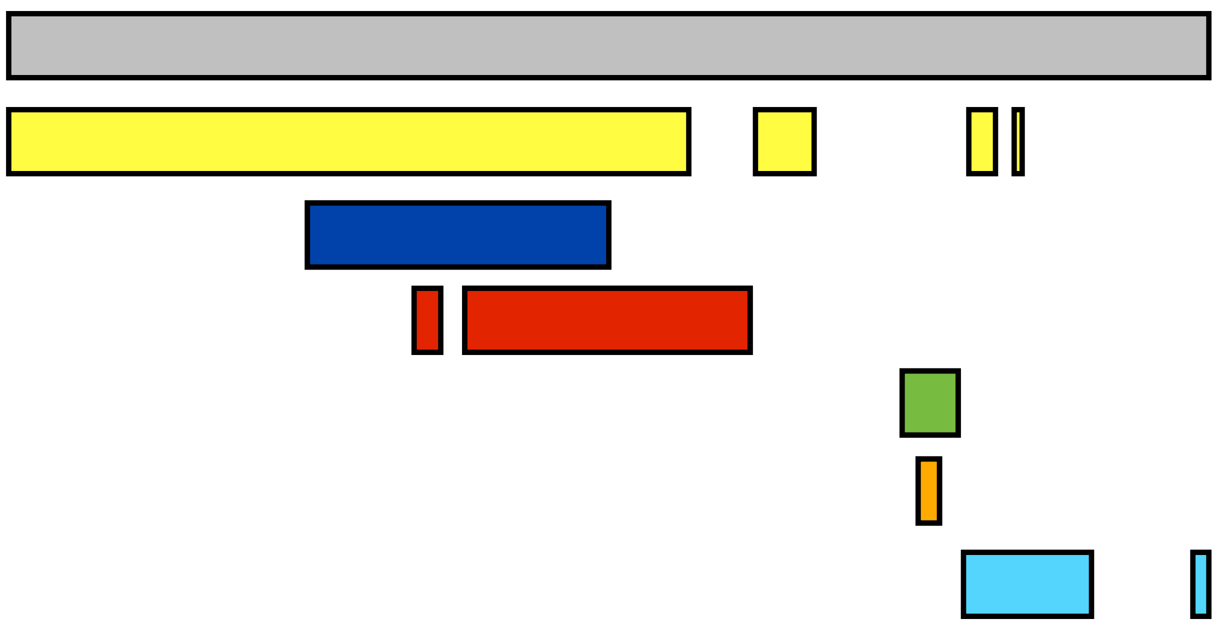
[Kingman 1978, Broderick, Pitman, Jordan 2013

De Finetti mixing measures

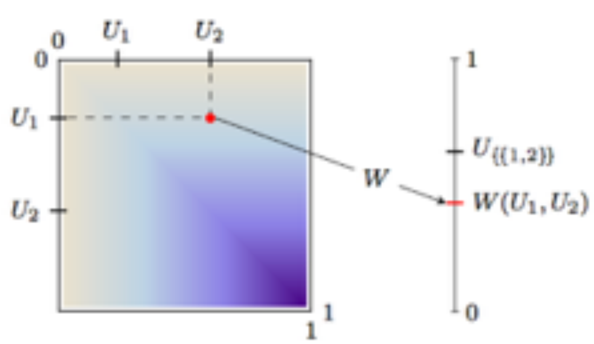
- Clustering: Kingman paintbox

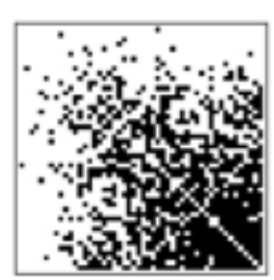


- Feature allocation: Feature paintbox



- Graphs/networks: Aldous-Hoover theorem



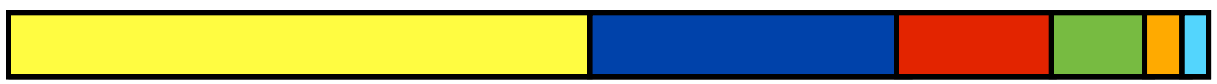


[Lloyd 2012]

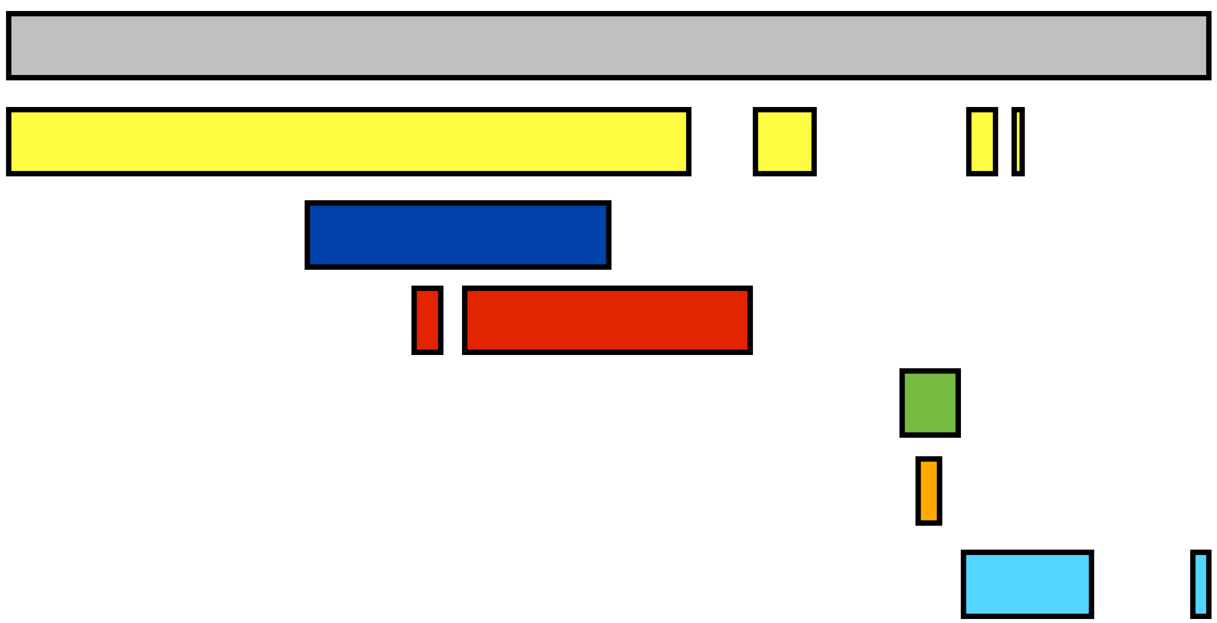
[Kingman 1978, Broderick, Pitman, Jordan 2013

De Finetti mixing measures

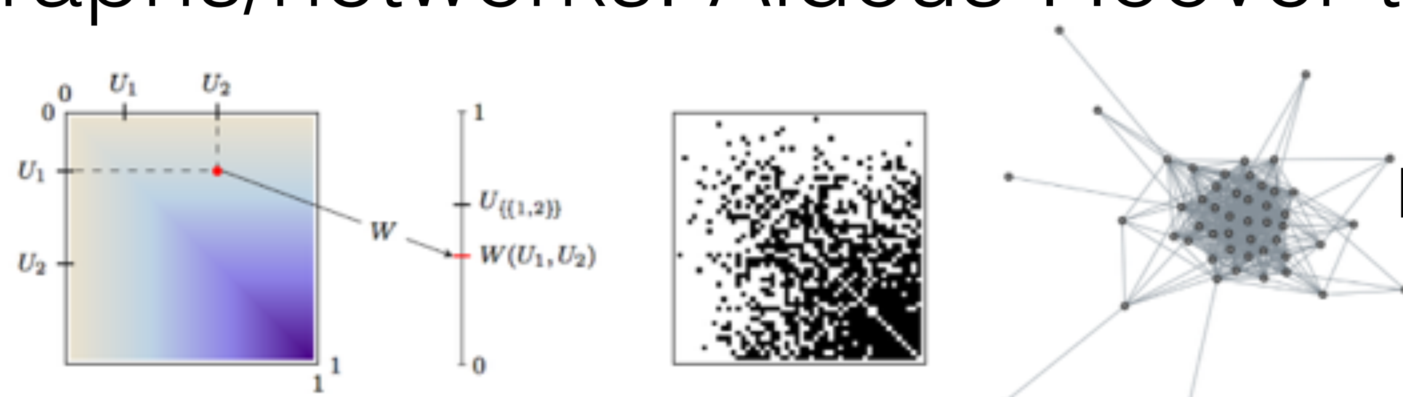
- Clustering: Kingman paintbox



- Feature allocation: Feature paintbox



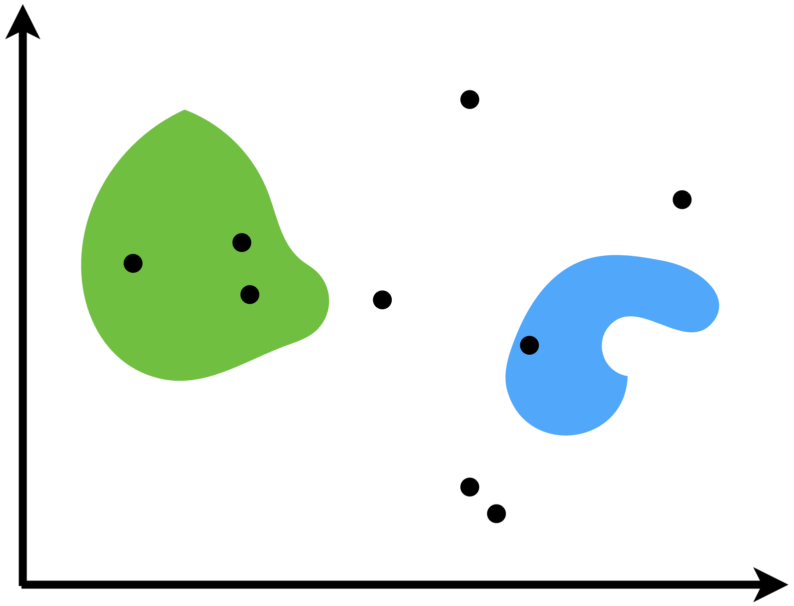
- Graphs/networks: Aldous-Hoover theorem



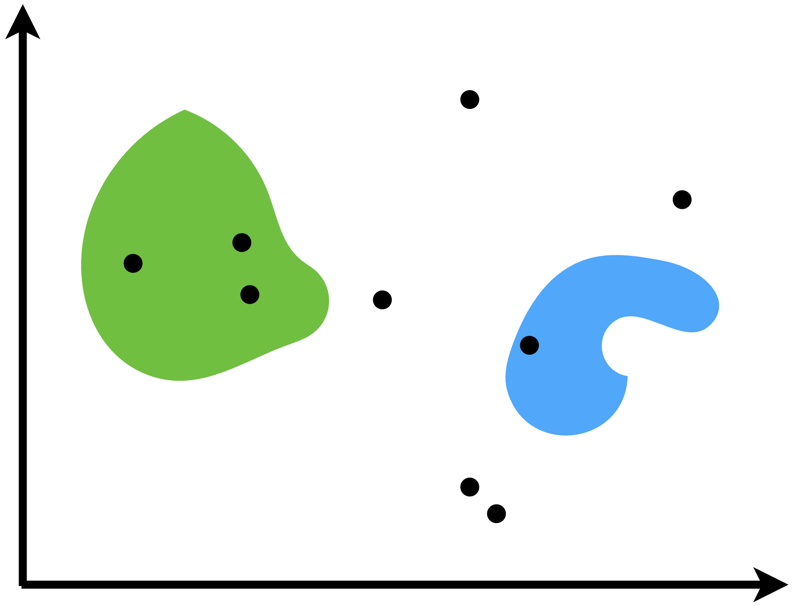
[Lloyd 2012]

[Kingman 1978, Broderick, Pitman, Jordan 2013, Aldous 1983, Orbanz, Roy 2015]

Poisson point processes



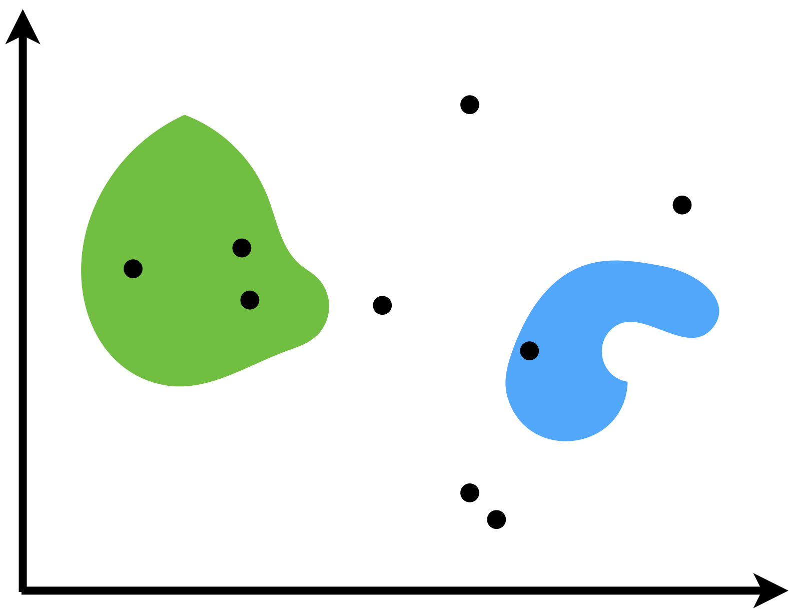
Poisson point processes



[Kingman 1992

Poisson point processes

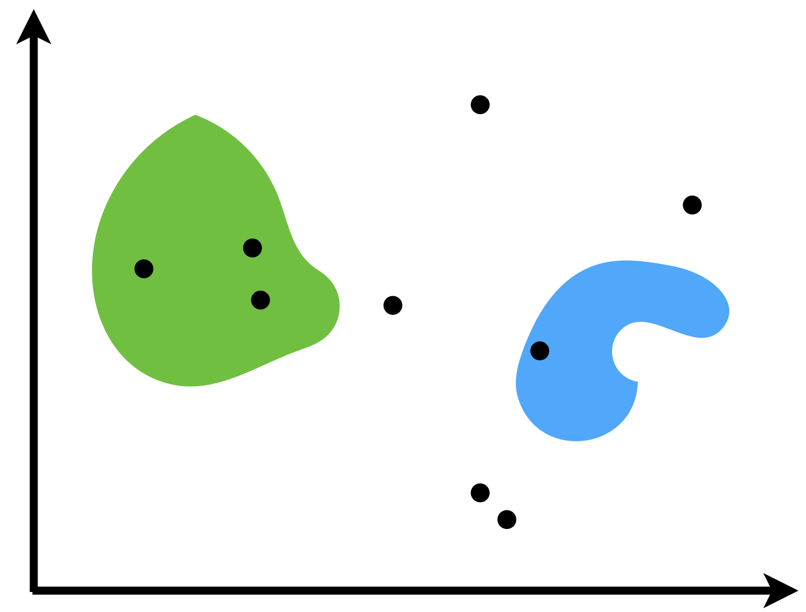
- Beta process, Bernoulli process (Indian buffet)



[Kingman 1992

Poisson point processes

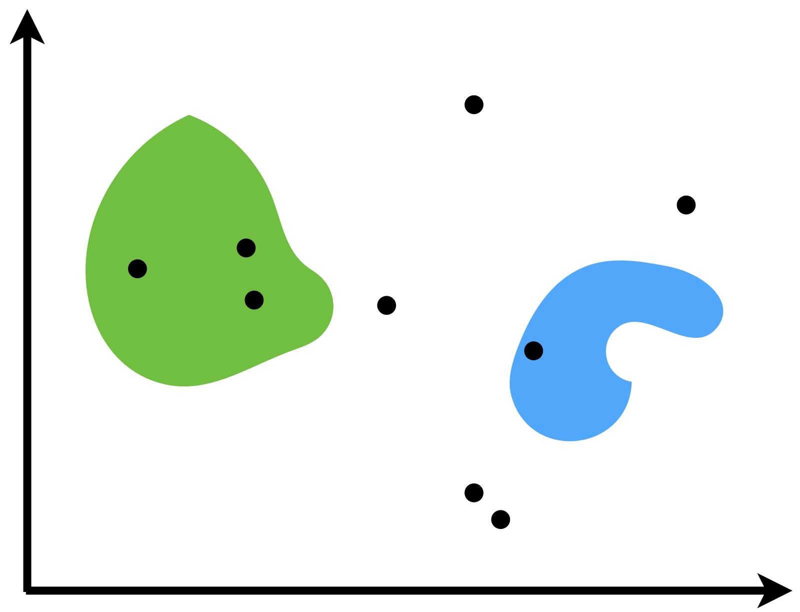
- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)



[Kingman 1992

Poisson point processes

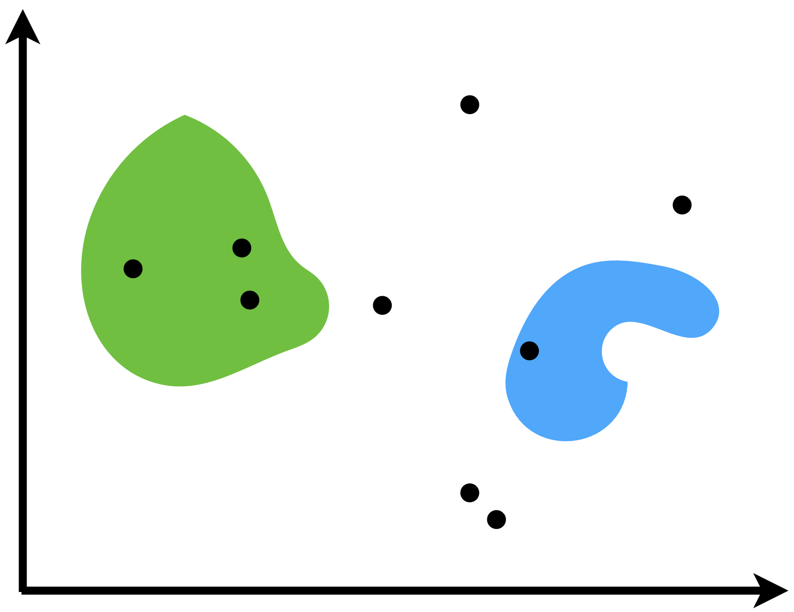
- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process



[Kingman 1992

Poisson point processes

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

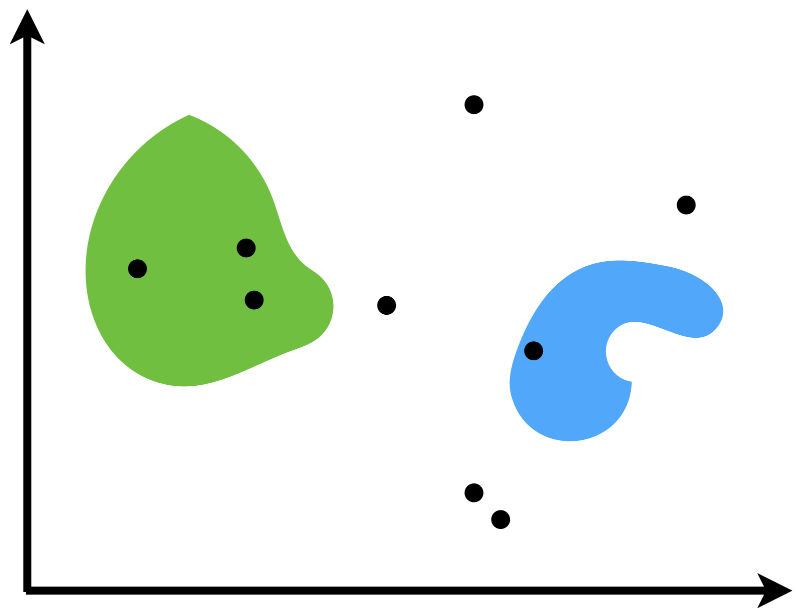


- Posteriors, conjugacy, and exponential families for completely random measures

[Kingman 1992

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

[Kingman 1992, Orbanz 2009, Broderick et al 2014, James 2014]

Nonparametric Bayes

Nonparametric Bayes

- Bayesian statistics that is not parametric

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

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

Nonparametric Bayes

- Bayesian statistics that is not parametric
- Bayesian

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

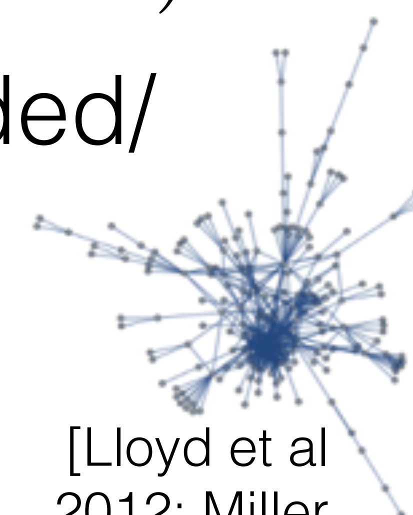
- Not parametric (i.e. not finite parameter, unbounded/ growing/infinite number of parameters)

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)



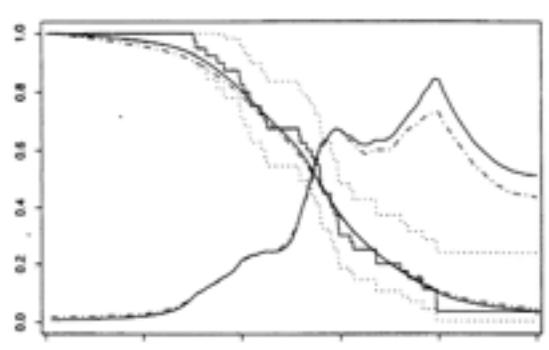
[Lloyd et al 2012; Miller et al, 2010]

[wikipedia.org]

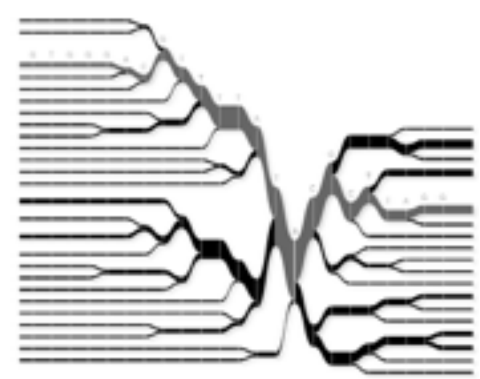
[Ed Bowlby, NOAA]

[Fox, et al 2014]

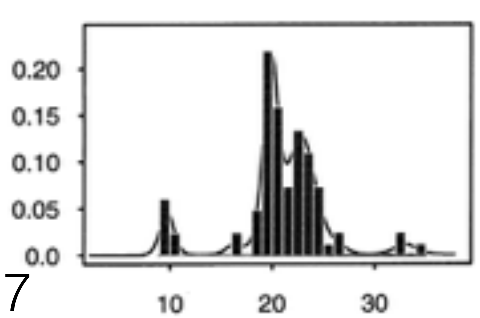
[Sudderth, Jordan 2009]



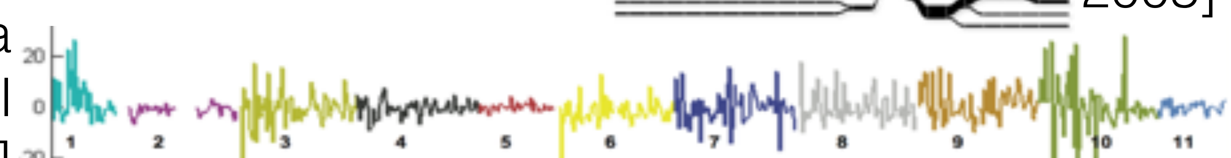
[Arjas, Gasbarra 1994]



[Ewens, 1972; Hartl, Clark 2003]



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



[Saria et al 2010]

References (Part III), page 1

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

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

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

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.

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

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.

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

References (Part III), page 2

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.

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

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.

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.