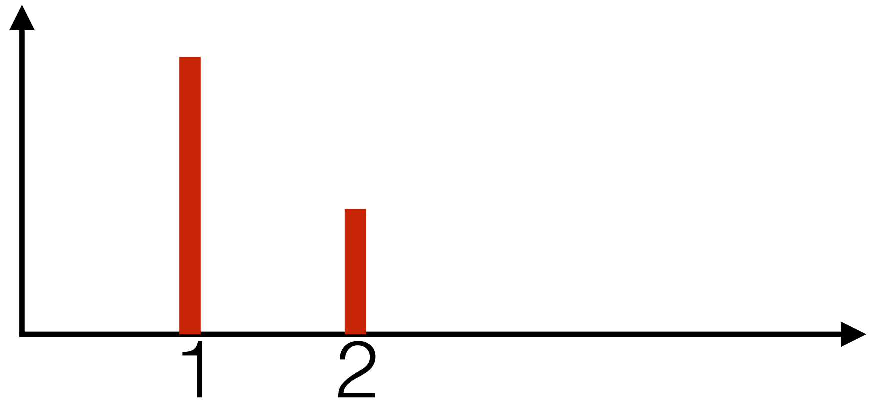
Nonparametric Bayes: Part II

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

Distributions

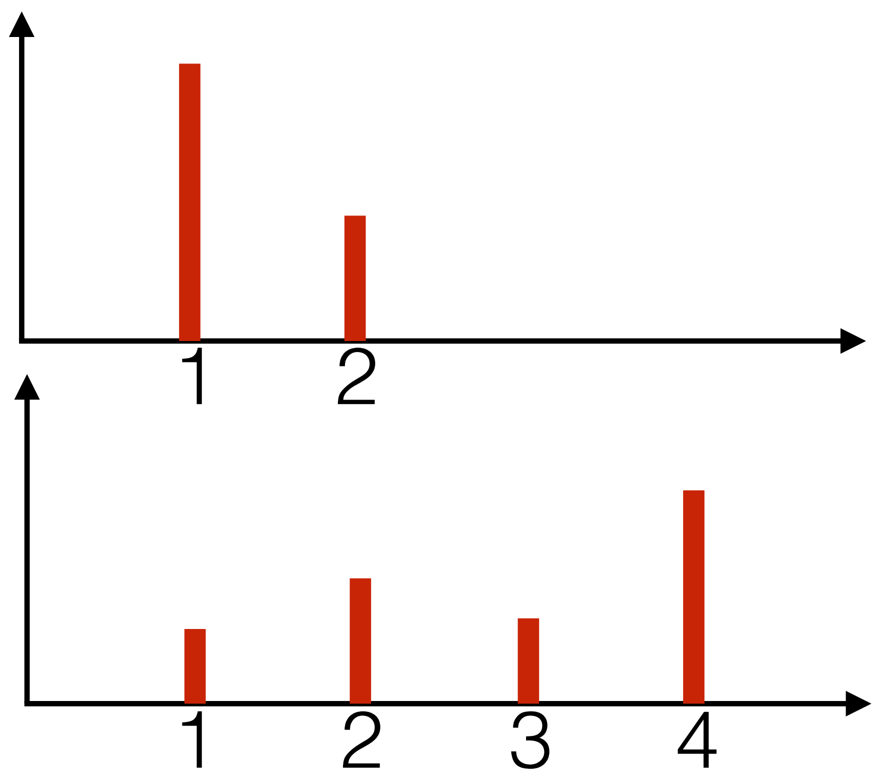
Distributions

- Beta → random distribution over $1, 2$



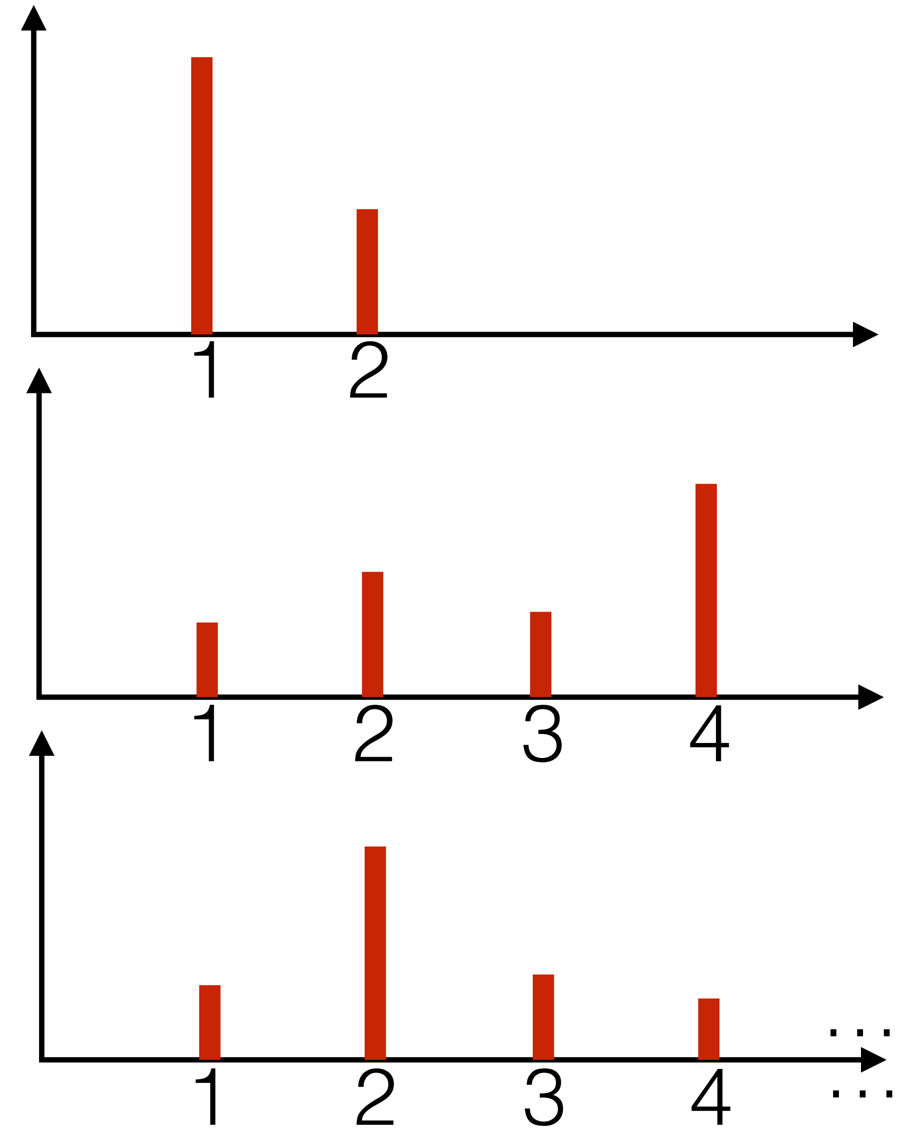
Distributions

- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$



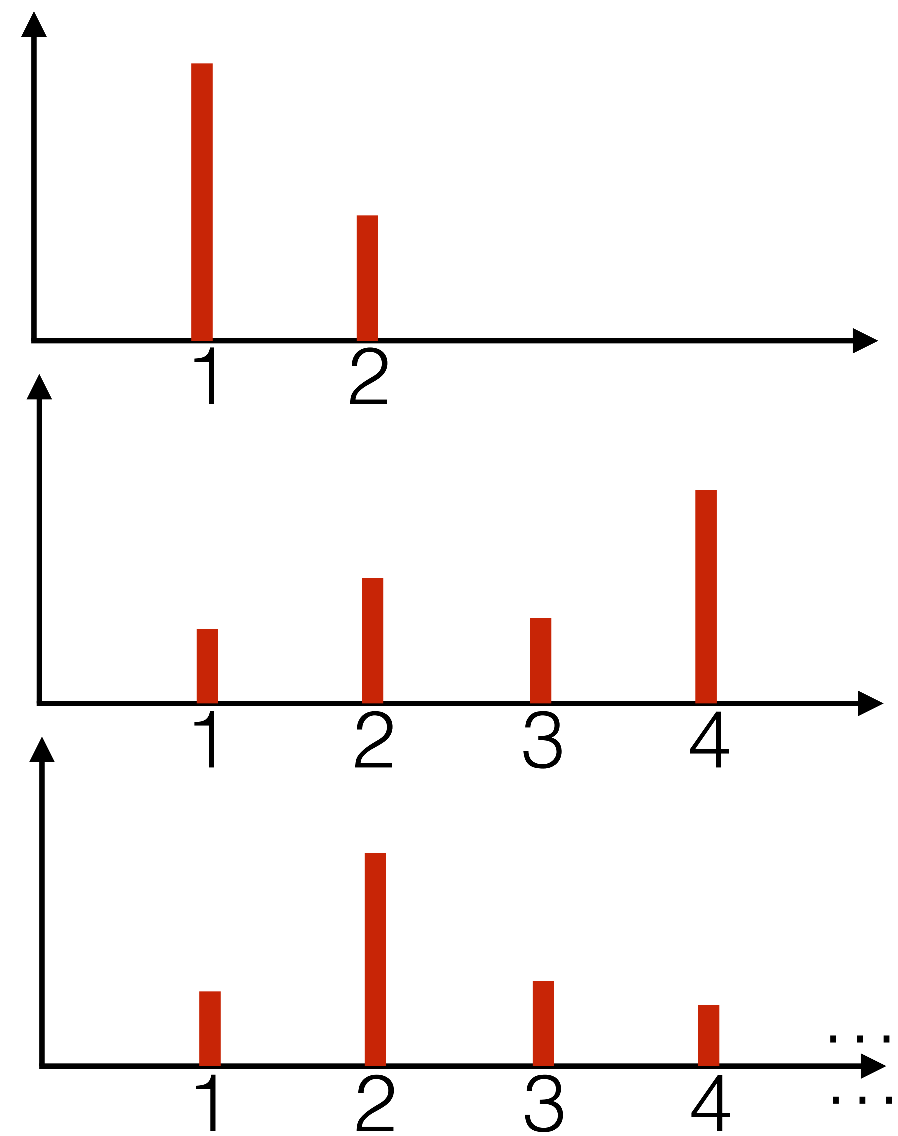
Distributions

- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



Distributions

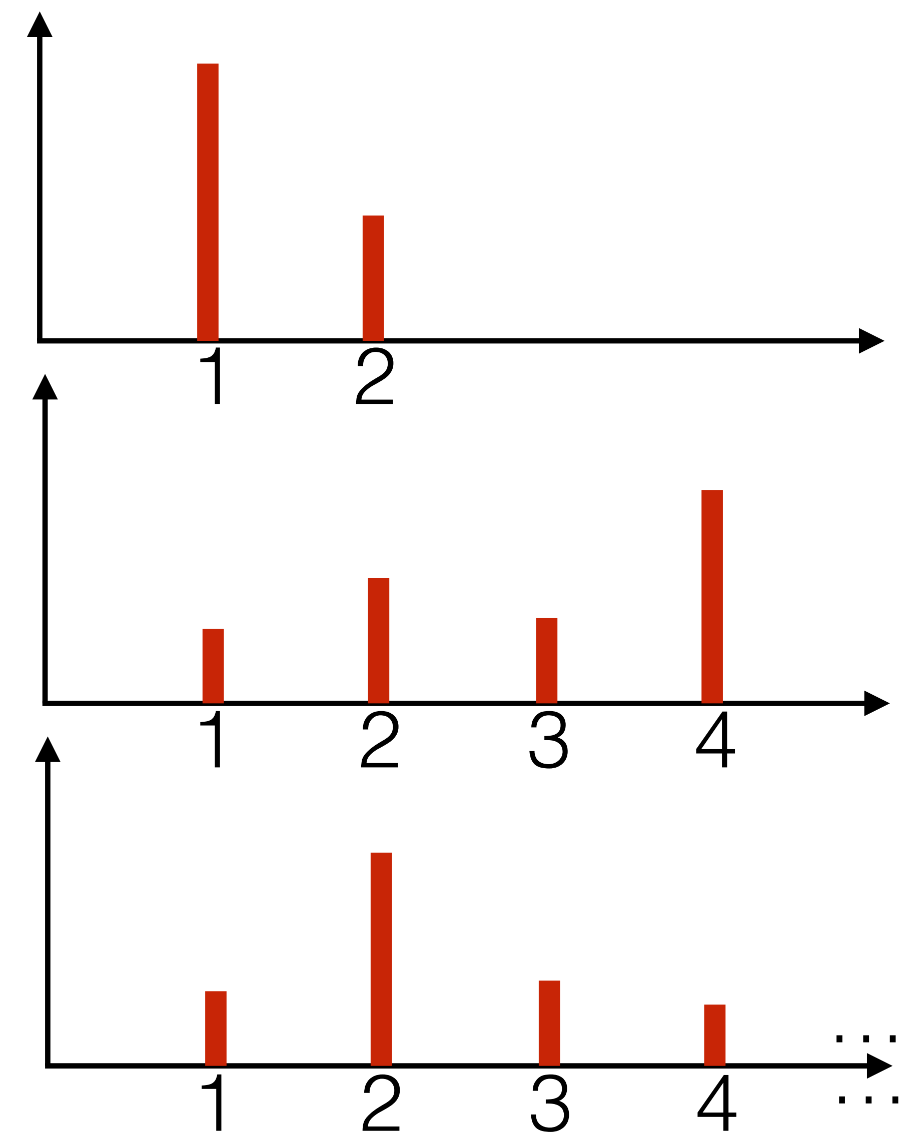
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



- Infinity of parameters: components
- Growing number of parameters: clusters

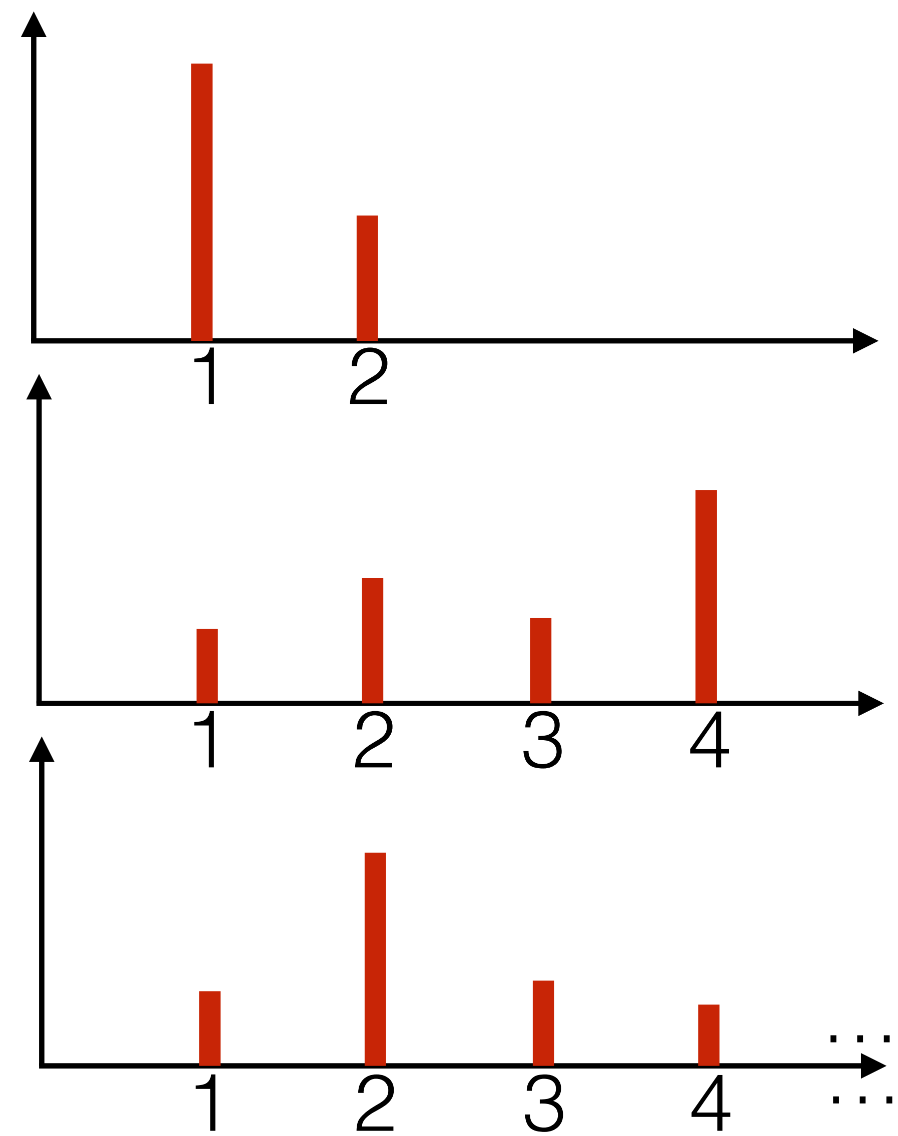
Distributions

- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



Distributions

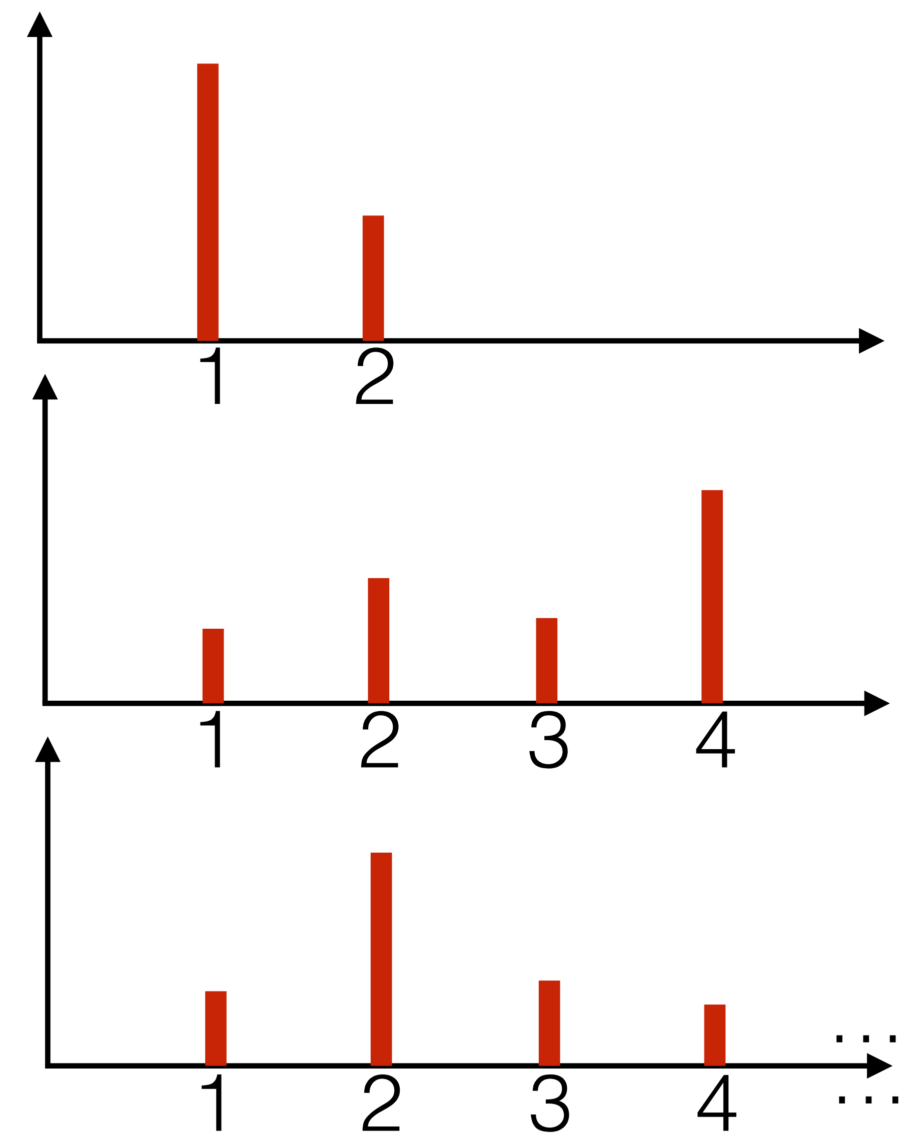
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



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

Distributions

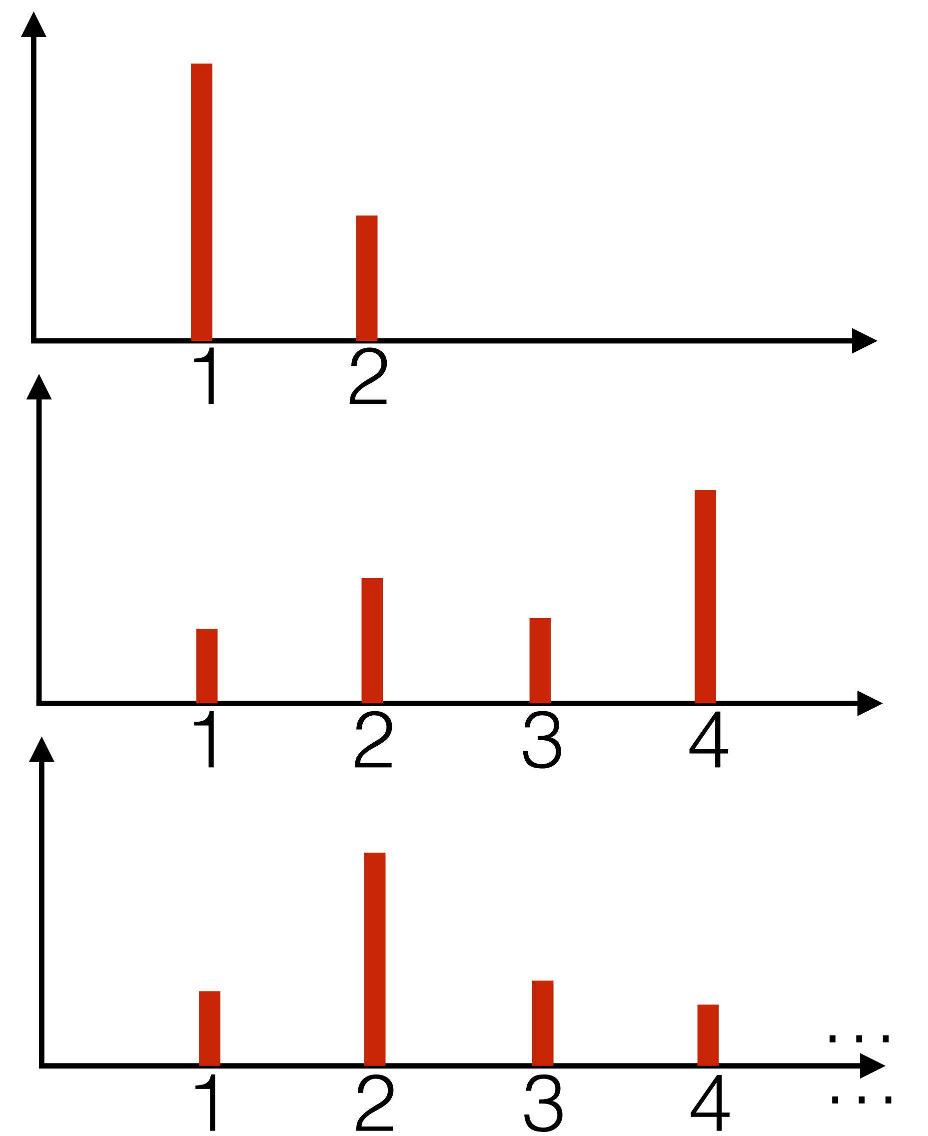
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$
$$\phi_k \overset{iid}{\sim} G_0$$

Distributions

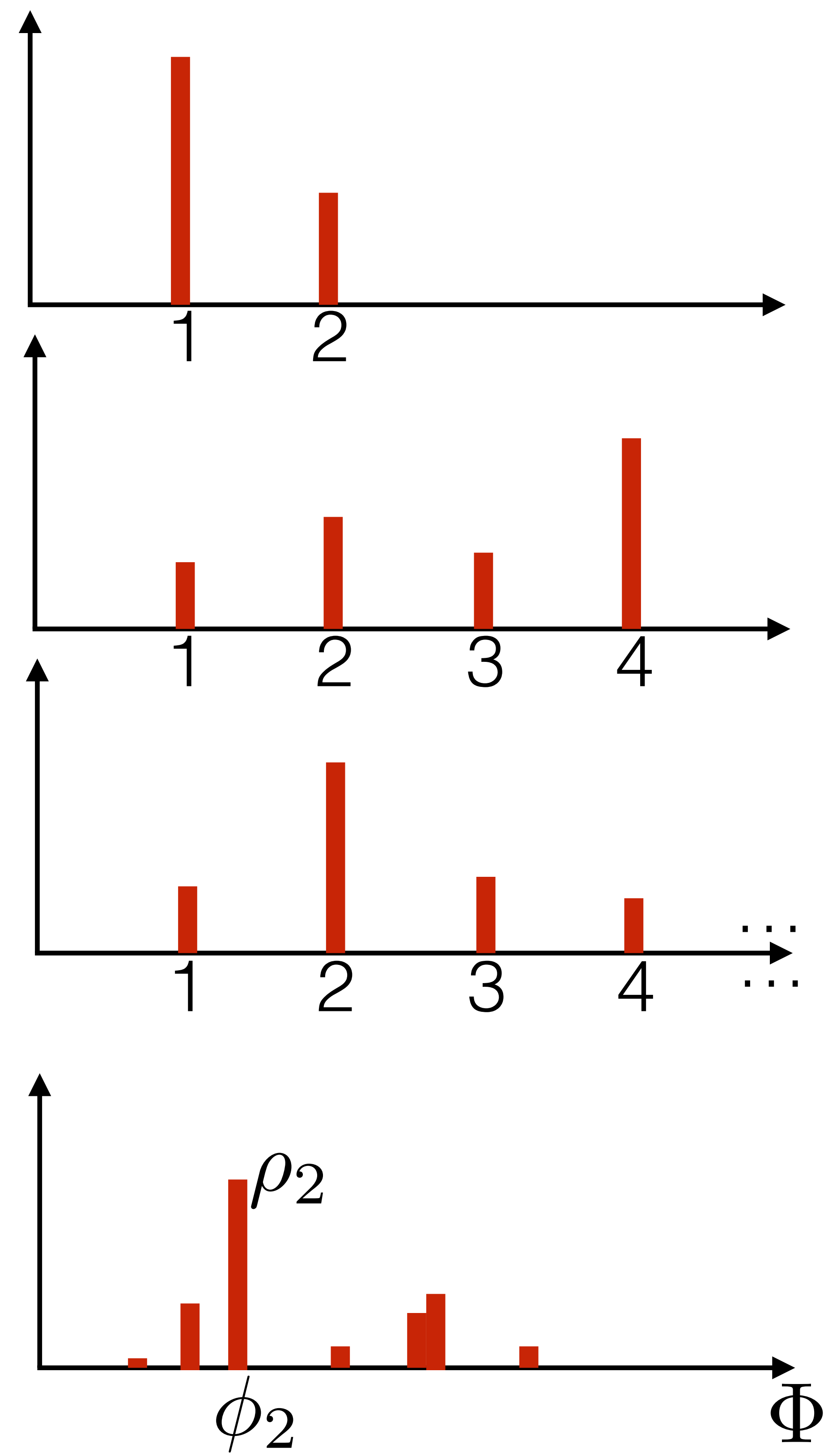
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$
$$\phi_k \overset{iid}{\sim} G_0$$
$$G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}$$

Distributions

- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$
$$\phi_k \overset{iid}{\sim} G_0$$
$$G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}$$

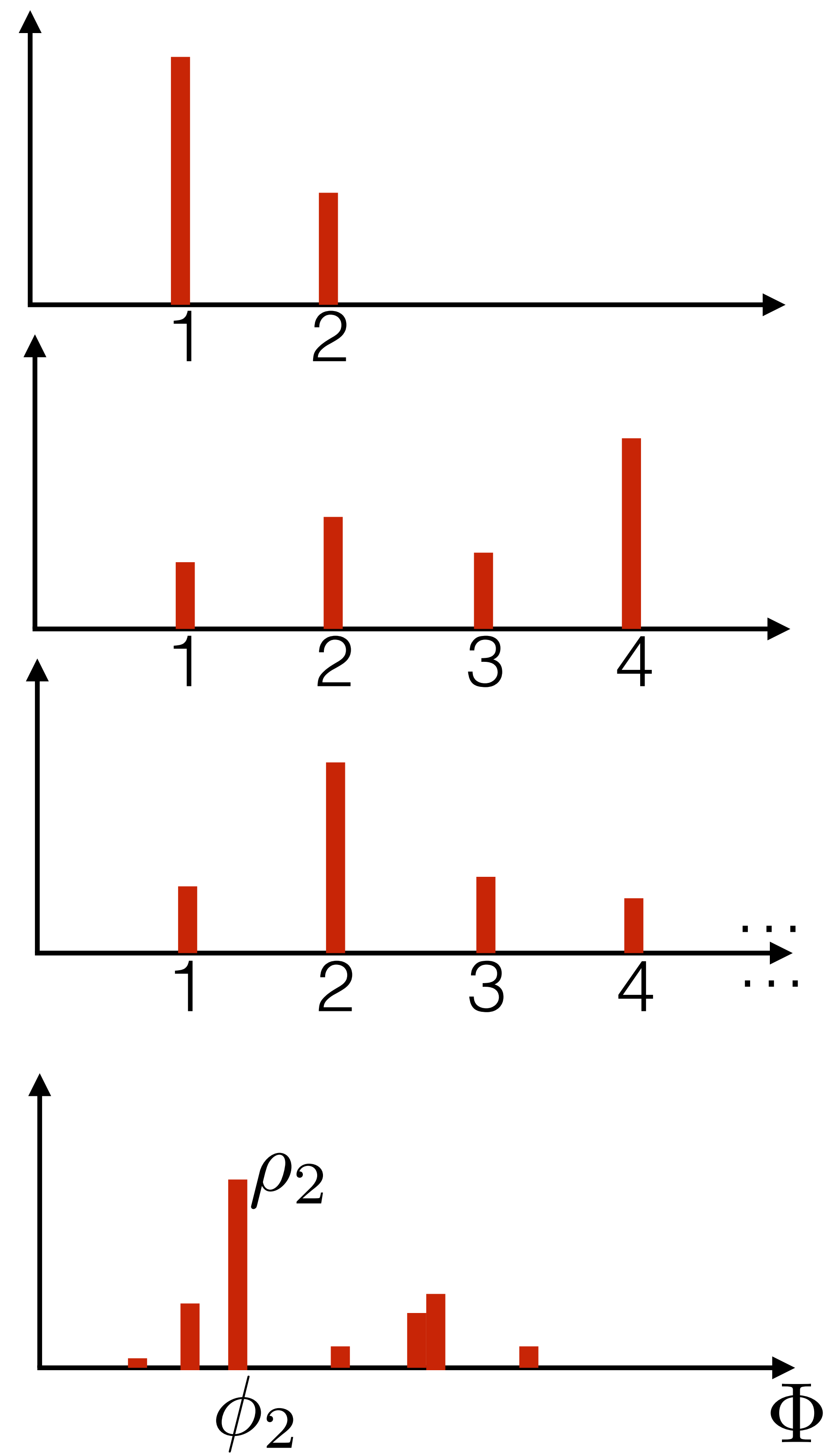
Distributions

- Beta → random distribution over $1, 2$

- Dirichlet → random distribution over $1, 2, \ldots, K$

- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$

- **Dirichlet process →** random distribution over Φ:
  $\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$
  $\phi_k \stackrel{iid}{\sim} G_0$
  $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}$



[Ferguson 1973]

Dirichlet process mixture model

Dirichlet process mixture model

- Gaussian mixture model

Dirichlet process mixture model

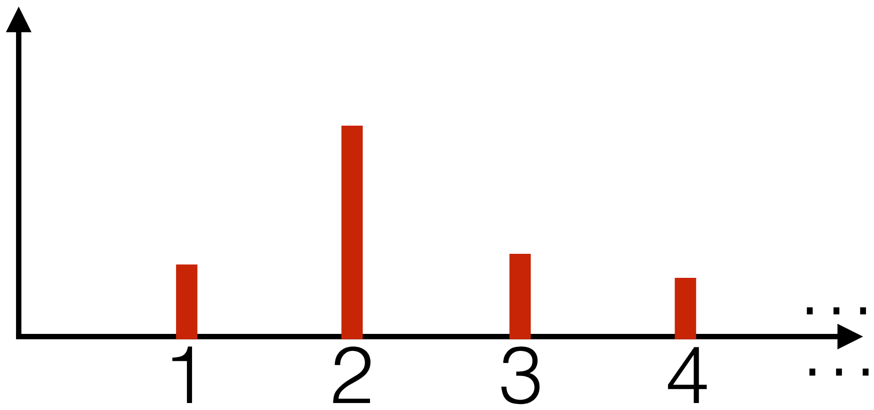
- Gaussian mixture model

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

Dirichlet process mixture model

- Gaussian mixture model

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

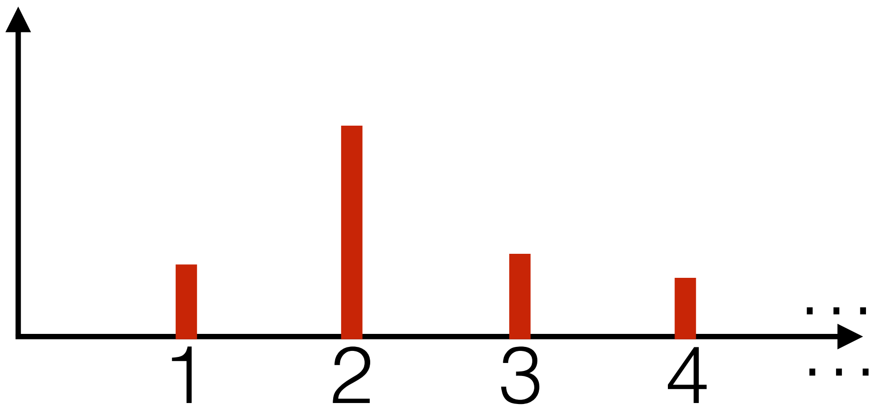


Dirichlet process mixture model

- Gaussian mixture model

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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

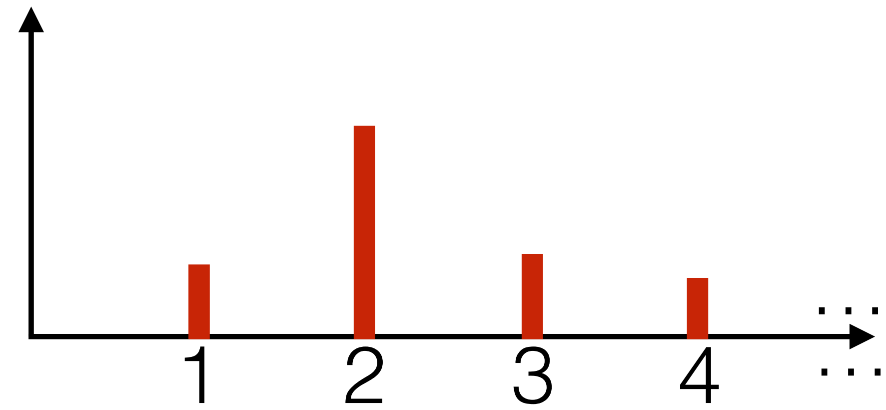


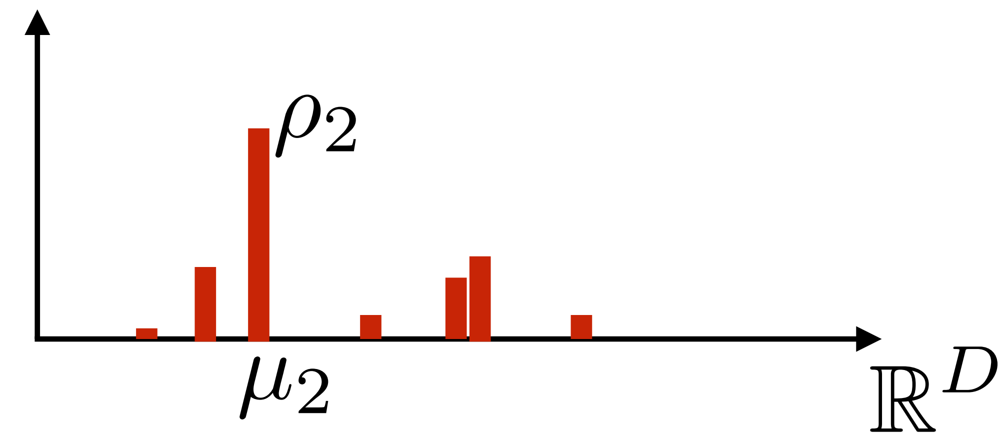
Dirichlet process mixture model

- Gaussian mixture model

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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$





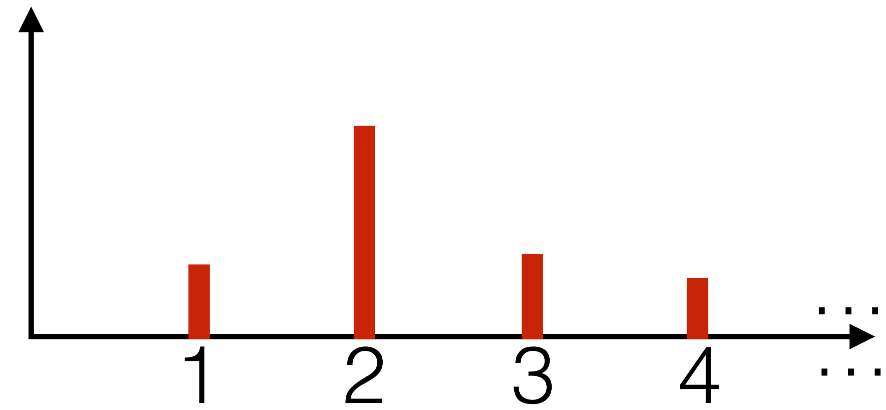
Dirichlet process mixture model

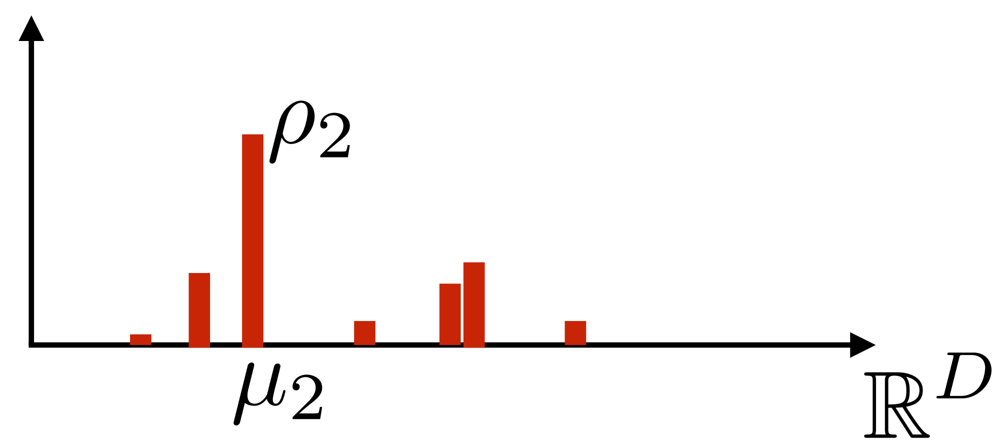
- Gaussian mixture model

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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

  - i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k}$





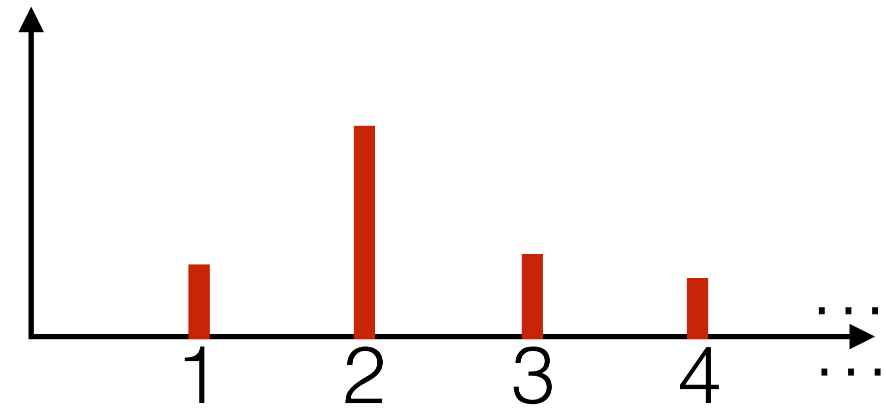
Dirichlet process mixture model

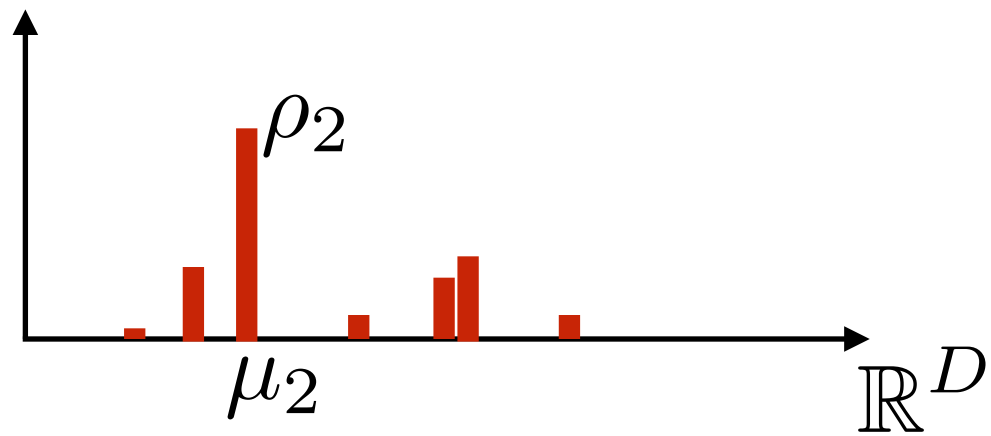
- Gaussian mixture model

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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

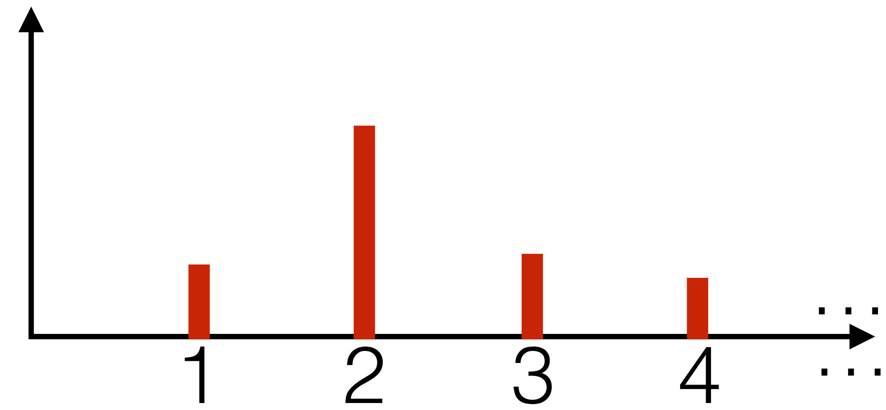
  - i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$





Dirichlet process mixture model

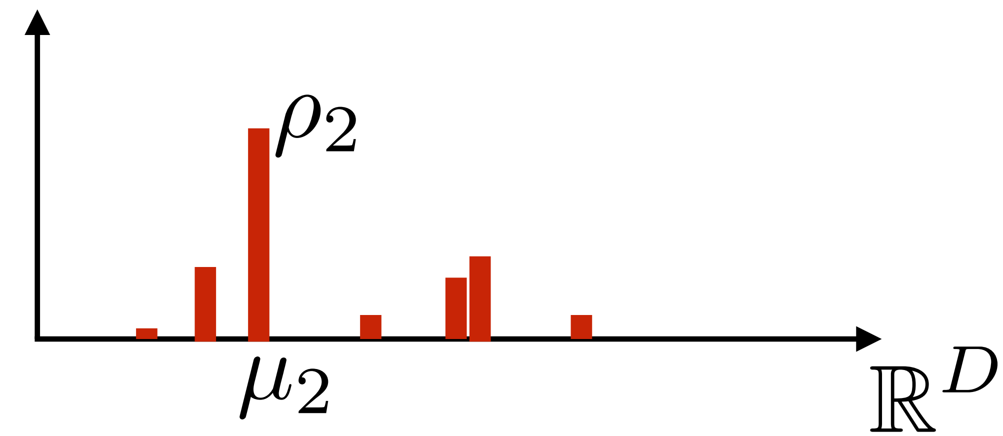
- Gaussian mixture model



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

$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

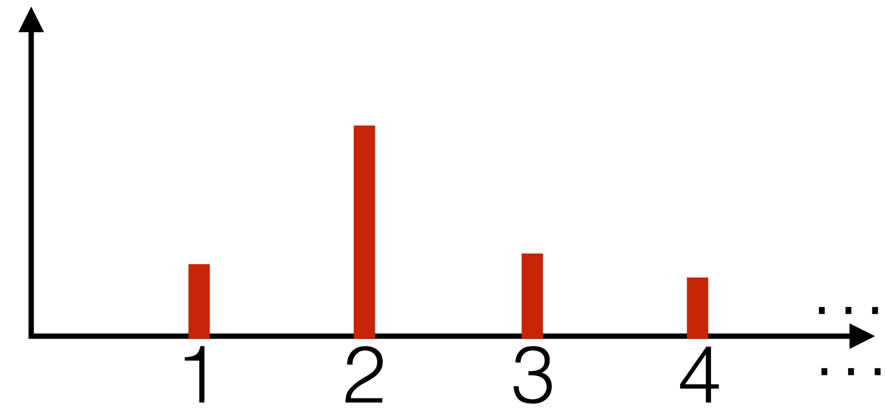
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \stackrel{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

Dirichlet process mixture model

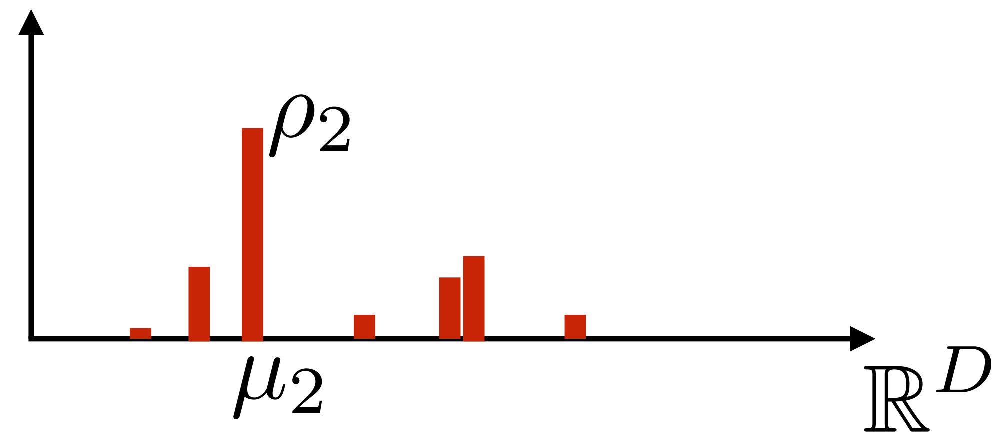
- Gaussian mixture model



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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$

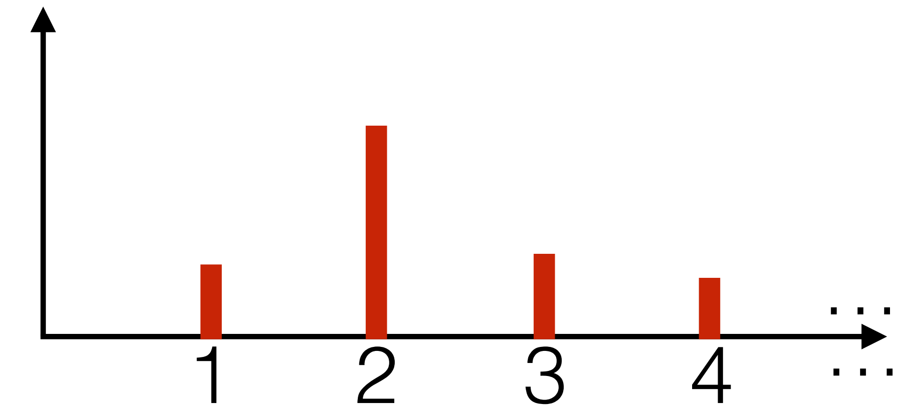


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

Dirichlet process mixture model

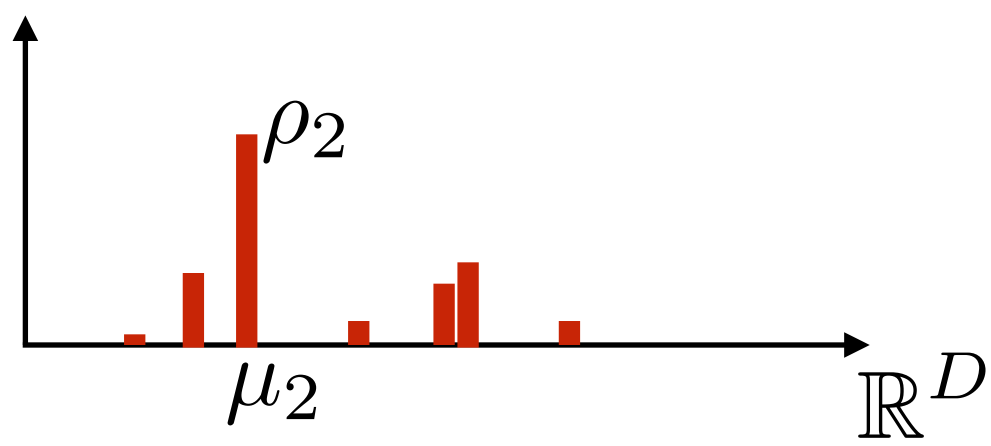
- Gaussian mixture model



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

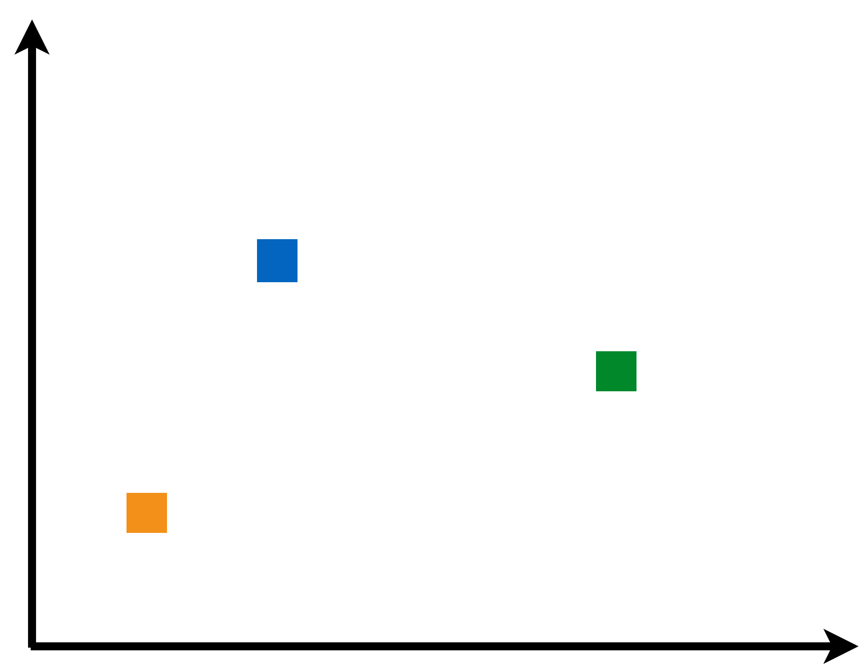
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

$$\mu_n^* = \mu_{z_n}$$

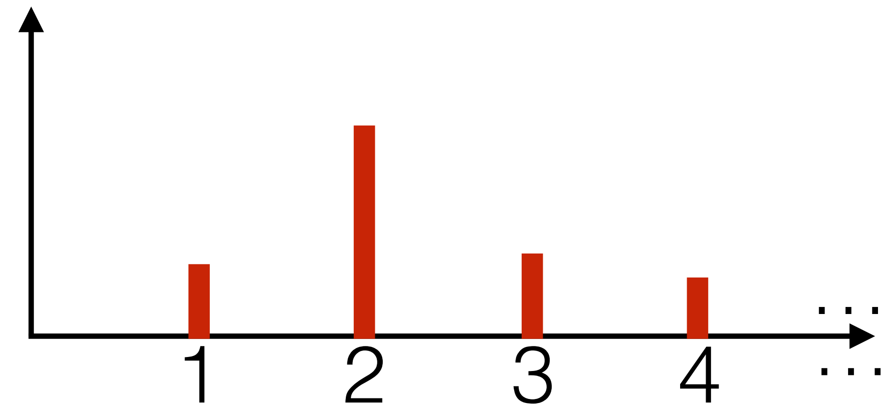


Dirichlet process mixture model

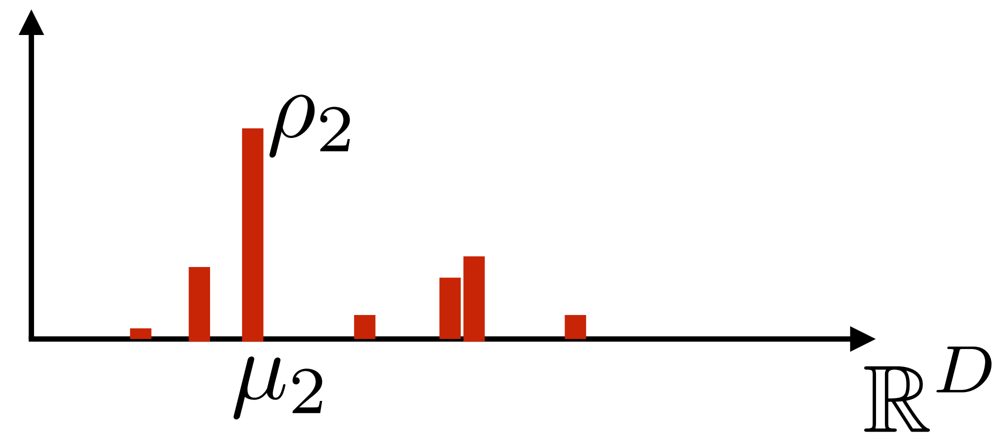
- Gaussian mixture model

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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$



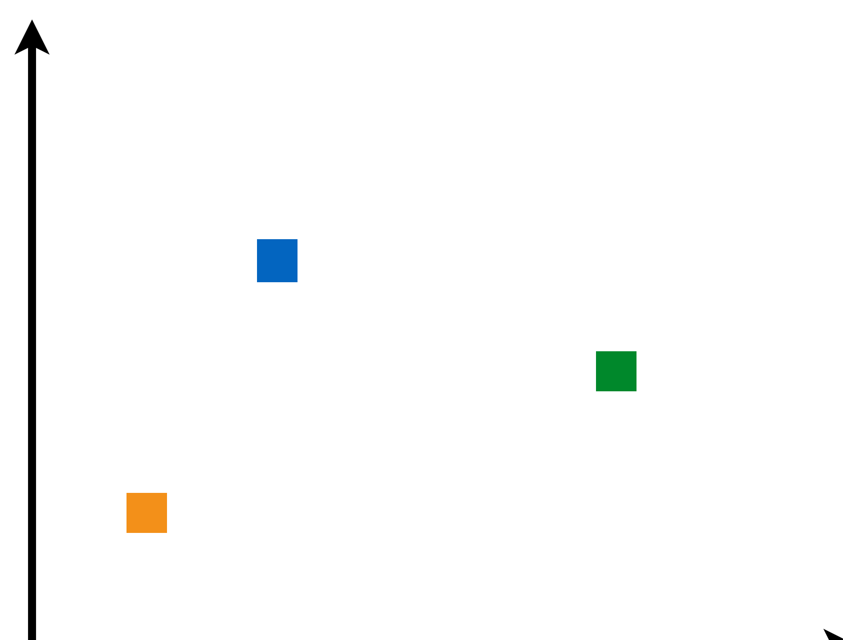
  - i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

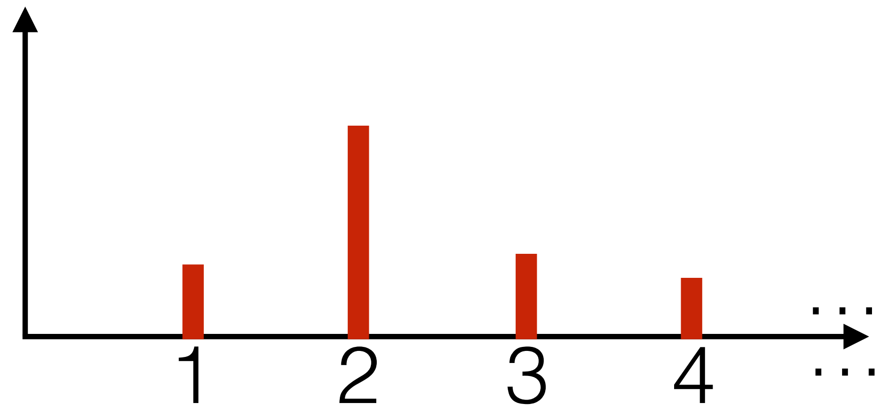
$$\mu_n^* = \mu_{z_n}$$

  - i.e. $\mu_n^* \overset{iid}{\sim} G$



Dirichlet process mixture model

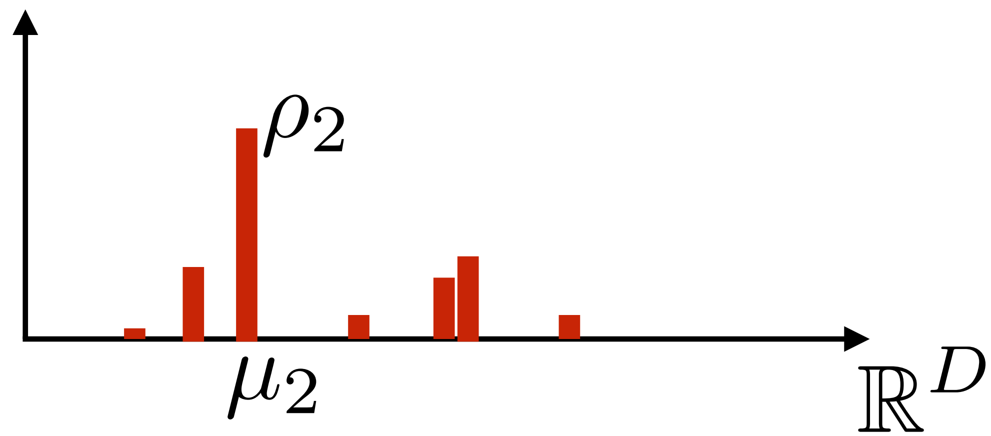
- Gaussian mixture model



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

$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

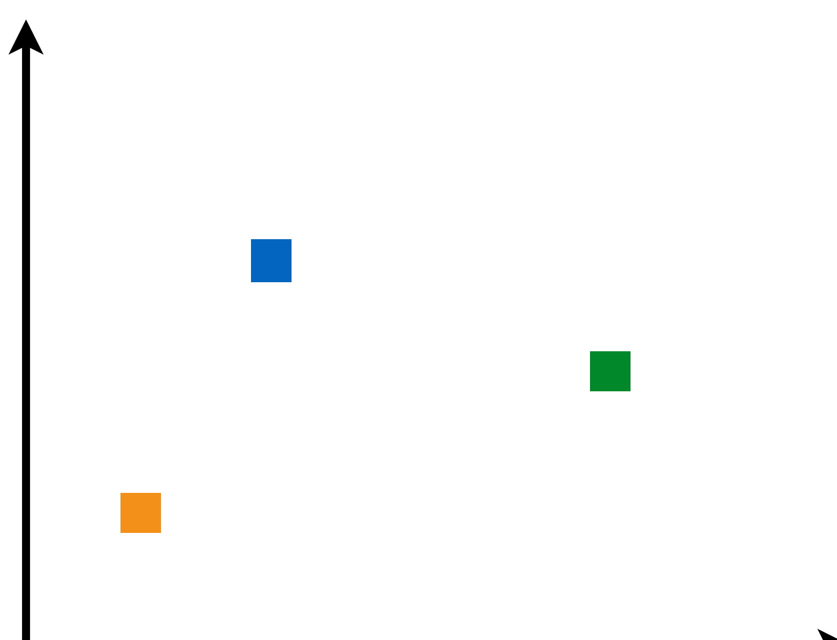
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \stackrel{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \stackrel{iid}{\sim} G$



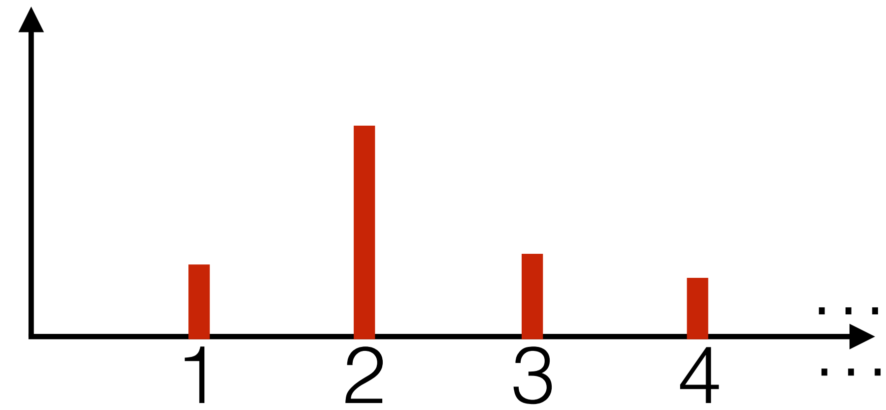
$$x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

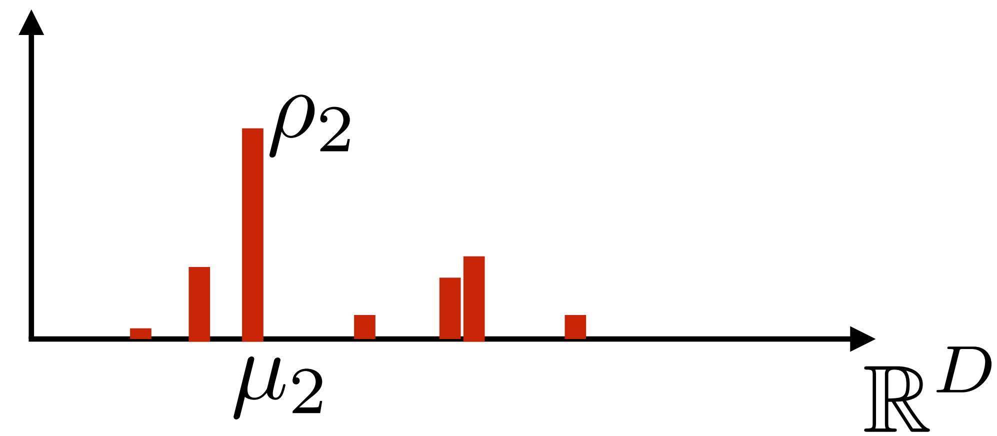
- Gaussian mixture model

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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$



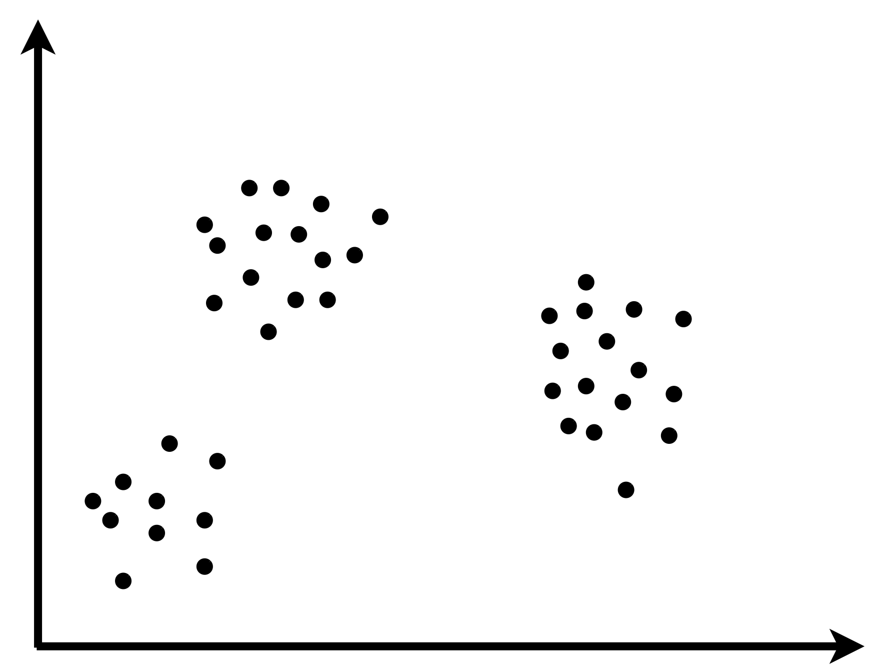
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

$$\mu_n^* = \mu_{z_n}$$

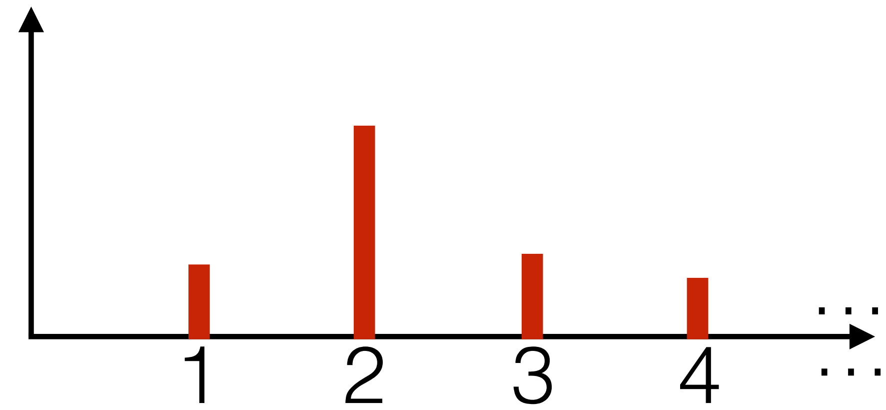
- i.e. $\mu_n^* \overset{iid}{\sim} G$



$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

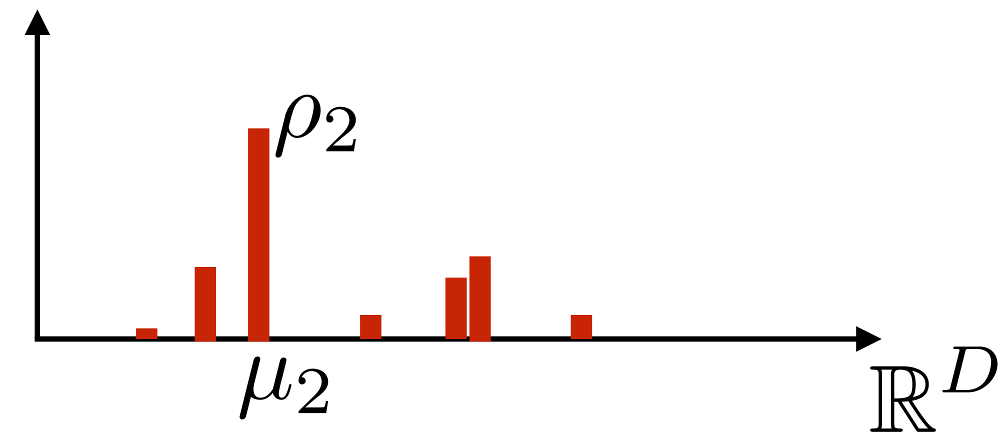
- Gaussian mixture model



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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

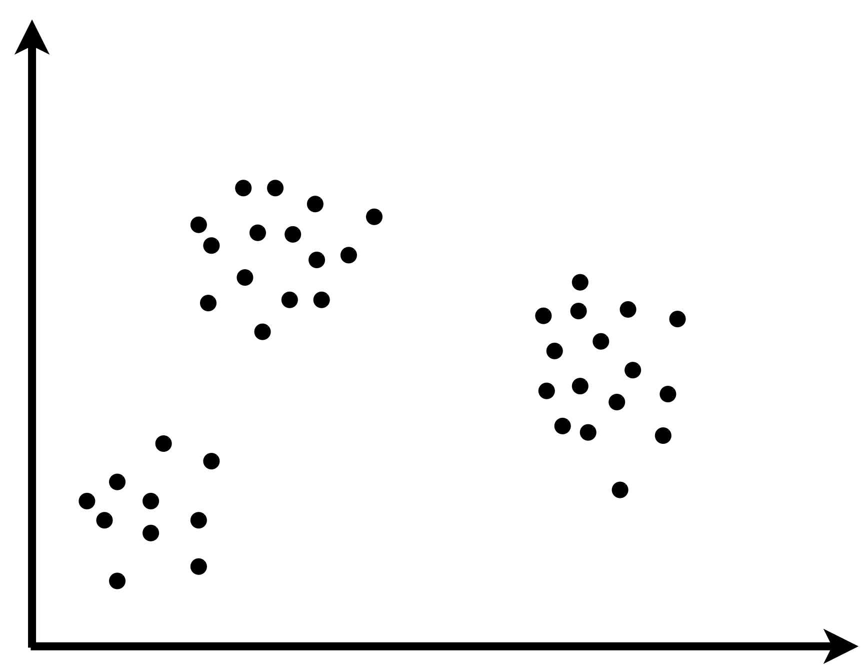
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \overset{iid}{\sim} G$



$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

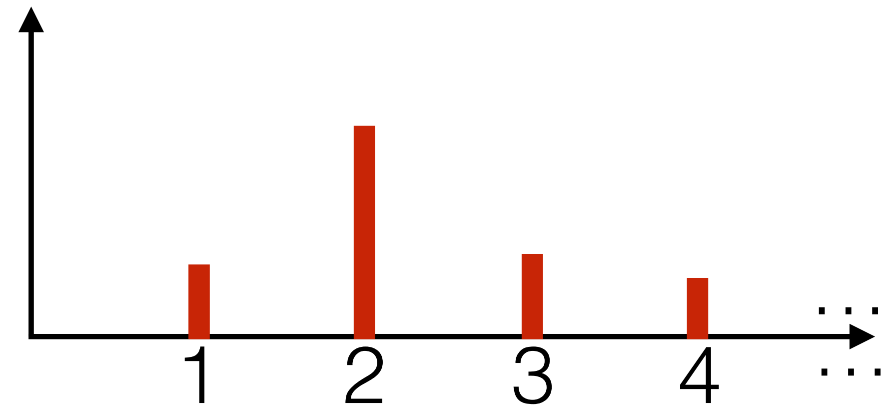
[demo]

Dirichlet process mixture model

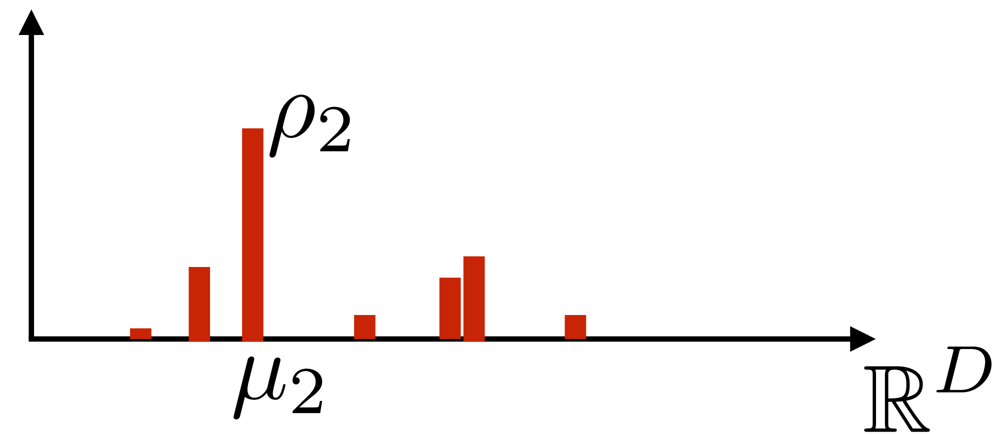
- More generally

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

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$



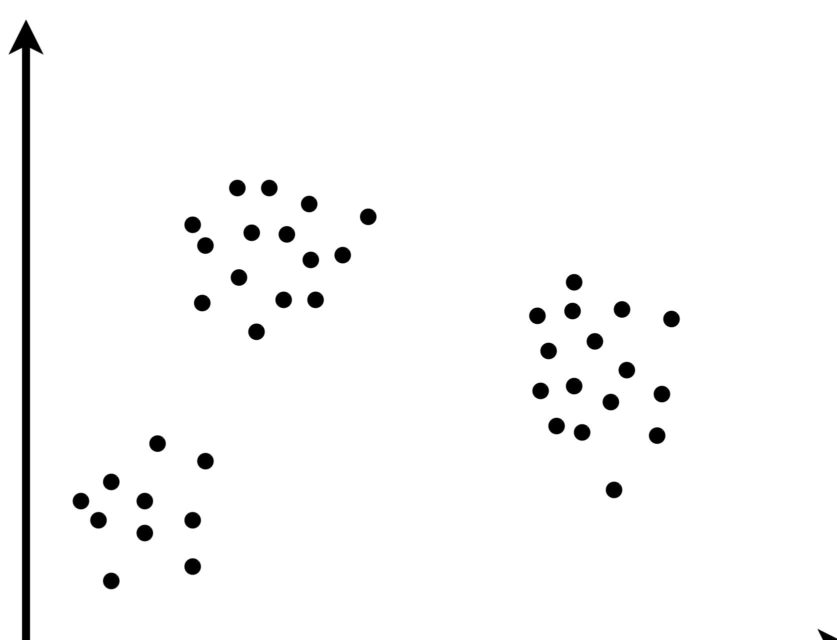
  - i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

$$\mu_n^* = \mu_{z_n}$$

  - i.e. $\mu_n^* \overset{iid}{\sim} G$



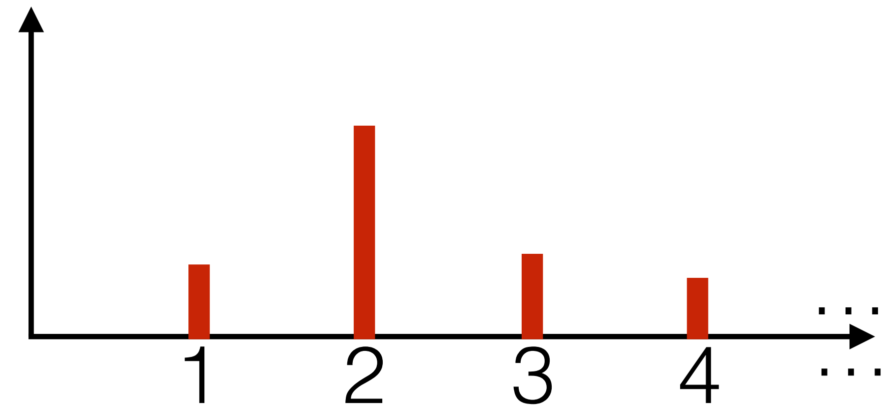
$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

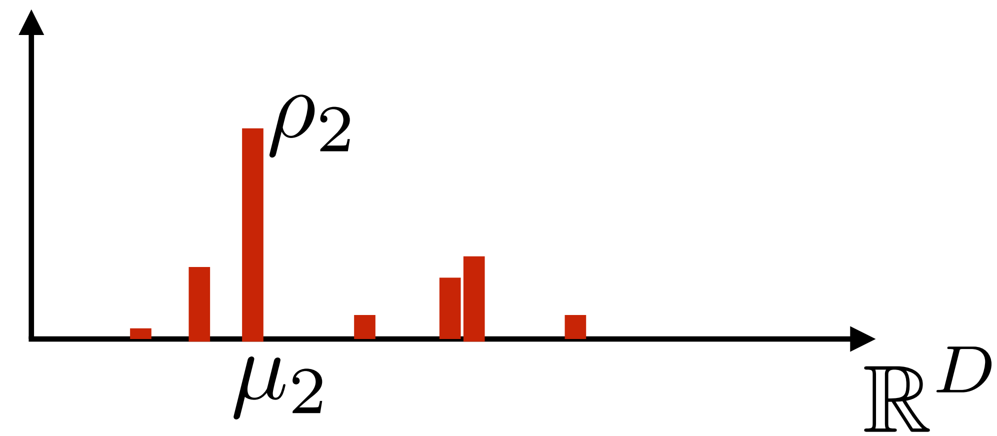
- More generally

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

$$\phi_k \overset{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$



- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$

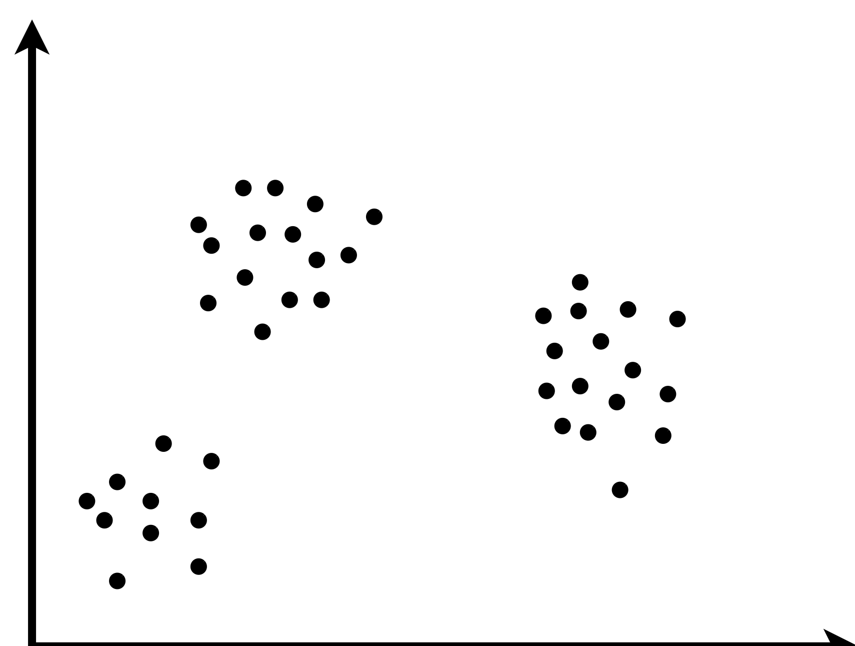


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \overset{iid}{\sim} G$

$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

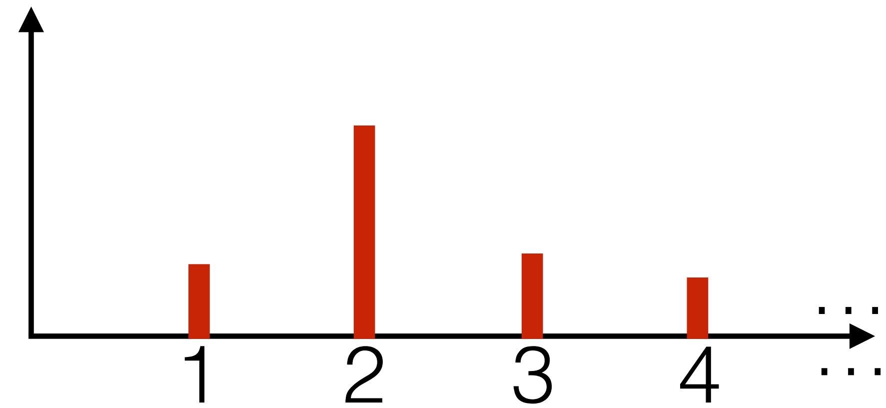


Dirichlet process mixture model

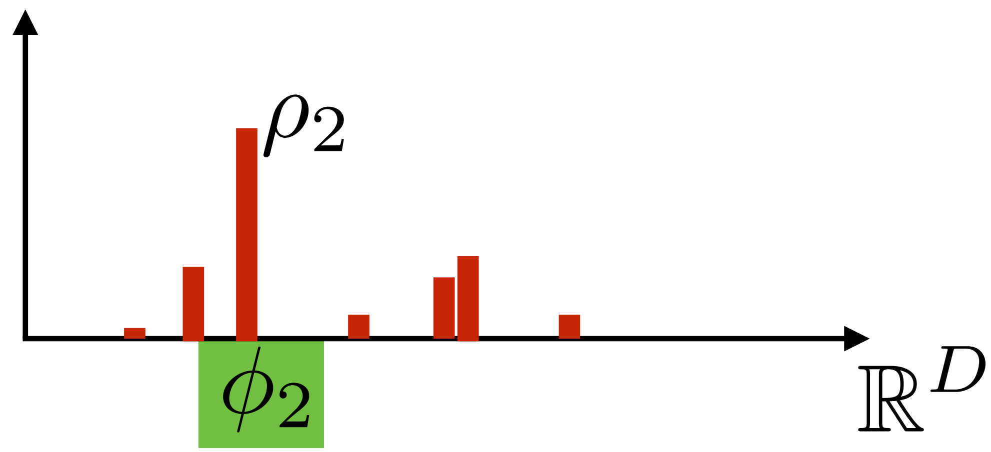
- More generally

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

$$\phi_k \overset{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$



- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$

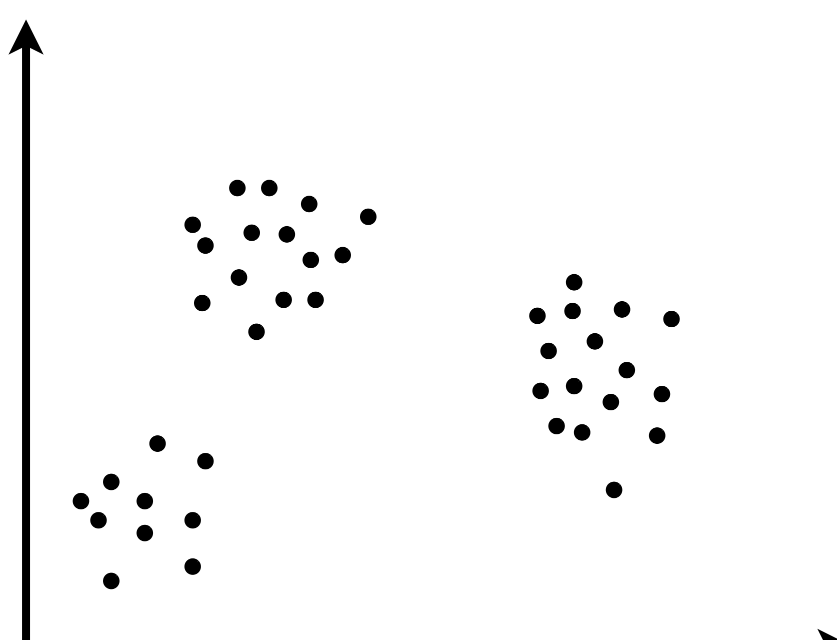


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \overset{iid}{\sim} G$

$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

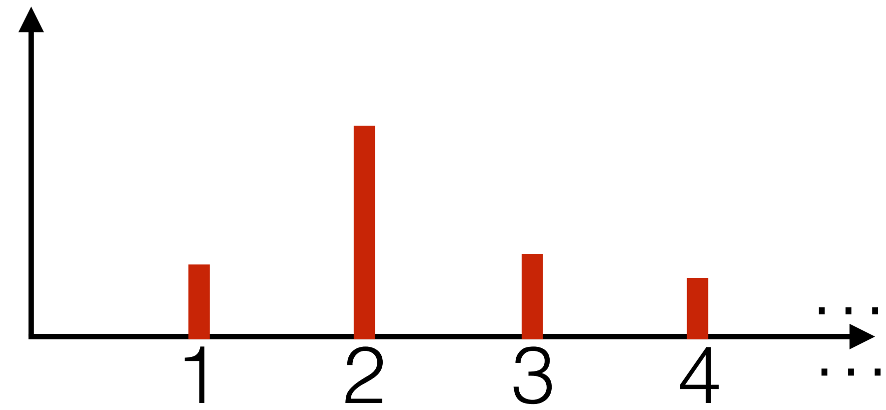


Dirichlet process mixture model

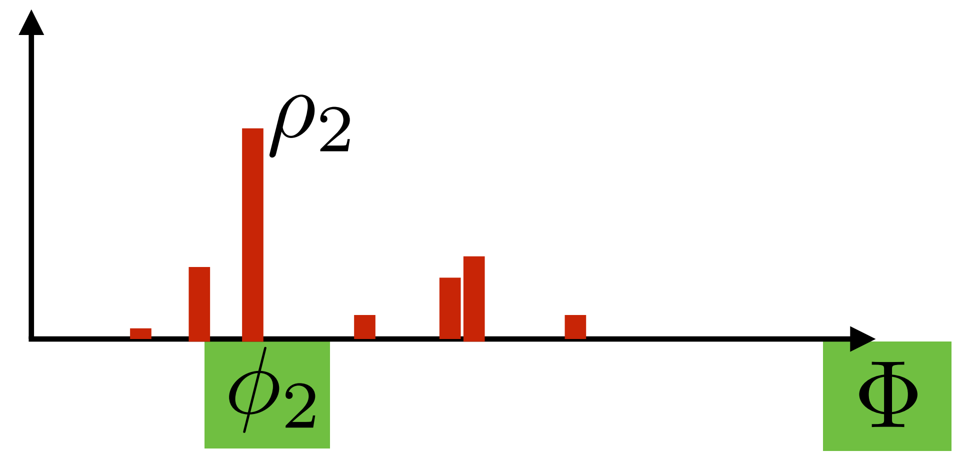
- More generally

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

$$\phi_k \stackrel{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$



- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \stackrel{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$

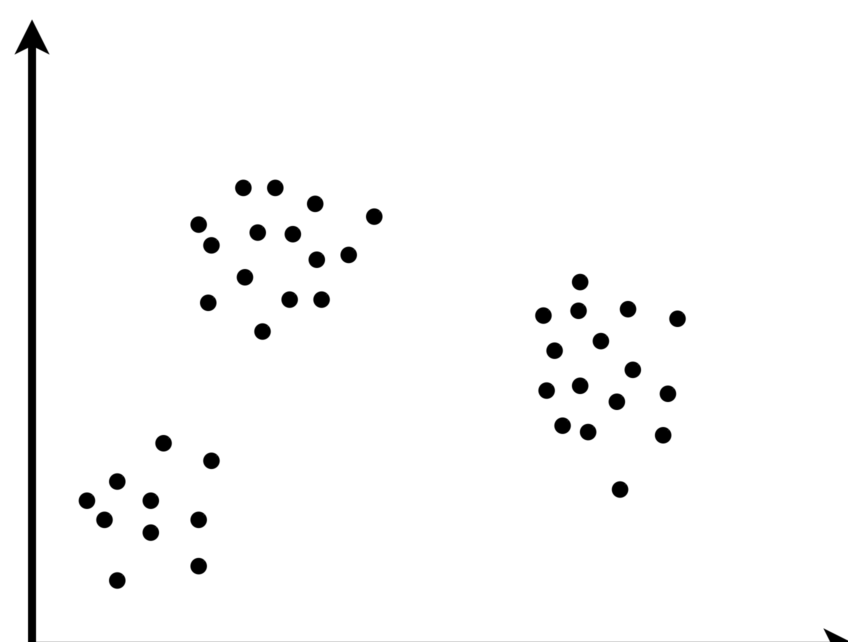


$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \stackrel{iid}{\sim} G$

$$x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

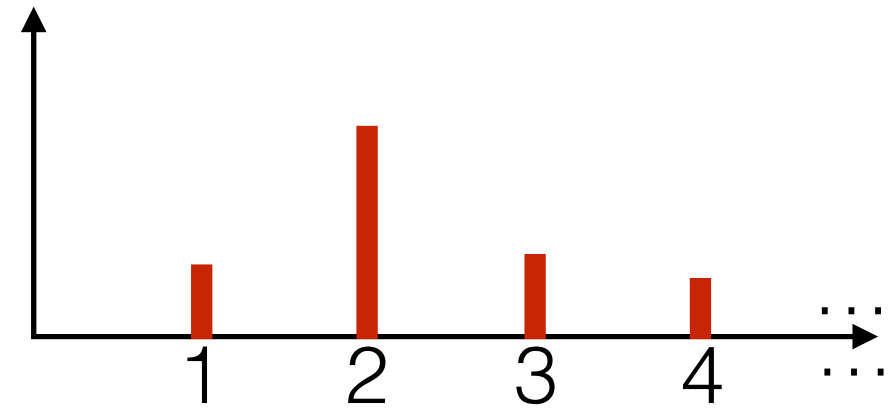


Dirichlet process mixture model

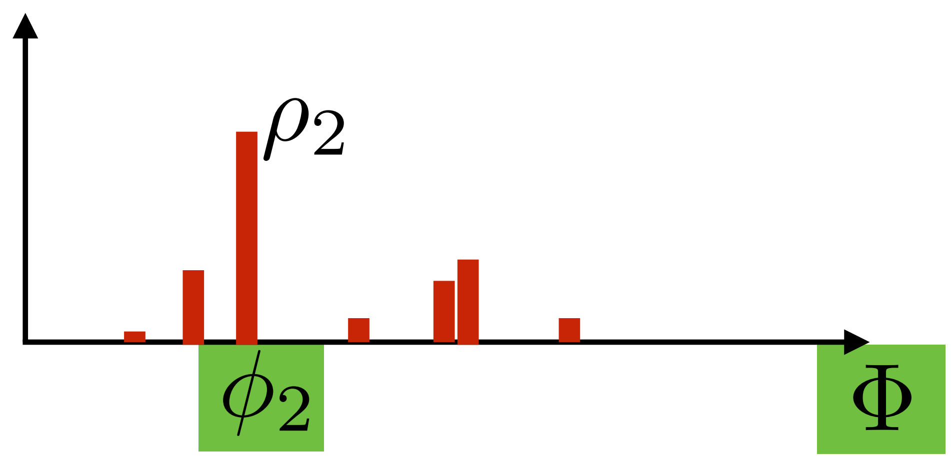
- More generally

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

$$\phi_k \overset{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$



- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$

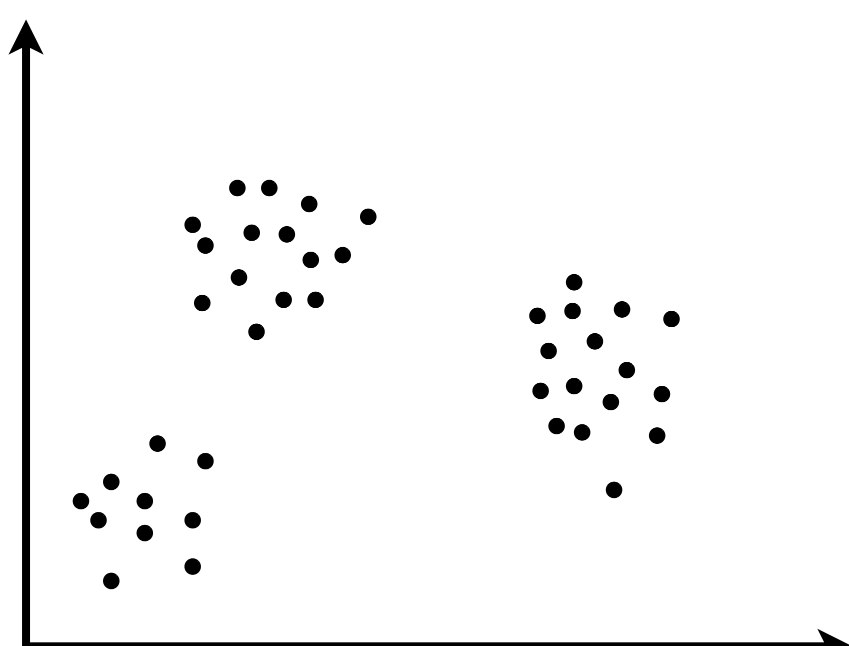


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \overset{iid}{\sim} G$

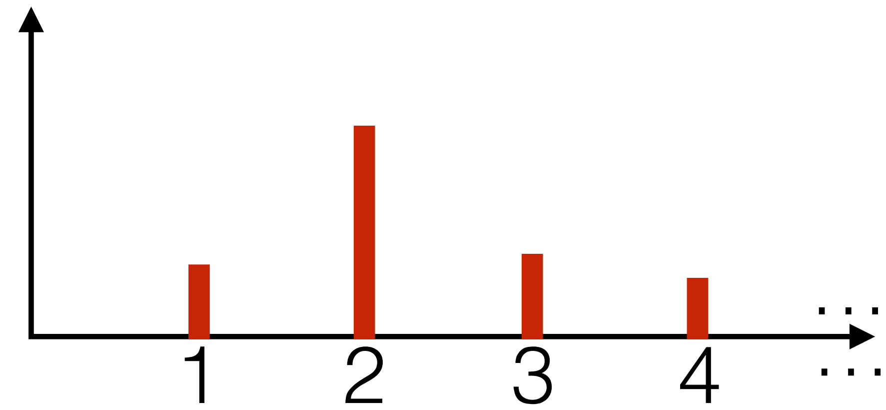
$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$



Dirichlet process mixture model

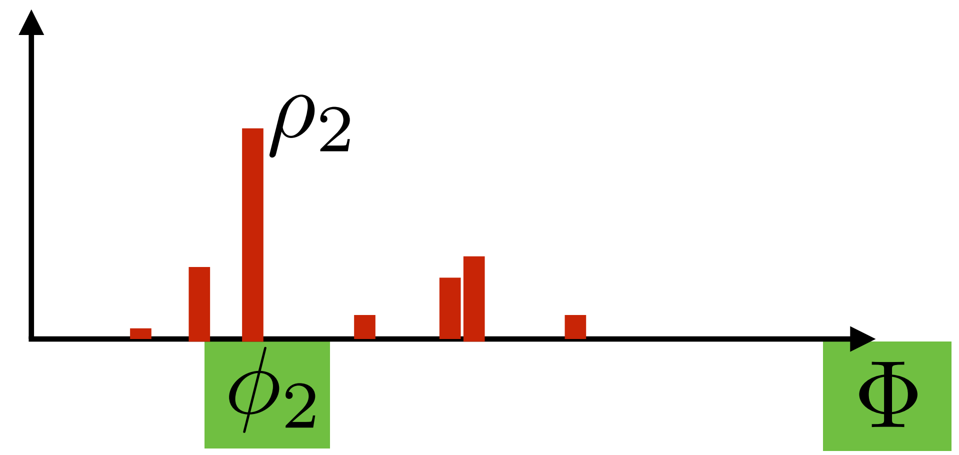
- More generally

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



$$\phi_k \stackrel{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$

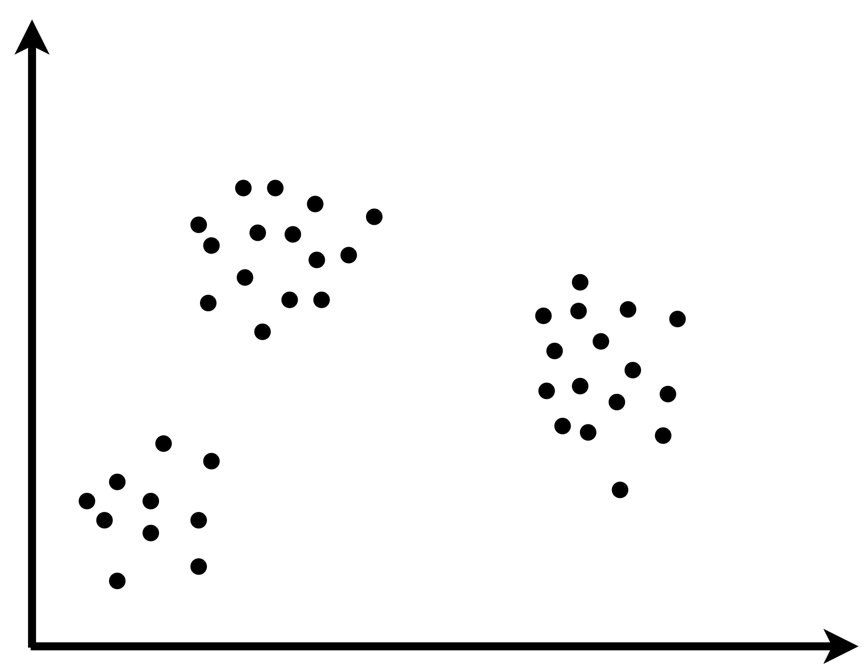
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \stackrel{d}{=} \text{DP}(\alpha, G_0)$



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

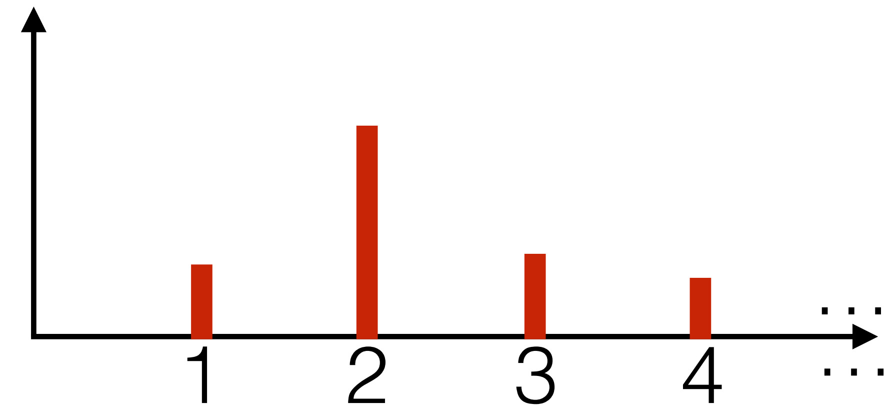
$$\mu_n^* = \mu_{z_n}$$

- i.e. $\mu_n^* \stackrel{iid}{\sim} G$



$$x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

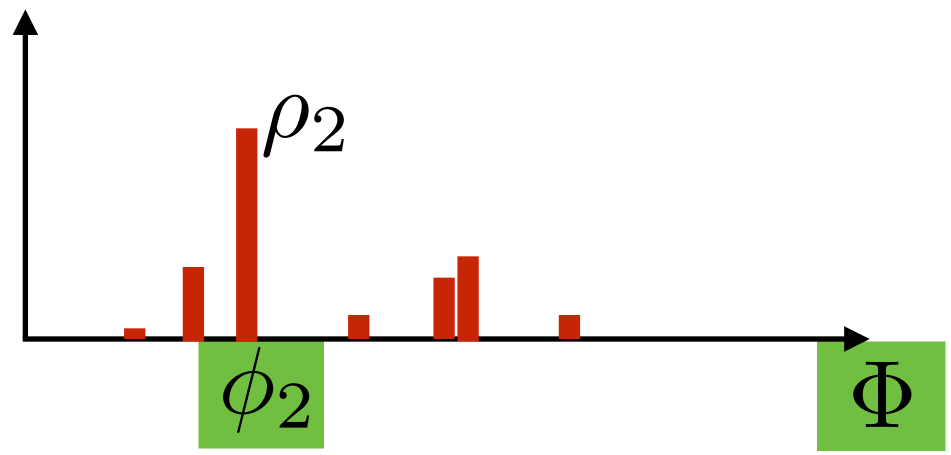


- More generally

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

$$\phi_k \overset{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$

- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, G_0)$

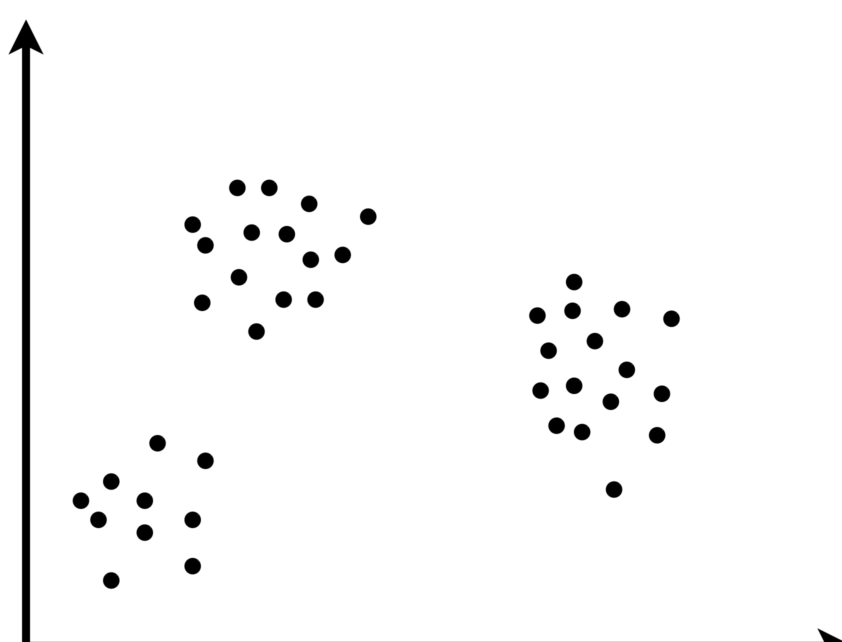


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

$$\theta_n = \phi_{z_n}$$

- i.e. $\mu_n^* \overset{iid}{\sim} G$

$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

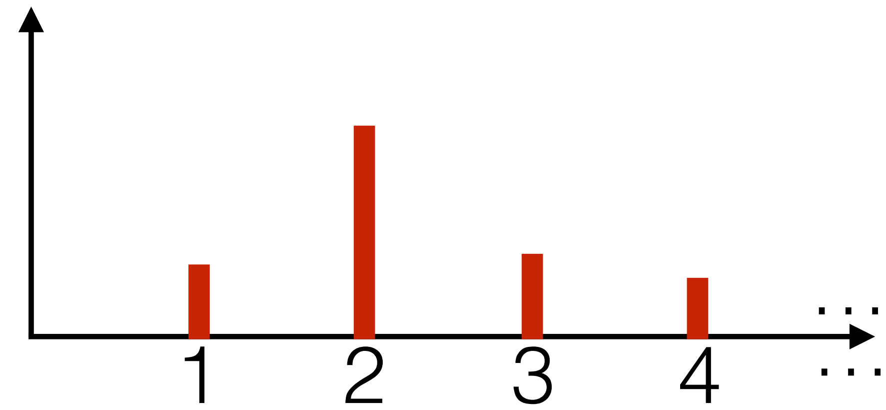


Dirichlet process mixture model

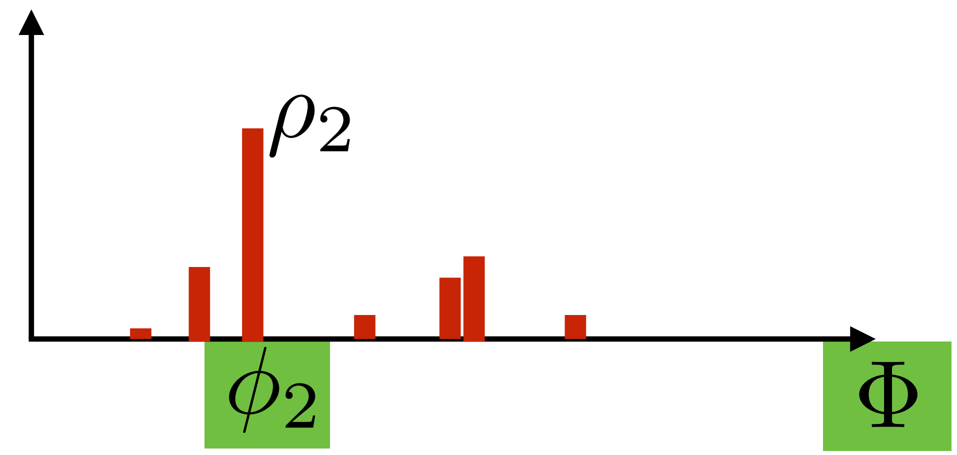
- More generally

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

$$\phi_k \overset{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$



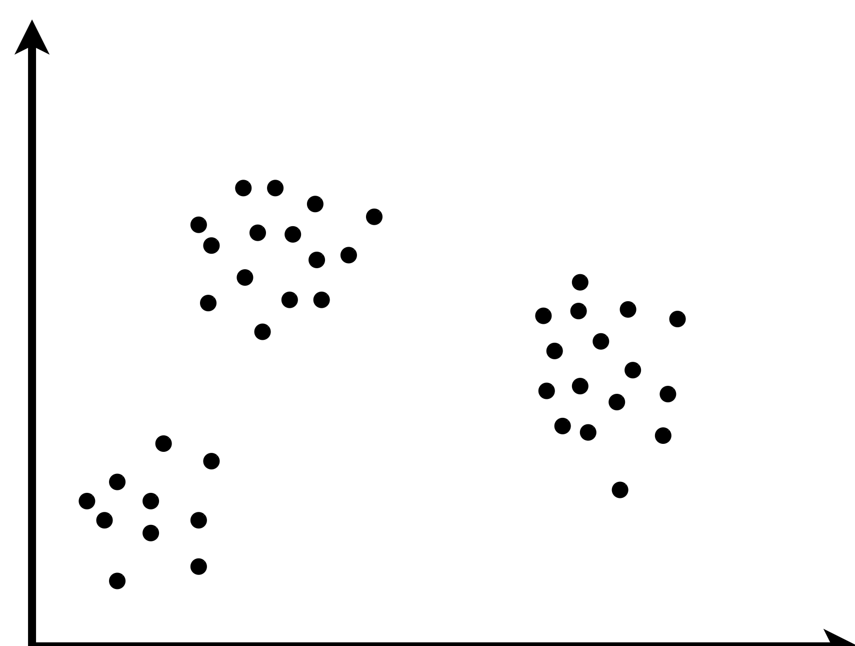
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \mathrm{DP}(\alpha, G_0)$



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

$$\theta_n = \phi_{z_n}$$

- i.e. $\theta_n \overset{iid}{\sim} G$

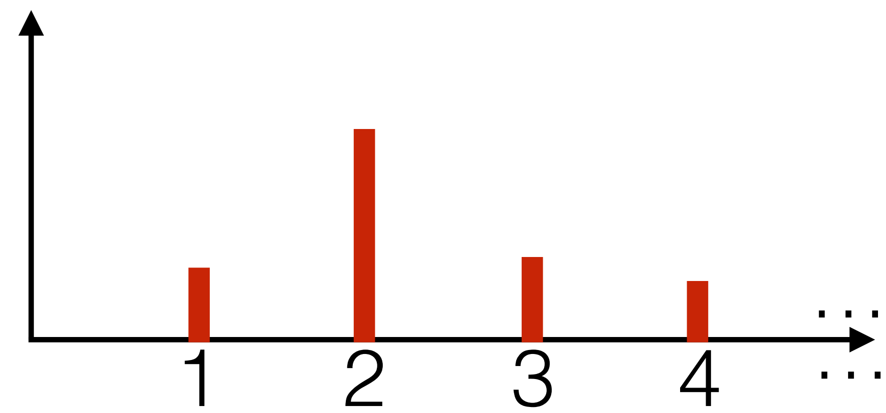


$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

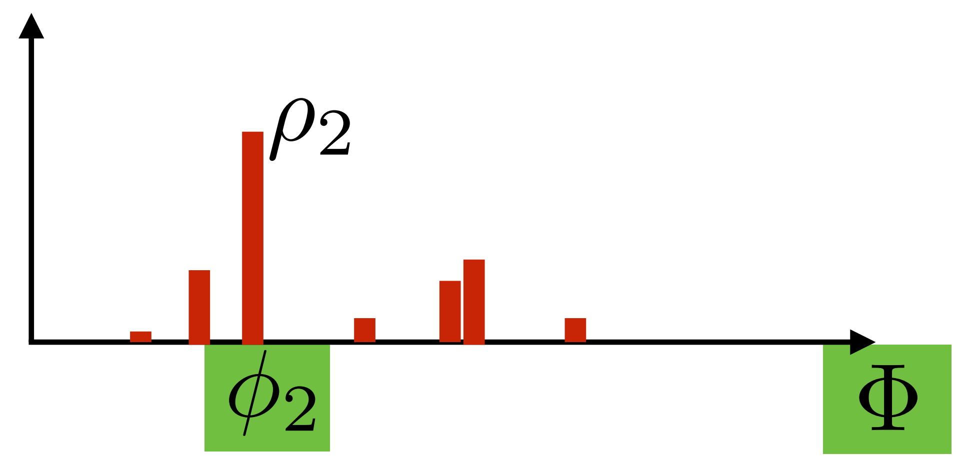
- More generally

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



$$\phi_k \overset{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$

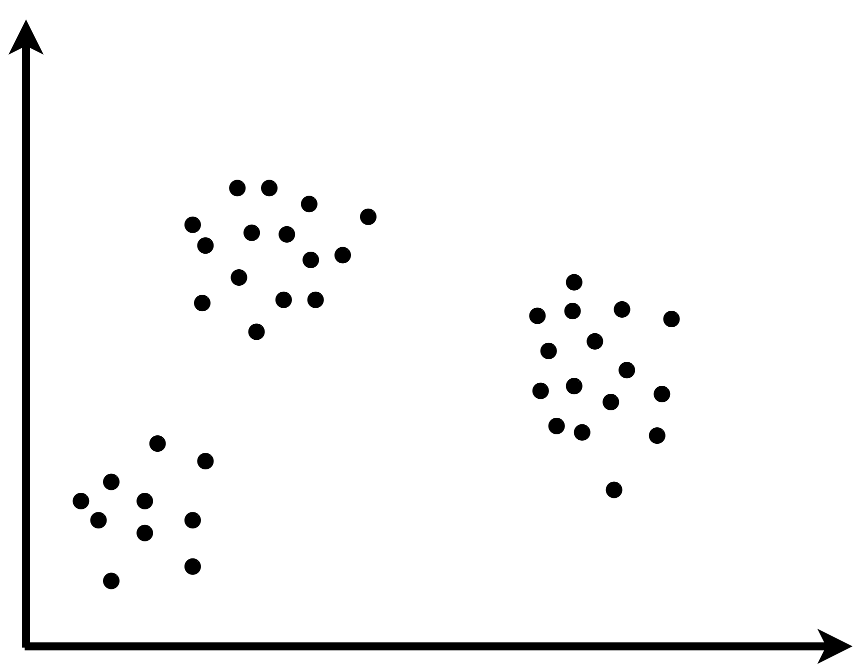
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \mathrm{DP}(\alpha, G_0)$



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

$$\theta_n = \phi_{z_n}$$

- i.e. $\theta_n \overset{iid}{\sim} G$



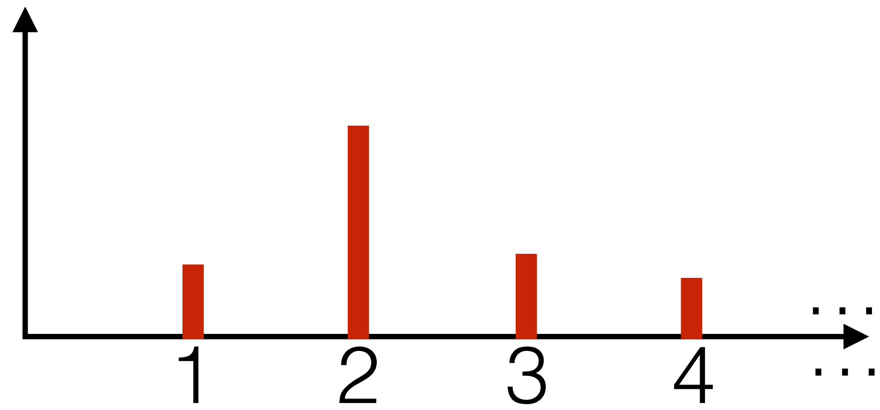
$$x_n \overset{indep}{\sim} F(\theta_n)$$

Dirichlet process mixture model

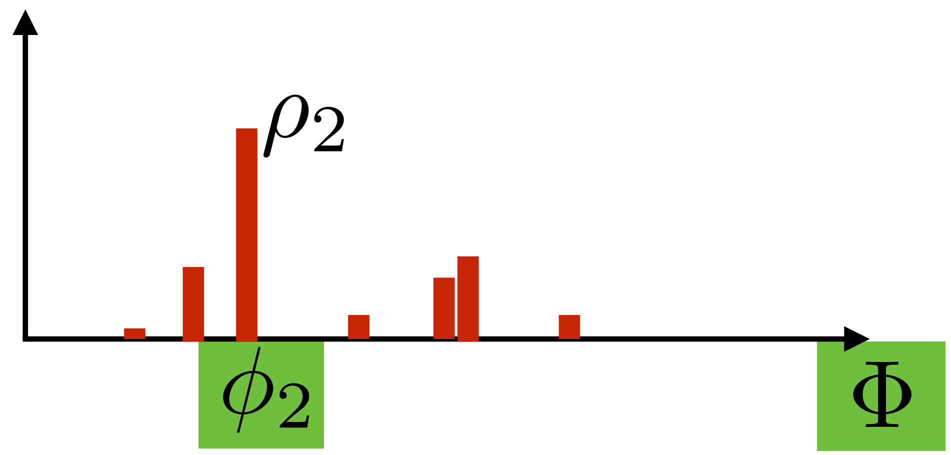
- More generally

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

$$\phi_k \overset{iid}{\sim} G_0 \qquad k = 1, 2, \ldots$$



- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, G_0)$

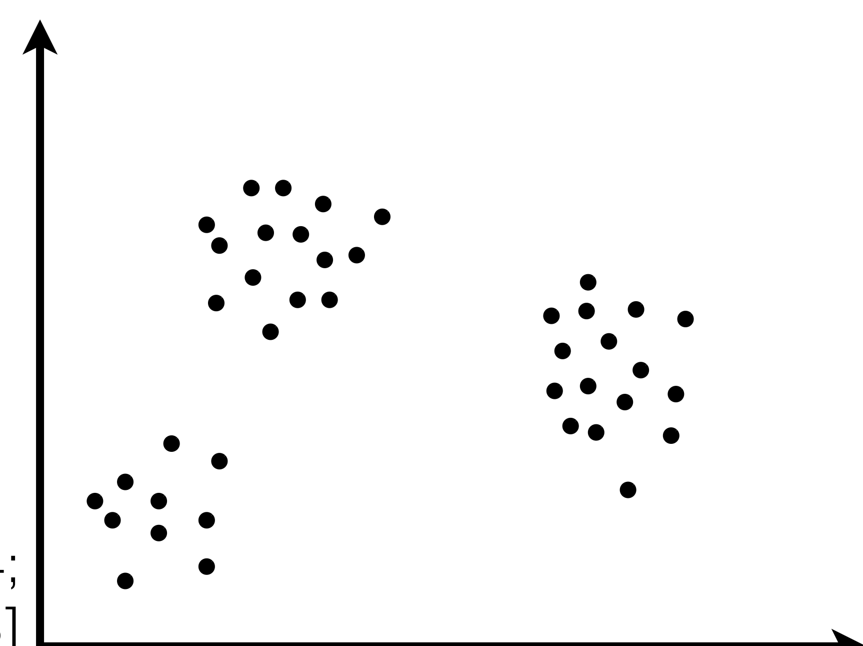


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

$$\theta_n = \phi_{z_n}$$

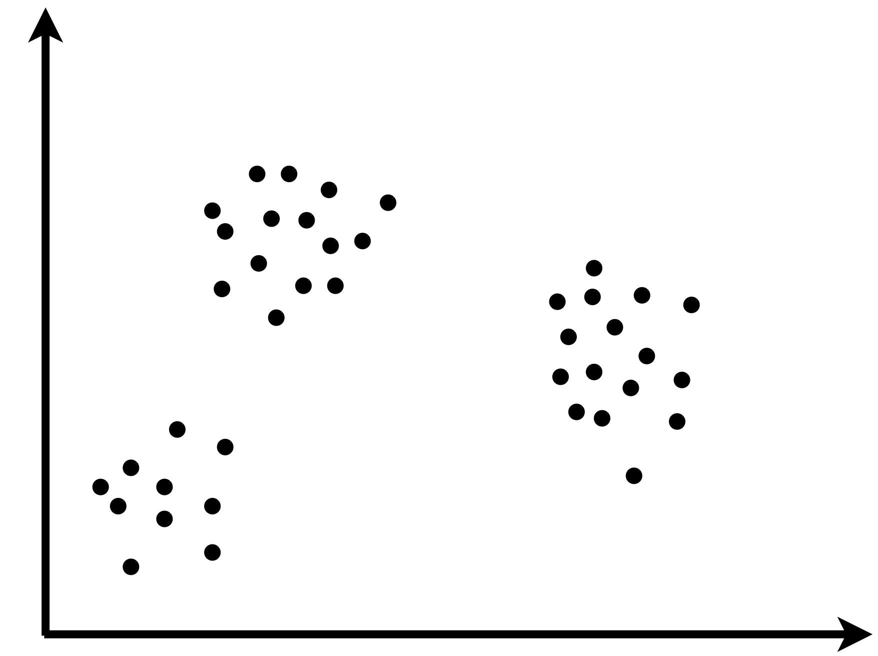
- i.e. $\theta_n \overset{iid}{\sim} G$

$$x_n \overset{indep}{\sim} F(\theta_n)$$



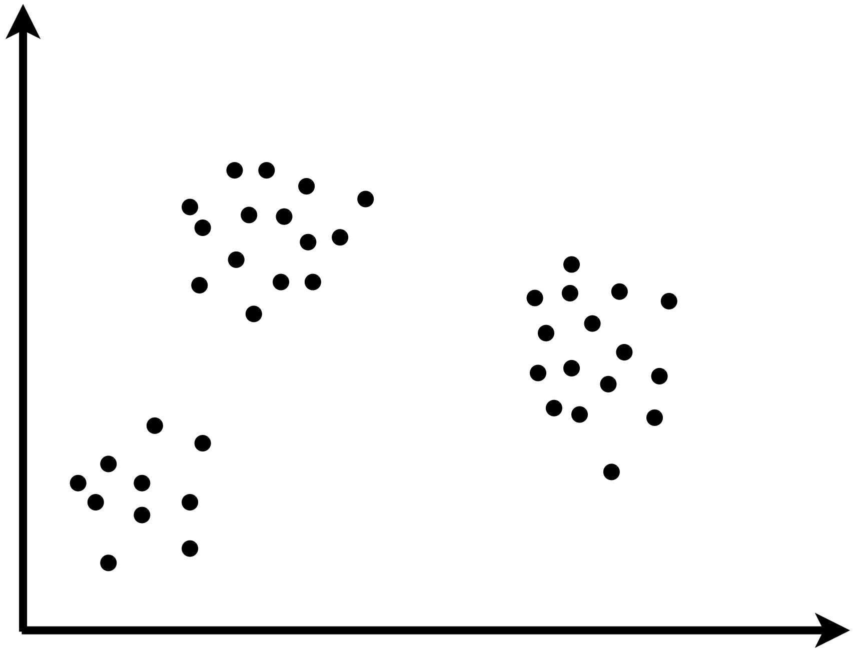
[Antoniak 1974; Ferguson 1983; West, Müller, Escobar 1994; Escobar, West 1995; MacEachern, Müller 1998]

DP or not DP, that is the question



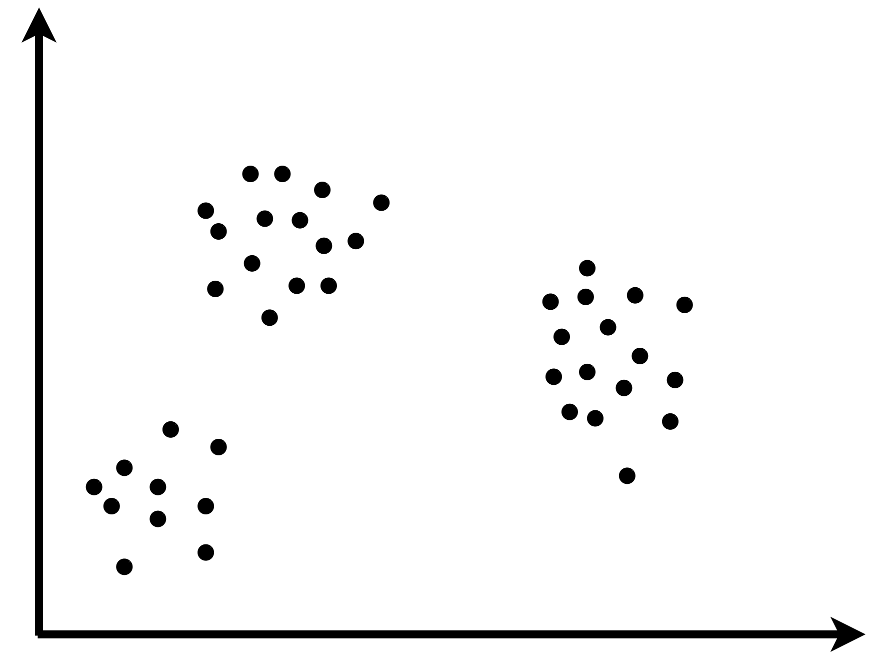
DP or not DP, that is the question

- GEM: 



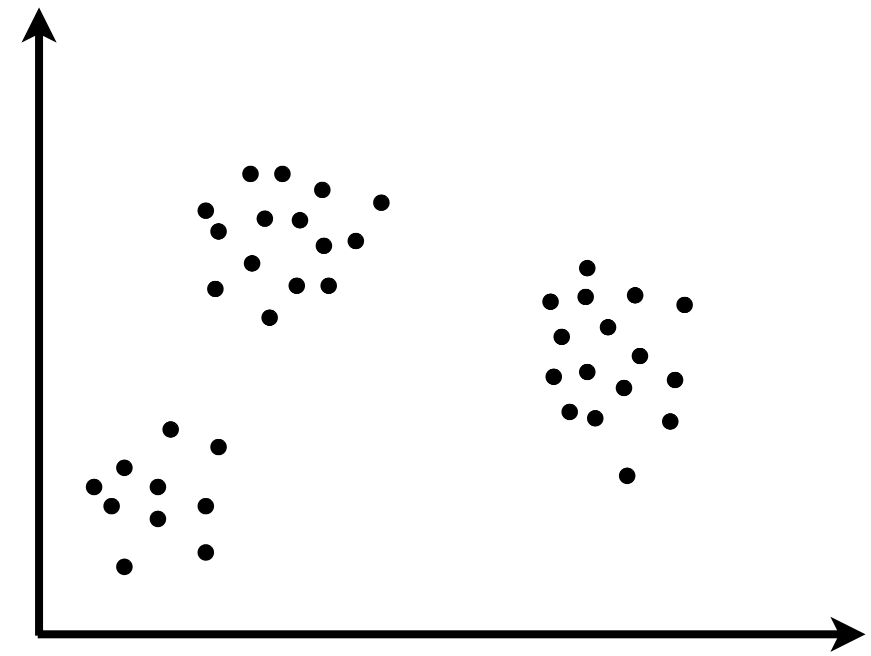
DP or not DP, that is the question

- GEM: 
- Compare to:



DP or not DP, that is the question

- GEM: 
- Compare to:
  - Finite (small *K*) mixture model

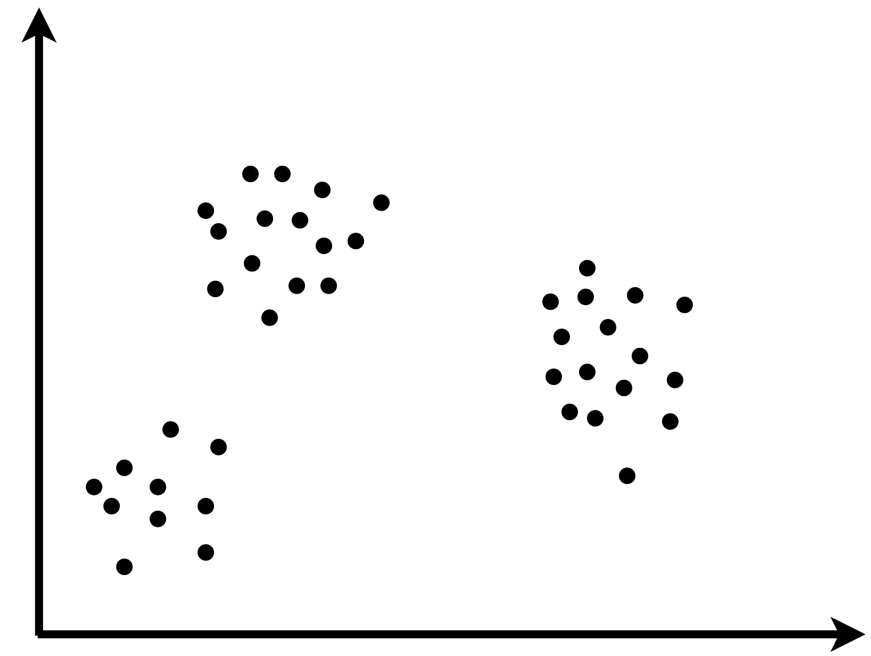


DP or not DP, that is the question

- GEM: 

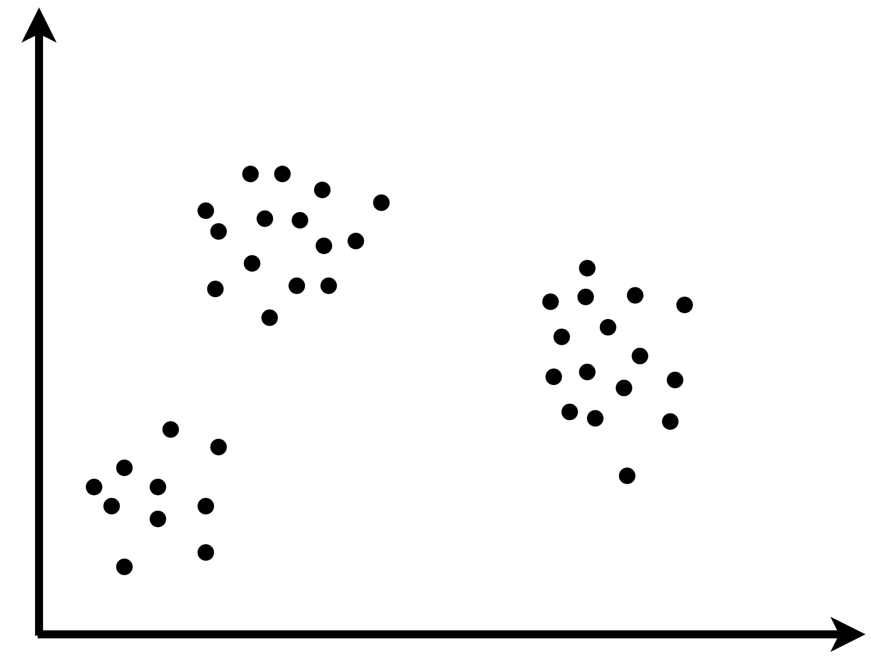
- Compare to:
  - Finite (small *K*) mixture model

  - Finite (large *K*) mixture model



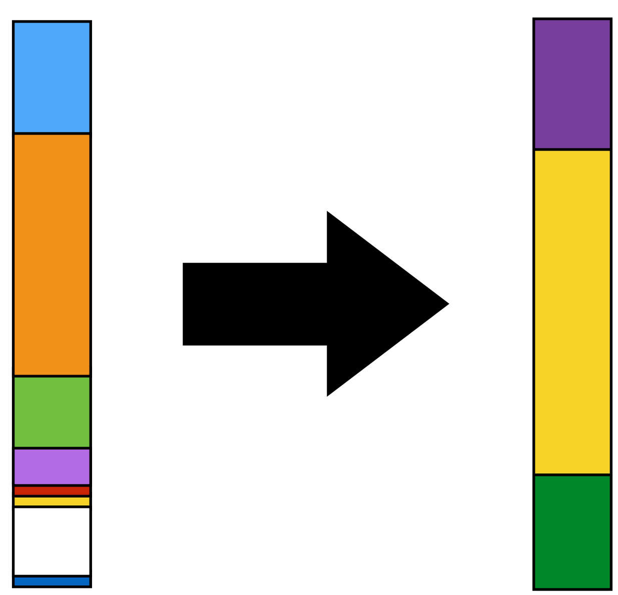
DP or not DP, that is the question

- GEM: 
- Compare to:
  - Finite (small *K*) mixture model



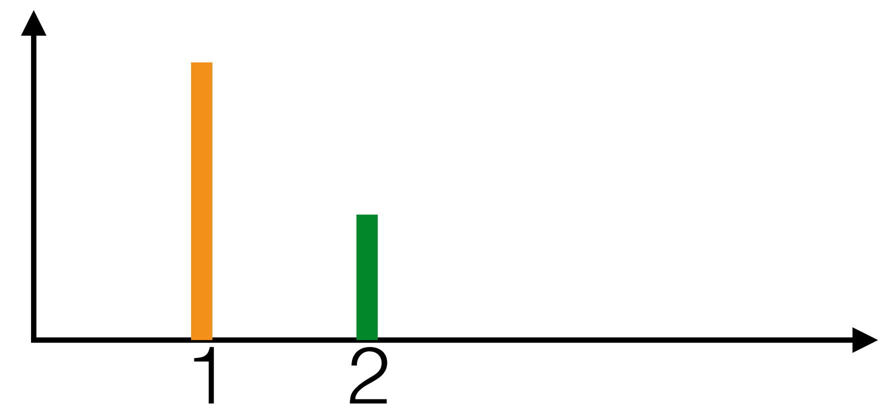
  - Finite (large *K*) mixture model

  - Time series



Marginal cluster assignments

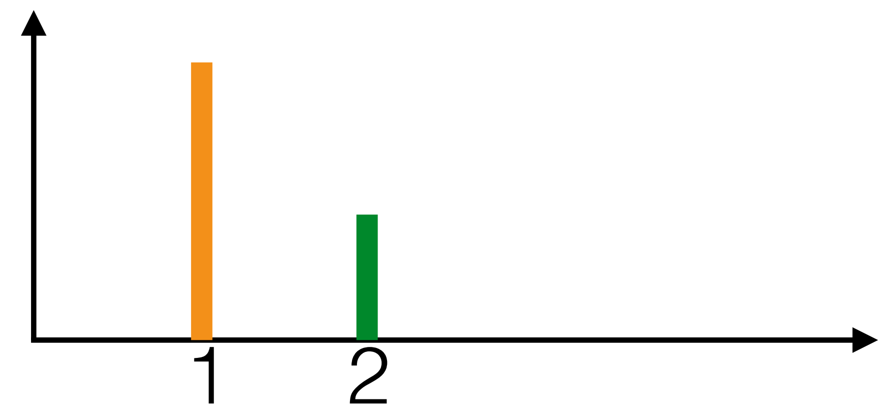
$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$



Marginal cluster assignments

- Integrate out the frequencies

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

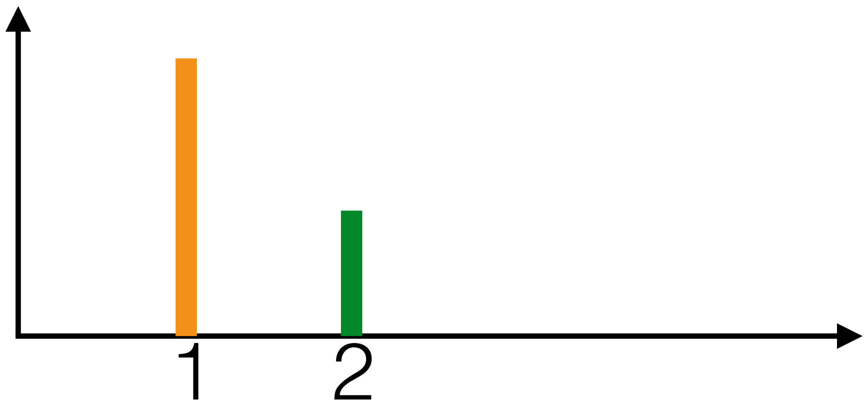


Marginal cluster assignments

- Integrate out the frequencies

$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

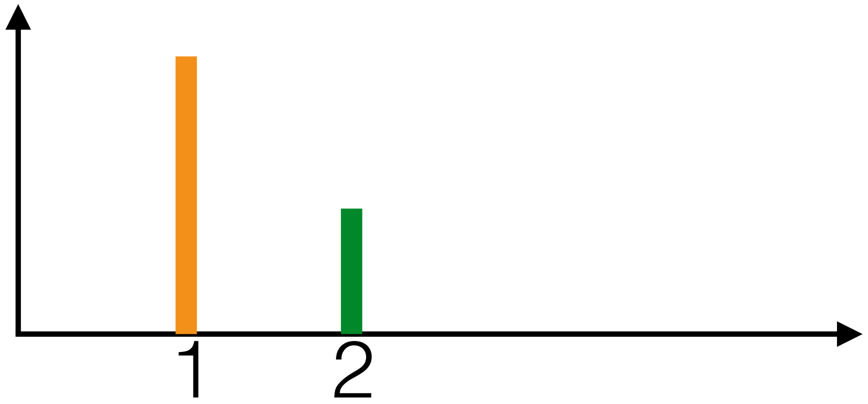


Marginal cluster assignments

- Integrate out the frequencies

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

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$
$$= \int p(z_n = 1, \rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$



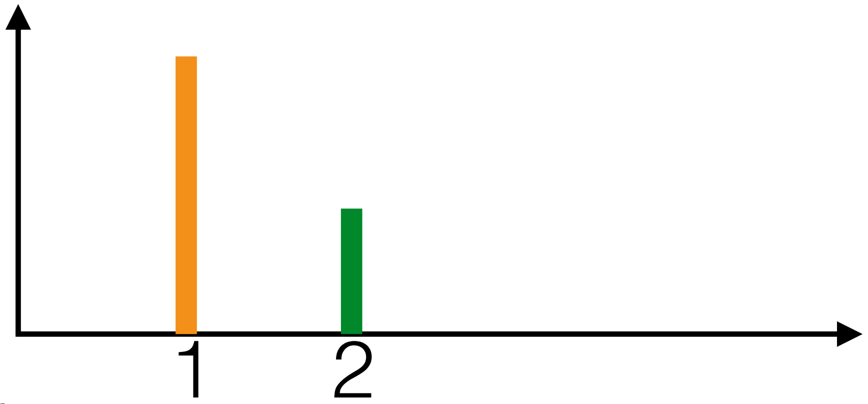
Marginal cluster assignments

- Integrate out the frequencies

$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

$$= \int p(z_n = 1 | \rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

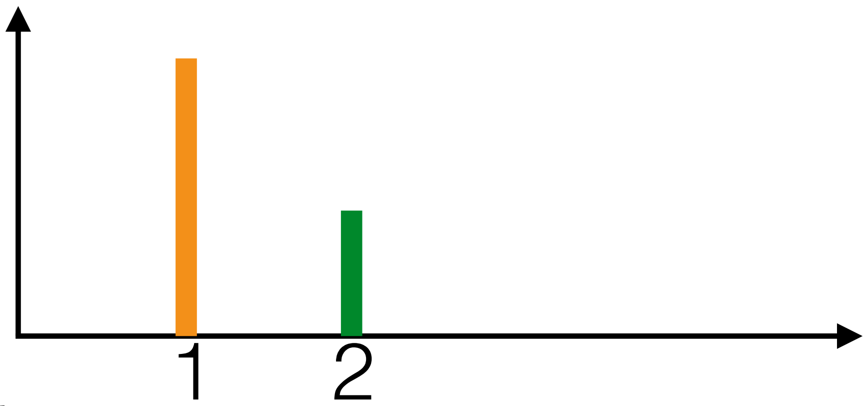


Marginal cluster assignments

- Integrate out the frequencies

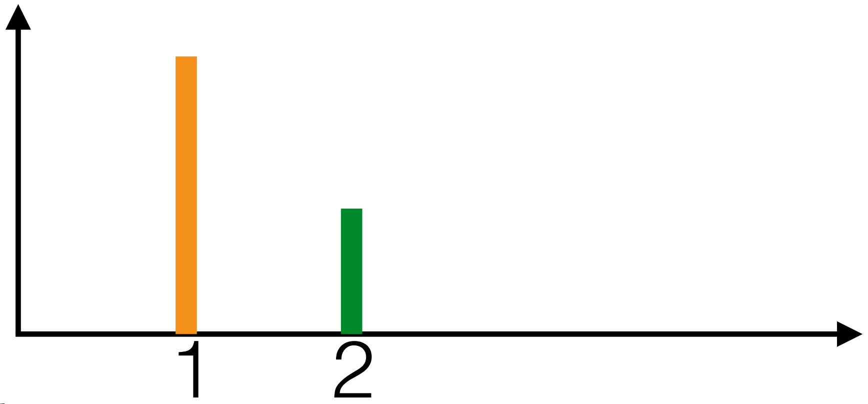
$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$
$$= \int p(z_n = 1 | \rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$



Marginal cluster assignments

- Integrate out the frequencies



$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

$$= \int p(z_n = 1 | \rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

$$= \int \rho_1$$

Marginal cluster assignments

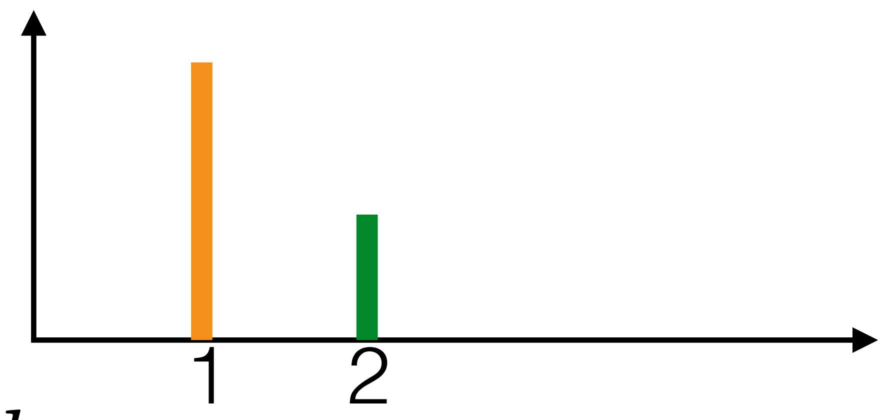
- Integrate out the frequencies

$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

$$= \int p(z_n = 1 | \rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

$$= \int \rho_1$$



Marginal cluster assignments

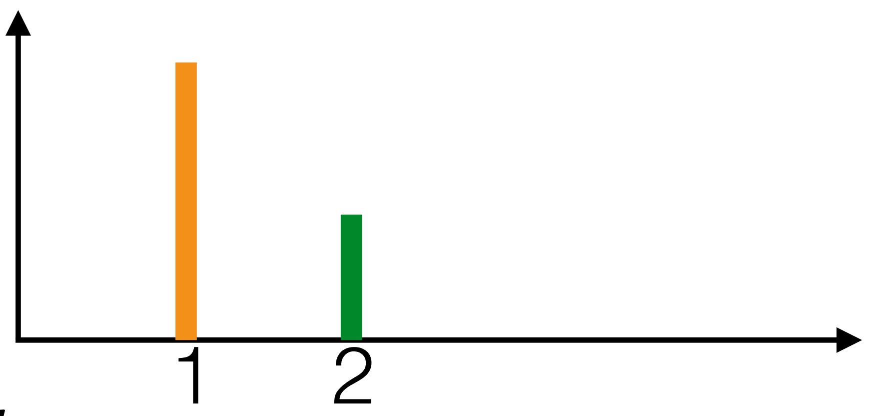
- Integrate out the frequencies

$$\rho_1 \sim \text{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

$$= \int p(z_n = 1|\rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

$$= \int \rho_1 \text{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$

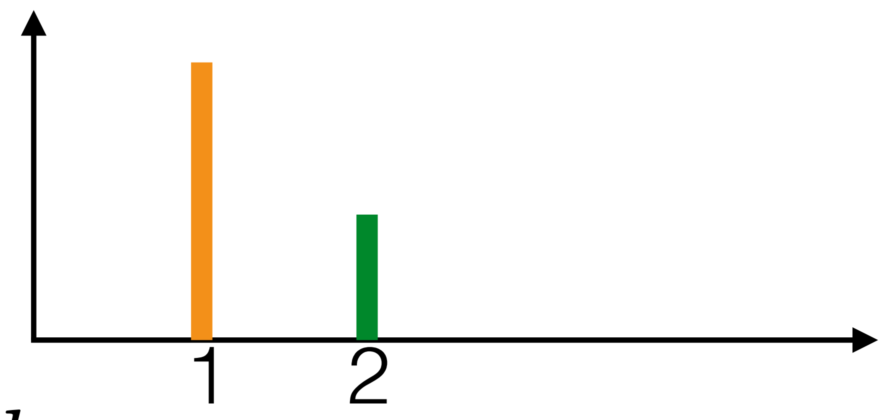


Marginal cluster assignments

- Integrate out the frequencies

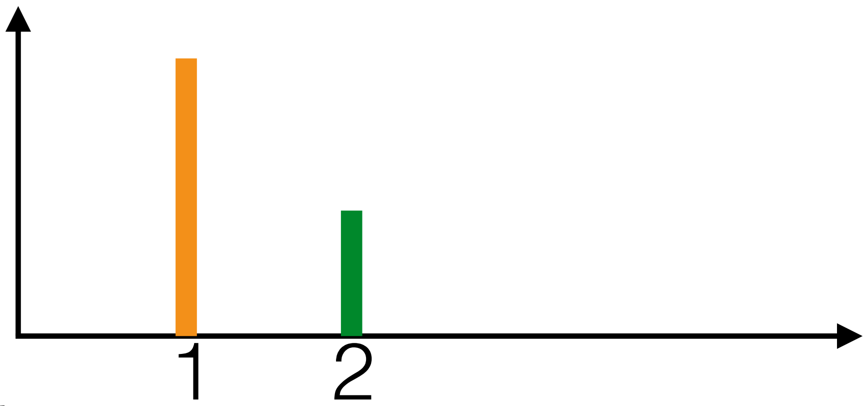
$$\rho_1 \sim \text{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$
$$= \int p(z_n = 1 | \rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$
$$= \int \rho_1 \text{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$



Marginal cluster assignments

- Integrate out the frequencies

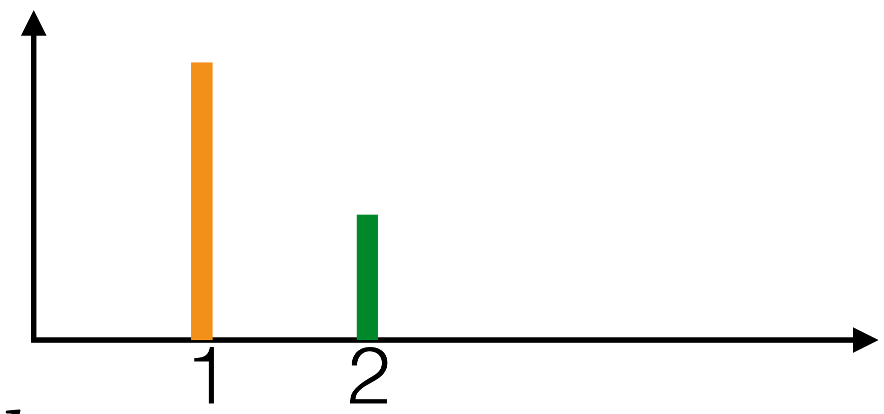


$$\rho_1 \sim \text{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$
$$= \int p(z_n = 1|\rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$
$$= \int \rho_1 \text{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

- Integrate out the frequencies



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

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

$$= \int p(z_n = 1|\rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

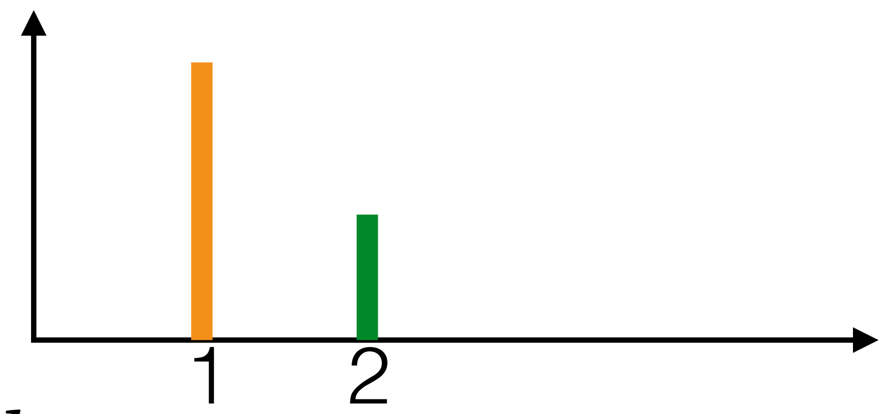
$$= \int \rho_1 \text{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

$$= \int \rho_1 \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \rho_1^{a_{1,n}-1} (1 - \rho_1)^{a_{2,n}-1} d\rho_1$$

Marginal cluster assignments

- Integrate out the frequencies



$$\rho_1 \sim \mathrm{Beta}(a_1, a_2),\, z_n \stackrel{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

$$= \int p(z_n = 1|\rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

$$= \int \rho_1 \mathrm{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$

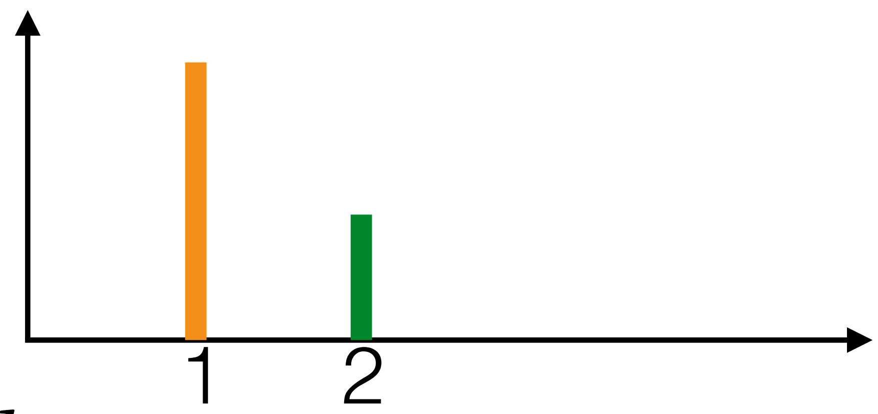
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

$$= \int \rho_1 \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \rho_1^{a_{1,n}-1} (1 - \rho_1)^{a_{2,n}-1} d\rho_1$$

$$= \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \frac{\Gamma(a_{1,n} + 1)\Gamma(a_{2,n})}{\Gamma(a_{1,n} + a_{2,n} + 1)}$$

Marginal cluster assignments

- Integrate out the frequencies



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

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$

$$= \int p(z_n = 1 | \rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

$$= \int \rho_1 \text{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

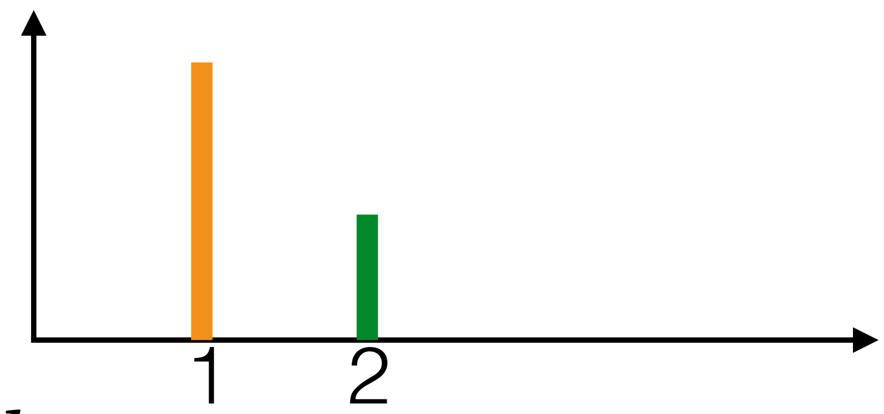
$$= \int \rho_1 \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \rho_1^{a_{1,n}-1} (1 - \rho_1)^{a_{2,n}-1} d\rho_1$$

$$= \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \frac{\Gamma(a_{1,n} + 1)\Gamma(a_{2,n})}{\Gamma(a_{1,n} + a_{2,n} + 1)}$$

Recall
$\Gamma(x+1) = x\Gamma(x)$

Marginal cluster assignments

- Integrate out the frequencies



$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

$$p(z_n = 1 | z_1, \ldots, z_{n-1})$$
$$= \int p(z_n = 1|\rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$
$$= \int \rho_1 \mathrm{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$
$$= \int \rho_1 \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \rho_1^{a_{1,n}-1} (1-\rho_1)^{a_{2,n}-1} d\rho_1$$
$$= \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \frac{\Gamma(a_{1,n}+1)\Gamma(a_{2,n})}{\Gamma(a_{1,n} + a_{2,n} + 1)}$$
$$= \frac{a_{1,n}}{a_{1,n} + a_{2,n}}$$

Recall
$\Gamma(x+1) = x\Gamma(x)$

Marginal cluster assignments

- Integrate out the frequencies

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

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

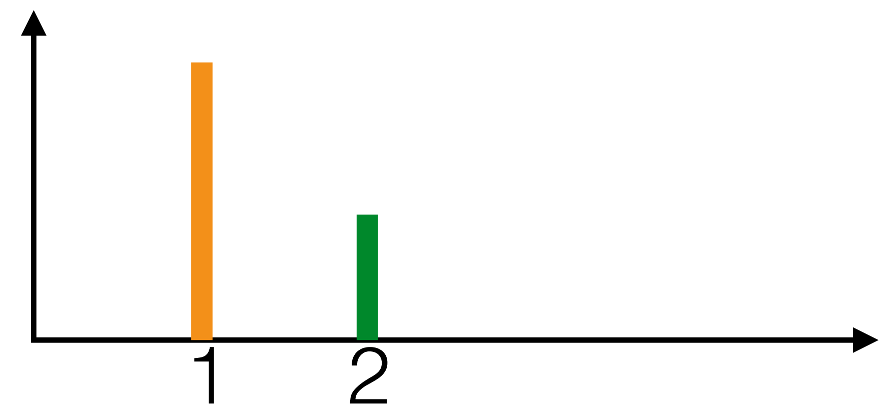
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

- Integrate out the frequencies

$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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



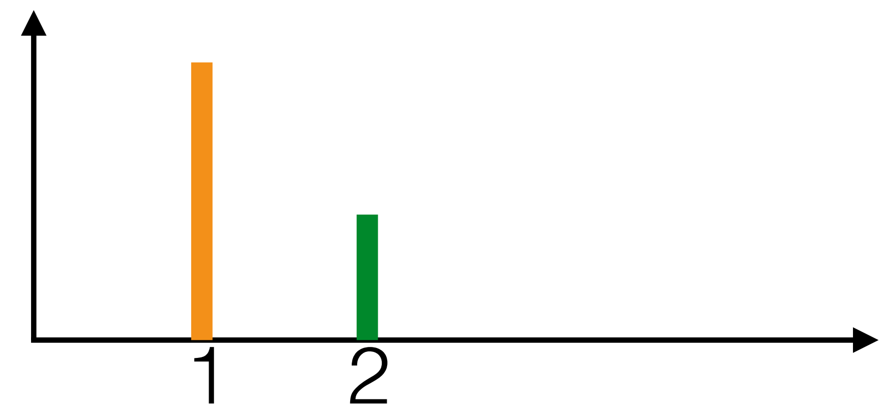
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

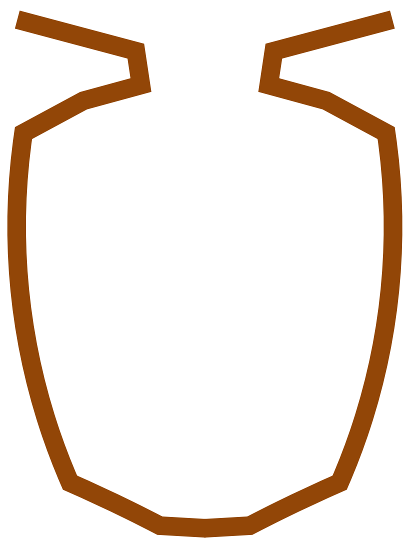
- Integrate out the frequencies

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

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\ a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

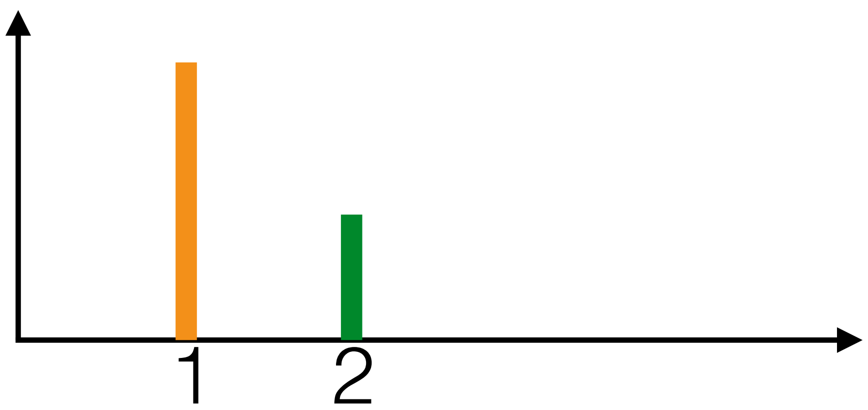


Marginal cluster assignments

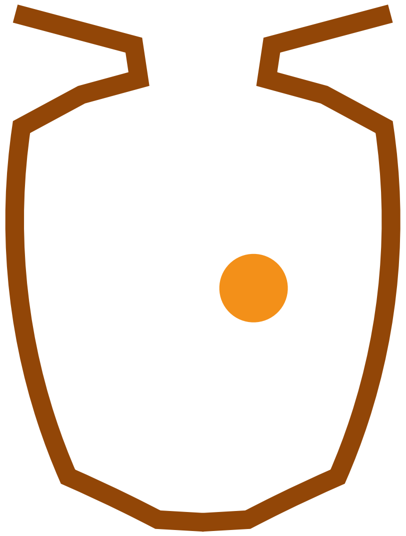
- Integrate out the frequencies

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

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

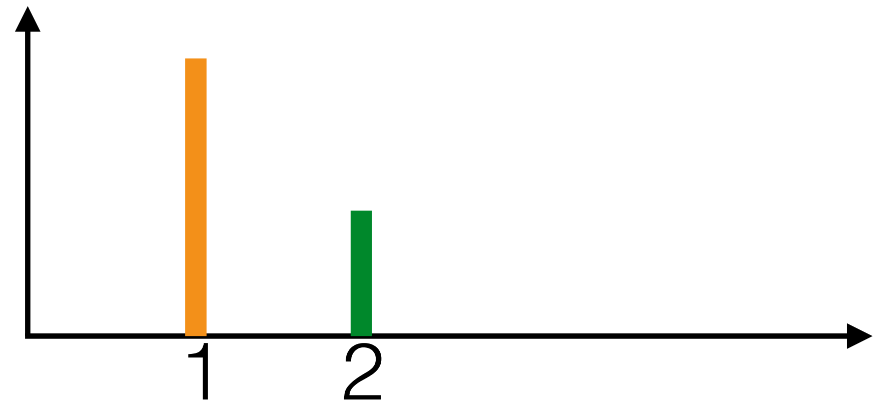


Marginal cluster assignments

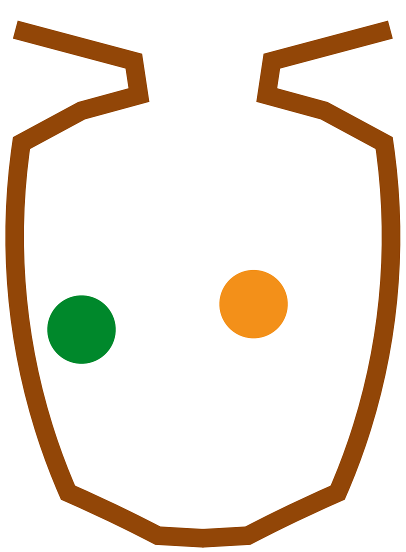
- Integrate out the frequencies

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

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

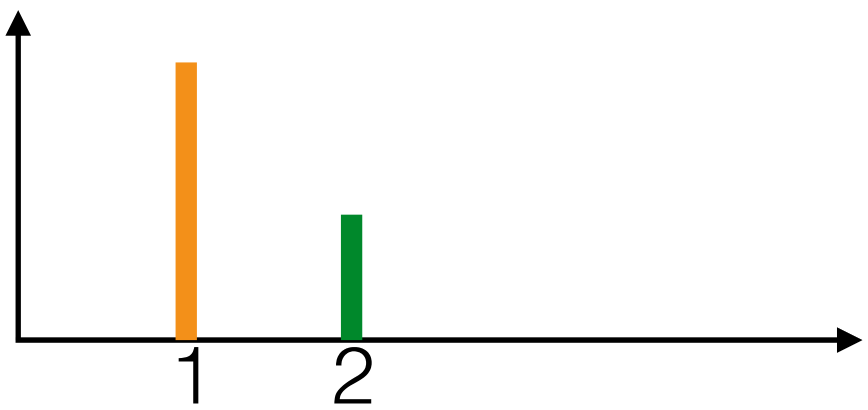


Marginal cluster assignments

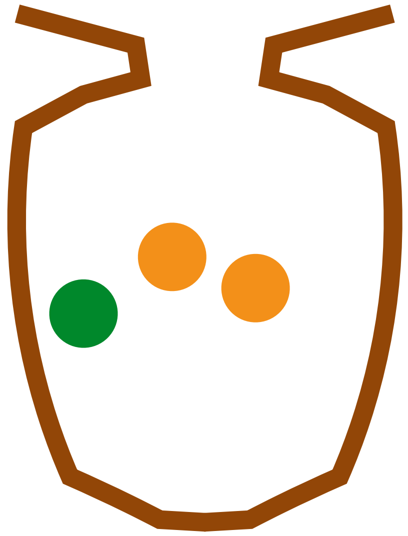
- Integrate out the frequencies

$$\rho_1 \sim \text{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

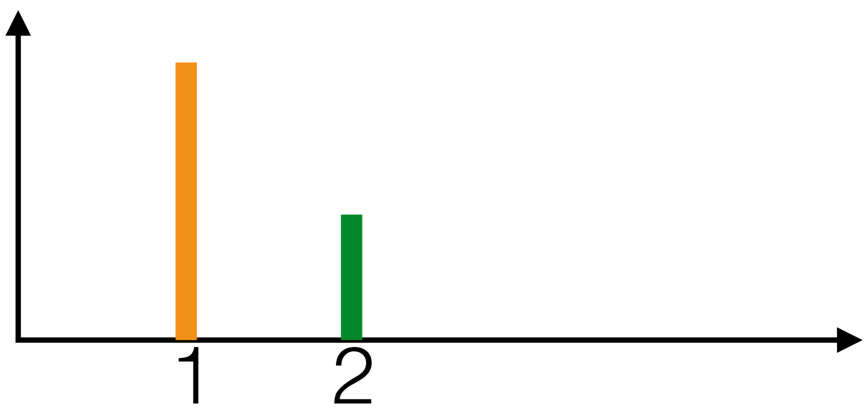


Marginal cluster assignments

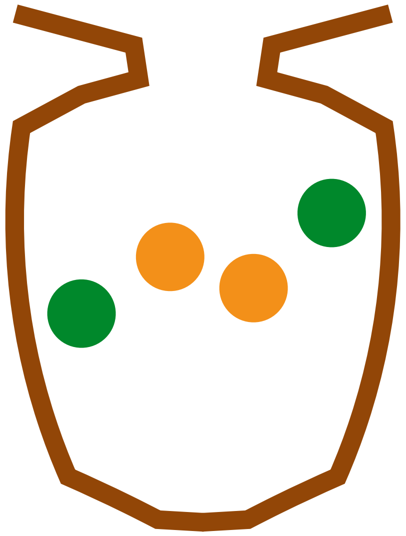
- Integrate out the frequencies

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

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



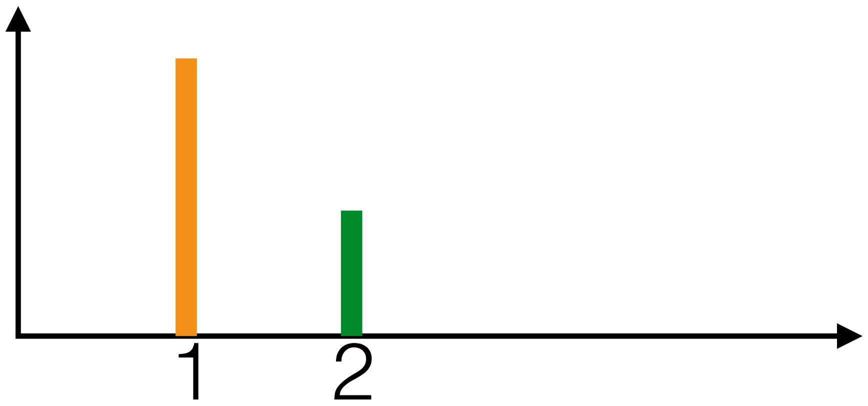
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$



Marginal cluster assignments

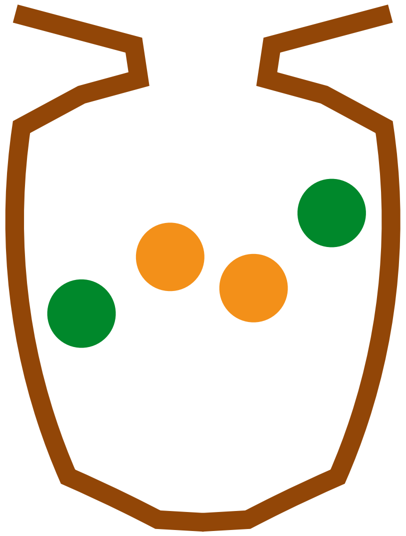
- Integrate out the frequencies

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



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

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

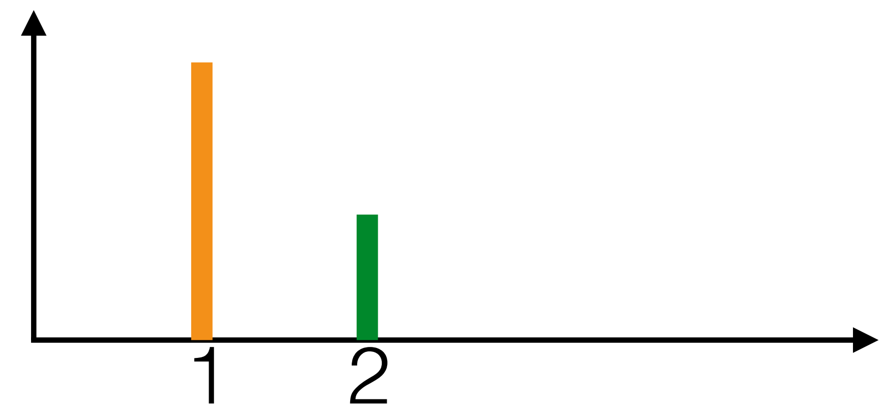


$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}}$$

Marginal cluster assignments

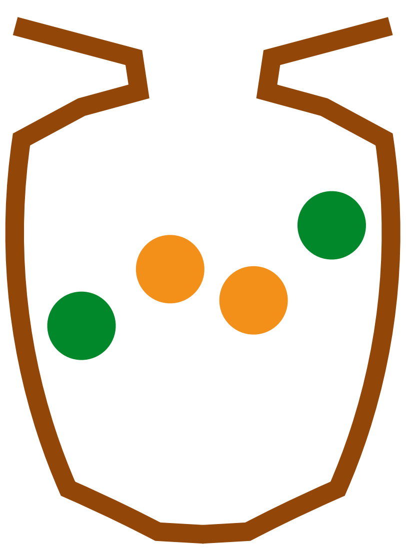
- Integrate out the frequencies

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



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

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$



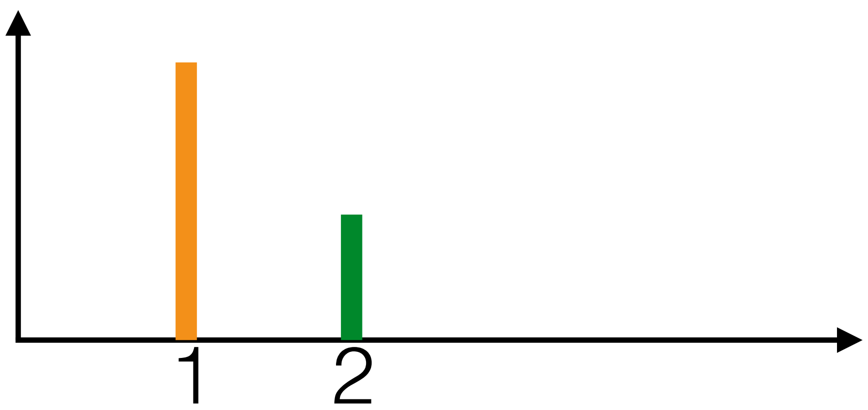
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}}$$

Marginal cluster assignments

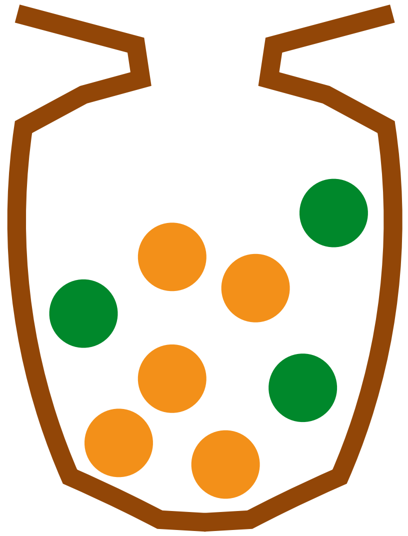
- Integrate out the frequencies

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

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

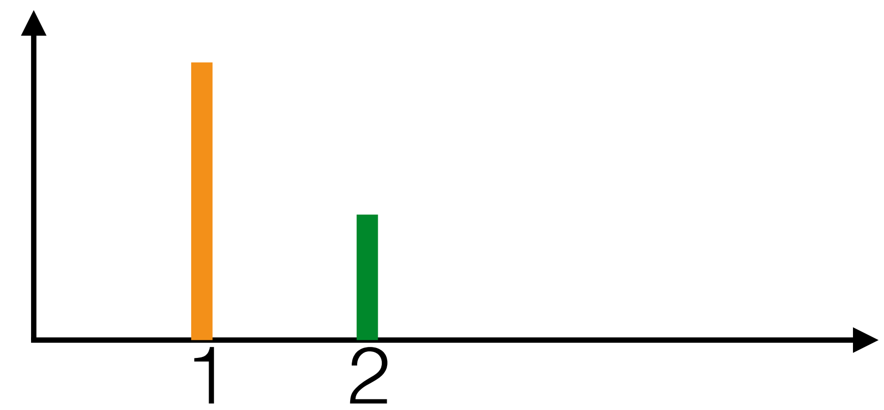


Marginal cluster assignments

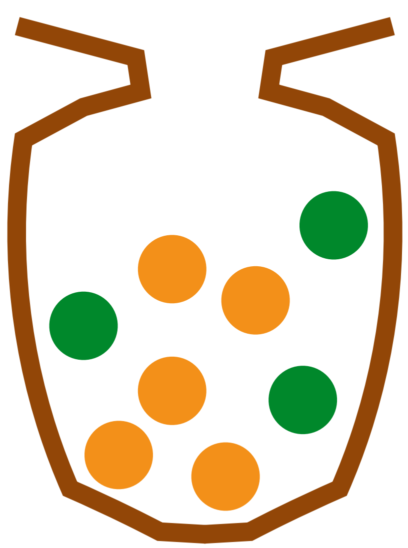
- Integrate out the frequencies

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

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

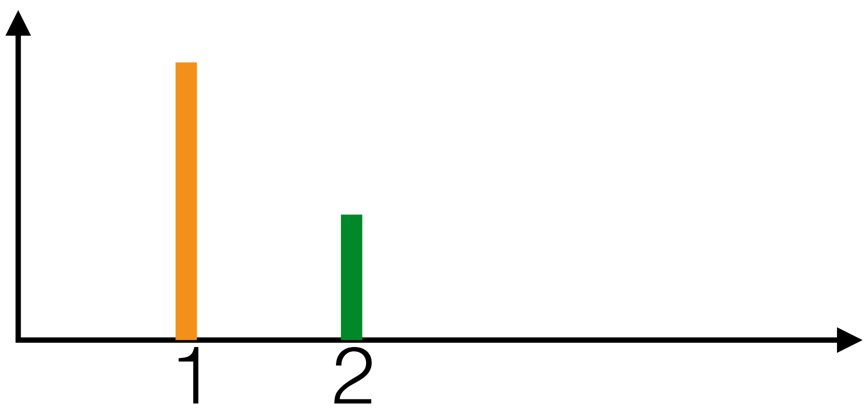


$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}}$$

Marginal cluster assignments

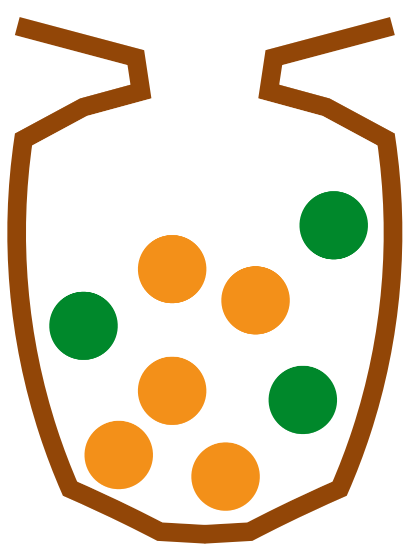
- Integrate out the frequencies

$$\rho_1 \sim \text{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$



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

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$



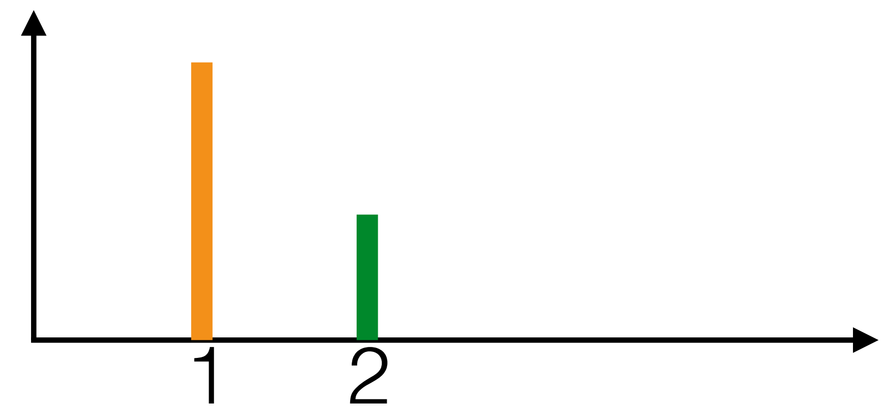
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}}$$

Marginal cluster assignments

- Integrate out the frequencies

$$\rho_1 \sim \mathrm{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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



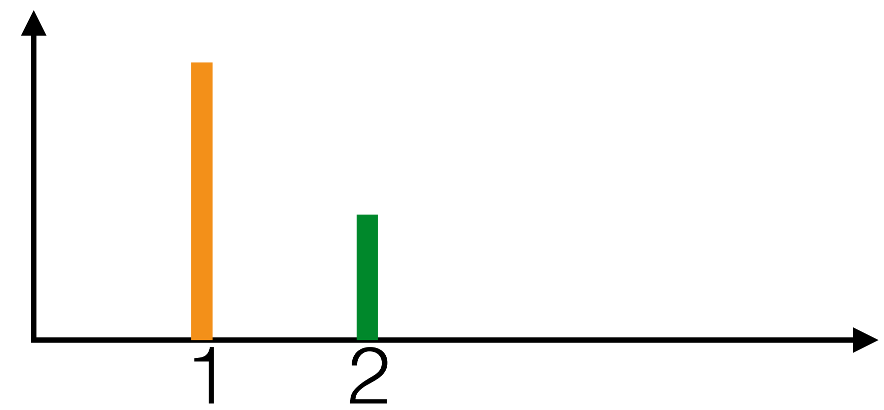
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

- Integrate out the frequencies

$$\rho_1 \sim \mathrm{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

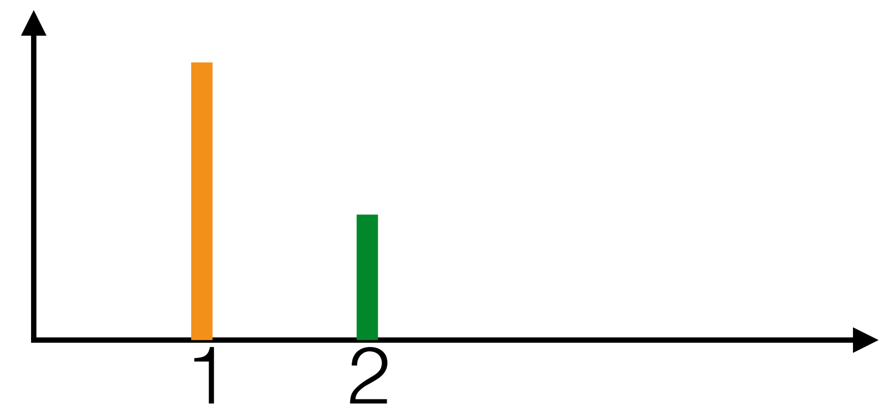
- Pólya urn

Marginal cluster assignments

- Integrate out the frequencies

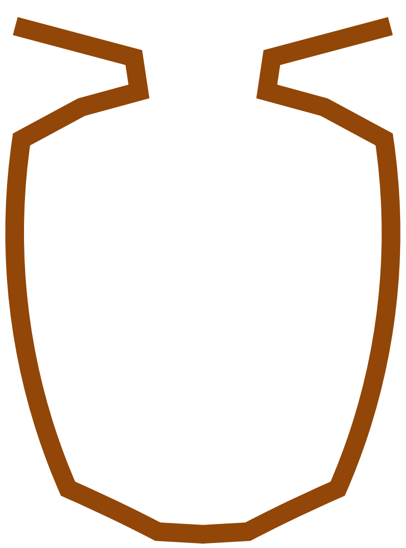
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\ a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn

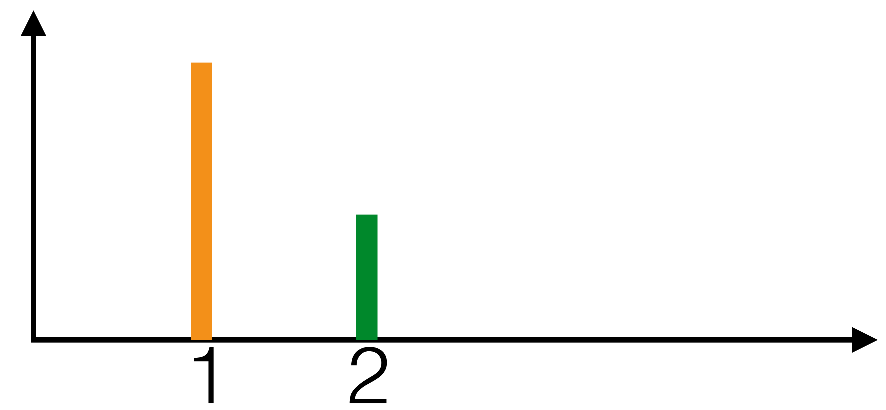


Marginal cluster assignments

- Integrate out the frequencies

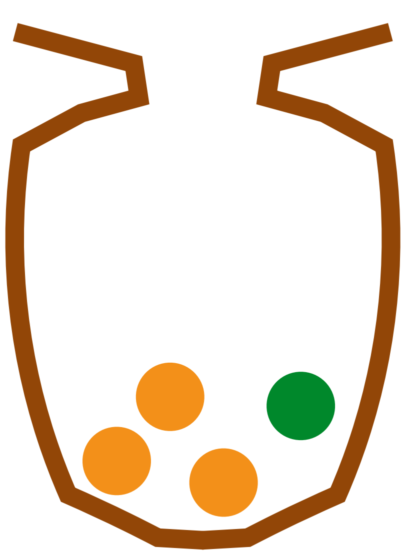
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn

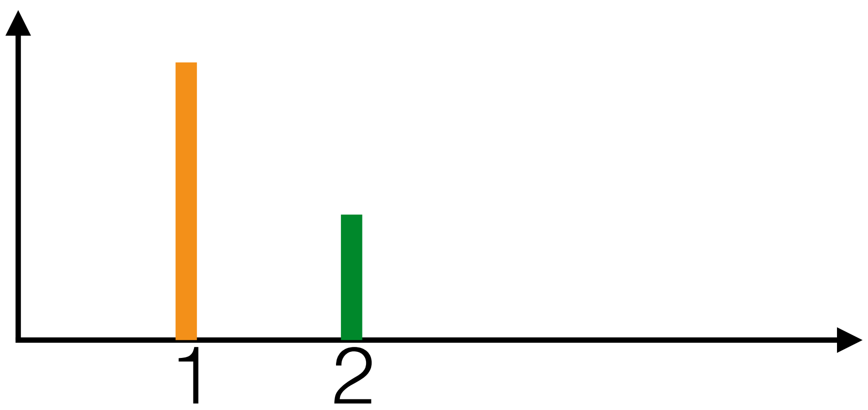


Marginal cluster assignments

- Integrate out the frequencies

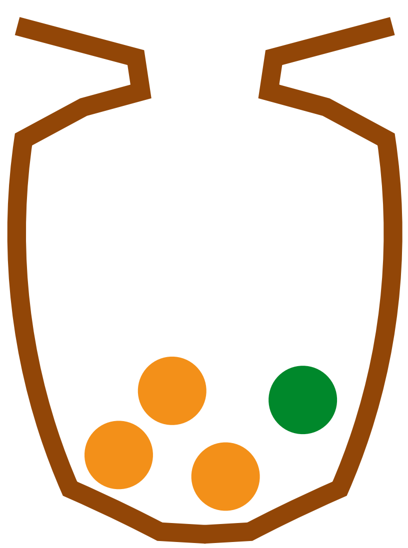
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
  - Choose any ball with equal probability

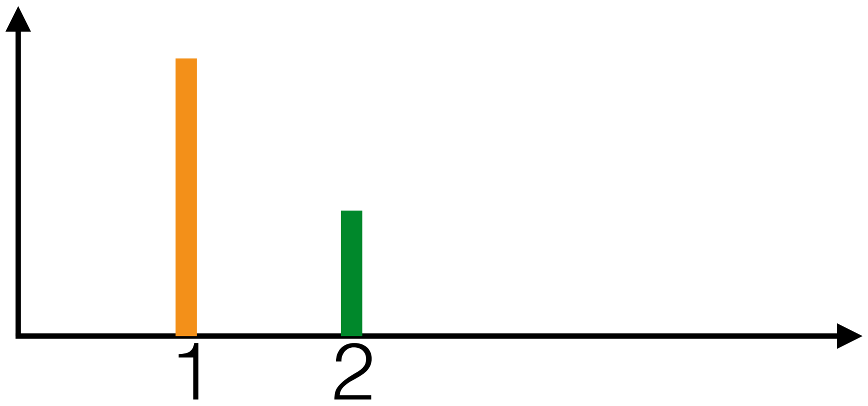


Marginal cluster assignments

- Integrate out the frequencies

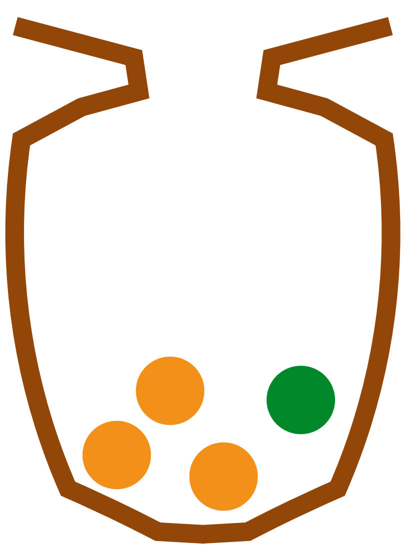
$$\rho_1 \sim \text{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color

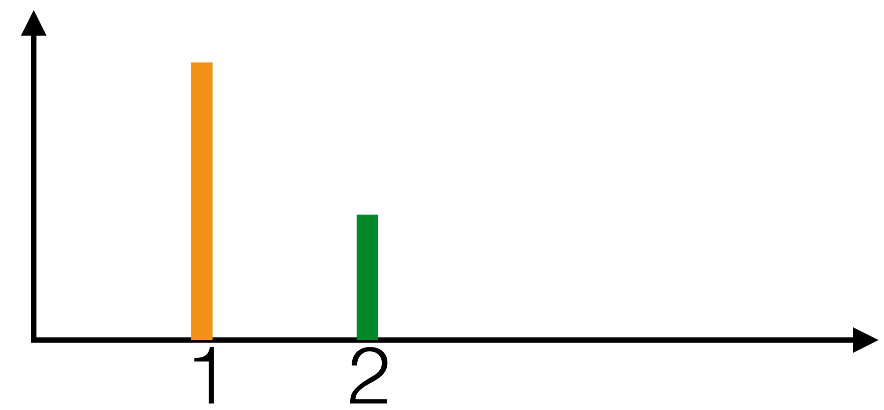


Marginal cluster assignments

- Integrate out the frequencies

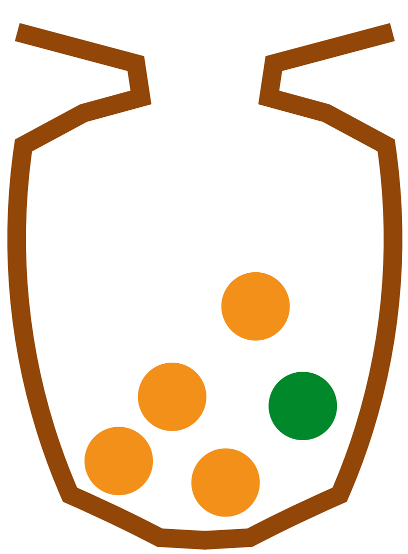
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
    - Choose any ball with equal probability
    - Replace and add ball of same color

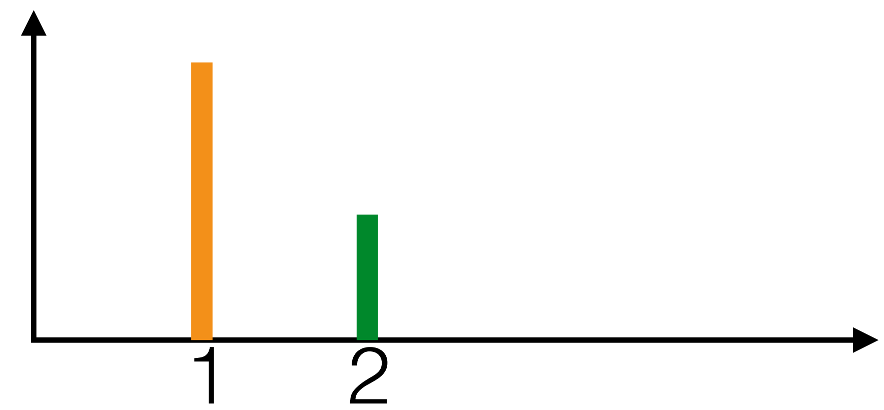


Marginal cluster assignments

- Integrate out the frequencies

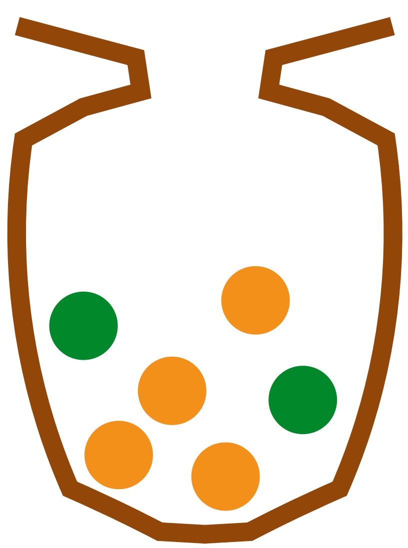
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color

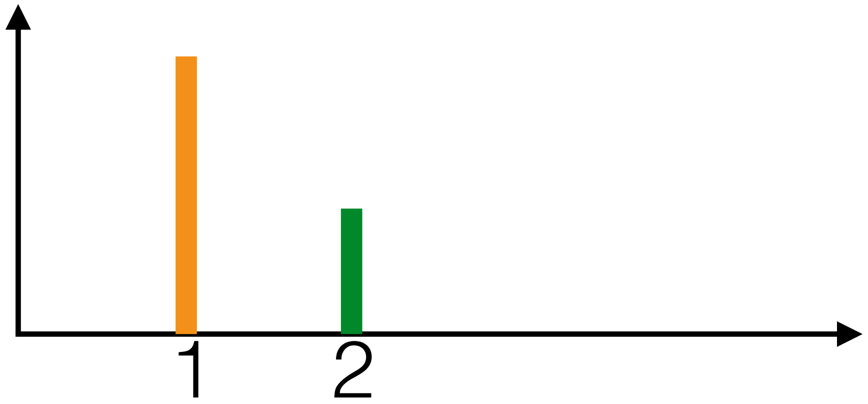


Marginal cluster assignments

- Integrate out the frequencies

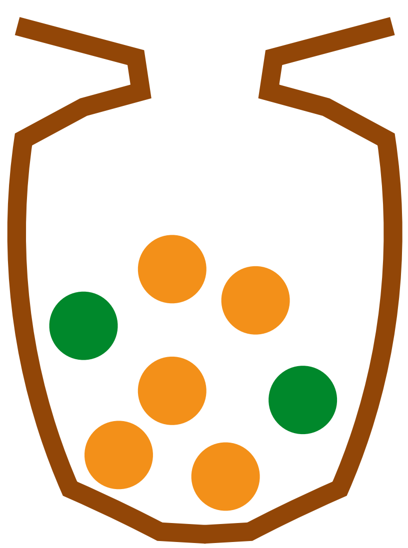
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

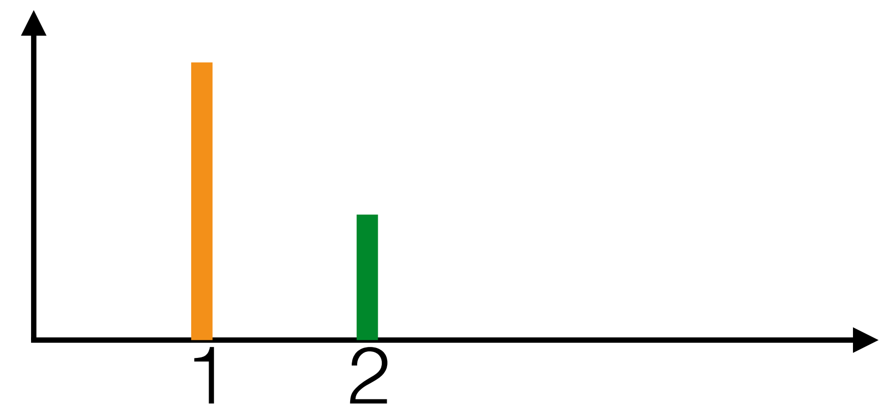
- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color



Marginal cluster assignments

- Integrate out the frequencies

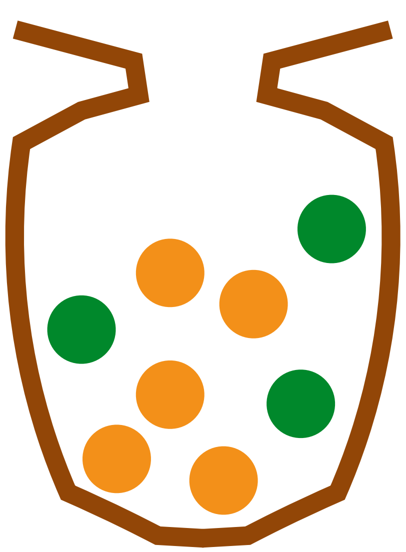
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$



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

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\ a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color

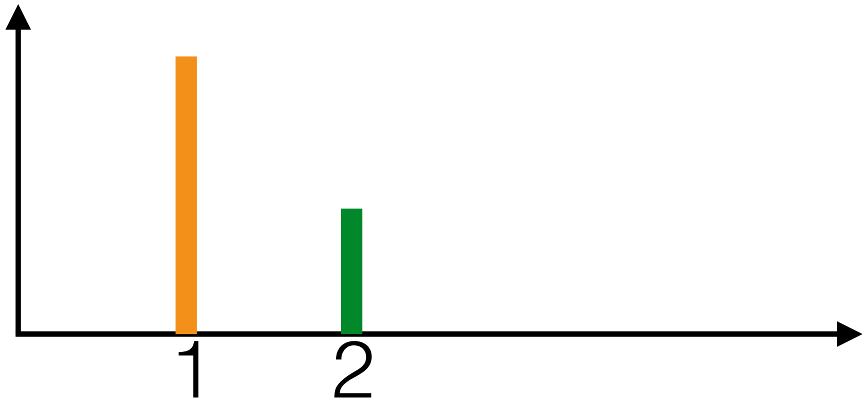


Marginal cluster assignments

- Integrate out the frequencies

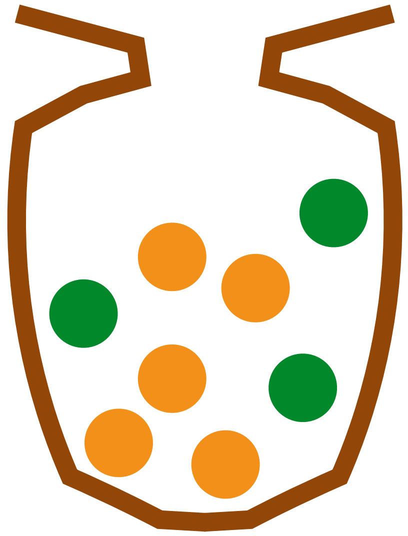
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\ a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color



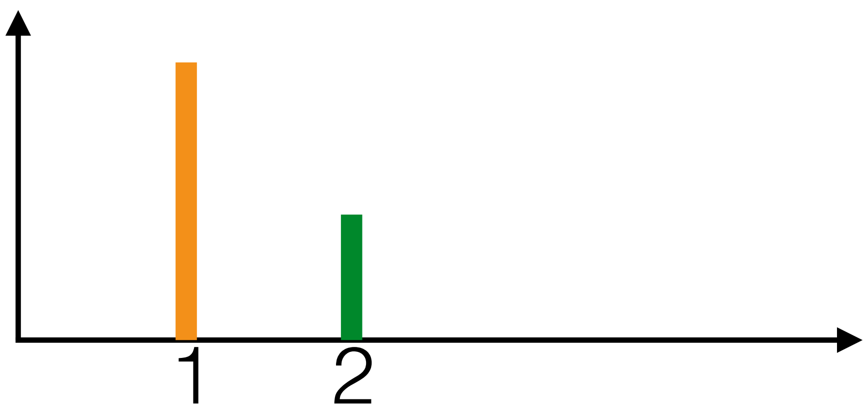
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}}$$

Marginal cluster assignments

- Integrate out the frequencies

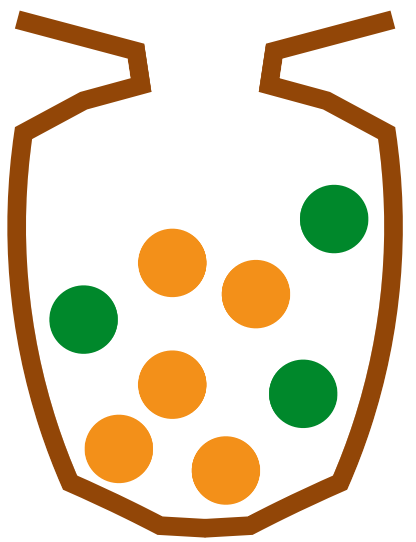
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\ a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

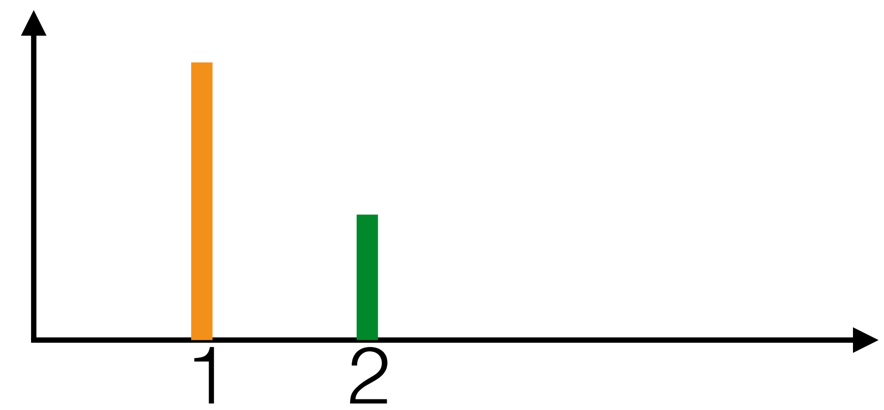
- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color



$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}}$$

Marginal cluster assignments

- Integrate out the frequencies

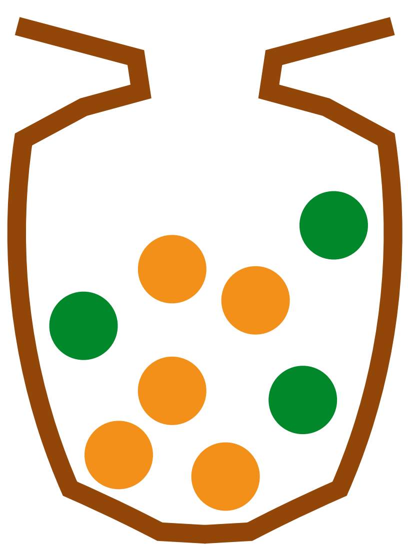


$$\rho_1 \sim \mathrm{Beta}(a_1, a_2),\, z_n \stackrel{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

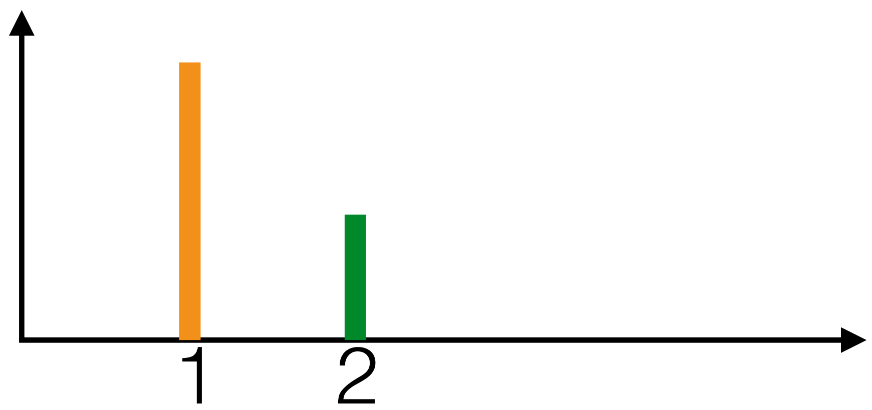
- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color



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

Marginal cluster assignments

- Integrate out the frequencies

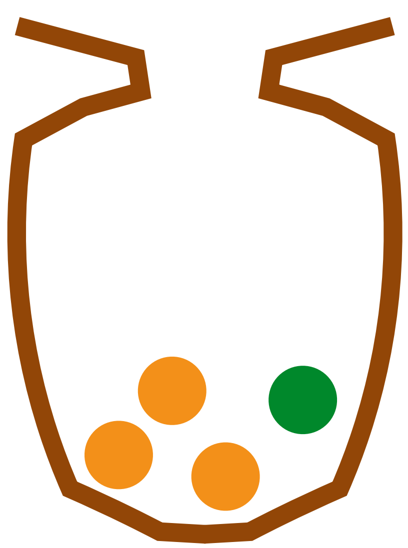


$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\ a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color

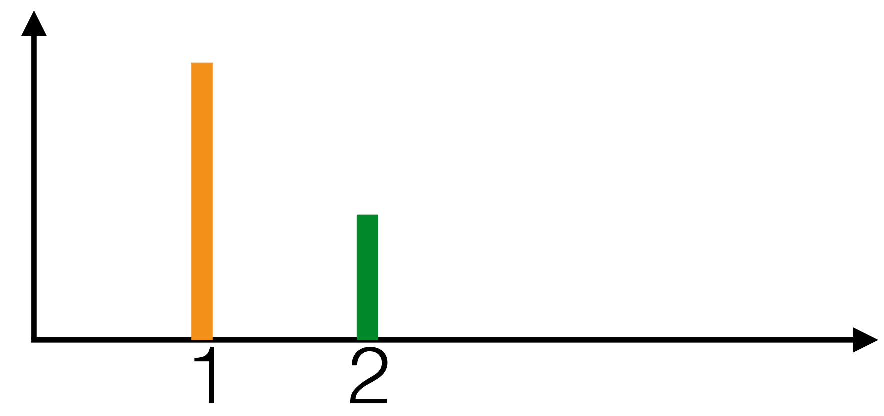


$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

Marginal cluster assignments

- Integrate out the frequencies

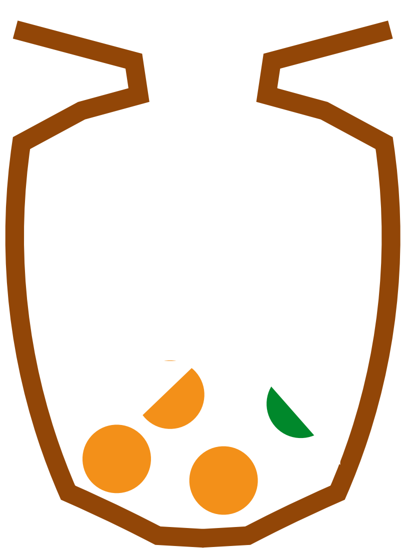
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$



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

$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\ a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color



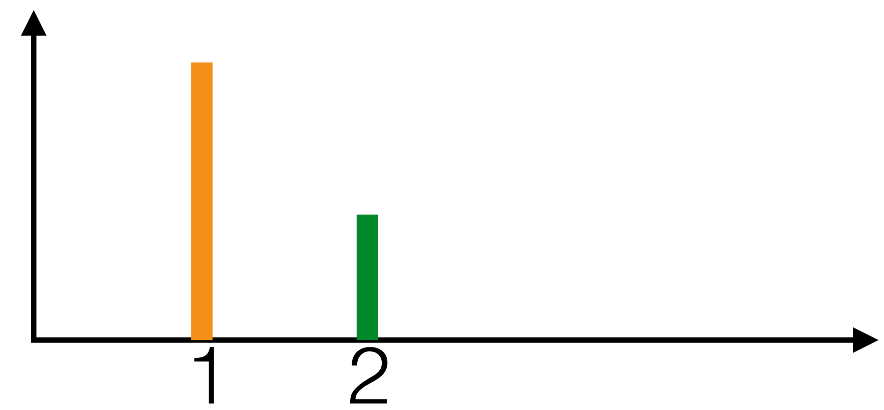
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

Marginal cluster assignments

- Integrate out the frequencies

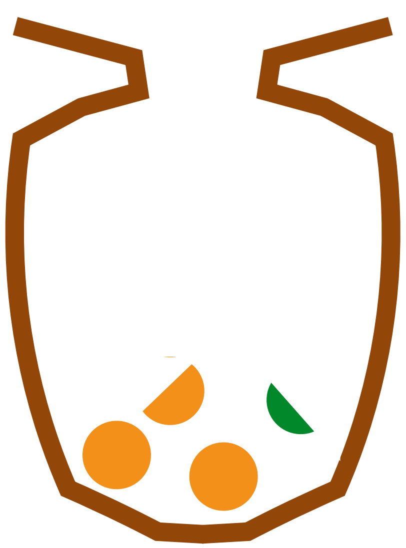
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

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



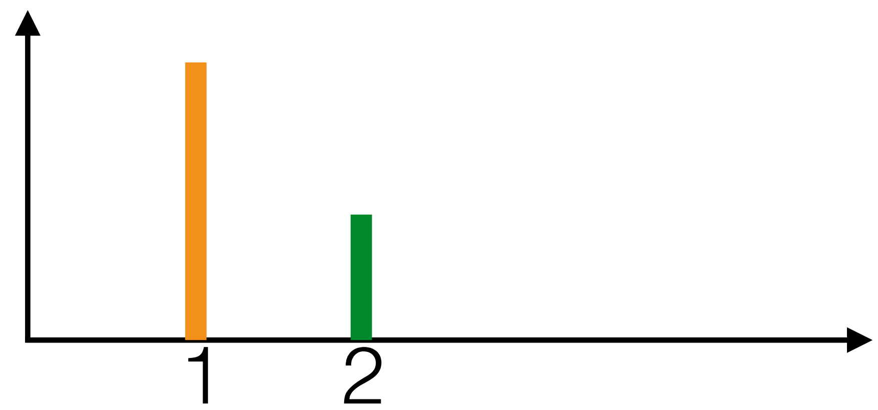
$$\lim_{n\to\infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

Marginal cluster assignments

- Integrate out the frequencies

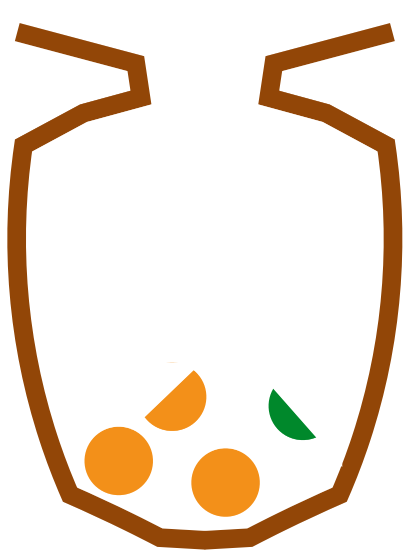
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

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

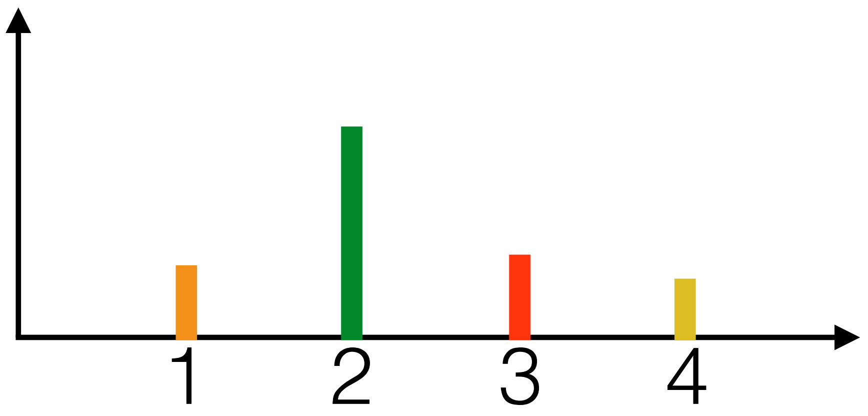


$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \stackrel{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

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

Marginal cluster assignments

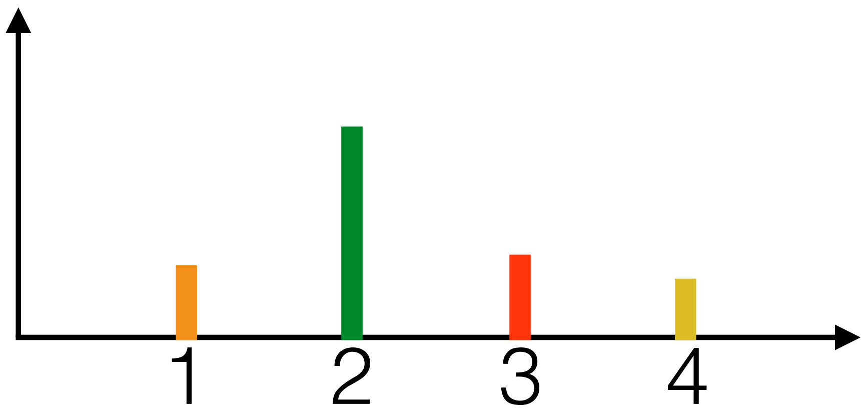
- Integrate out the frequencies



Marginal cluster assignments

- Integrate out the frequencies

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$

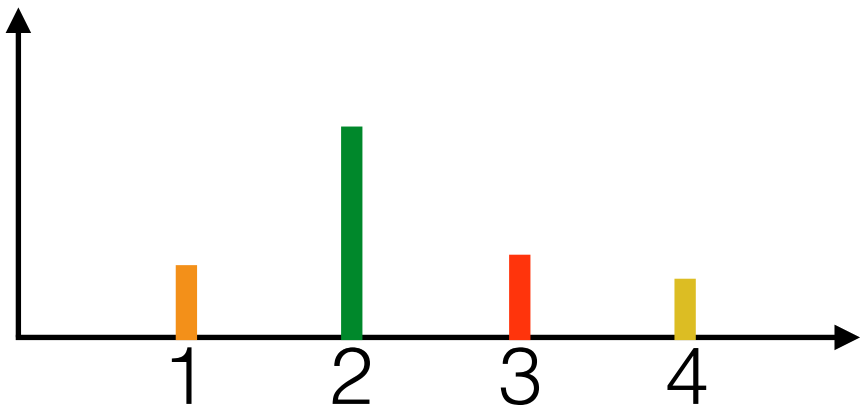


Marginal cluster assignments

- Integrate out the frequencies

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$

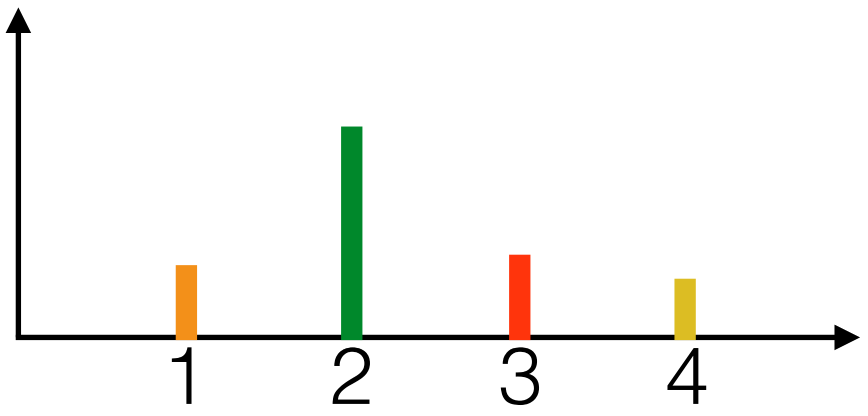
$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$



Marginal cluster assignments

- Integrate out the frequencies

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_{1:K})$$



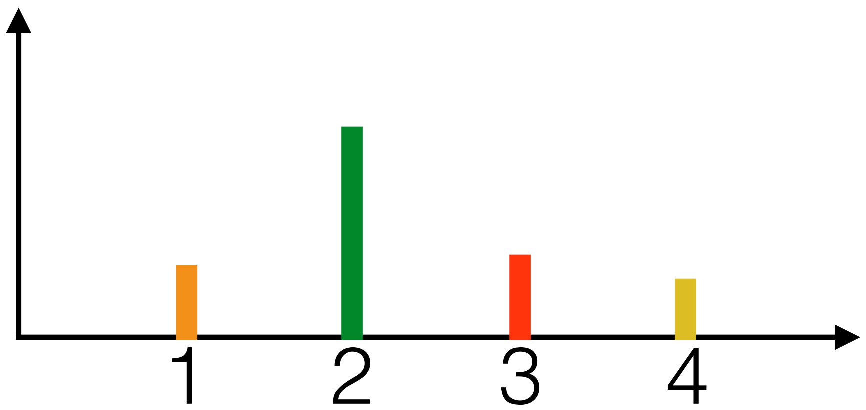
$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$

$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

Marginal cluster assignments

- Integrate out the frequencies

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_{1:K})$$



$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$

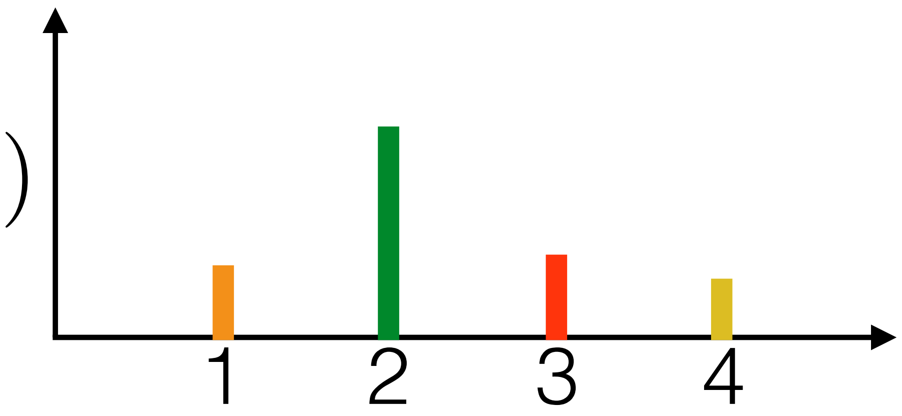
$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

- multivariate Pólya urn

Marginal cluster assignments

- Integrate out the frequencies

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$



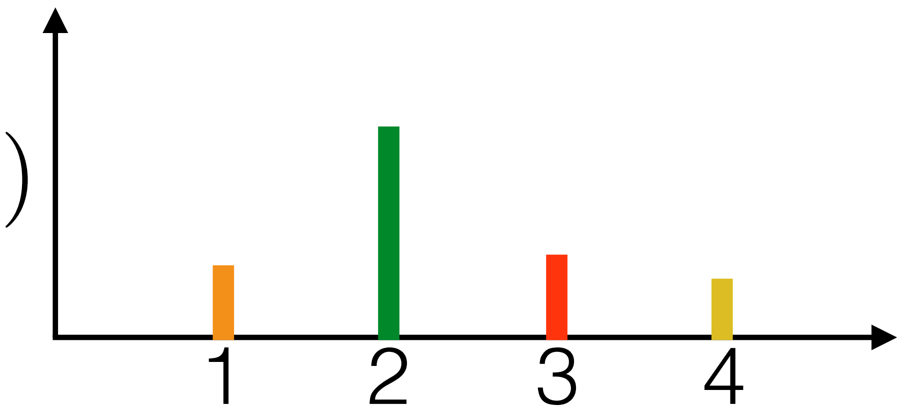
$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$

$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

- multivariate Pólya urn

Marginal cluster assignments

- Integrate out the frequencies



$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$

$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$

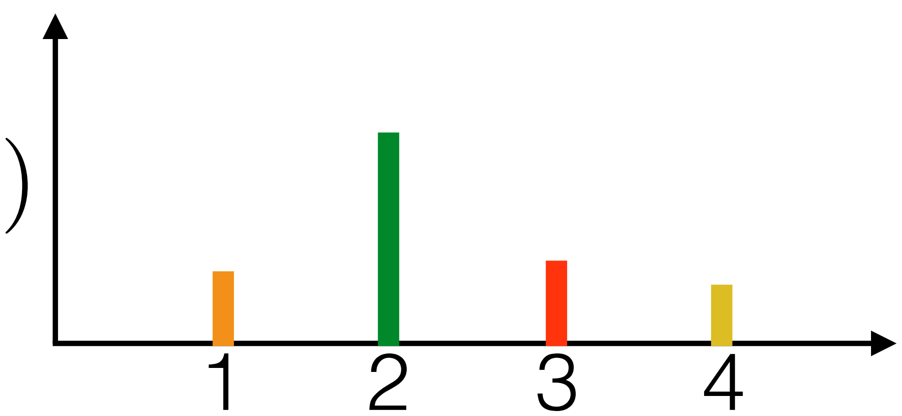
$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

- multivariate Pólya urn
  - Choose any ball with prob proportional to its mass

Marginal cluster assignments

- Integrate out the frequencies

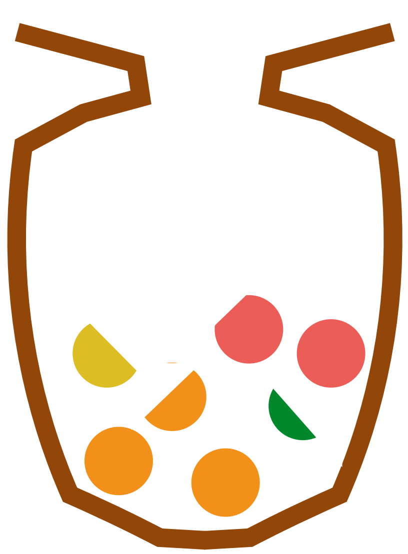
$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$



$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^K a_{j,n}}$$

$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

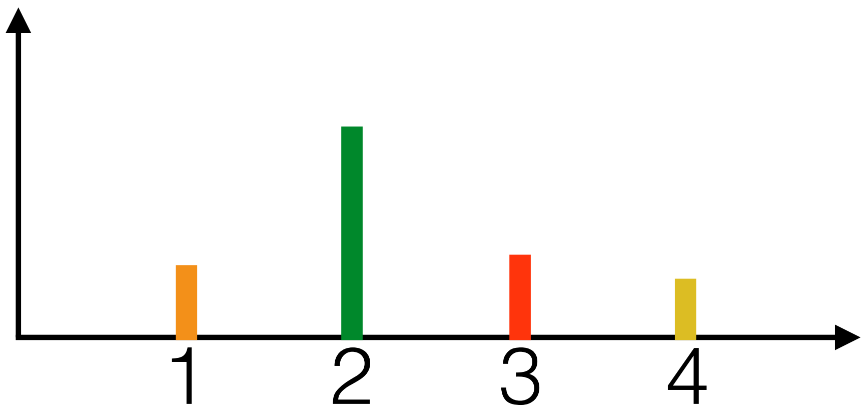
- multivariate Pólya urn
  - Choose any ball with prob proportional to its mass
  - Replace and add ball of same color



Marginal cluster assignments

- Integrate out the frequencies

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$



$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$

$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

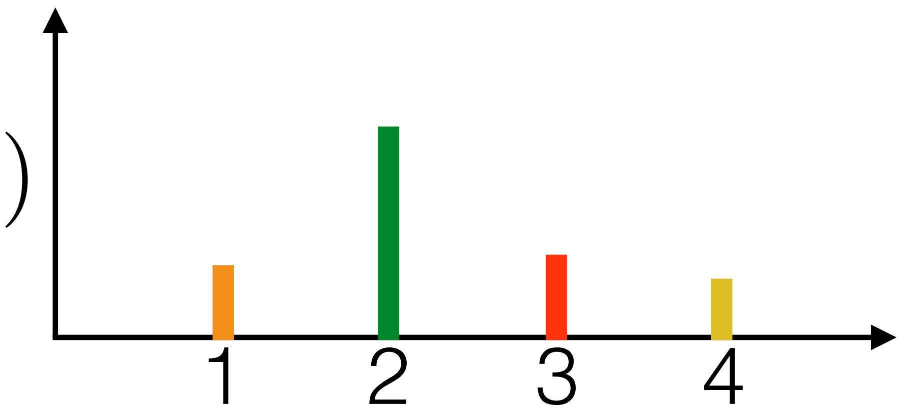
- multivariate Pólya urn
  - Choose any ball with prob proportional to its mass
  - Replace and add ball of same color

$$\lim_{n \to \infty} \frac{(\#\ \text{orange}, \#\ \text{green}, \#\ \text{red}, \#\ \text{yellow})}{\#\ \text{total}}$$

Marginal cluster assignments

- Integrate out the frequencies

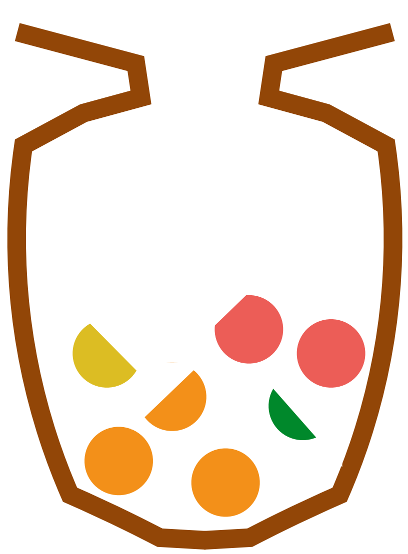
$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$



$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$

$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

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



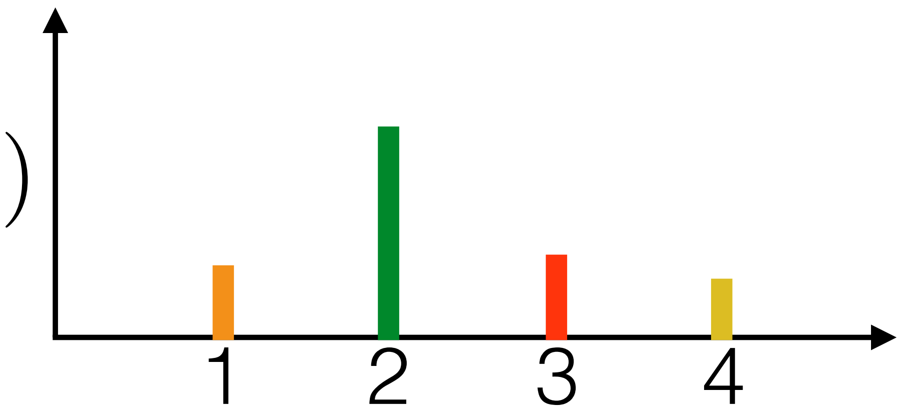
$$\lim_{n\to\infty} \frac{(\#\text{ orange},\, \#\text{ green},\, \#\text{ red},\, \#\text{ yellow})}{\#\text{ total}}$$

$$\to (\rho_{\text{orange}}, \rho_{\text{green}}, \rho_{\text{red}}, \rho_{\text{yellow}})$$

Marginal cluster assignments

- Integrate out the frequencies

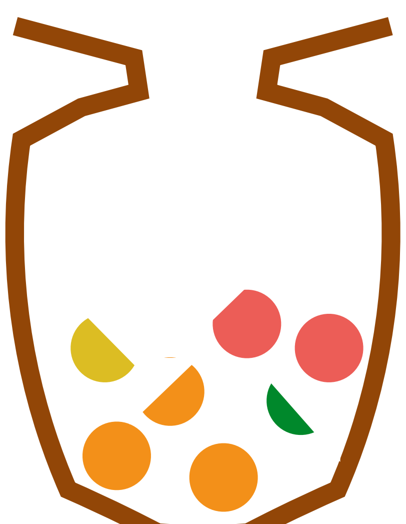
$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}),\, z_n \stackrel{iid}{\sim} \text{Cat}(\rho_{1:K})$$



$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$

$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

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



$$\lim_{n\to\infty} \frac{(\#\text{ orange},\, \#\text{ green},\, \#\text{ red},\, \#\text{ yellow})}{\#\text{ total}}$$

$$\to (\rho_{\text{orange}}, \rho_{\text{green}}, \rho_{\text{red}}, \rho_{\text{yellow}})$$

$$\stackrel{d}{=} \text{Dirichlet}(a_{\text{orange}}, a_{\text{green}}, a_{\text{red}}, a_{\text{yellow}})$$

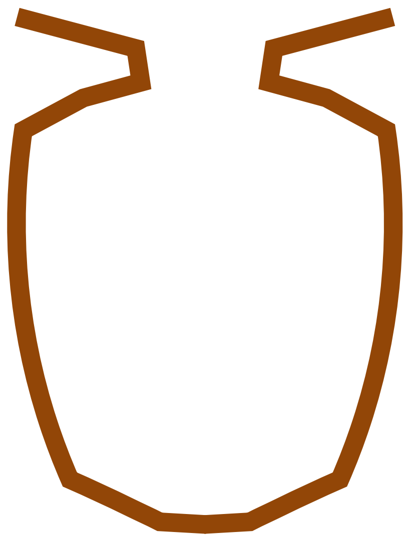
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

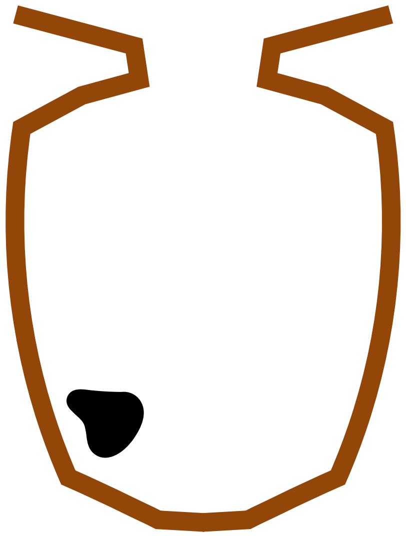
- Hoppe urn / Blackwell-MacQueen urn



[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

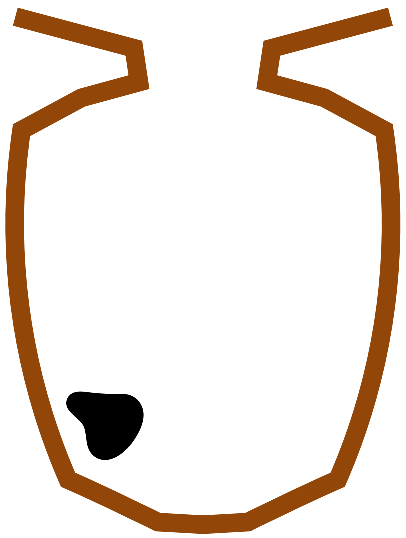
- Hoppe urn / Blackwell-MacQueen urn



[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

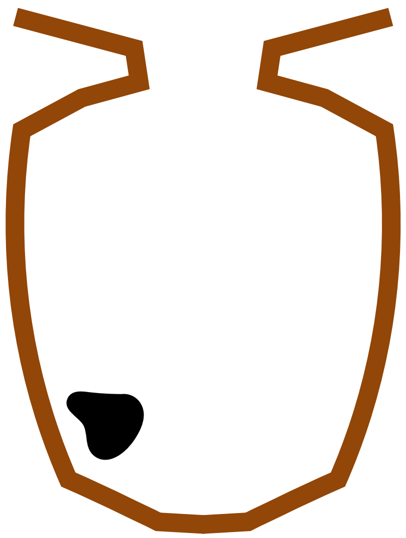


  - Choose ball with prob proportional to its mass

[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

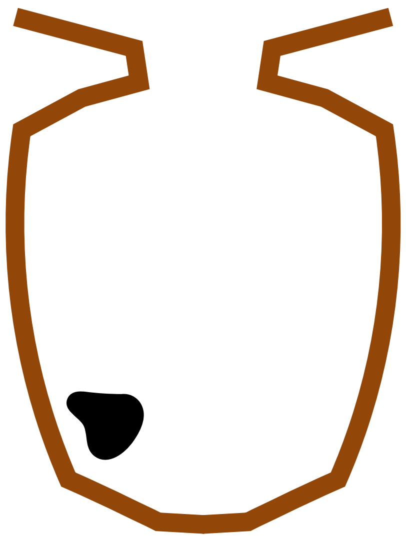


  - Choose ball with prob proportional to its mass
    - If black, replace and add ball of new color

[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

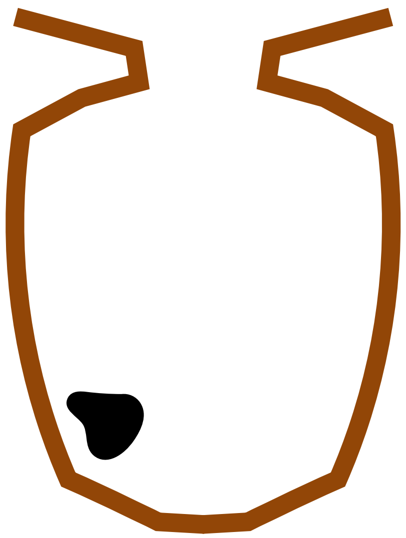


- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

[Blackwell, MacQueen 1973; Hoppe 1984]

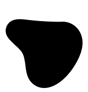
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



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

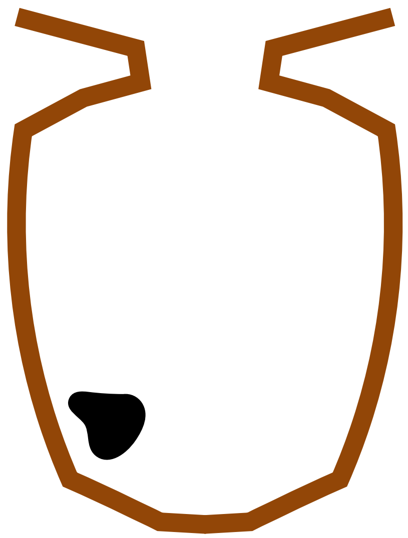
Step 0



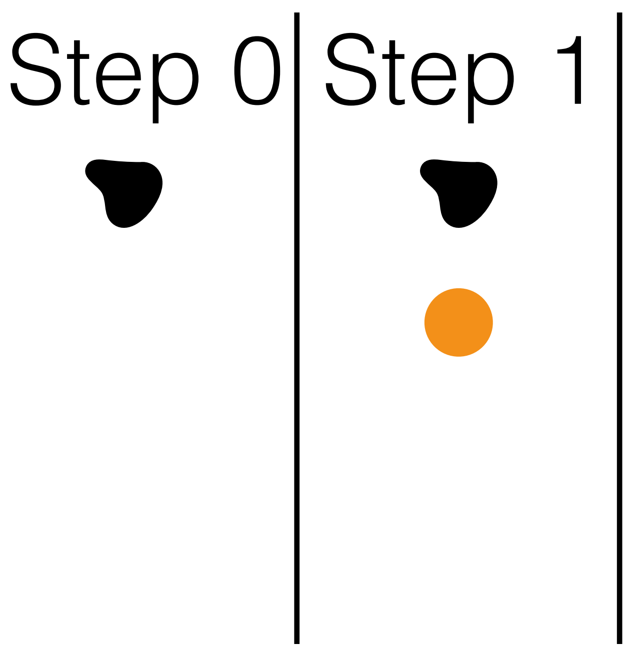
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



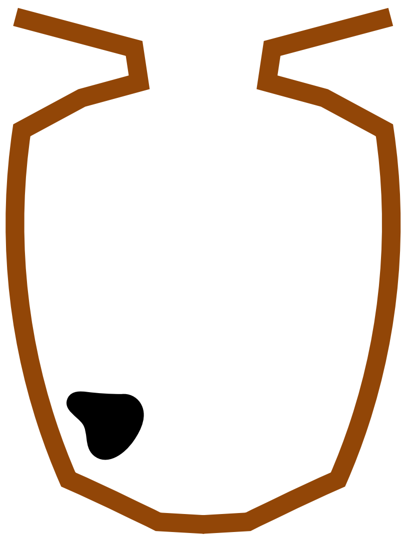
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



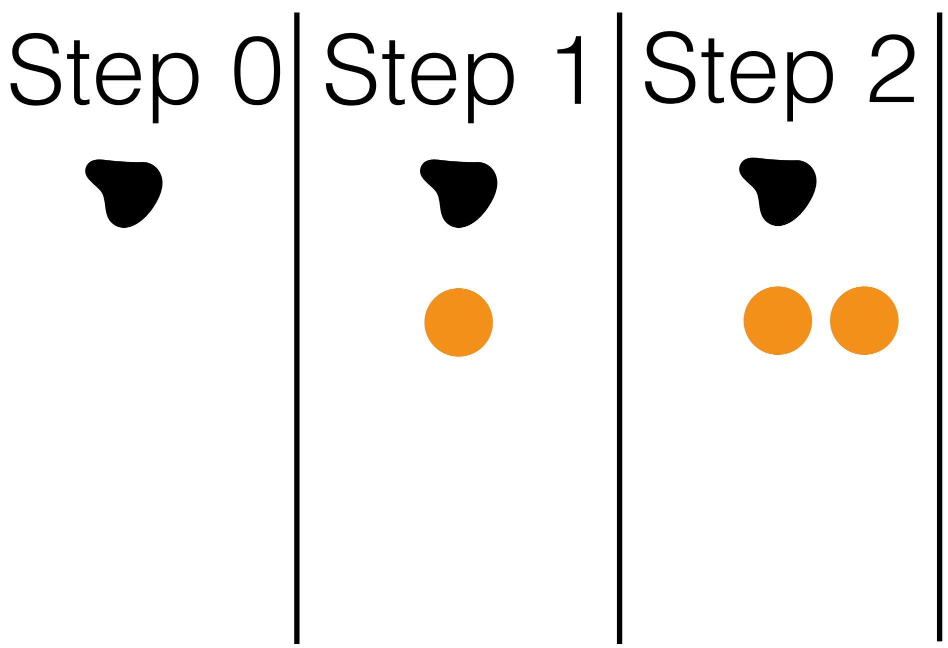
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



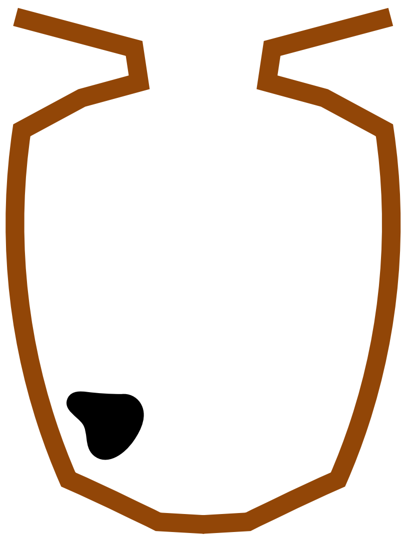
  - Choose ball with prob proportional to its mass
    - If black, replace and add ball of new color
    - Else, replace and add ball of same color



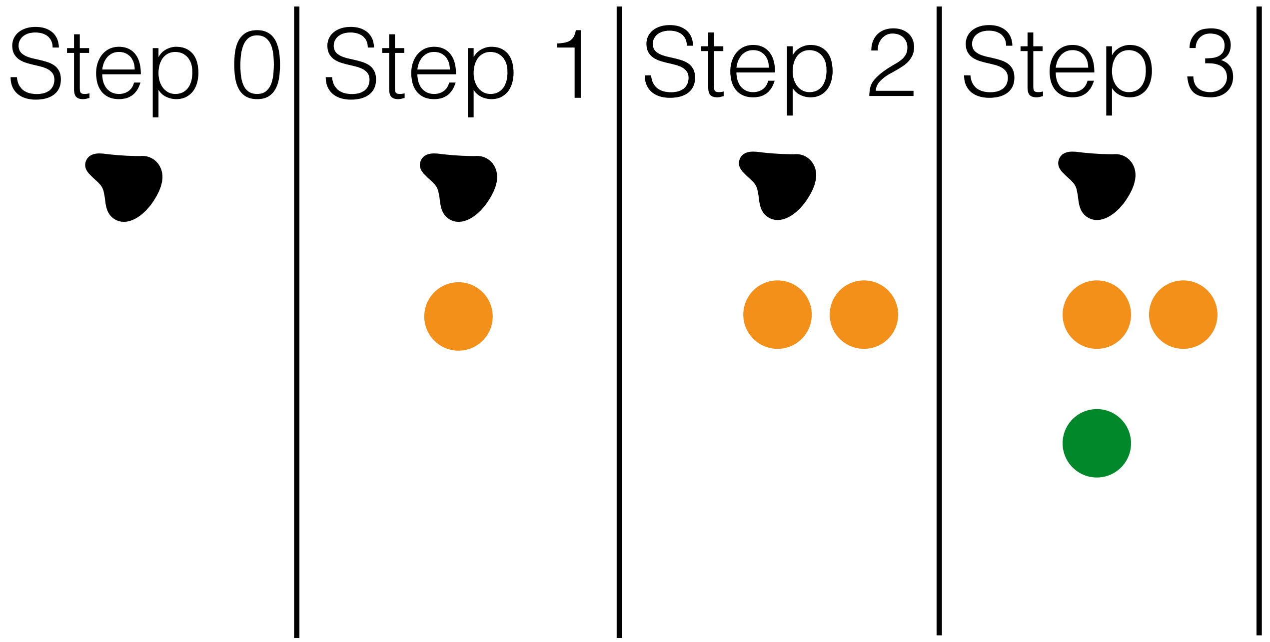
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



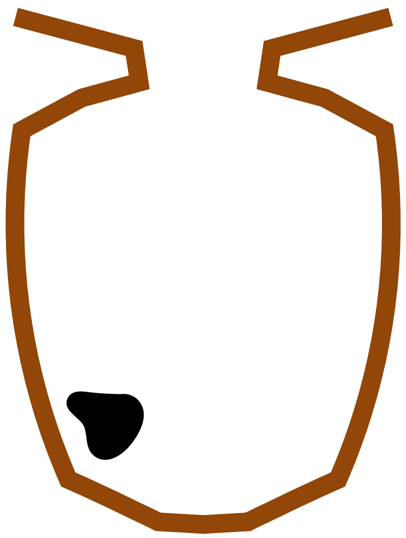
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



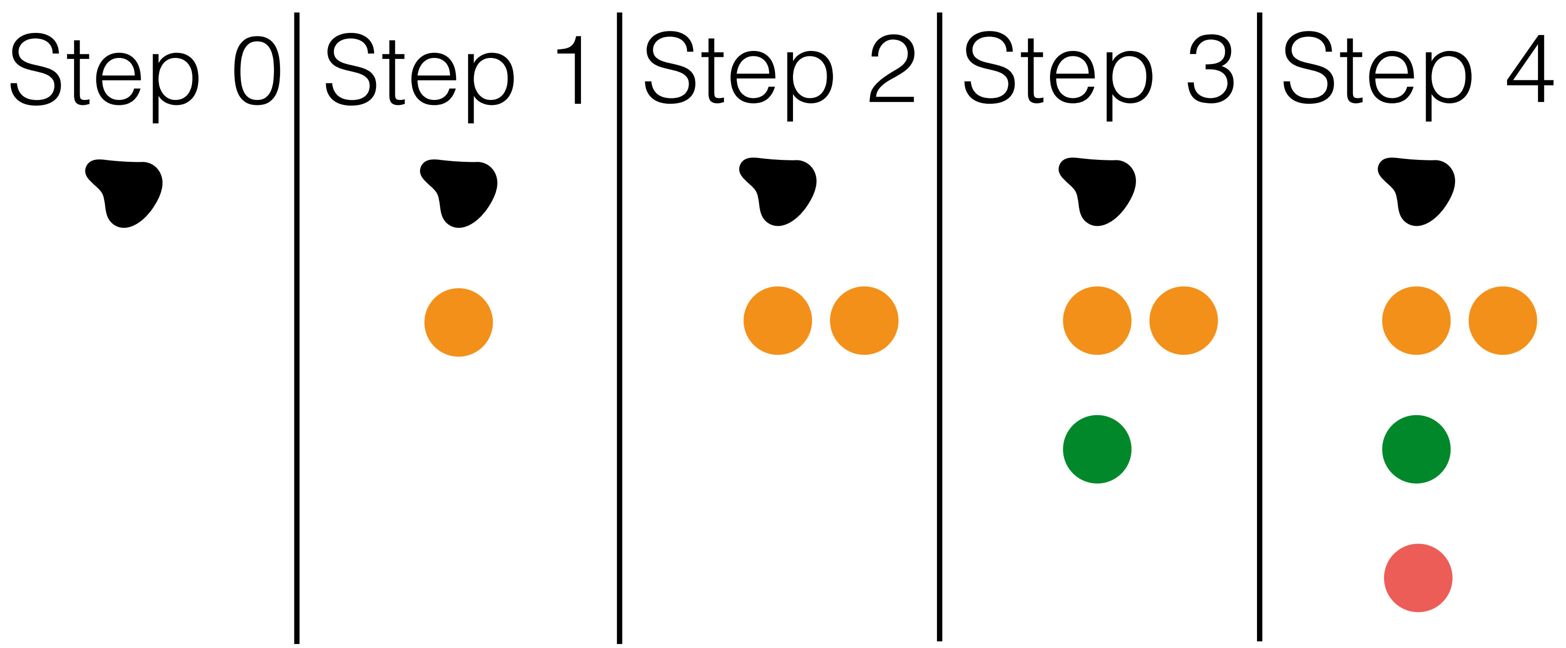
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



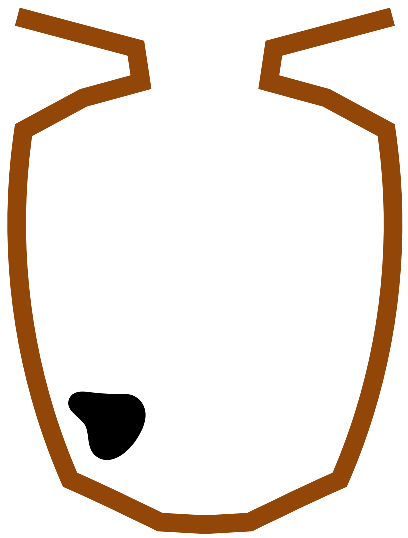
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



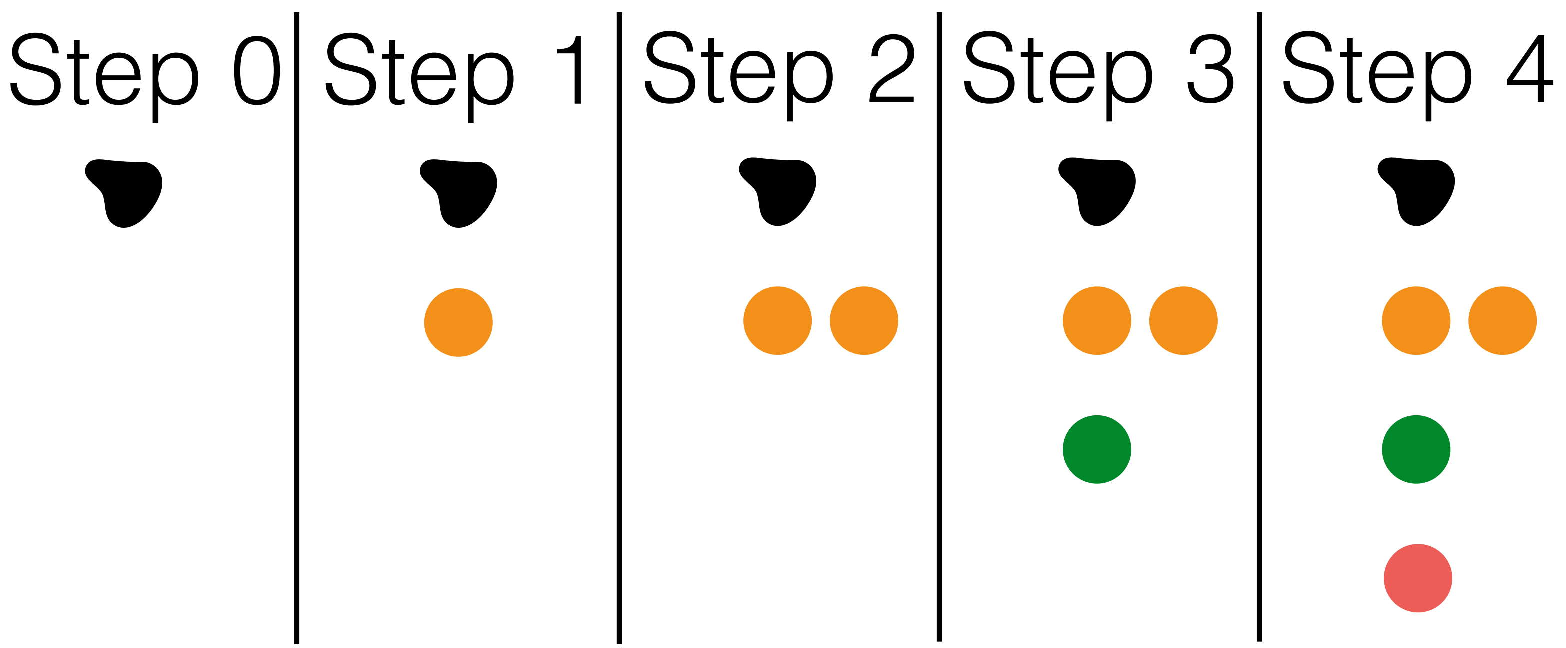
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



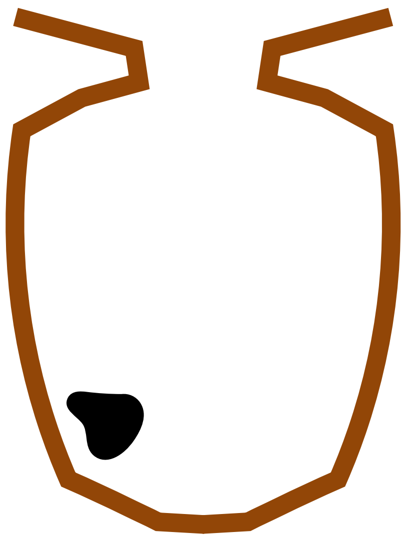
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



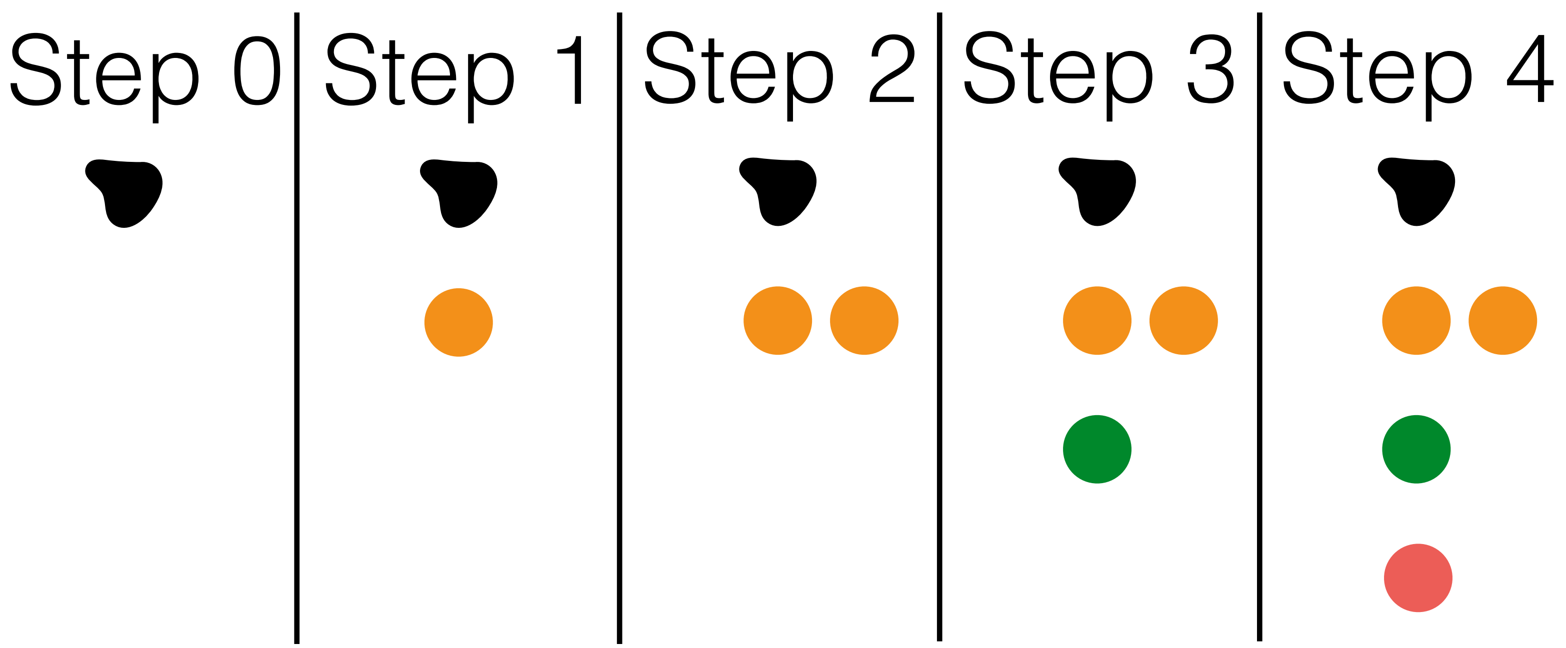
$$(\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$$

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



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

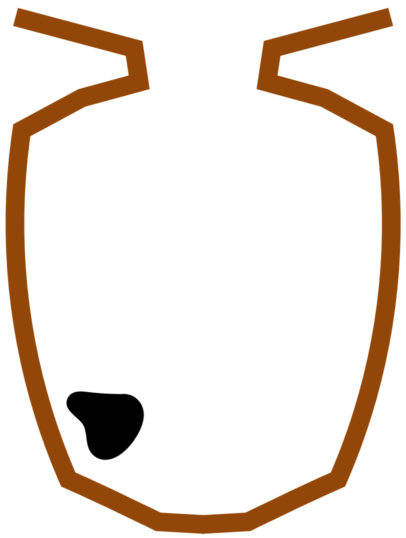


$(\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

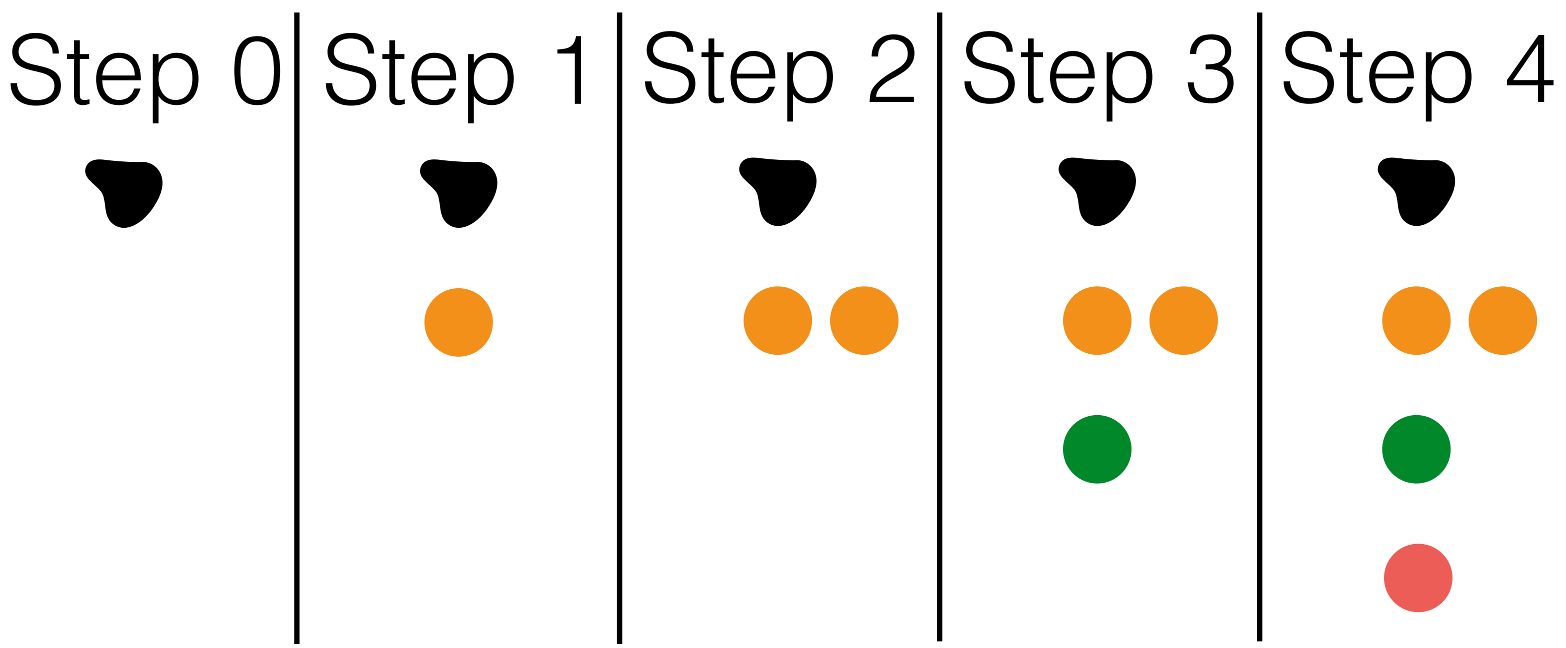
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



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

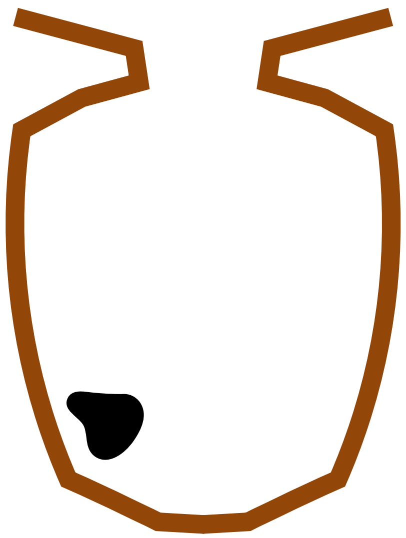


$$(\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$$

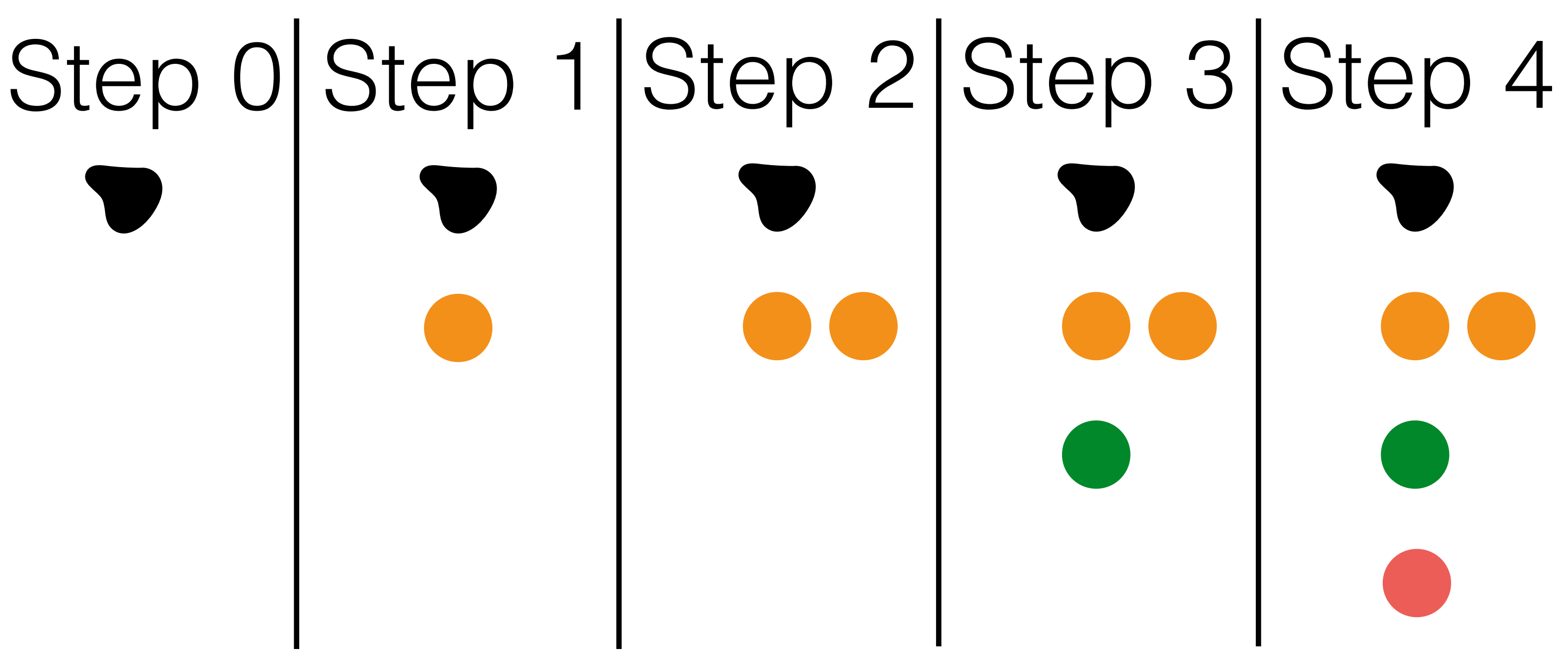
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



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



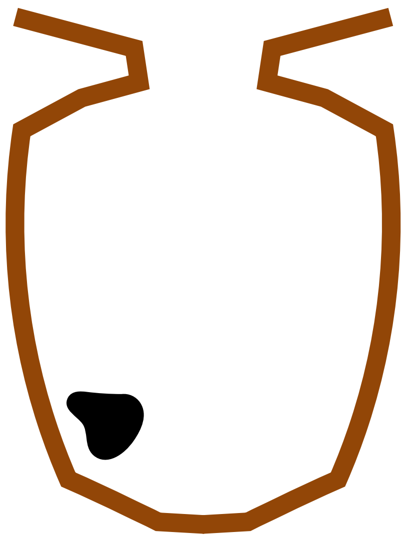
$$V_k \overset{iid}{\sim} \text{Beta}(1, \alpha)$$

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

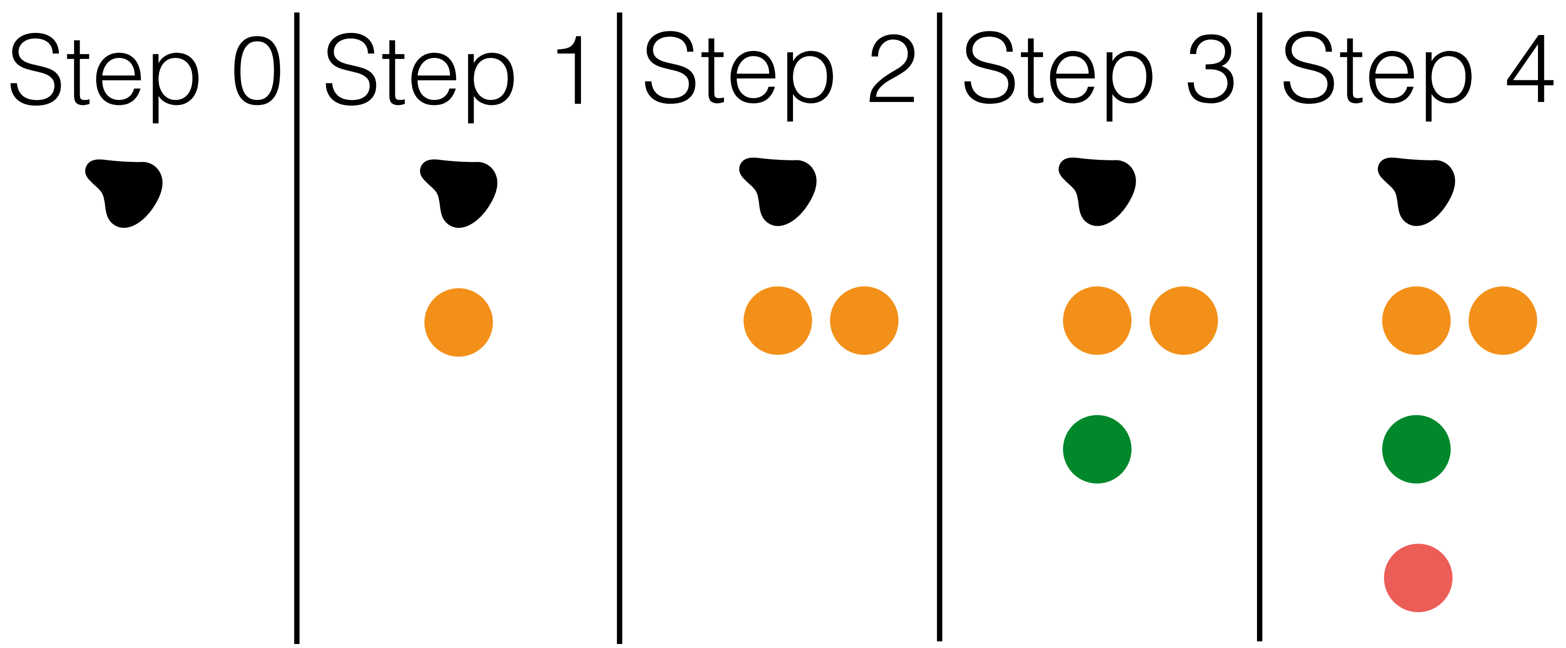
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



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



$$V_k \stackrel{iid}{\sim} \mathrm{Beta}(1, \alpha)$$

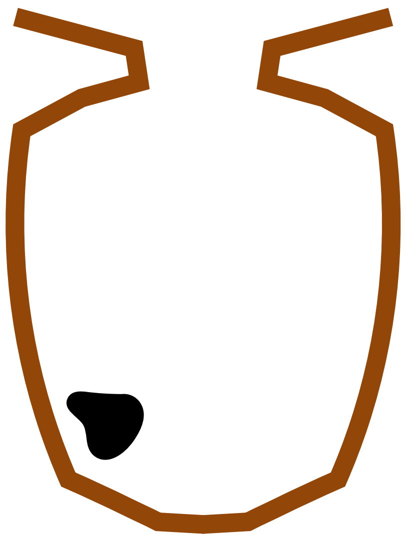
$$\rho_1 = V_1$$

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

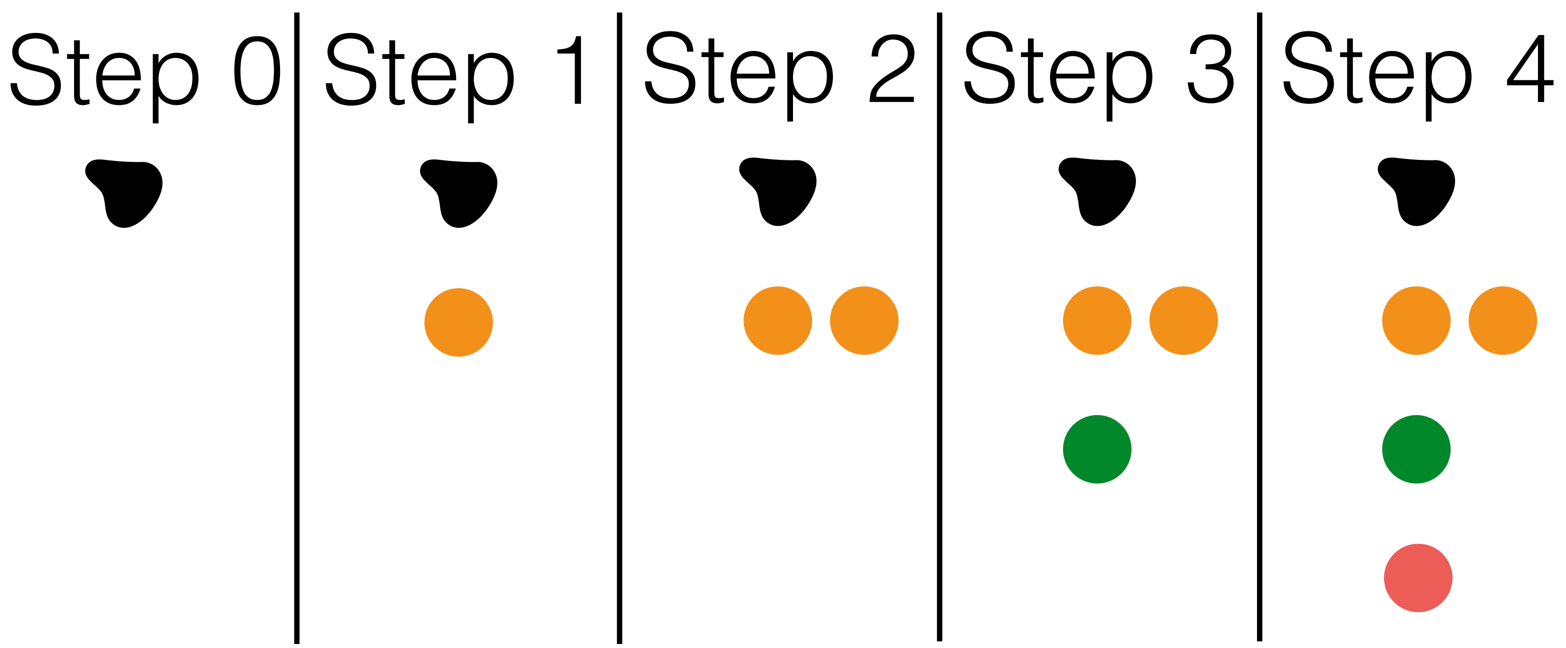
- not orange: $(\#\mathrm{green}, \#\mathrm{other}) = \mathrm{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\mathrm{red}, \#\mathrm{other}) = \mathrm{PolyaUrn}(1, \alpha)$

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



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



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

$$\rho_1 = V_1$$

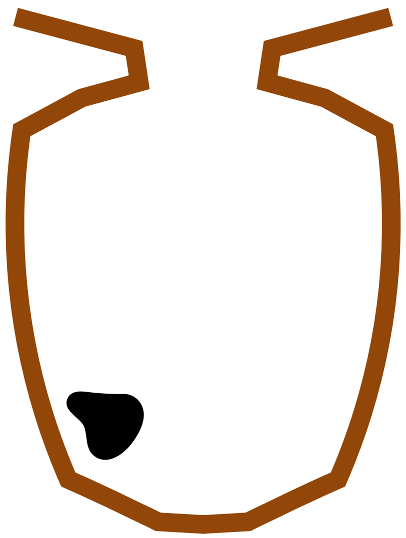
$$\rho_2 = (1 - V_1)V_2$$

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

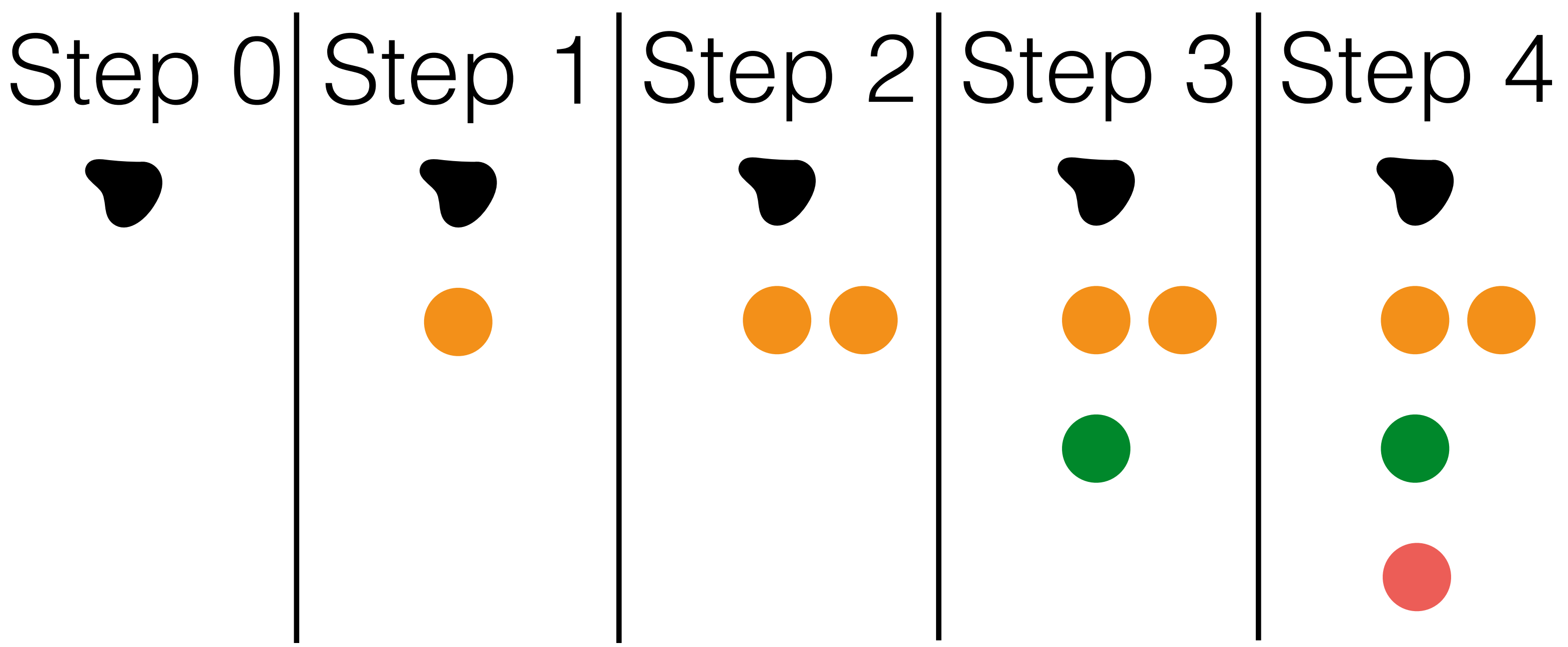
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



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



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

$$\rho_1 = V_1$$

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

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

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

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

Chinese restaurant process

Chinese restaurant process

- Same thing we just did

Chinese restaurant process

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

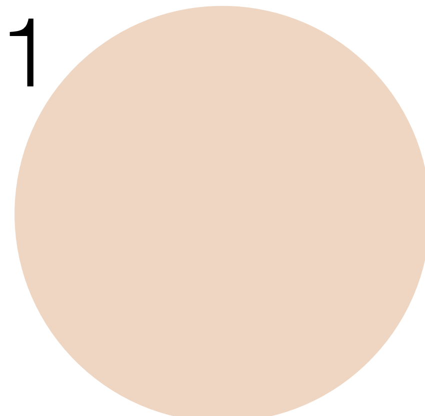
Chinese restaurant process

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

Chinese restaurant process

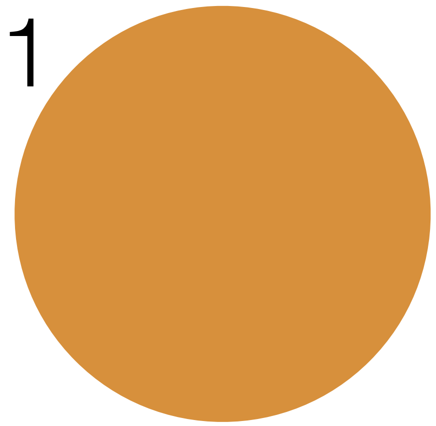
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



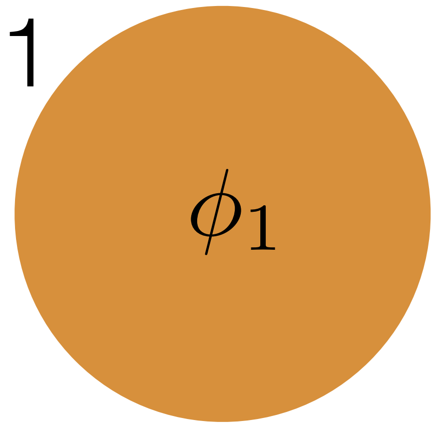
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

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

Chinese restaurant process

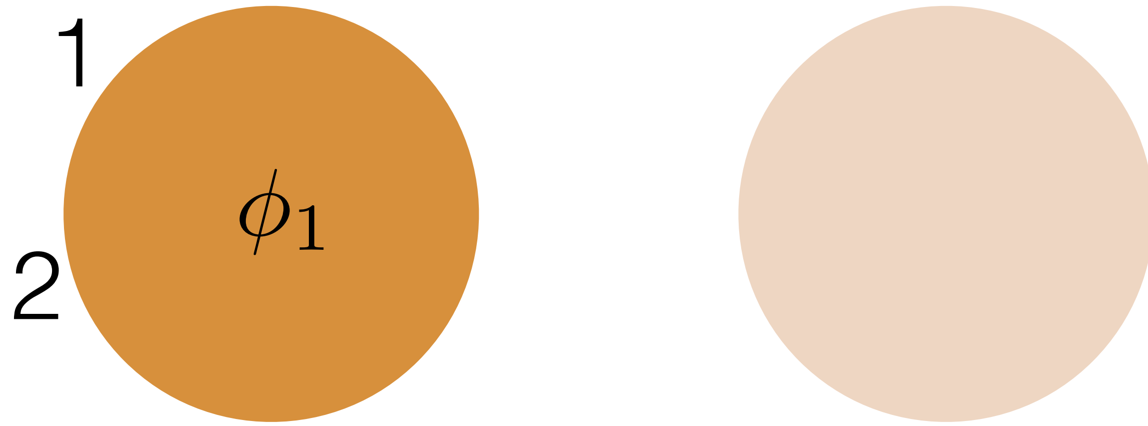


- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process

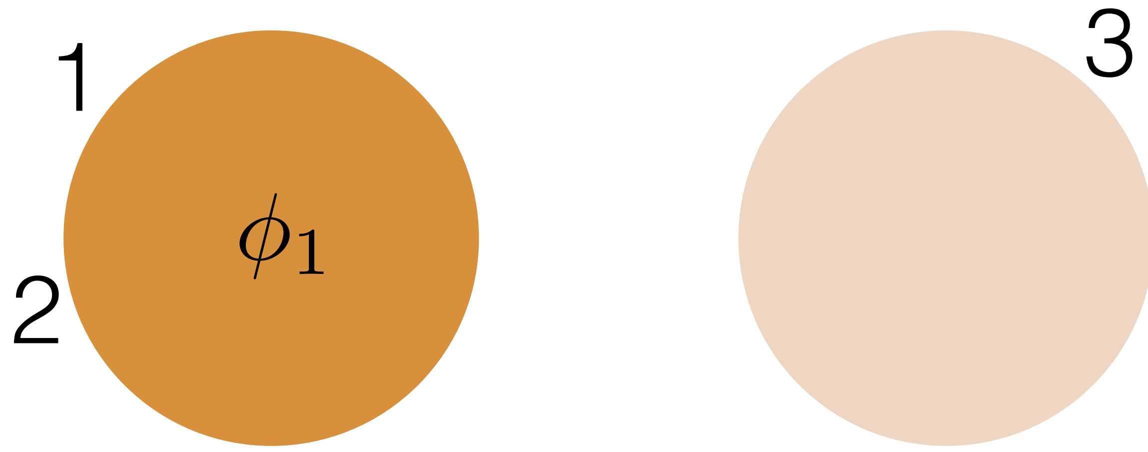
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

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

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

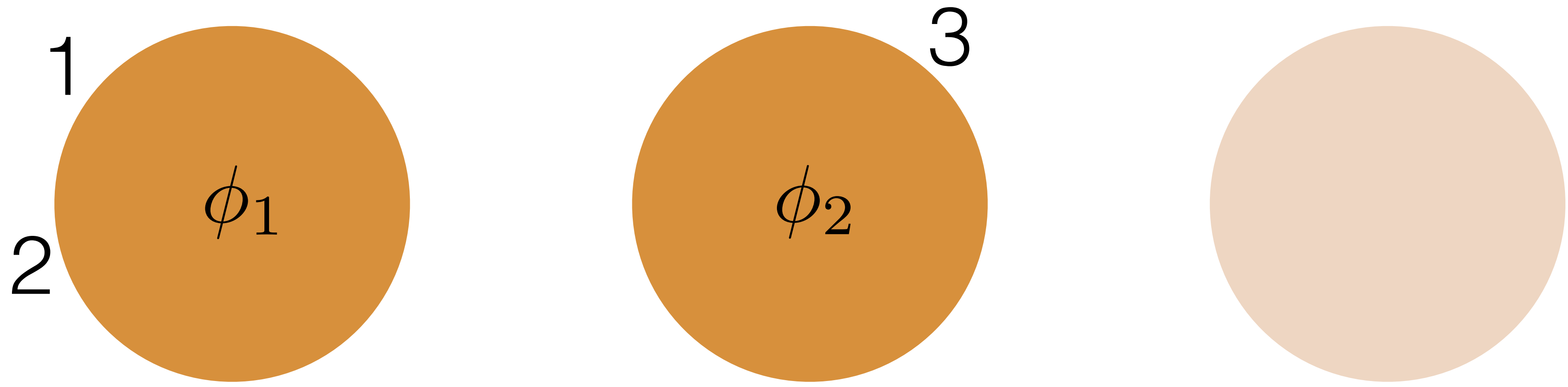
Chinese restaurant process

- 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

Chinese restaurant process

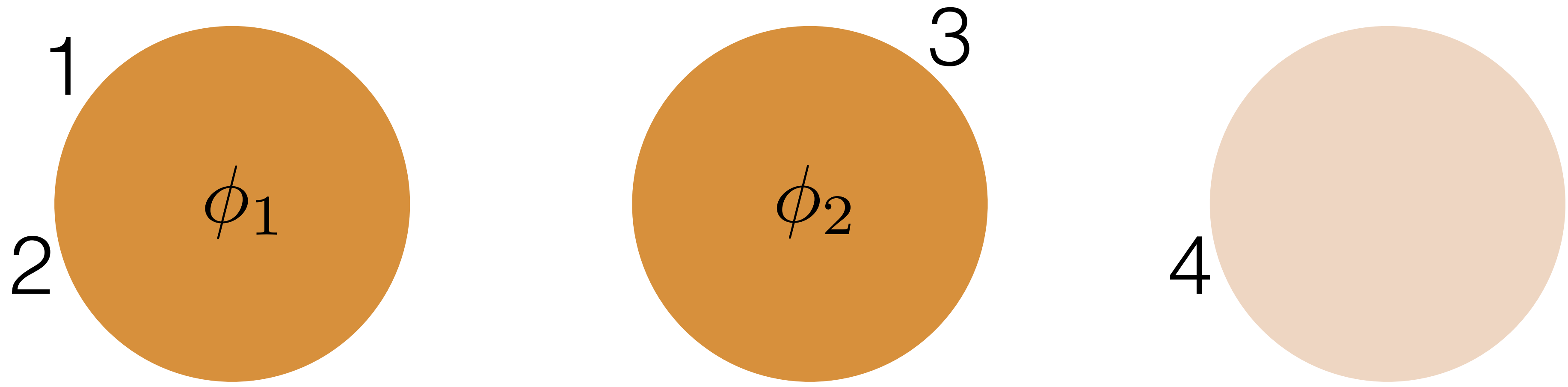
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



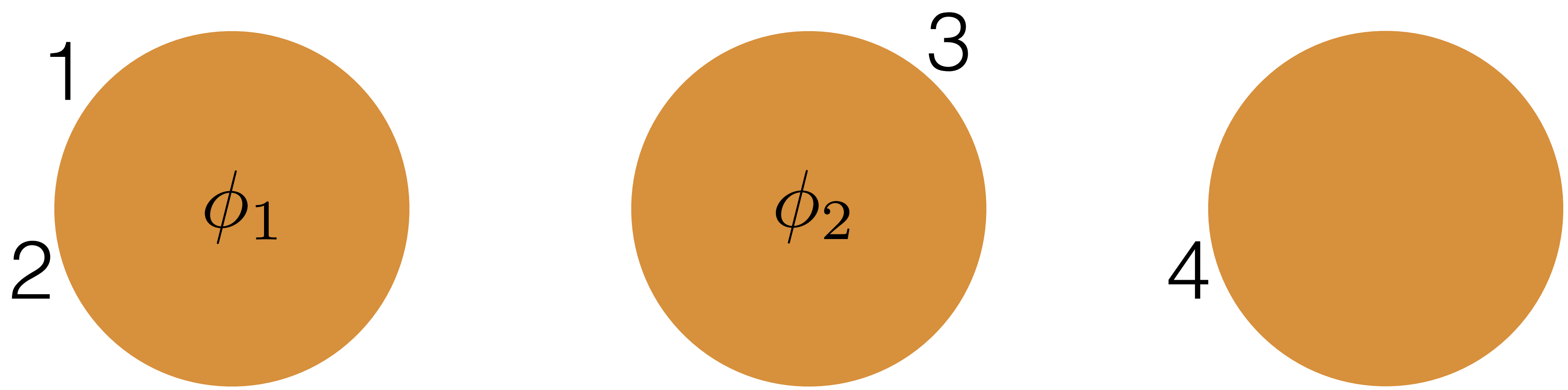
- Same thing we just did
- Each customer walks into the restaurant
    - Sits at existing table with prob proportional to # people there
    - Forms new table with prob proportional to α

Chinese restaurant process



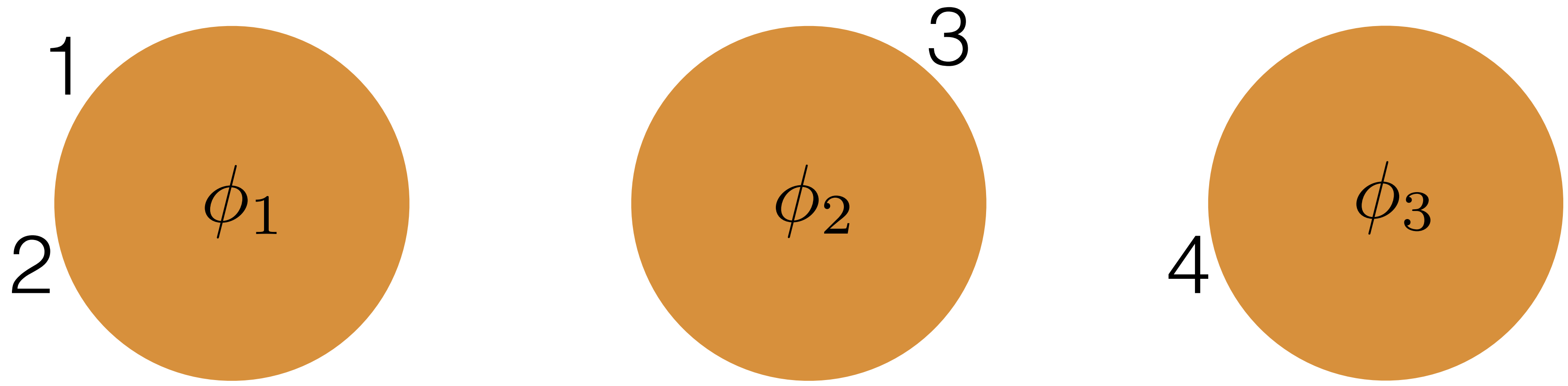
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



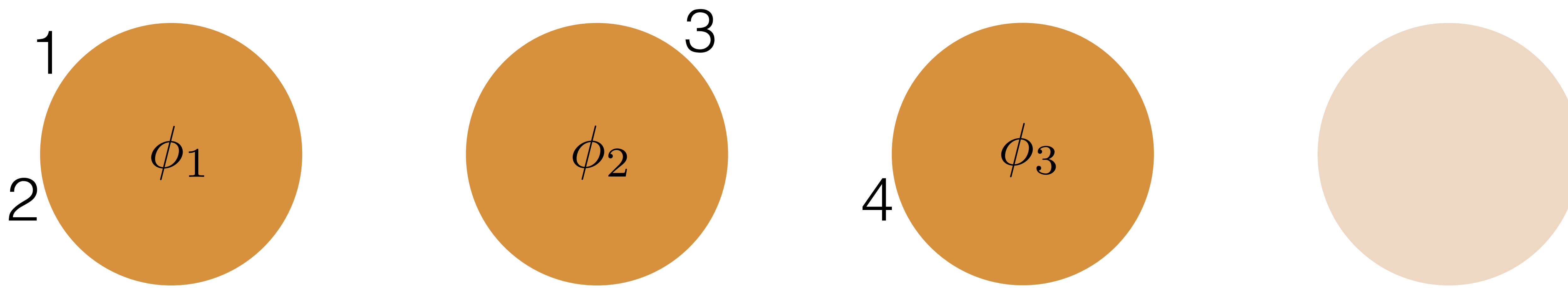
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



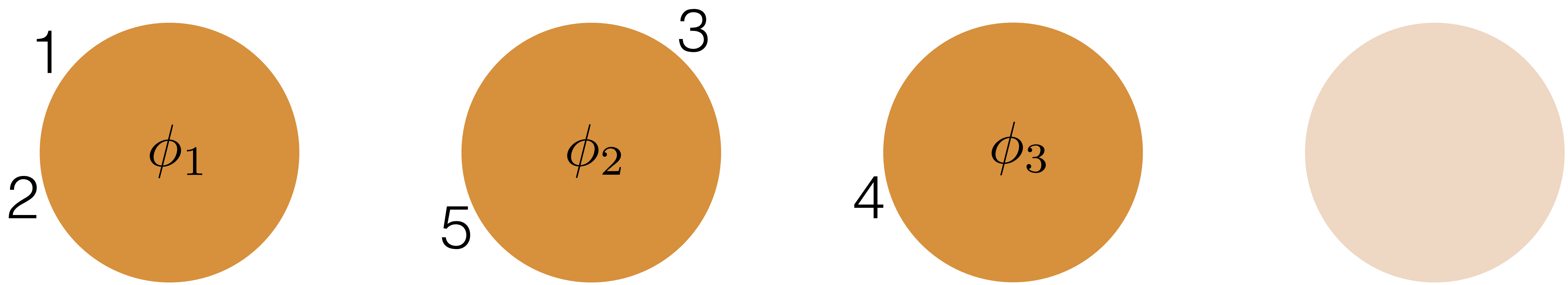
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



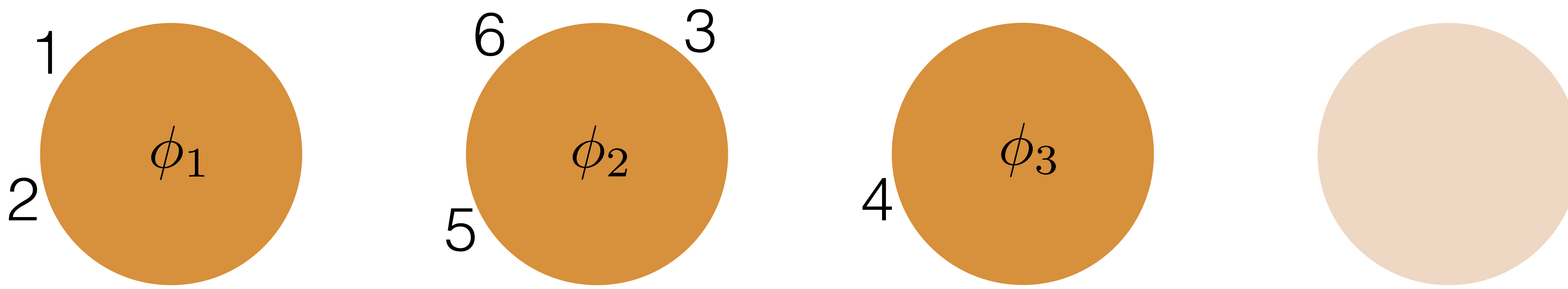
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



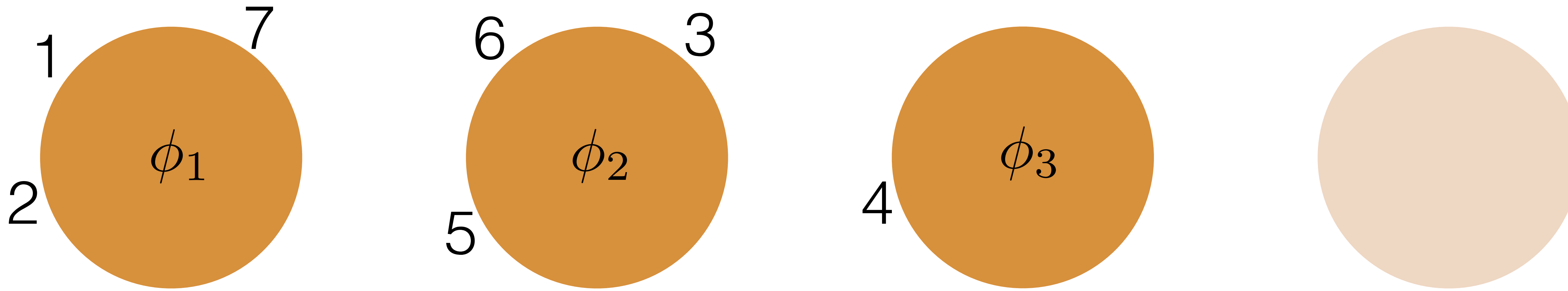
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



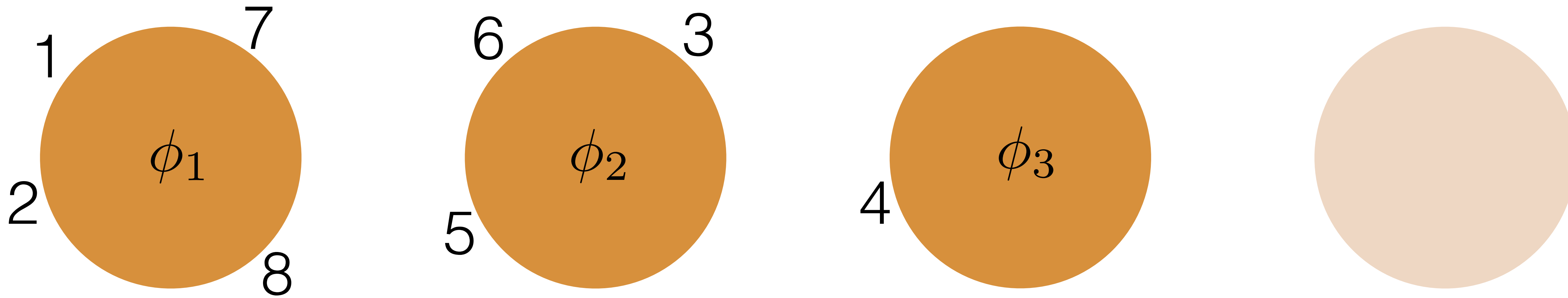
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



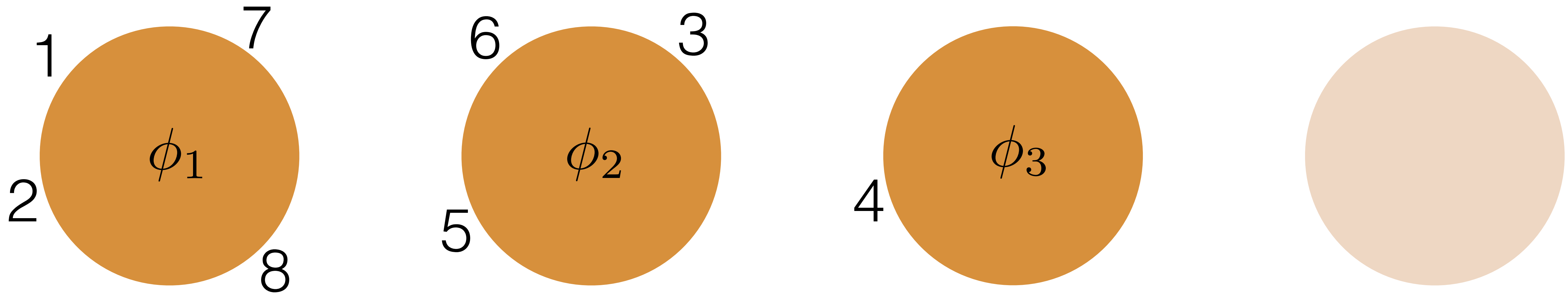
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

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

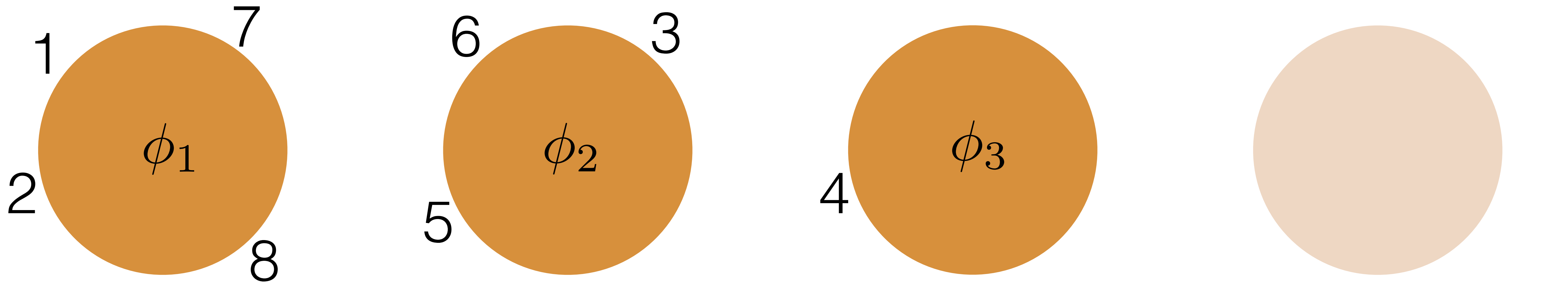
Chinese restaurant process



- 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

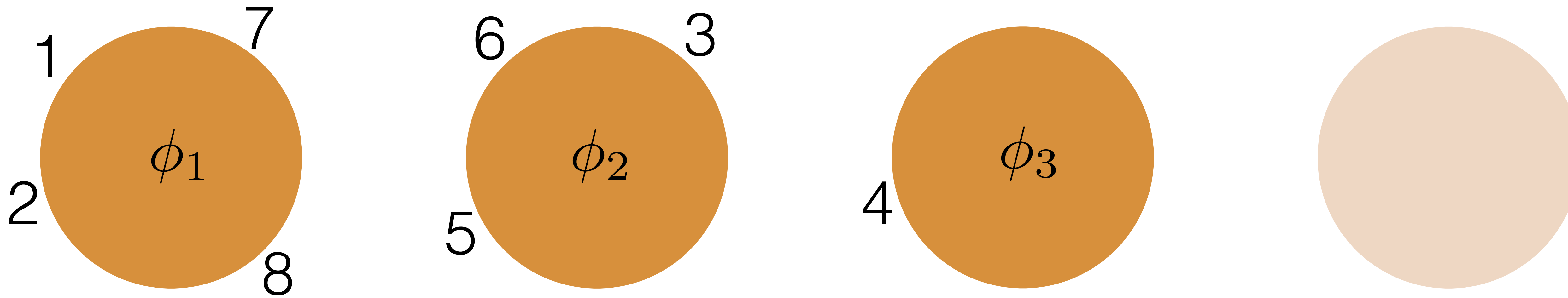
[Aldous 1983]

Chinese restaurant process



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

Chinese restaurant process

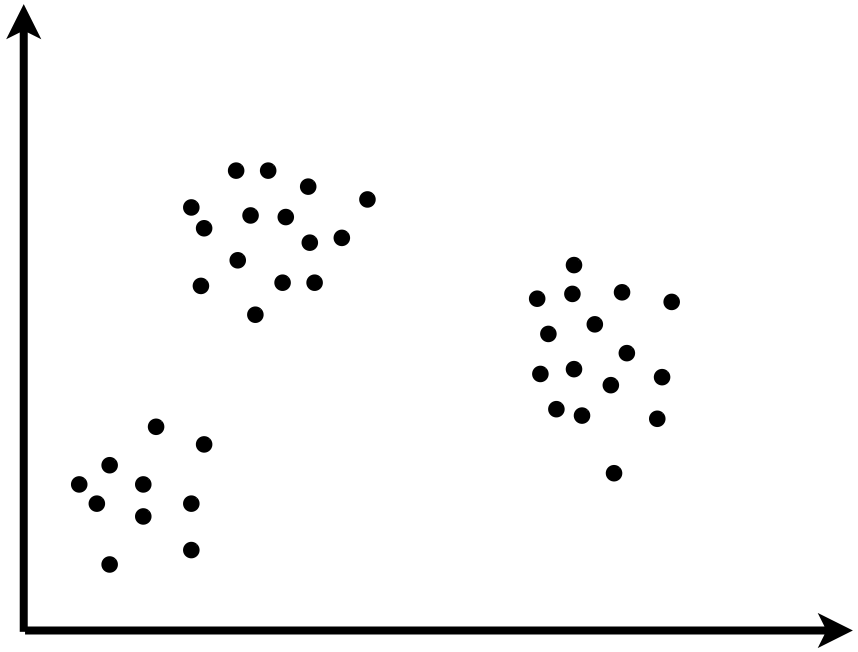


- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior

We've seen: Dirichlet process, Chinese restaurant process

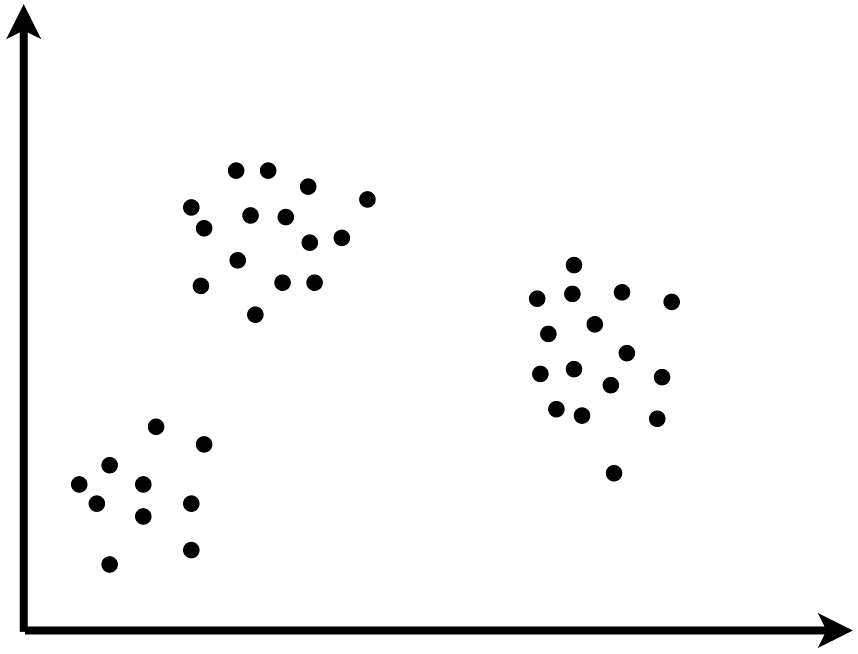
- Infinity of parameters (components)
- Growing number of parameters (clusters)

Inference



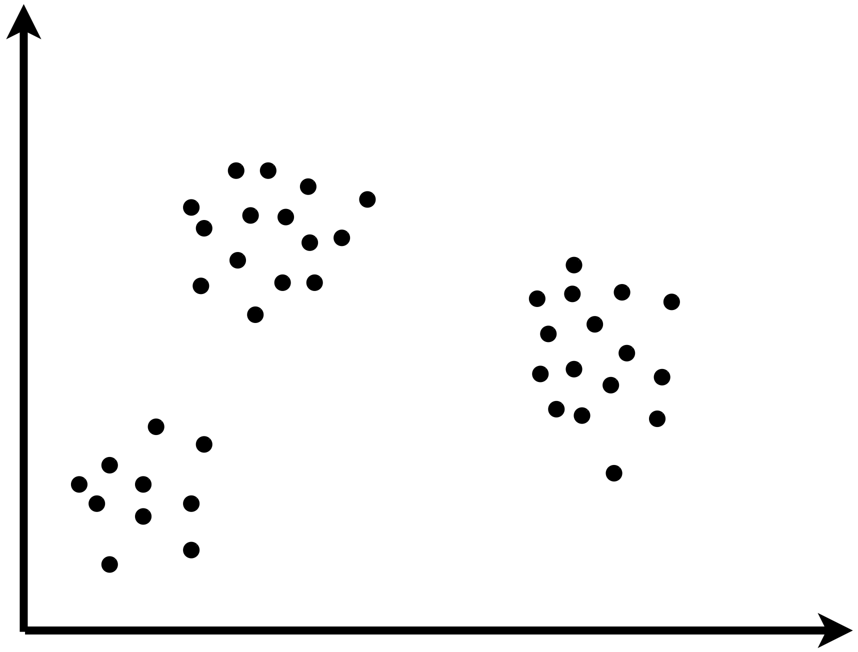
Inference

- DPMM



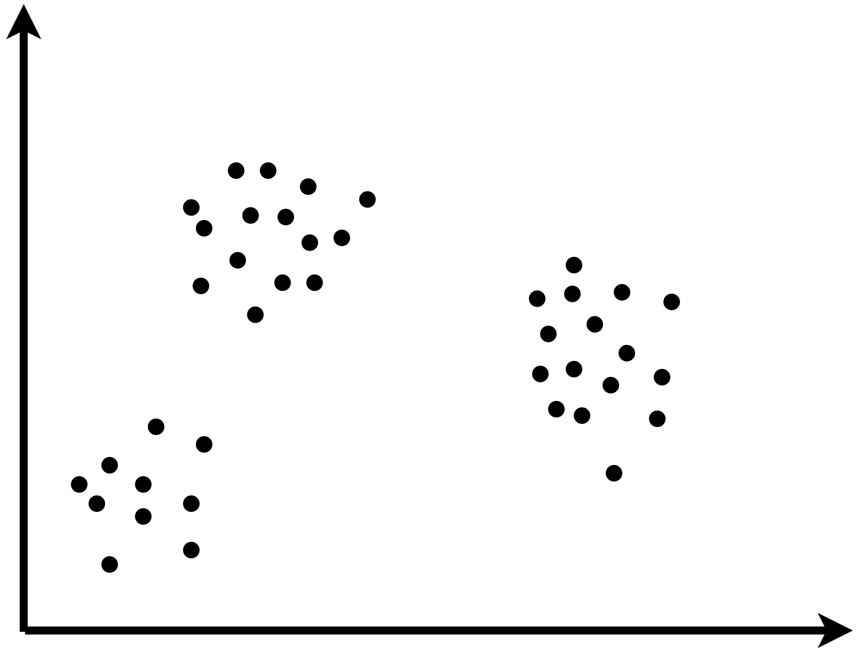
Inference

- DPMM; goal is a posterior over:



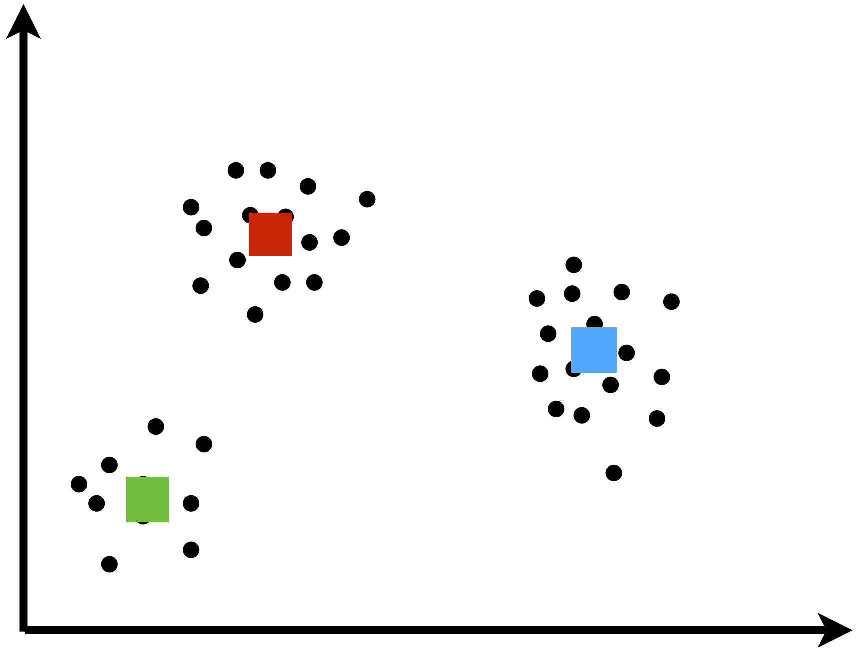
Inference

- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)



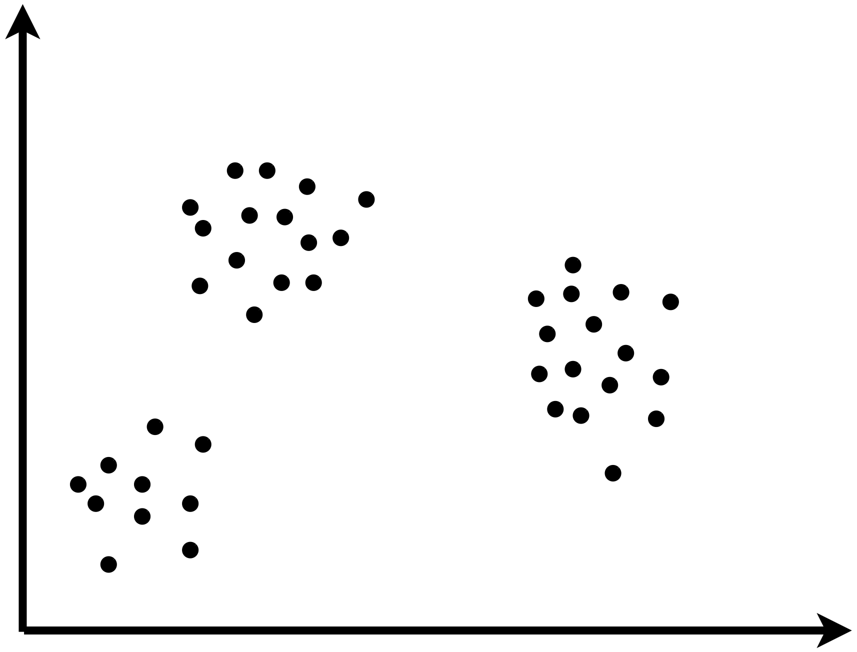
Inference

- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)



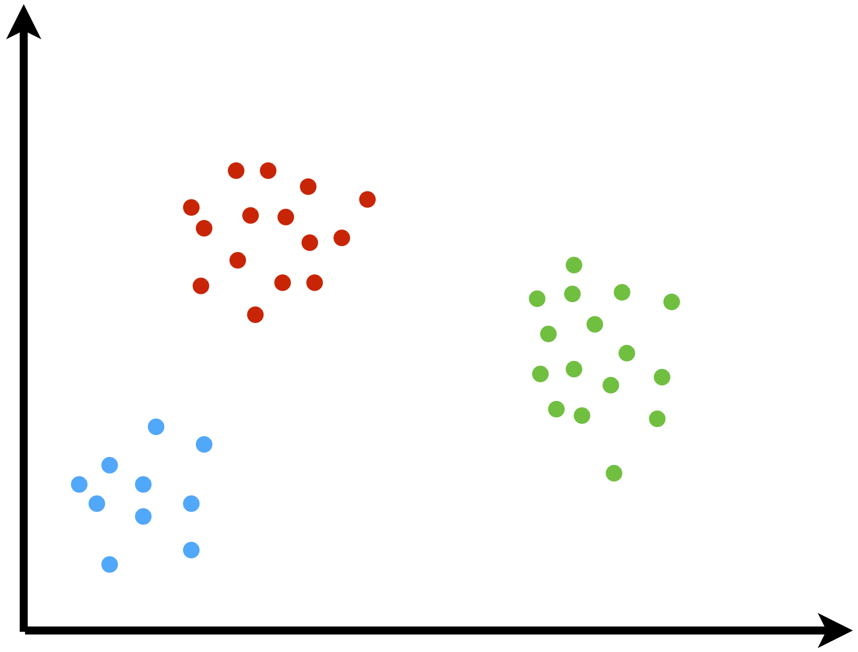
Inference

- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters

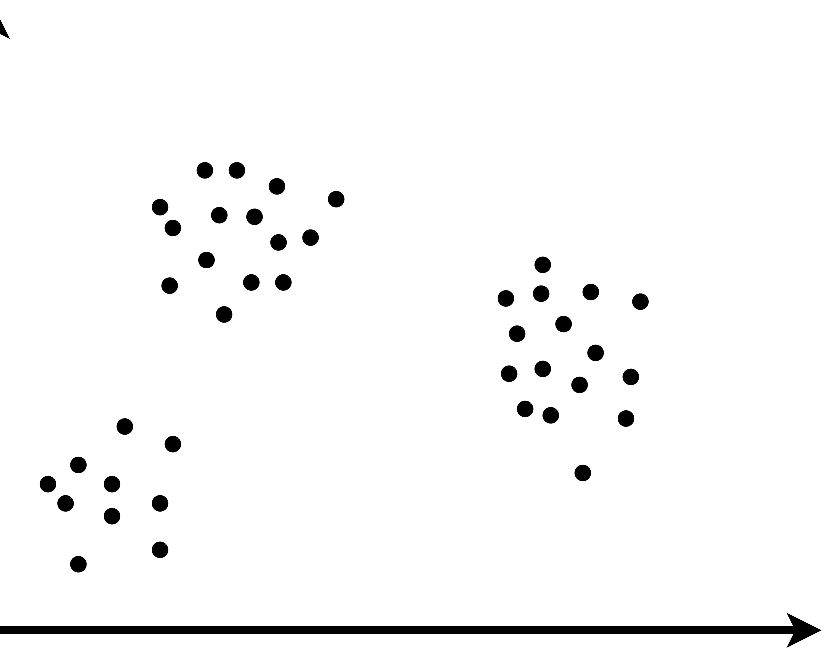


Inference

- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters

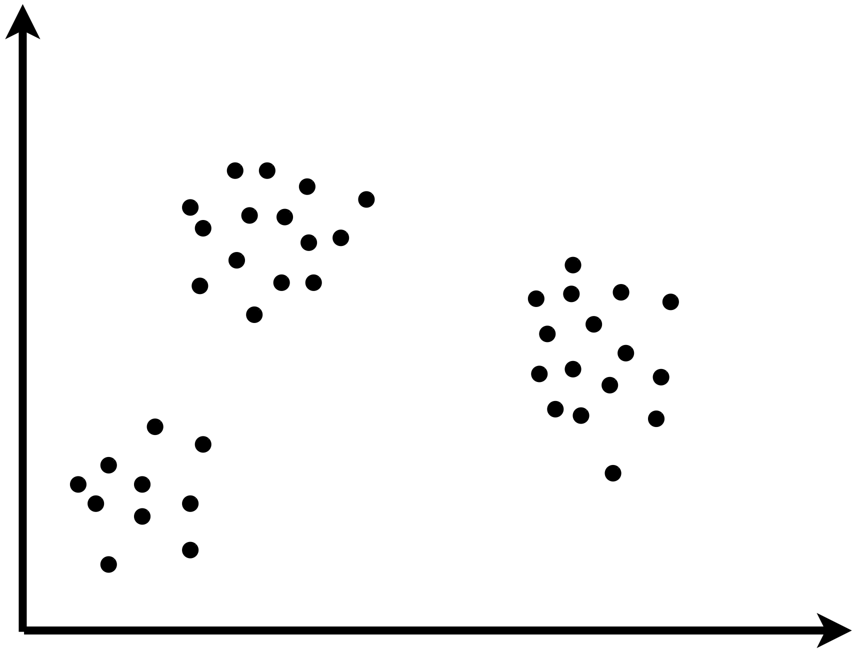


Inference



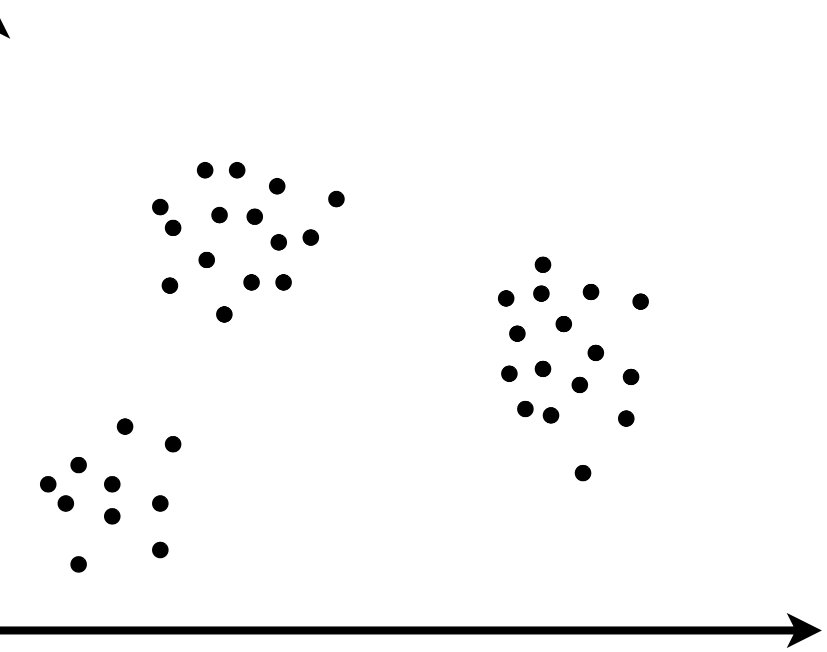
- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods

Inference



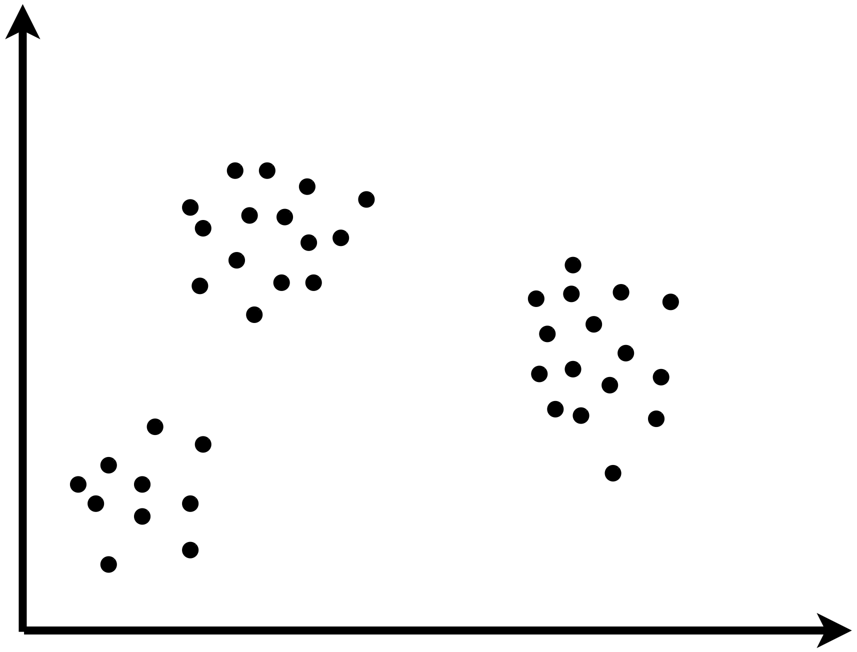
- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo

Inference



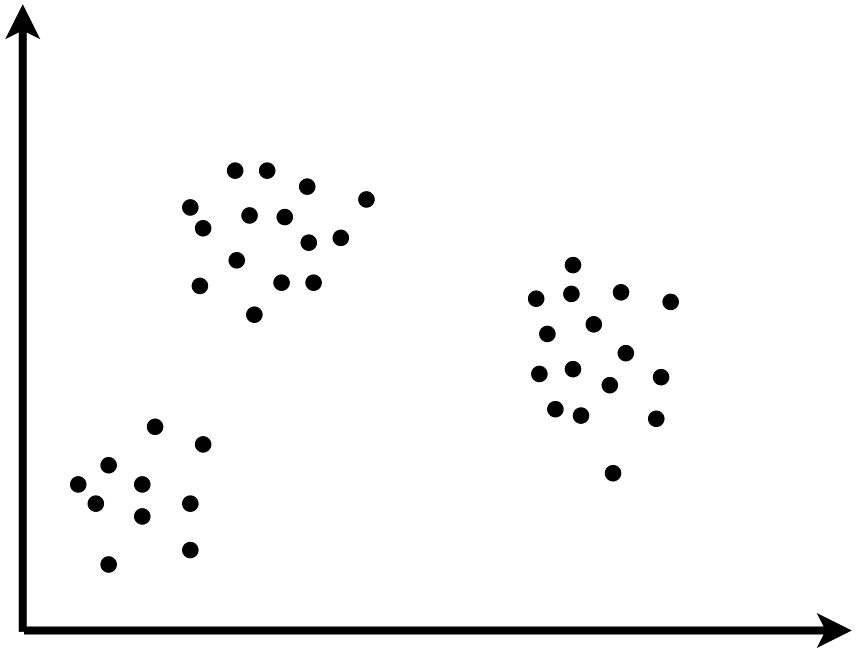
- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo
  - Variational Bayes

Inference



- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo
  - Variational Bayes
- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$

Inference

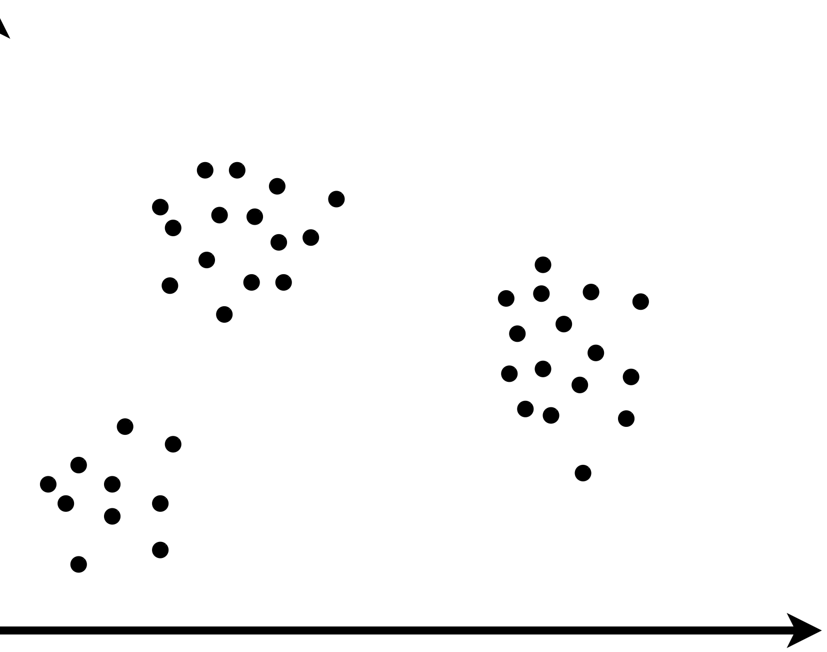


- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo
  - Variational Bayes
- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - t th step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

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

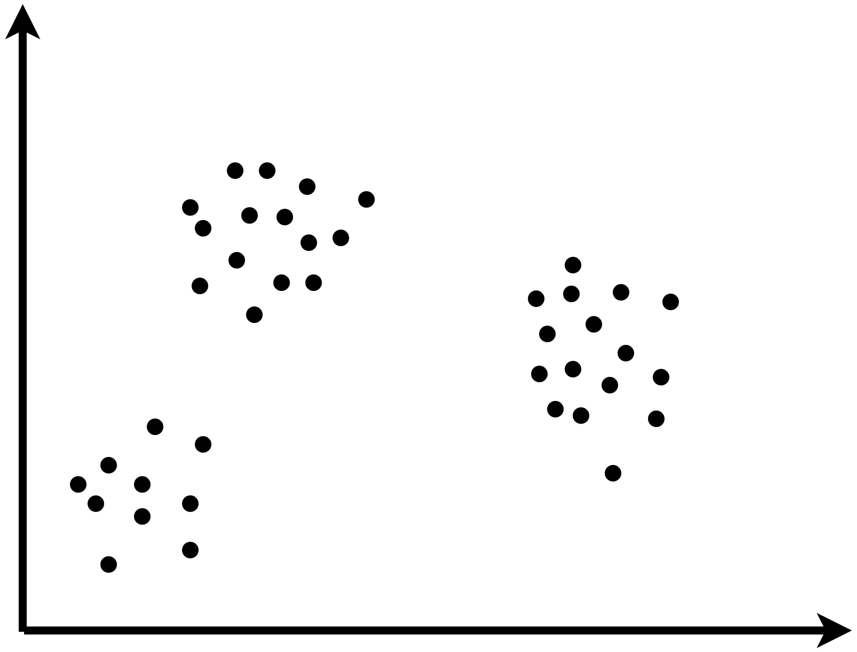
$$v_3^{(t)} \sim p(v_3 | v_1^{(t)}, v_2^{(t)})$$

Inference



- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo
  - Variational Bayes
- 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: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$ $\quad v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$ $\quad v_3^{(t)} \sim p(v_3 | v_1^{(t)}, v_2^{(t)})$
- Gibbs with the CRP

Inference



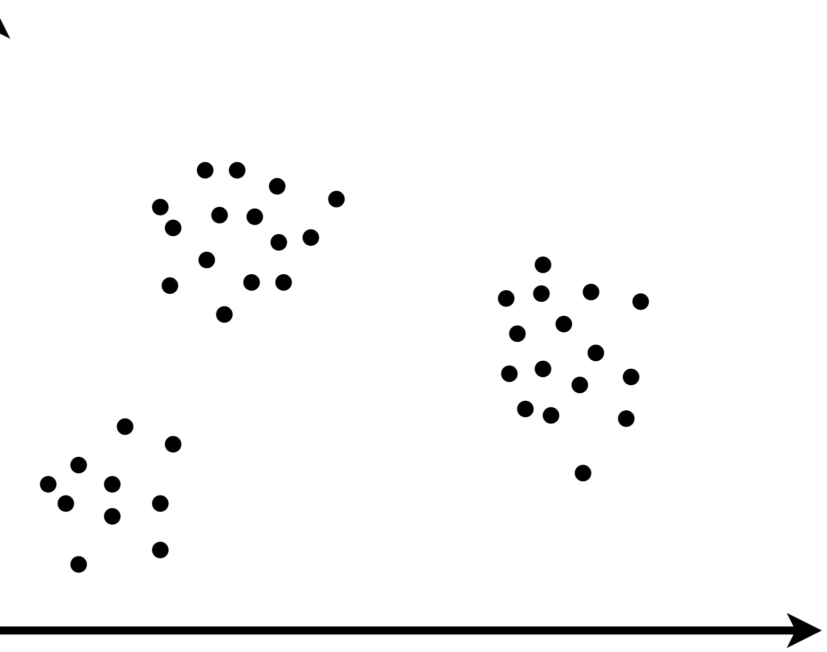
- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo
  - Variational Bayes
- Gibbs sampling review: target distribution $p(v_1, v_2, v_3)$
  - Start: $v_1^{(0)}, v_2^{(0)}, v_3^{(0)}$
  - t th step: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

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

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

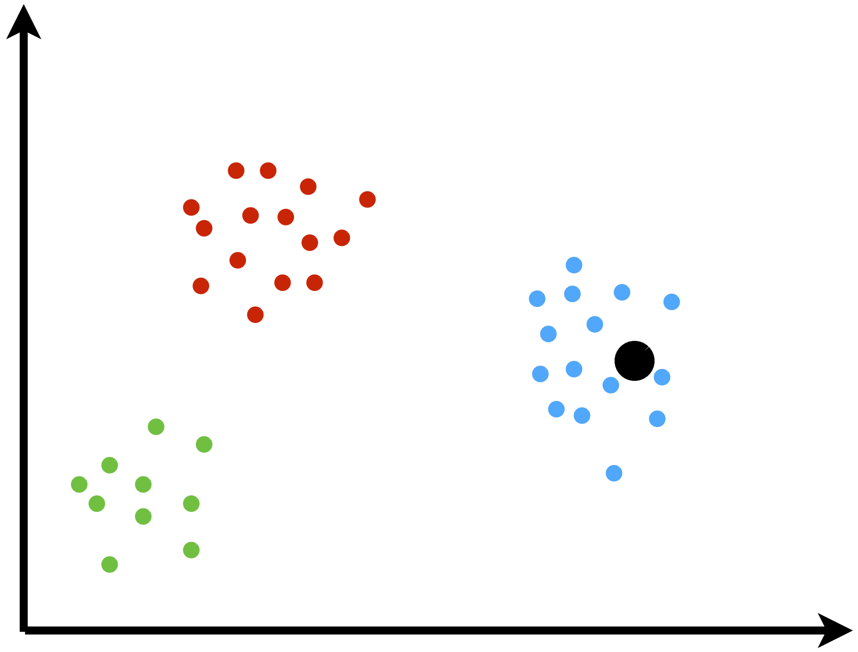
- Gibbs with the CRP: use exchangeability

Inference



- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo
  - Variational Bayes
- 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: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$ $\quad v_2^{(t)} \sim p(v_2 | v_1^{(t)}, v_3^{(t-1)})$ $\quad v_3^{(t)} \sim p(v_3 | v_1^{(t)}, v_2^{(t)})$
- Gibbs with the CRP: use exchangeability
  - Treat every data point like the final customer

Inference



- DPMM; goal is a posterior over:
  - Cluster parameters (center, shape)
  - Assignments of data points to clusters
- Two principal posterior approximation methods
  - Markov Chain Monte Carlo
  - Variational Bayes
- 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: $v_1^{(t)} \sim p(v_1 | v_2^{(t-1)}, v_3^{(t-1)})$

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

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

- Gibbs with the CRP: use exchangeability
  - Treat every data point like the final customer

More Markov Chain Monte Carlo

More Markov Chain Monte Carlo

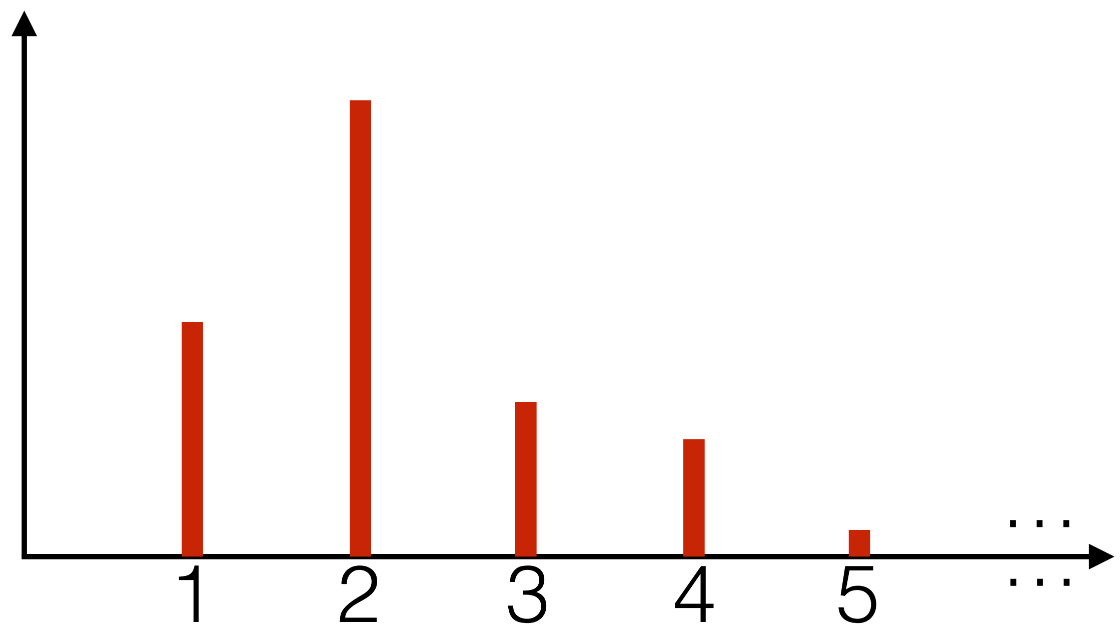
- Slice sampling

More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals

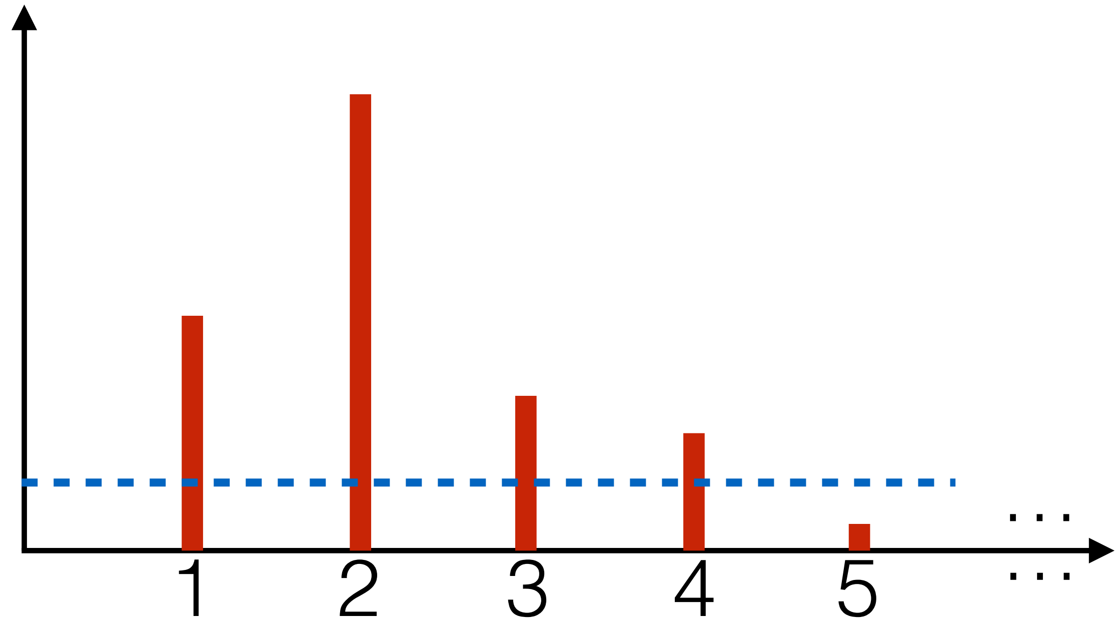
More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals



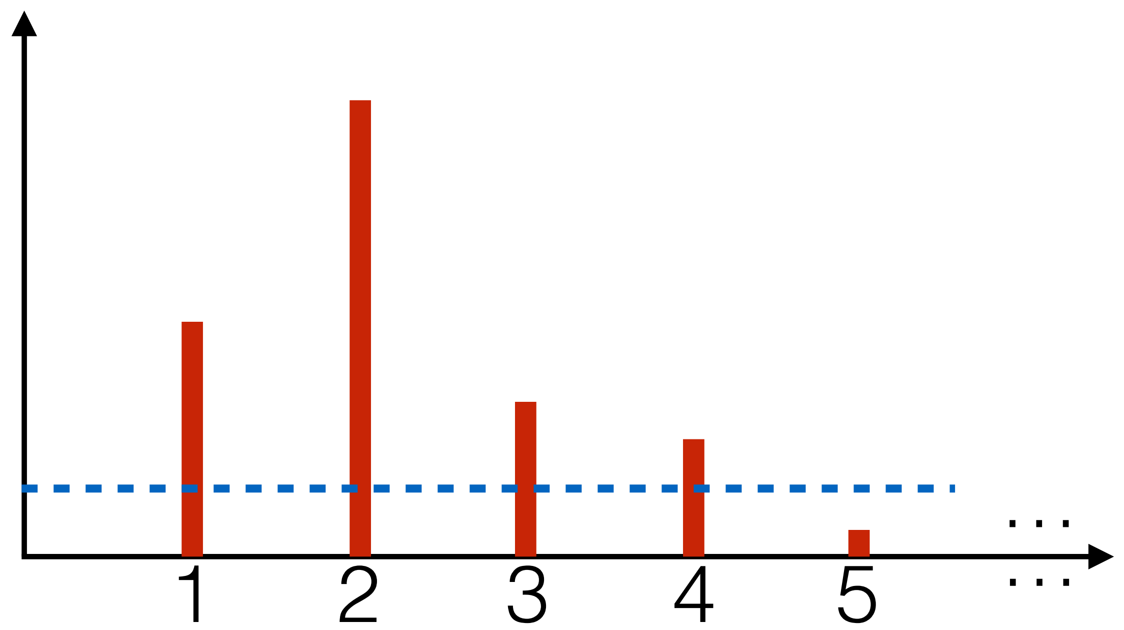
More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals



More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals

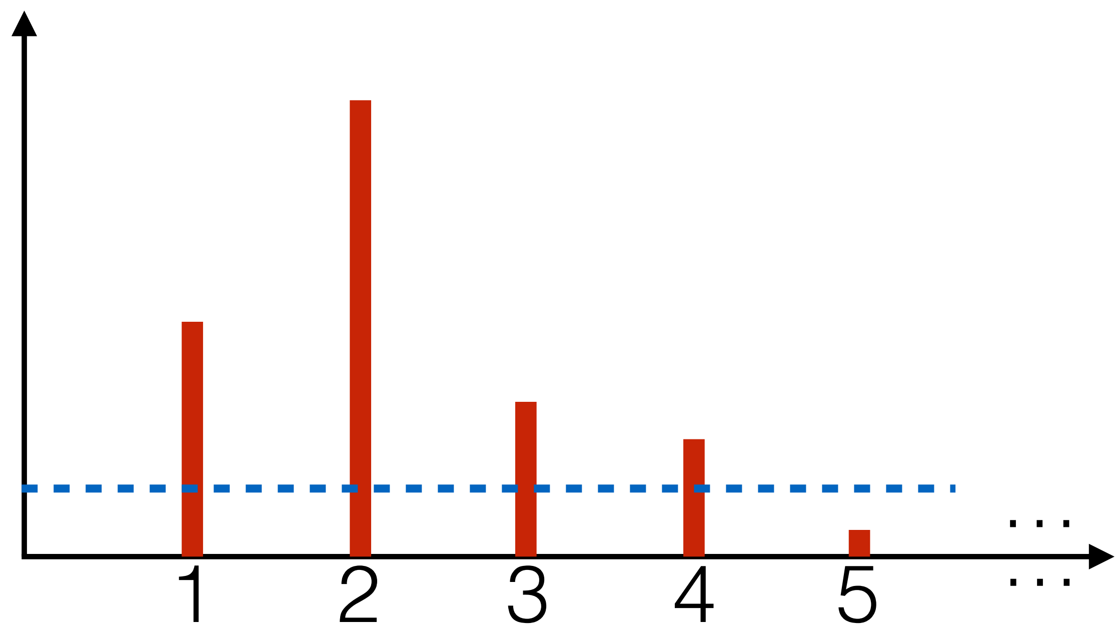


- Approximate with truncated distribution

[Ishwaran, James 2001; Campbell*, Huggins*, Broderick 2016]

More Markov Chain Monte Carlo

- Slice sampling
  - auxiliary variable ➔ finite conditionals

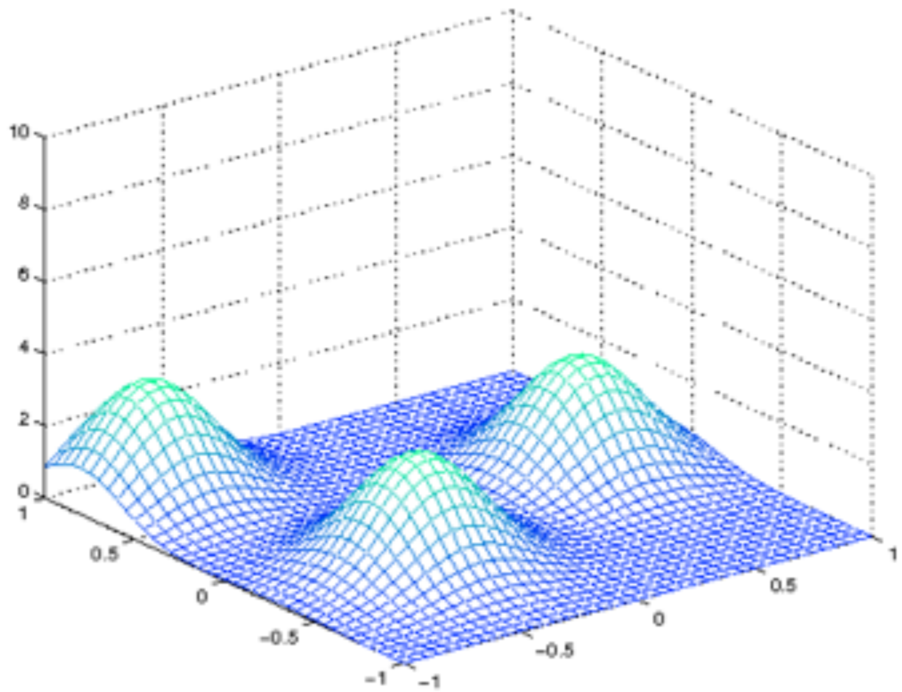


- Approximate with truncated distribution
  - E.g., Hamiltonian Monte Carlo

[Ishwaran, James 2001; Campbell*, Huggins*, Broderick 2016]

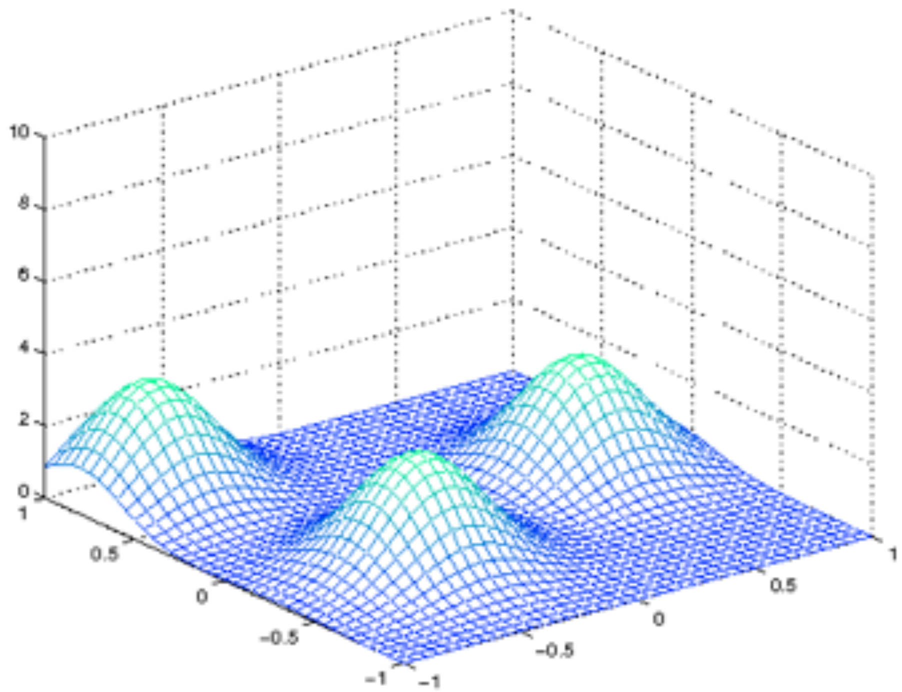
Variational Bayes

Variational Bayes



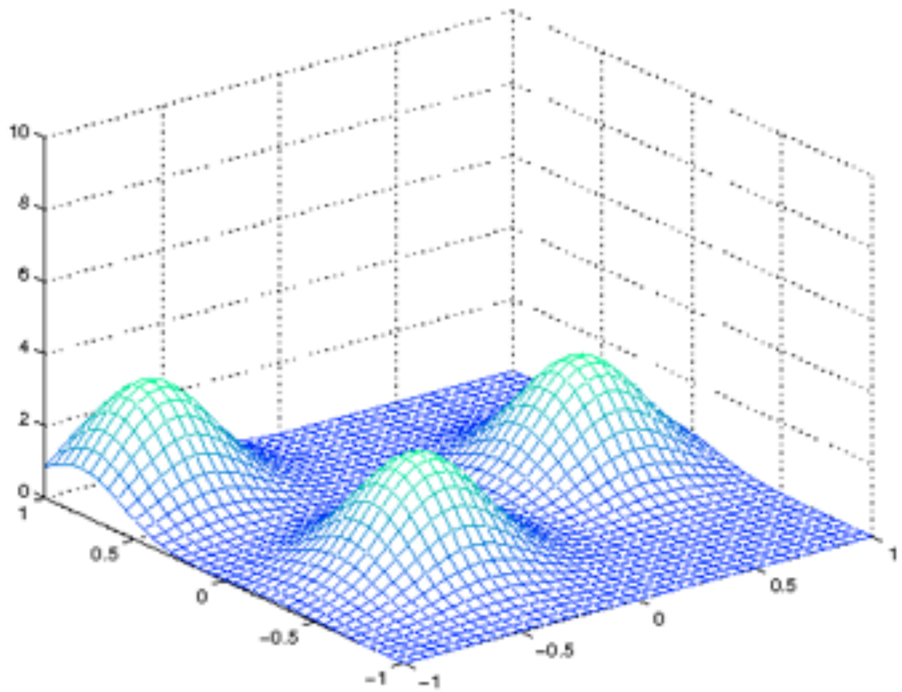
- Variational Bayes (VB)

Variational Bayes

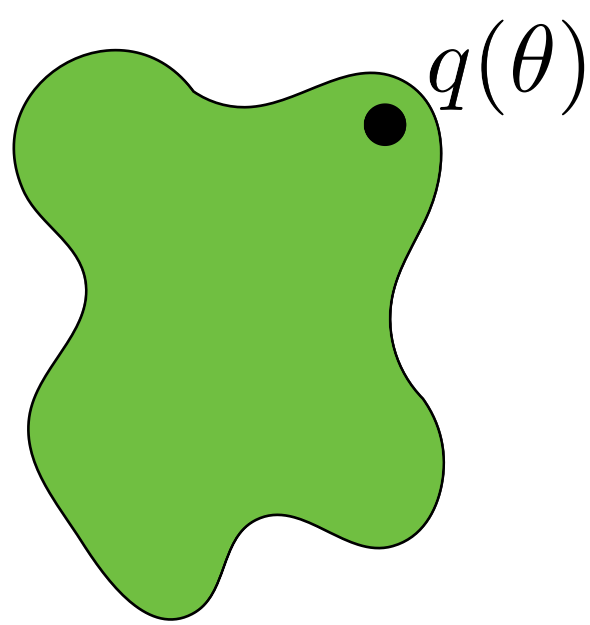


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$

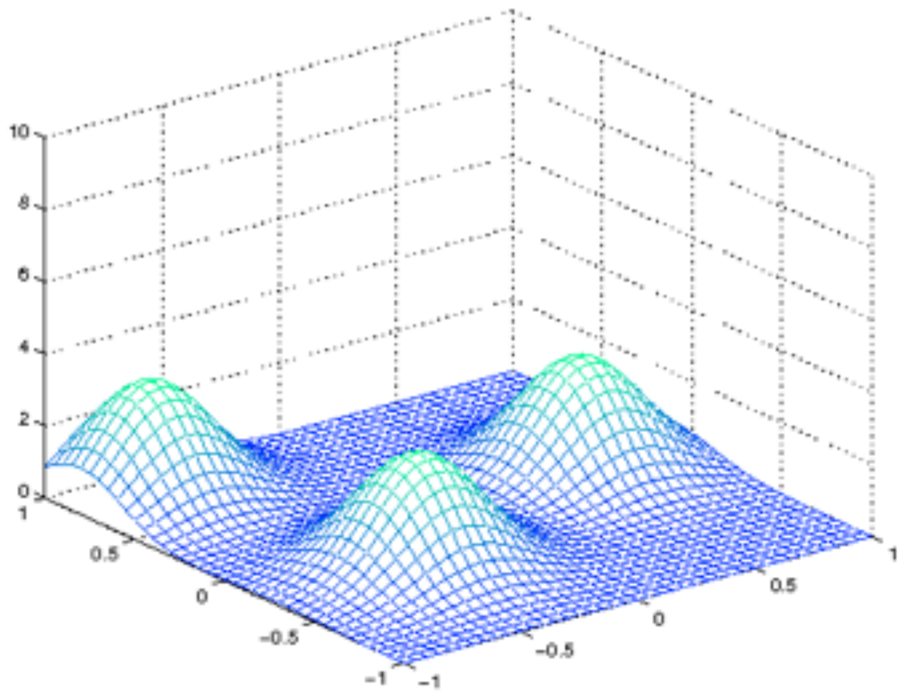
Variational Bayes

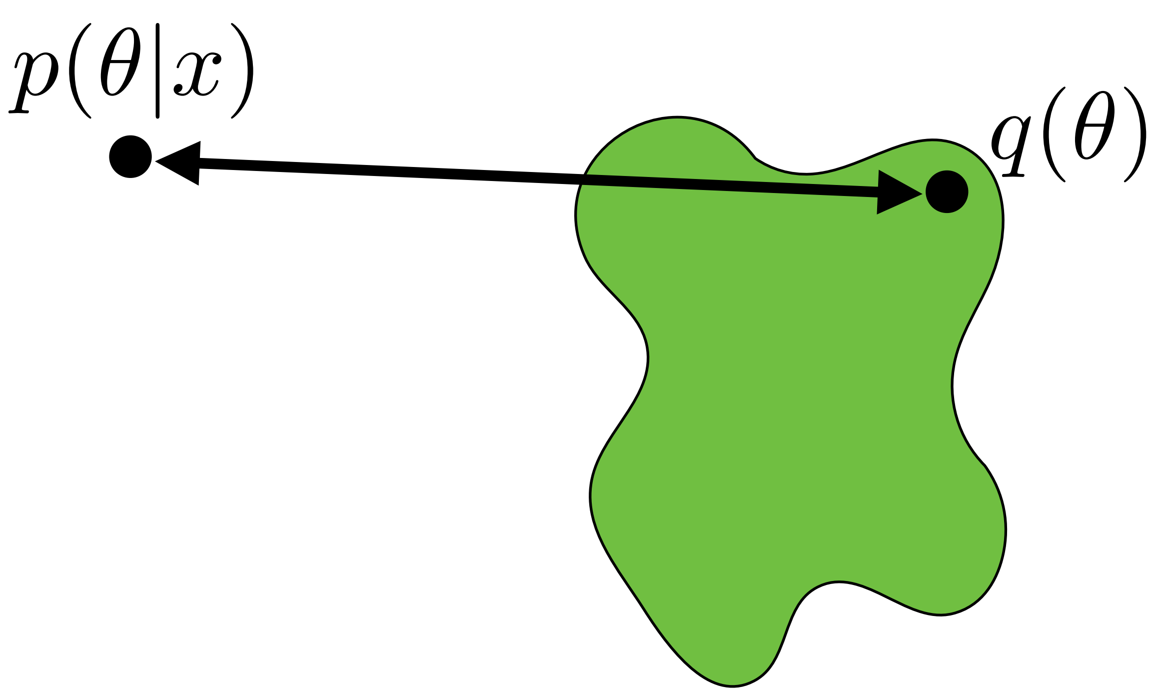


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$



Variational Bayes

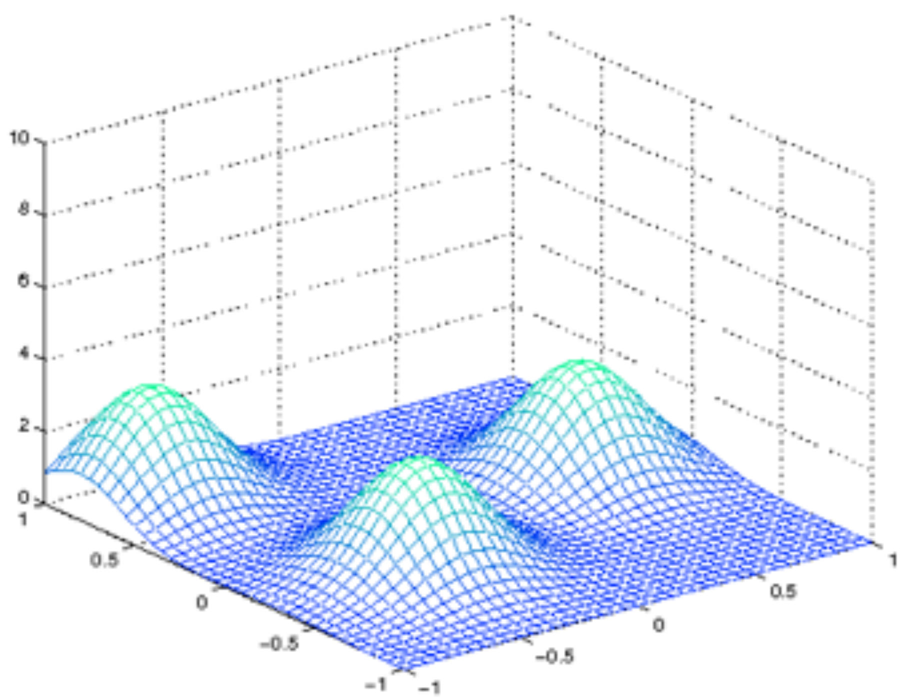


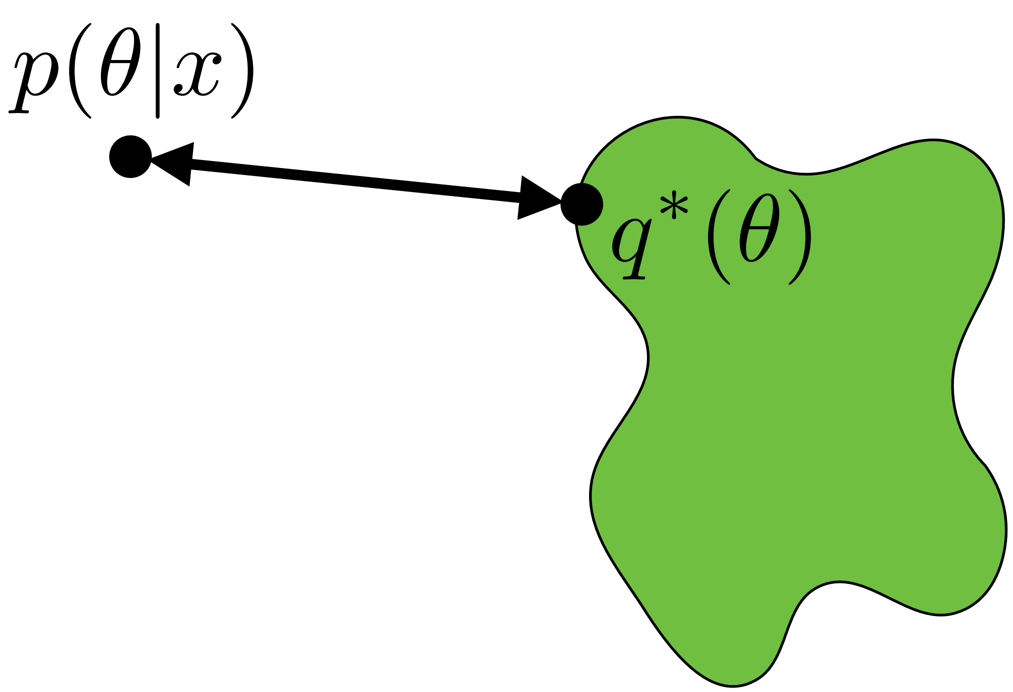


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$

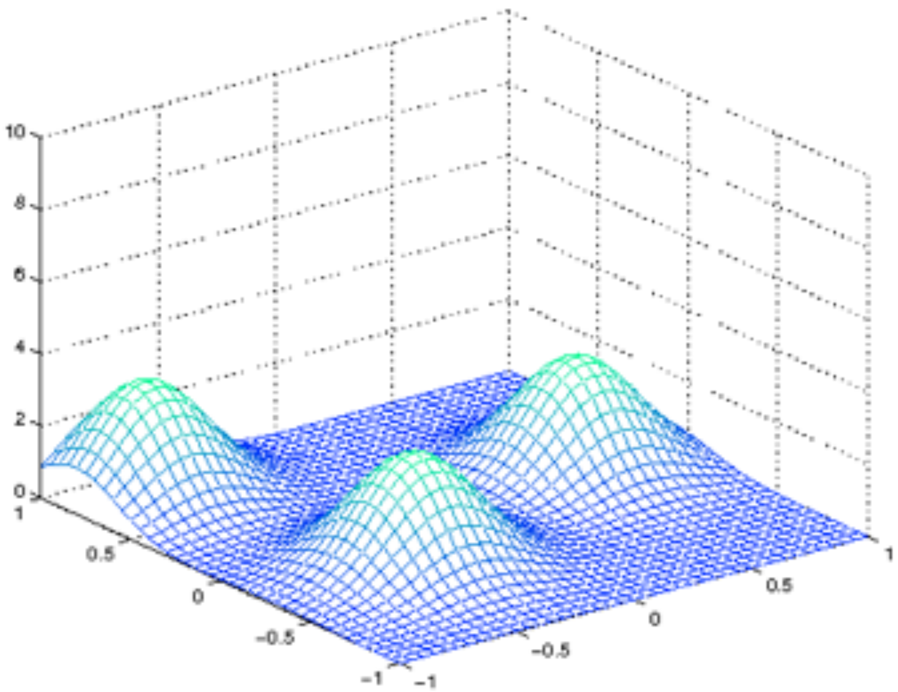
Variational Bayes

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



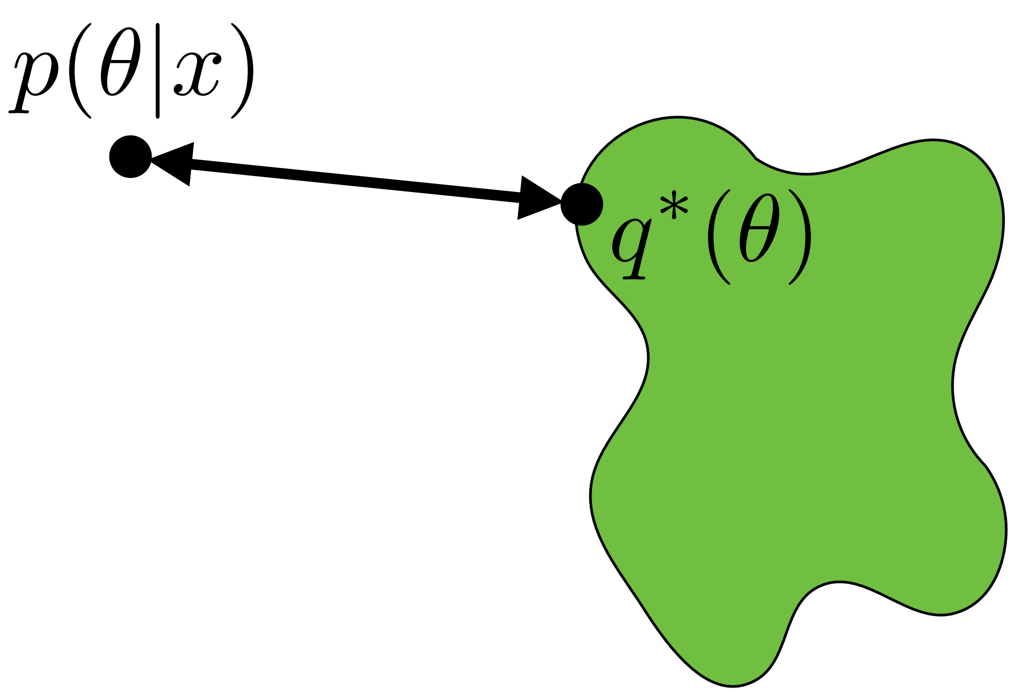


Variational Bayes

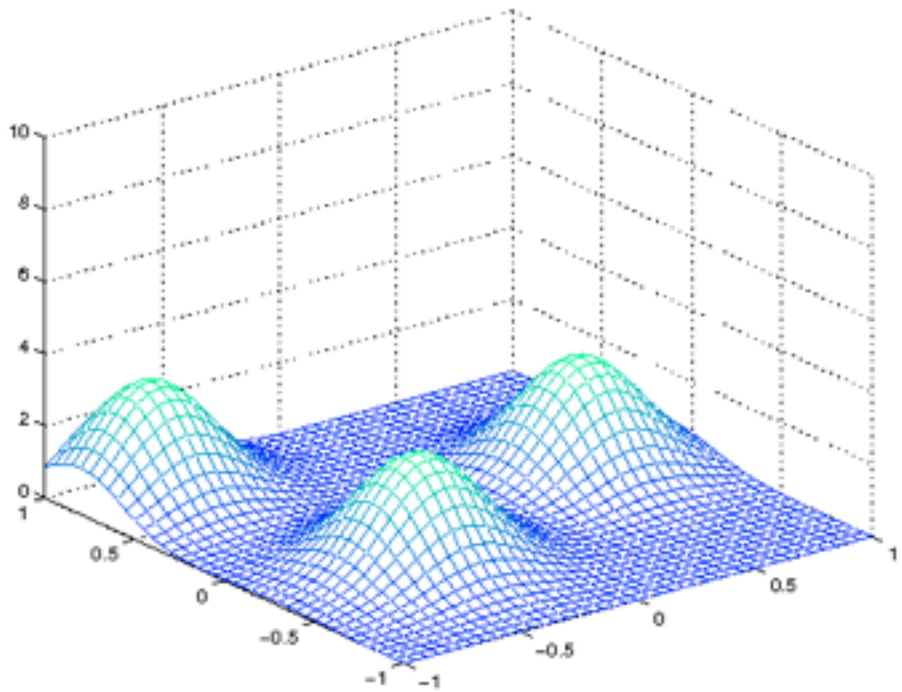


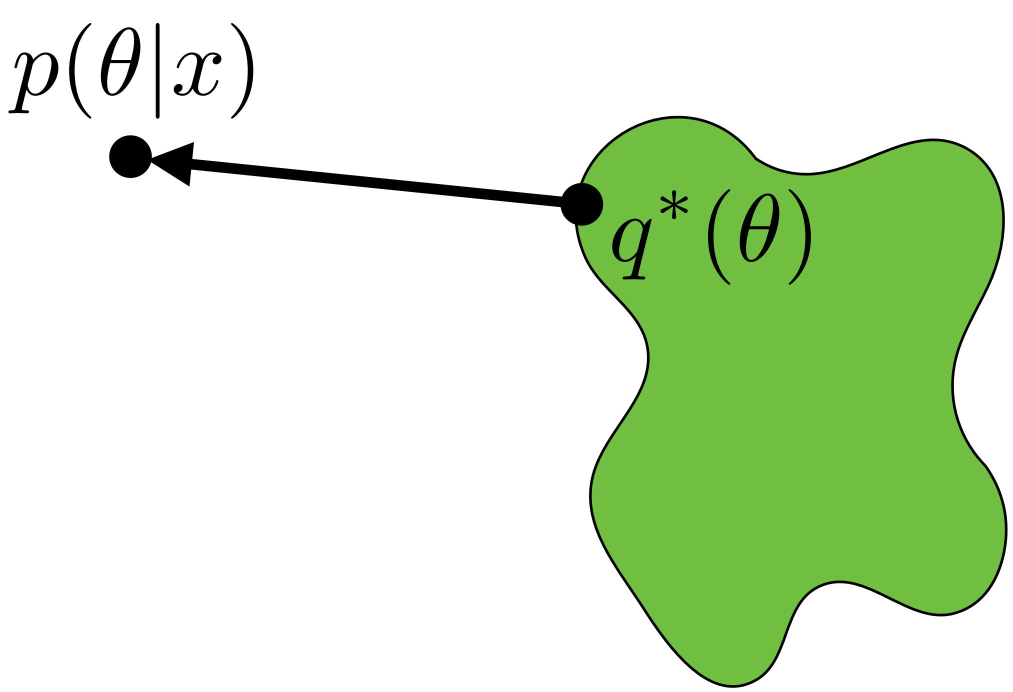
- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:

$$KL(q \| p(\cdot|x))$$



Variational Bayes

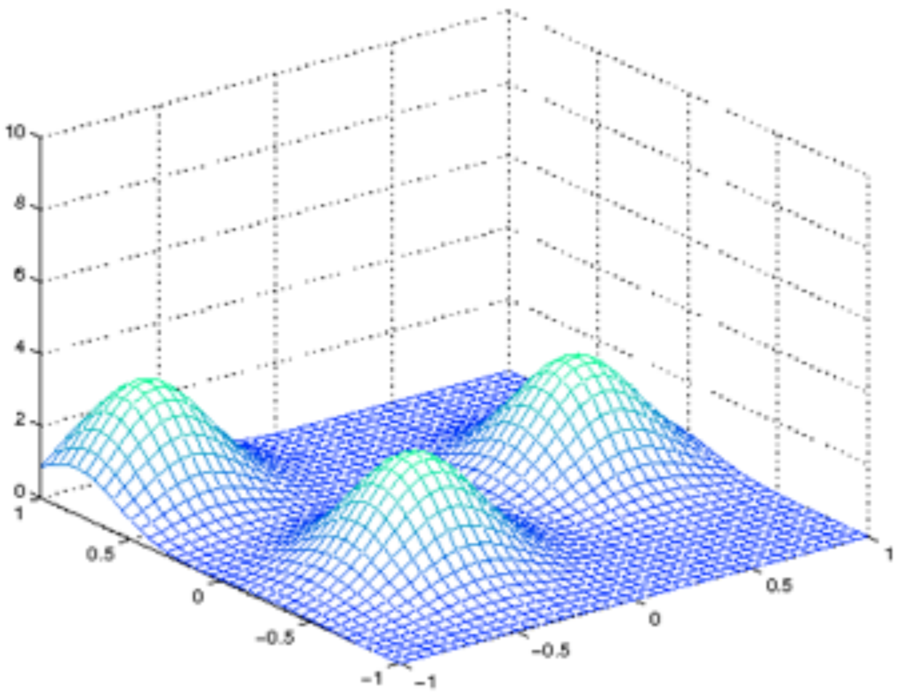


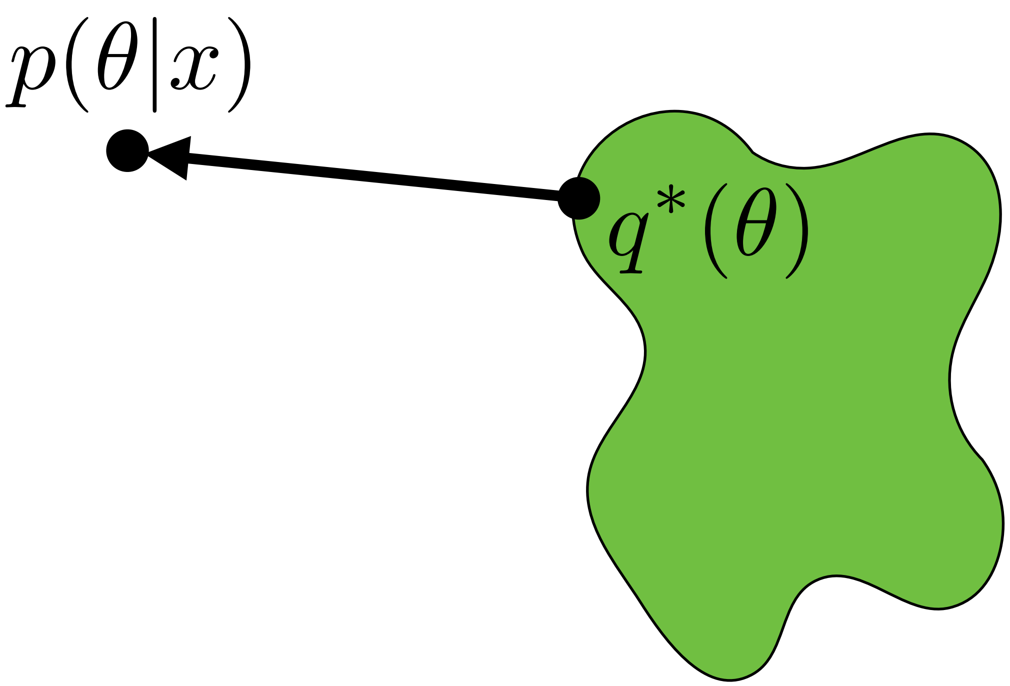


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:

$$KL(q\|p(\cdot|x))$$

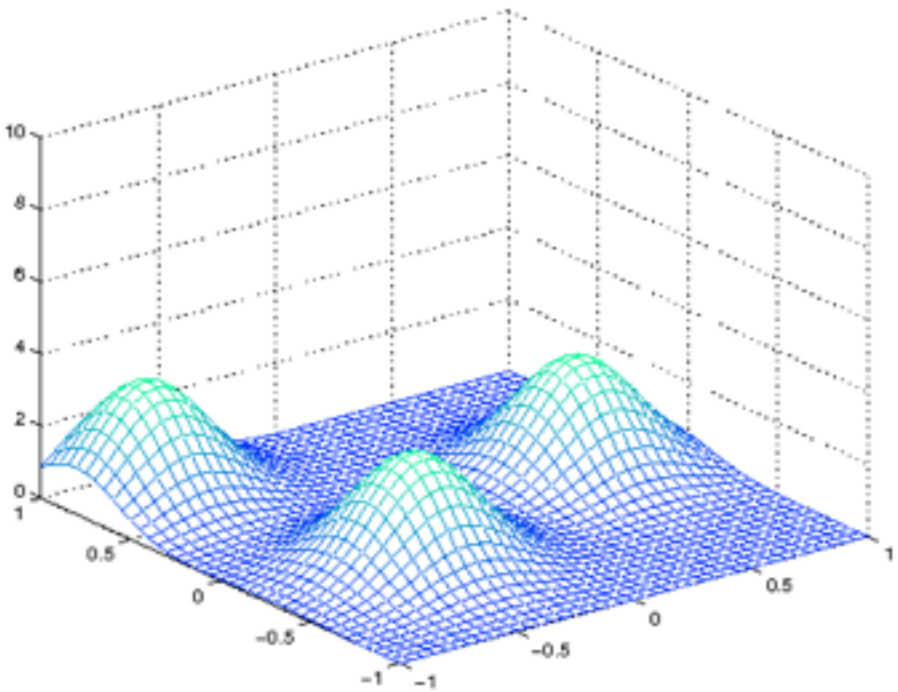
Variational Bayes

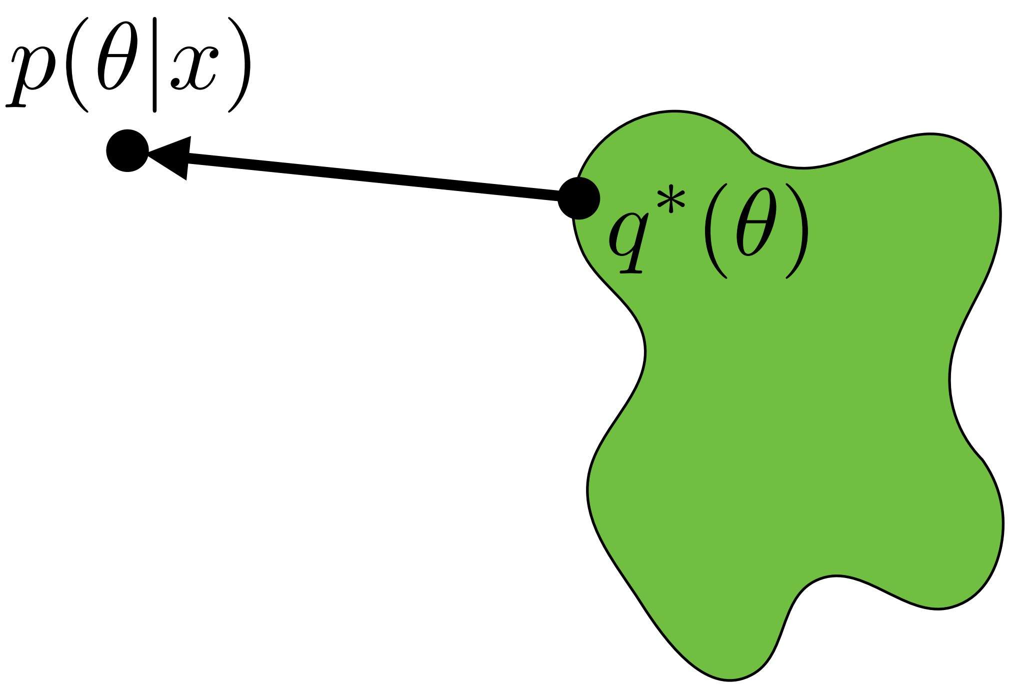




- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation

Variational Bayes

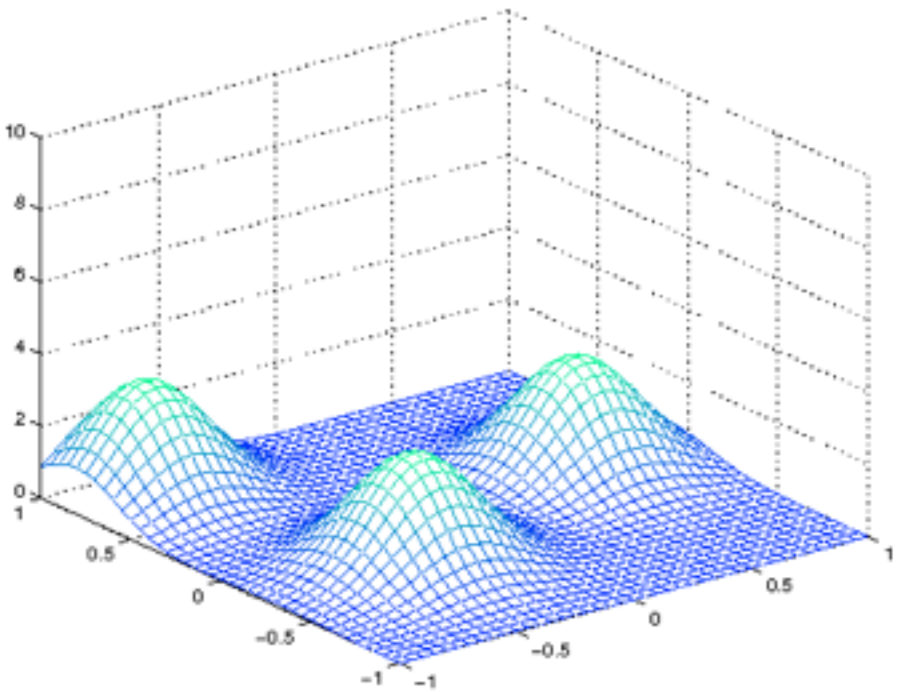


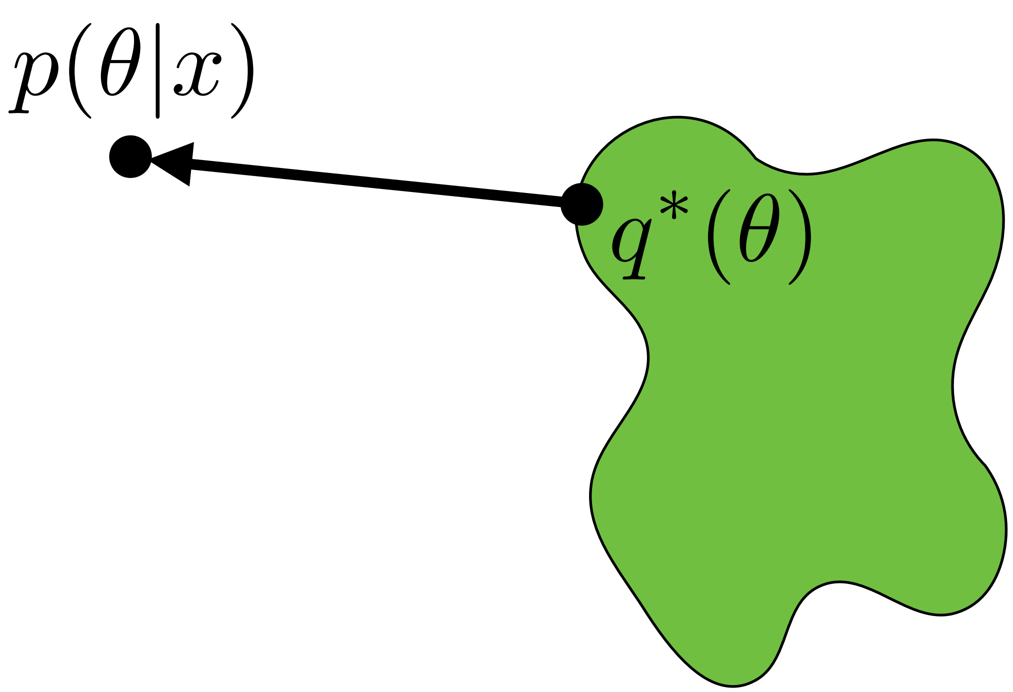


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation

- VB practical success

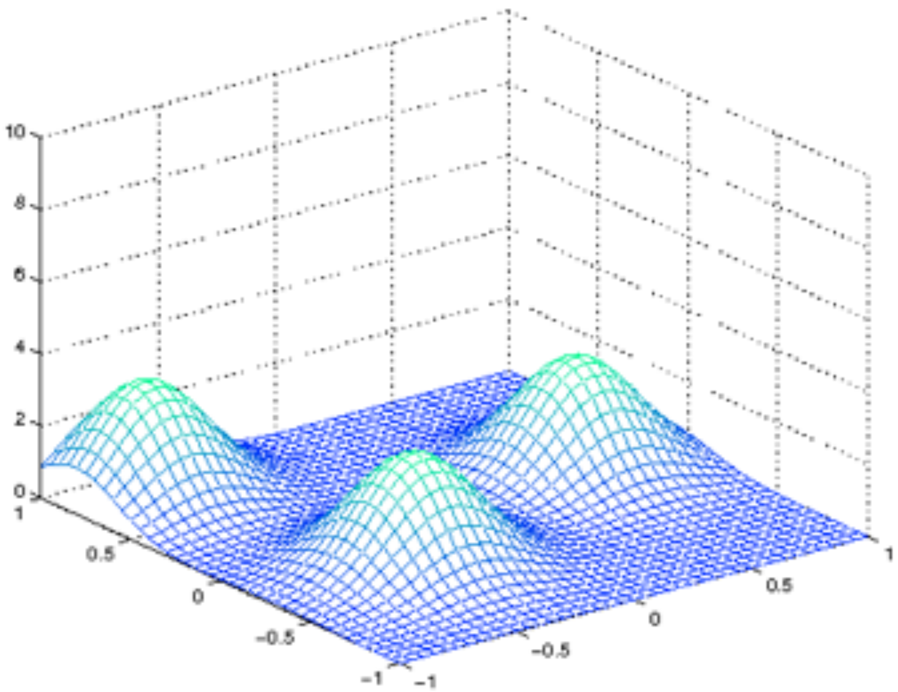
Variational Bayes

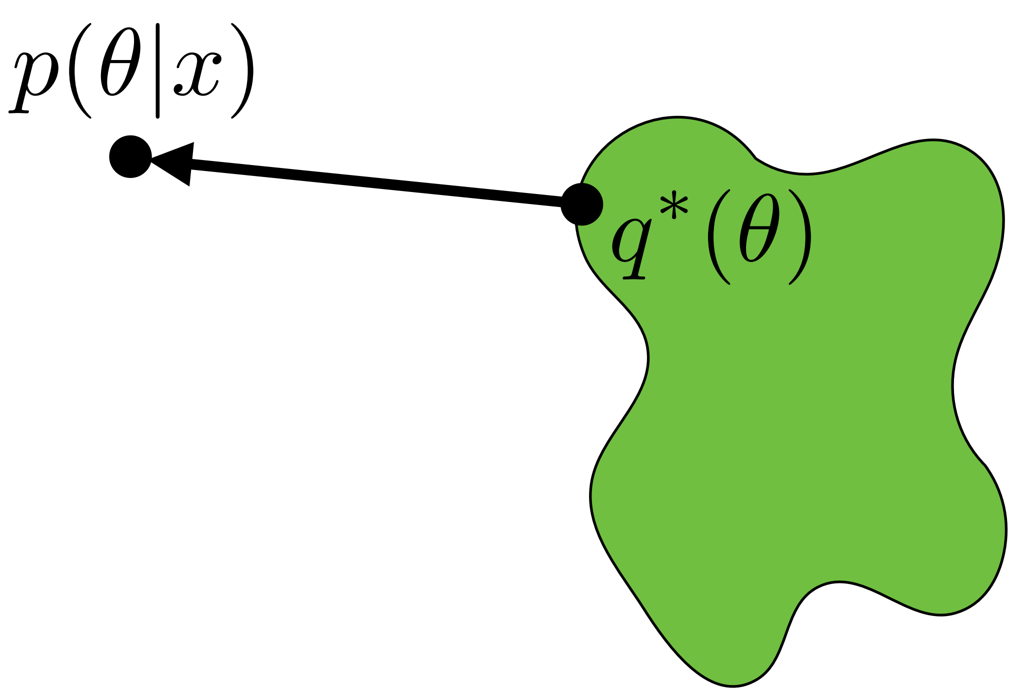




- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation
- VB practical success
  - point estimates and prediction

Variational Bayes

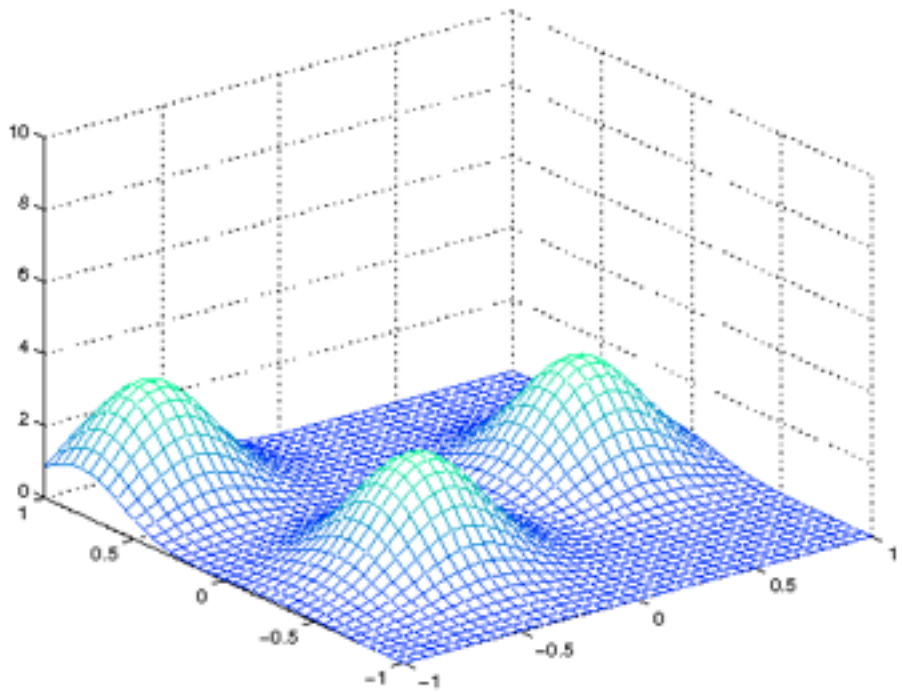


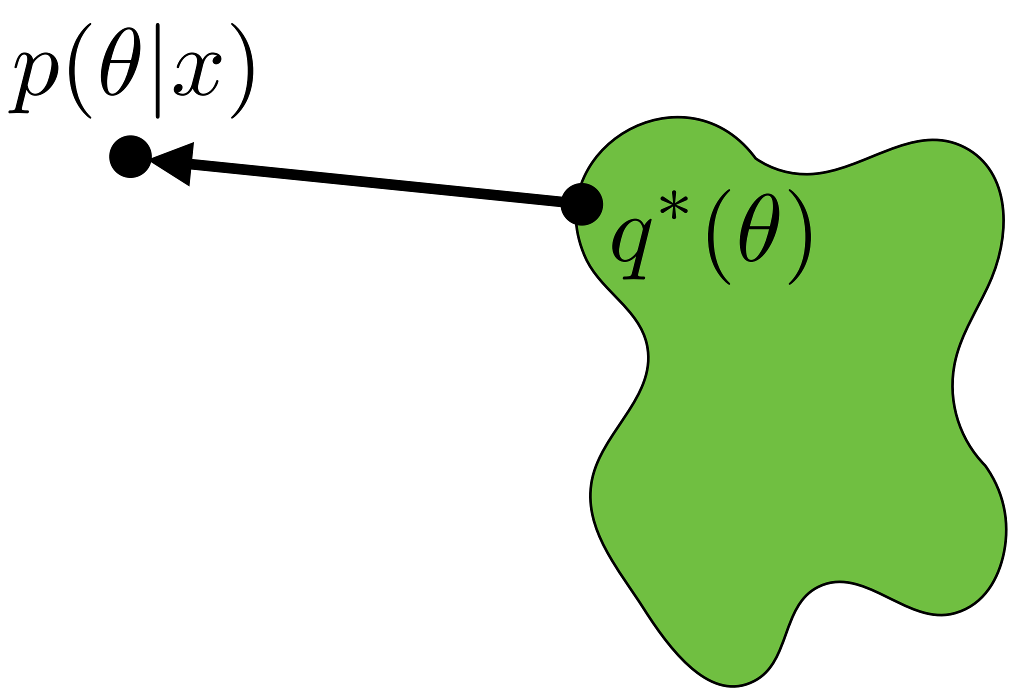


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
    $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation
- VB practical success
  - point estimates and prediction
  - fast, streaming,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

[Broderick, Boyd, Wibisono, Wilson, Jordan 2013]

Variational Bayes



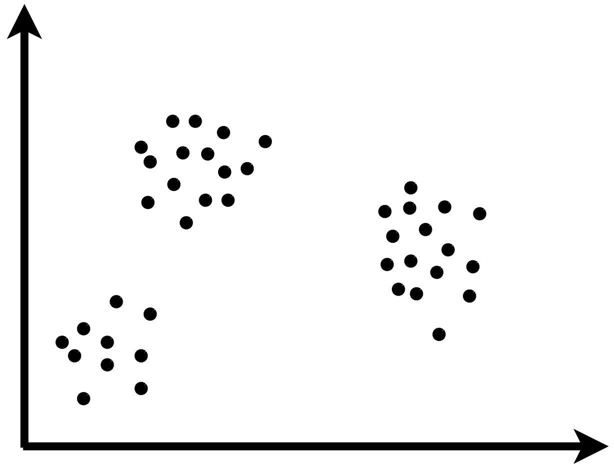


- Variational Bayes (VB)
  - Approximation $q^*(\theta)$ for posterior $p(\theta|x)$
  - "Close": Minimize Kullback-Liebler (KL) divergence:
  $$KL(q\|p(\cdot|x))$$
  - "Nice": factorizes, exponential family, truncation

- VB practical success
  - point estimates and prediction
  - fast, streaming, distributed
  - Linear response VB (LRVB) for accurate covariance

[Broderick, Boyd, Wibisono, Wilson, Jordan 2013; Giordano, Broderick, Jordan 2015]

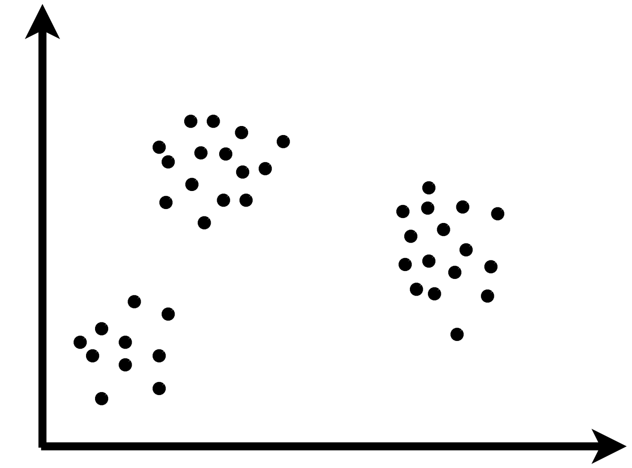
Exercises



Exercises

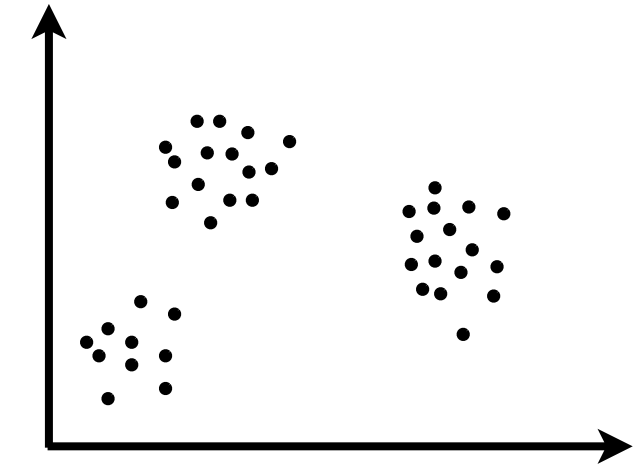
- Code a DPMM simulator using a:
  (1) CRP, (2) DP

Exercises



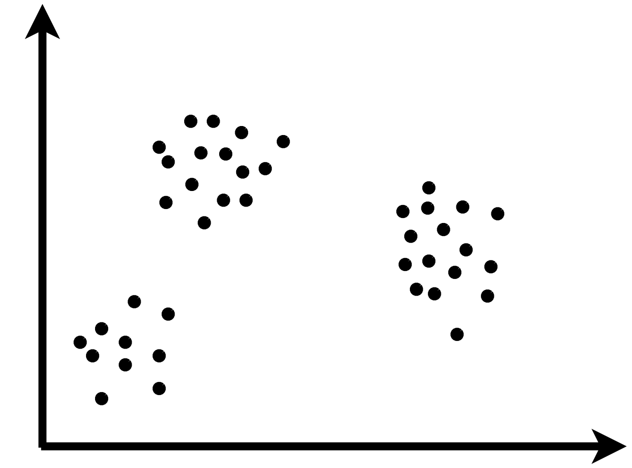
- Code a DPMM simulator using a:
  (1) CRP, (2) DP
- What is the expected number of clusters generated by a CRP(α) after *N* data points?

Exercises



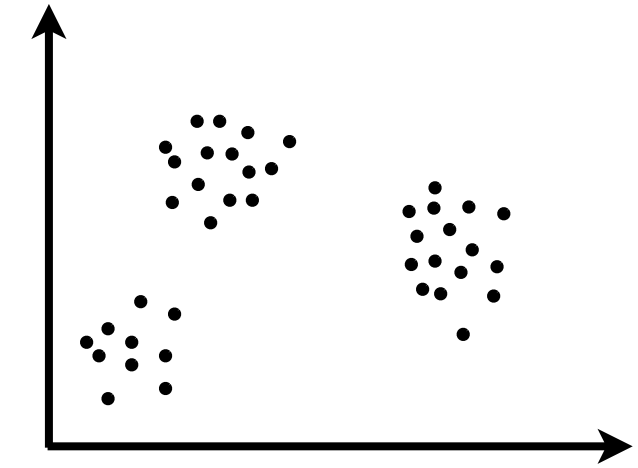
- Code a DPMM simulator using a:
  (1) CRP, (2) DP
- What is the expected number of clusters generated by a CRP(α) after *N* data points?
- What do you think about the answer to the previous question when it comes to real-life data modeling?

Exercises



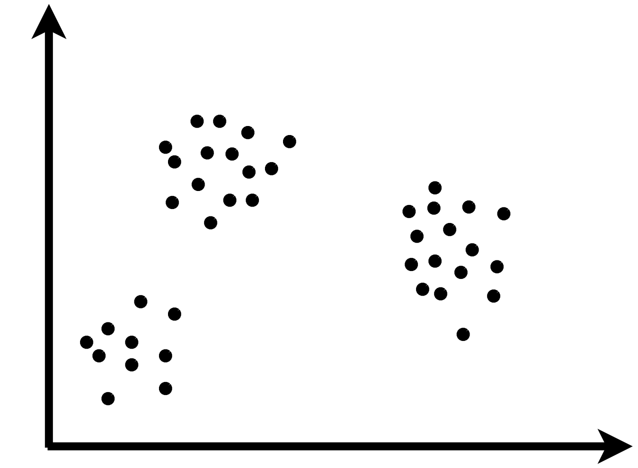
- Code a DPMM simulator using a:
  (1) CRP, (2) DP
- What is the expected number of clusters generated by a CRP(α) after N data points?
- What do you think about the answer to the previous question when it comes to real-life data modeling?
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]

Exercises



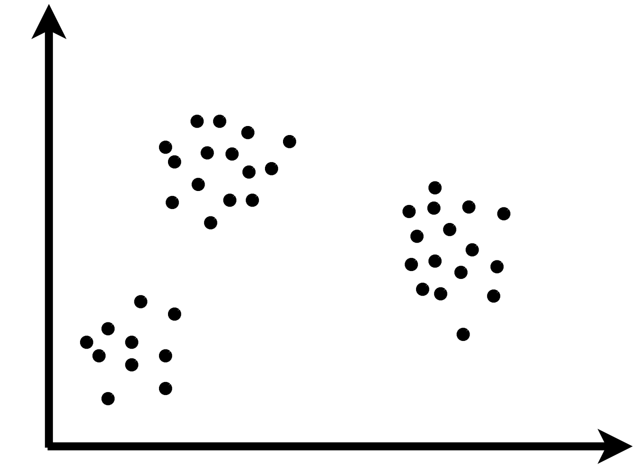
- Code a DPMM simulator using a:
  (1) CRP, (2) DP
- What is the expected number of clusters generated by a CRP(α) after N data points?
- What do you think about the answer to the previous question when it comes to real-life data modeling?
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
- Read [Neal 2000] and code a DPMM Gibbs sampler

Exercises



- Code a DPMM simulator using a:
  (1) CRP, (2) DP
- What is the expected number of clusters generated by a CRP(α) after N data points?
- What do you think about the answer to the previous question when it comes to real-life data modeling?
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
- Read [Neal 2000] and code a DPMM Gibbs sampler
- Read [Walker 2007; Kalli, Griffin, Walker 2011] and code a DPMM slice sampler

Exercises



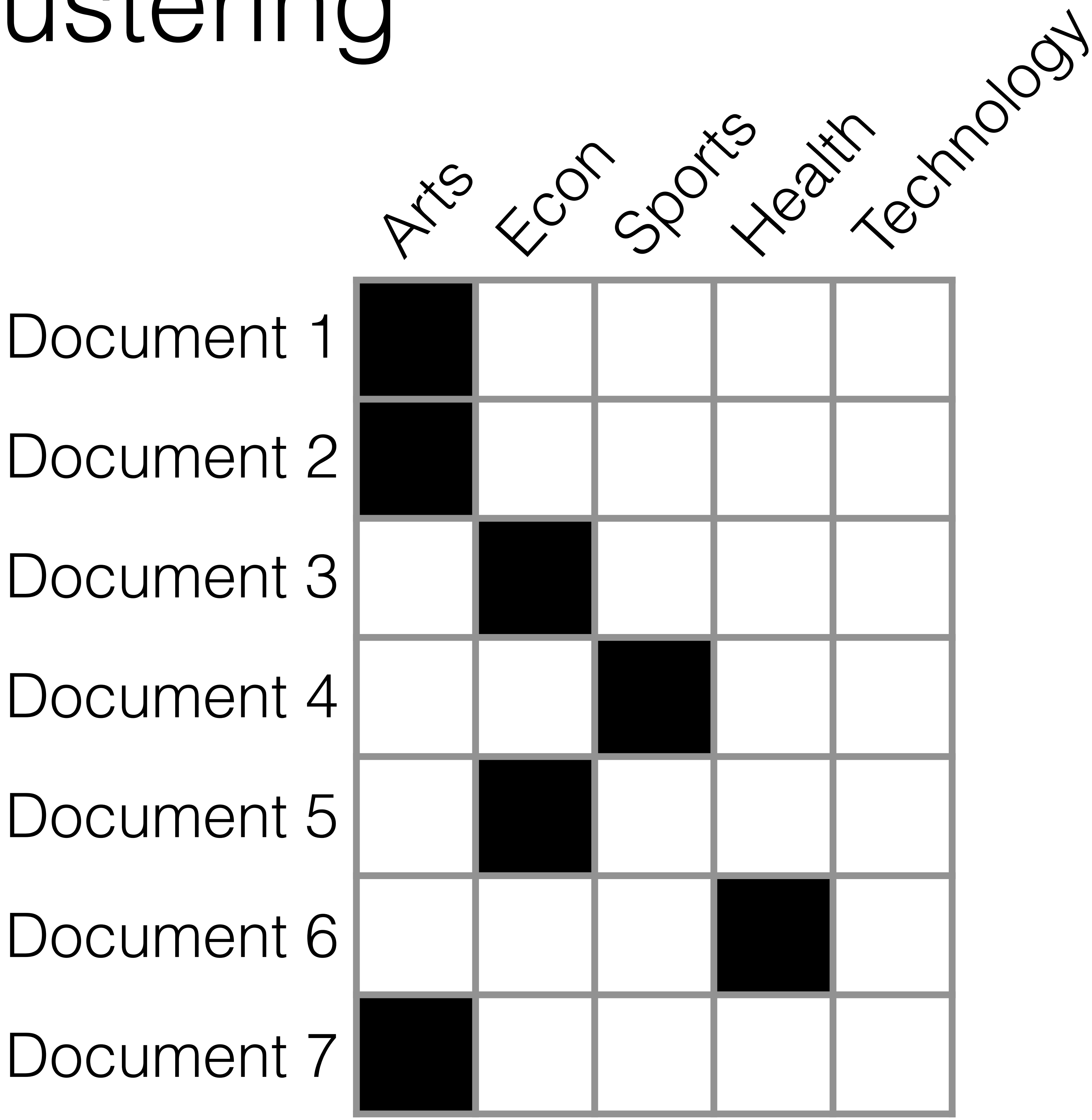
- Code a DPMM simulator using a:
  (1) CRP, (2) DP
- What is the expected number of clusters generated by a CRP(α) after *N* data points?
- What do you think about the answer to the previous question when it comes to real-life data modeling?
- Review Gibbs sampling, slice sampling [Neal 2003], variational Bayes [Bishop 2006]
- Read [Neal 2000] and code a DPMM Gibbs sampler
- Read [Walker 2007; Kalli, Griffin, Walker 2011] and code a DPMM slice sampler
- Read [Blei, Jordan 2006] and code variational inference for the DPMM

Resources online

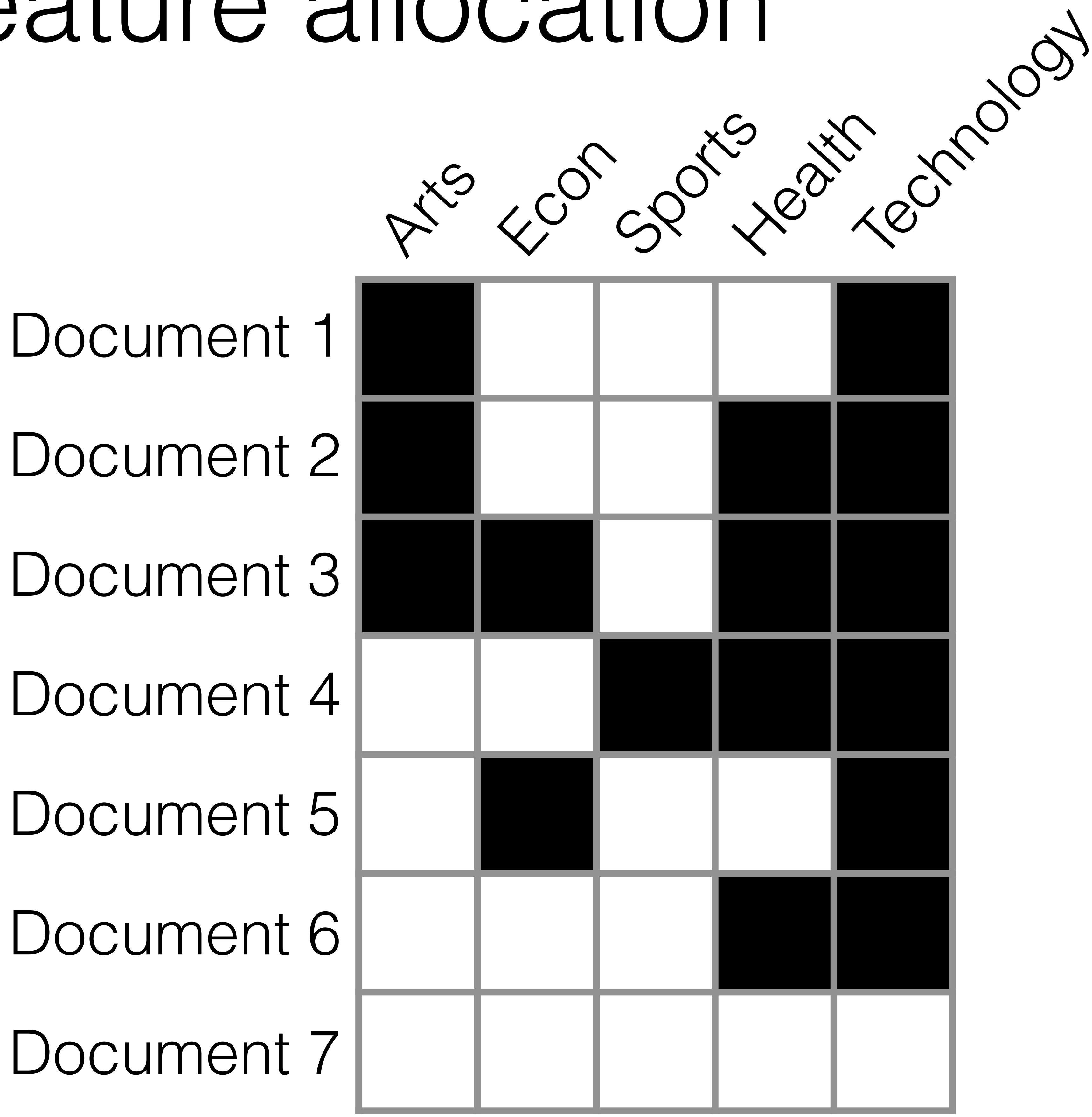
www.tamarabroderick.com/tutorials.html

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

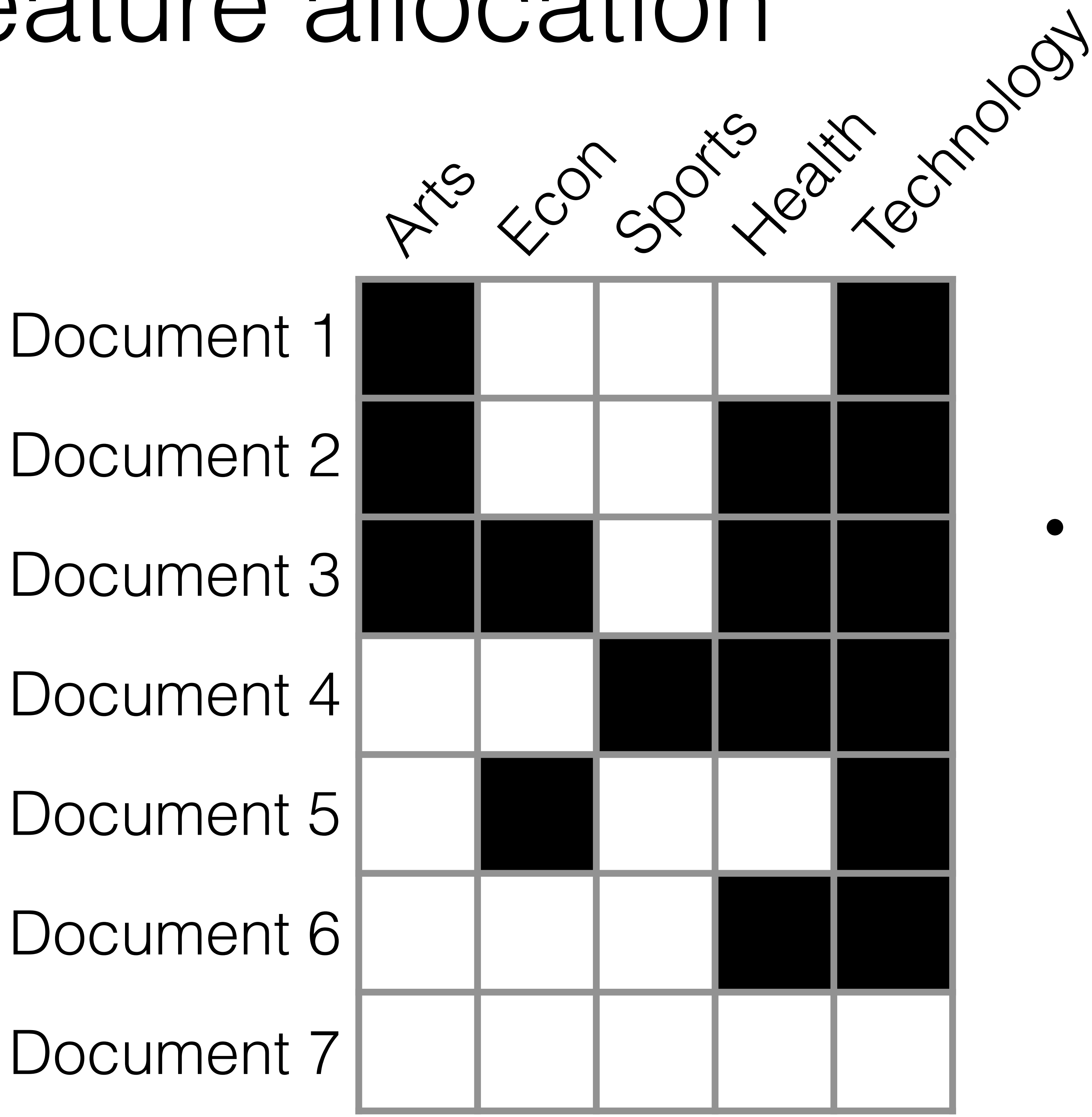
Clustering


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

Feature allocation

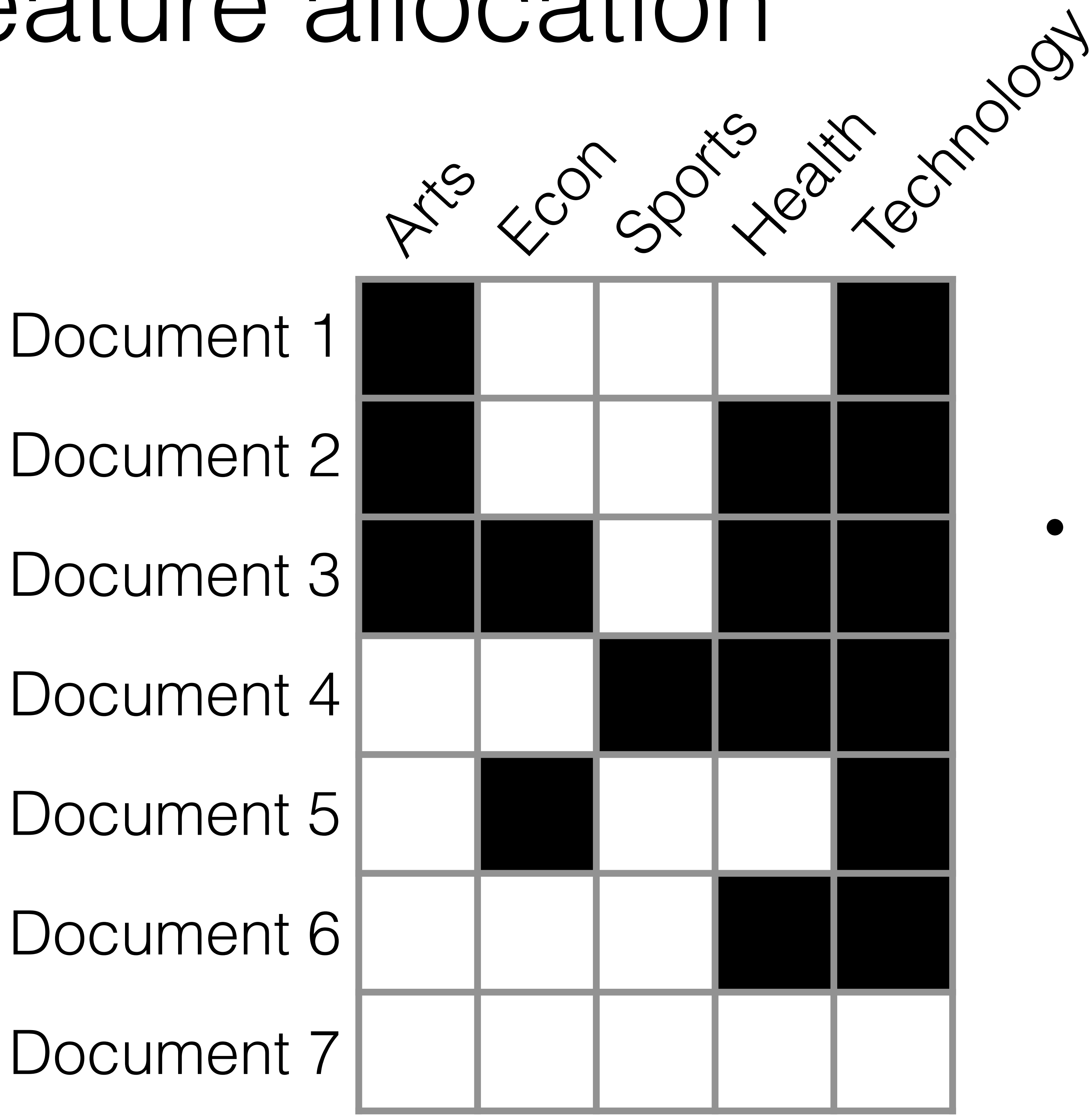


Feature allocation



- Indian buffet process

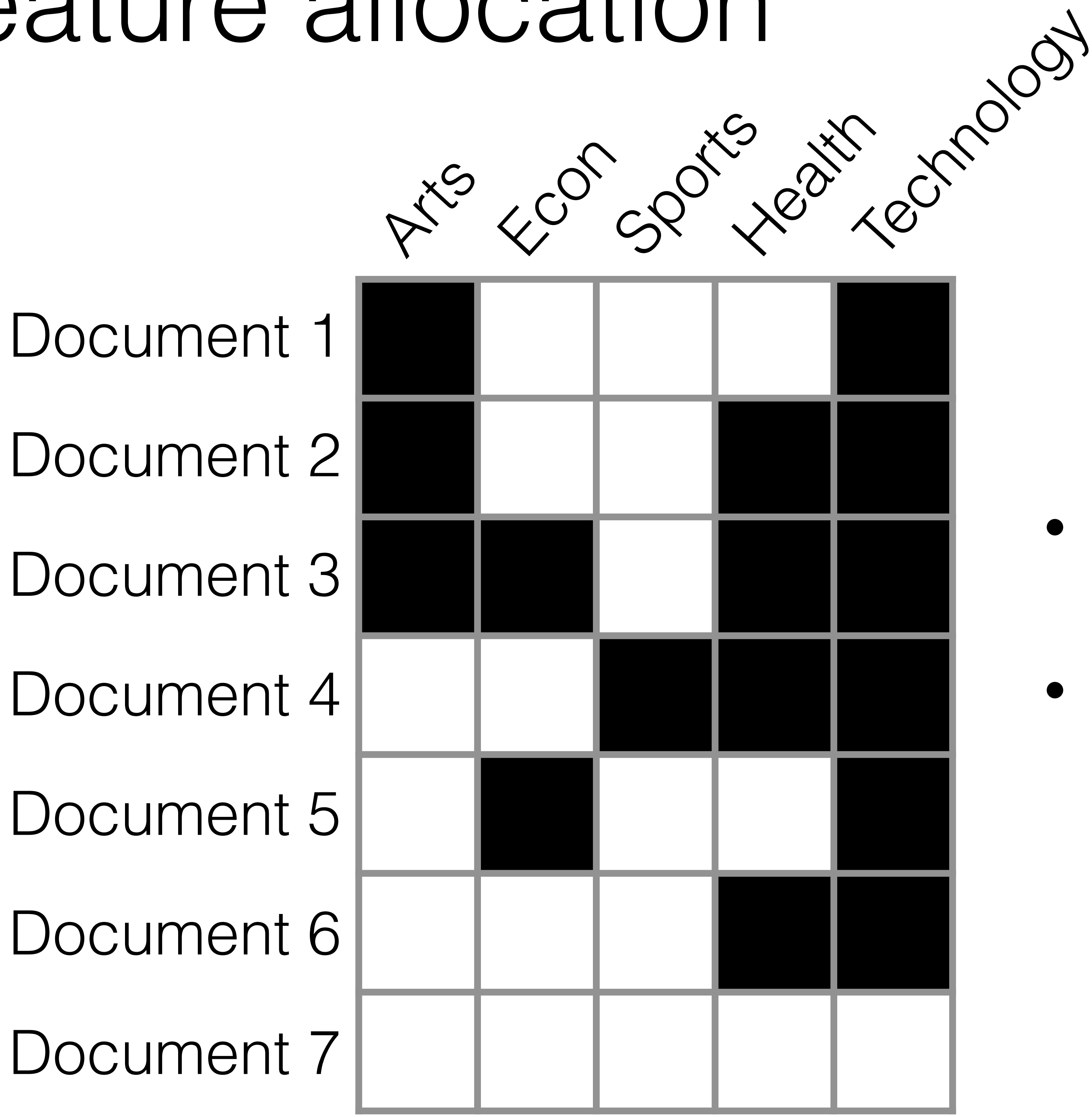
Feature allocation



- Indian buffet process

[Griffiths, Ghahramani 2005

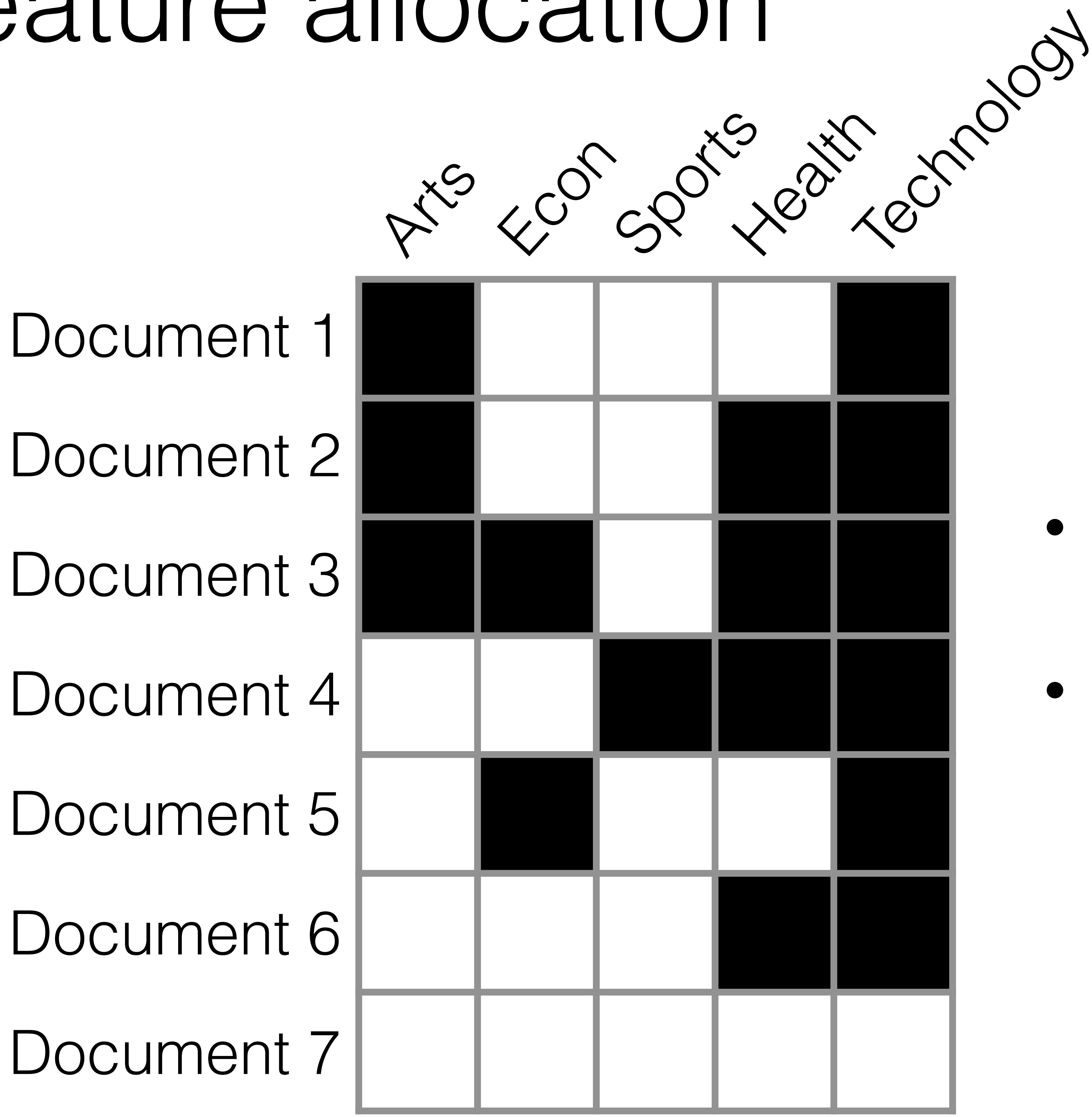
Feature allocation



- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005

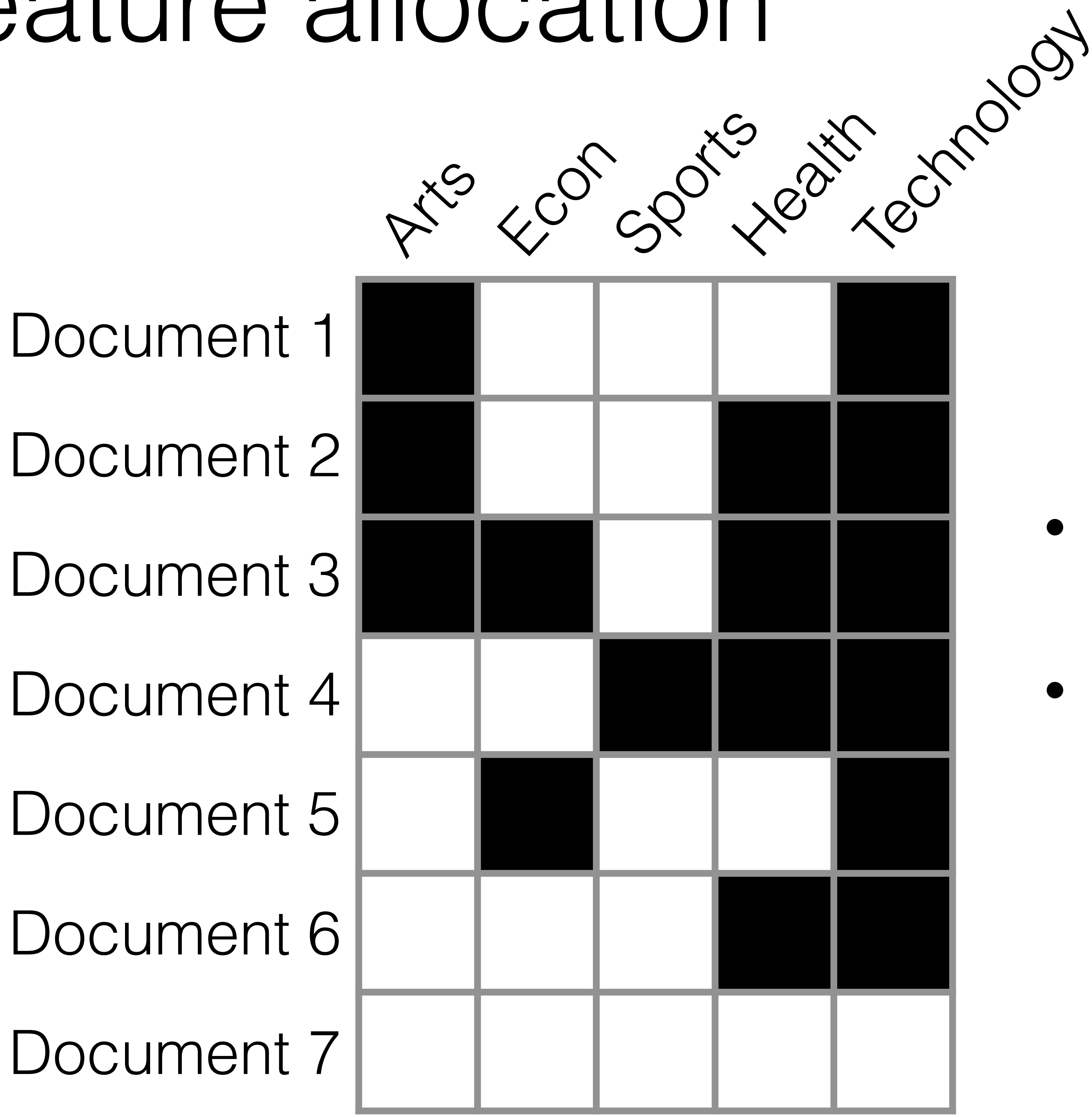
Feature allocation



- Indian buffet process
- Beta process

[Griffiths, Ghahramani 2005, Hjort 1990

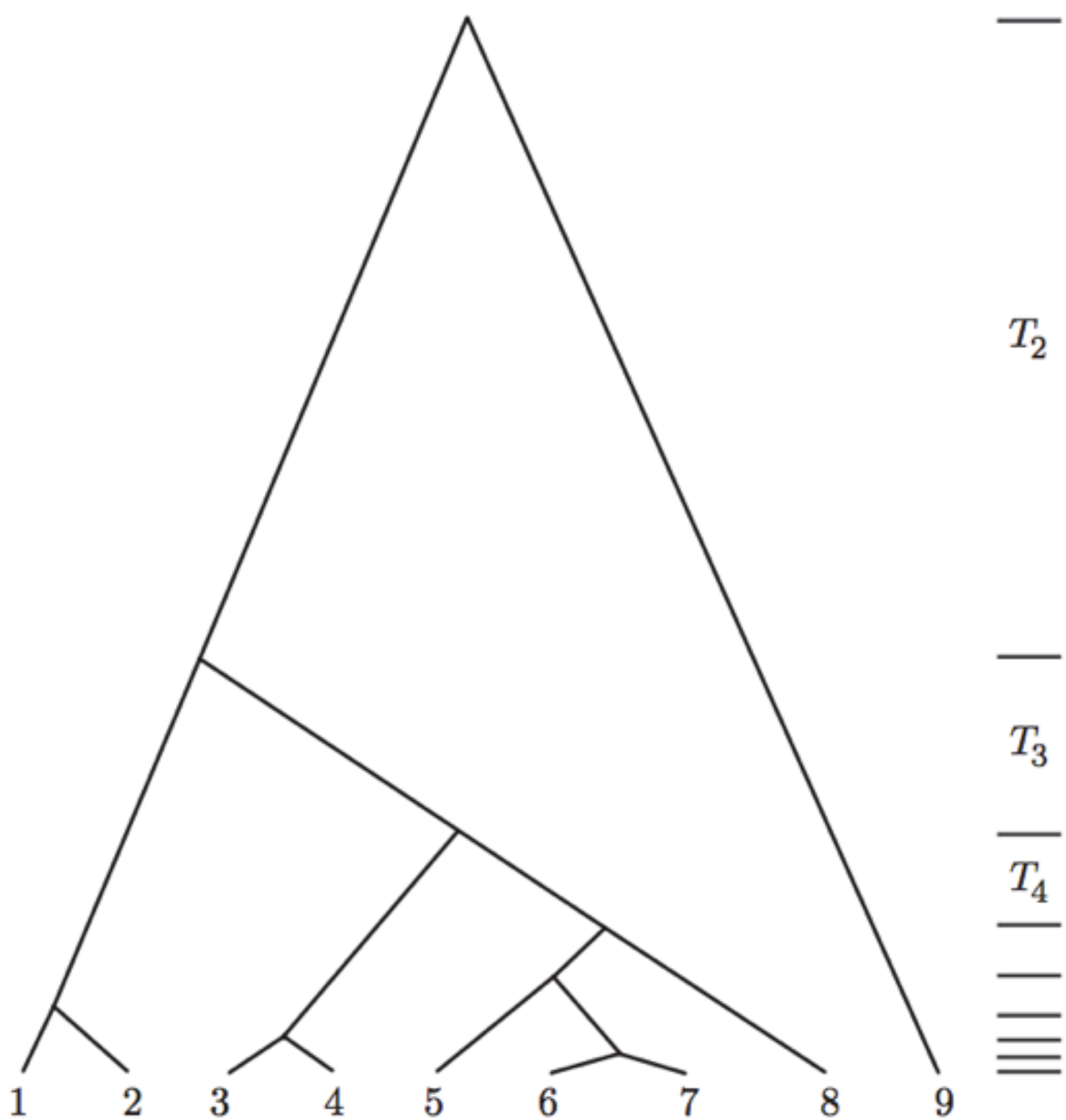
Feature allocation



- Indian buffet process
- Beta process

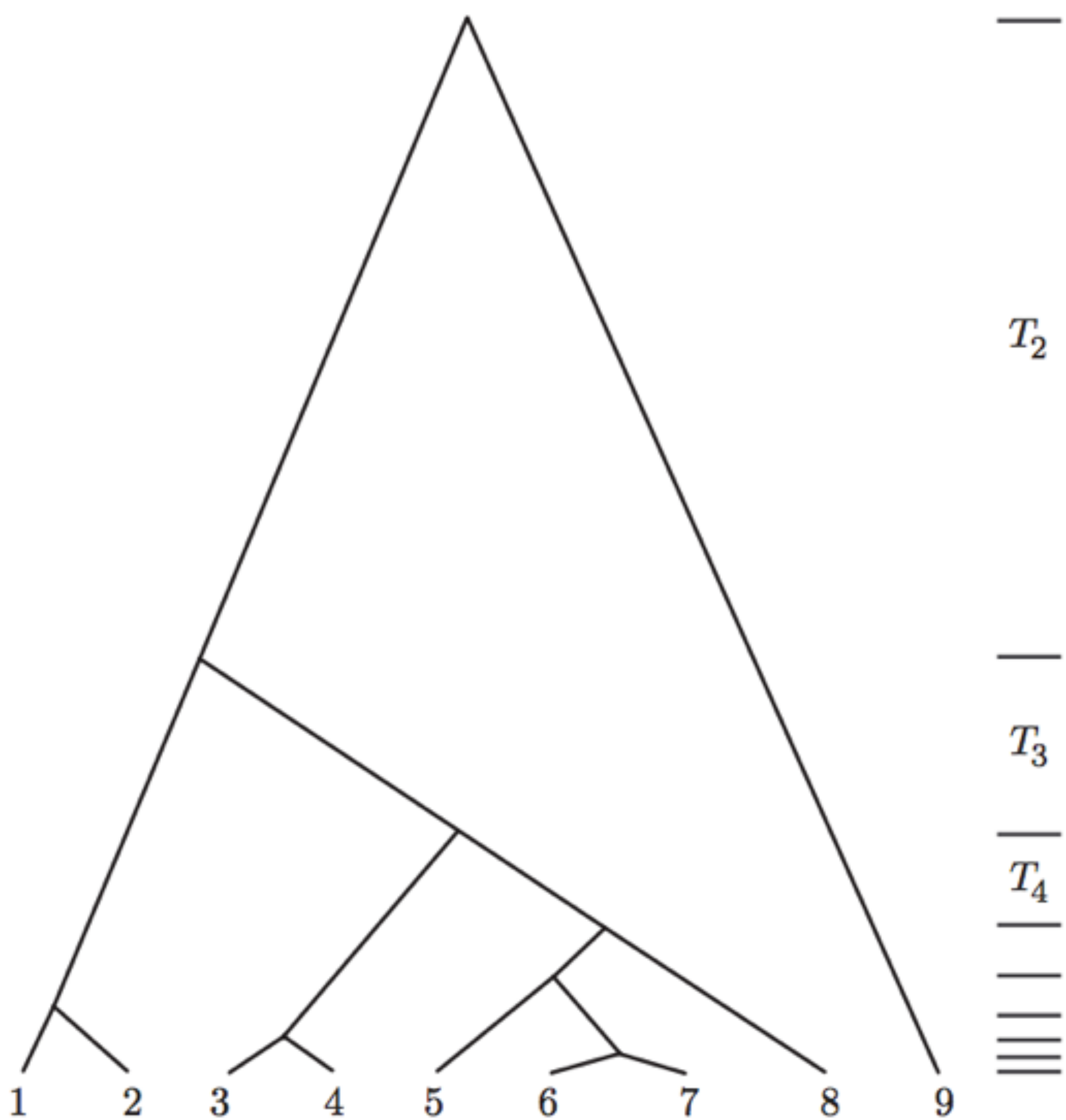
[Griffiths, Ghahramani 2005, Hjort 1990, Kim 1999, Thibaux, Jordan 2007, Broderick, Jordan, Pitman 2013]

Genealogy, trees, beyond trees



[Wakeley 2008]

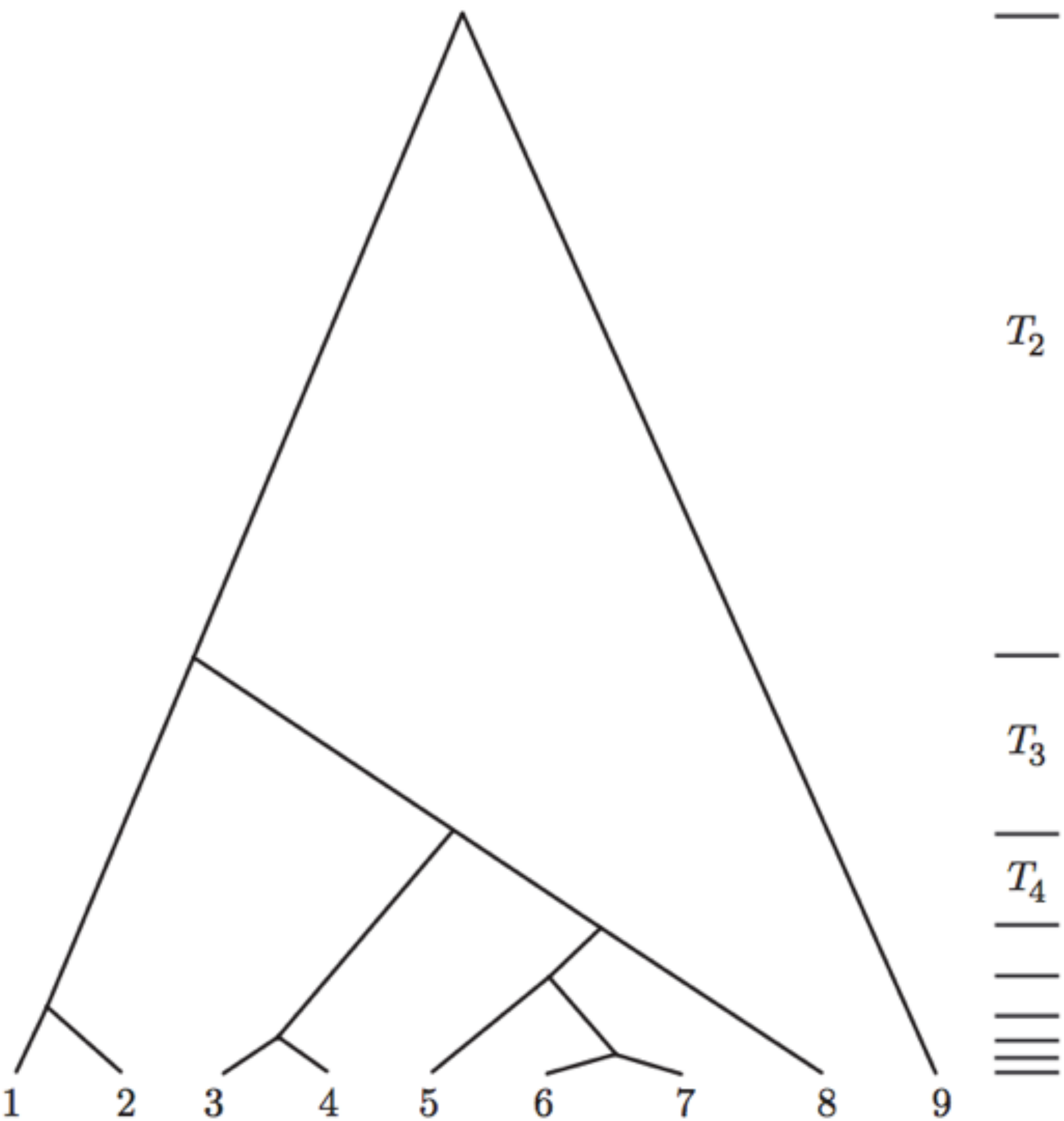
Genealogy, trees, beyond trees



[Wakeley 2008]

- Kingman coalescent

Genealogy, trees, beyond trees

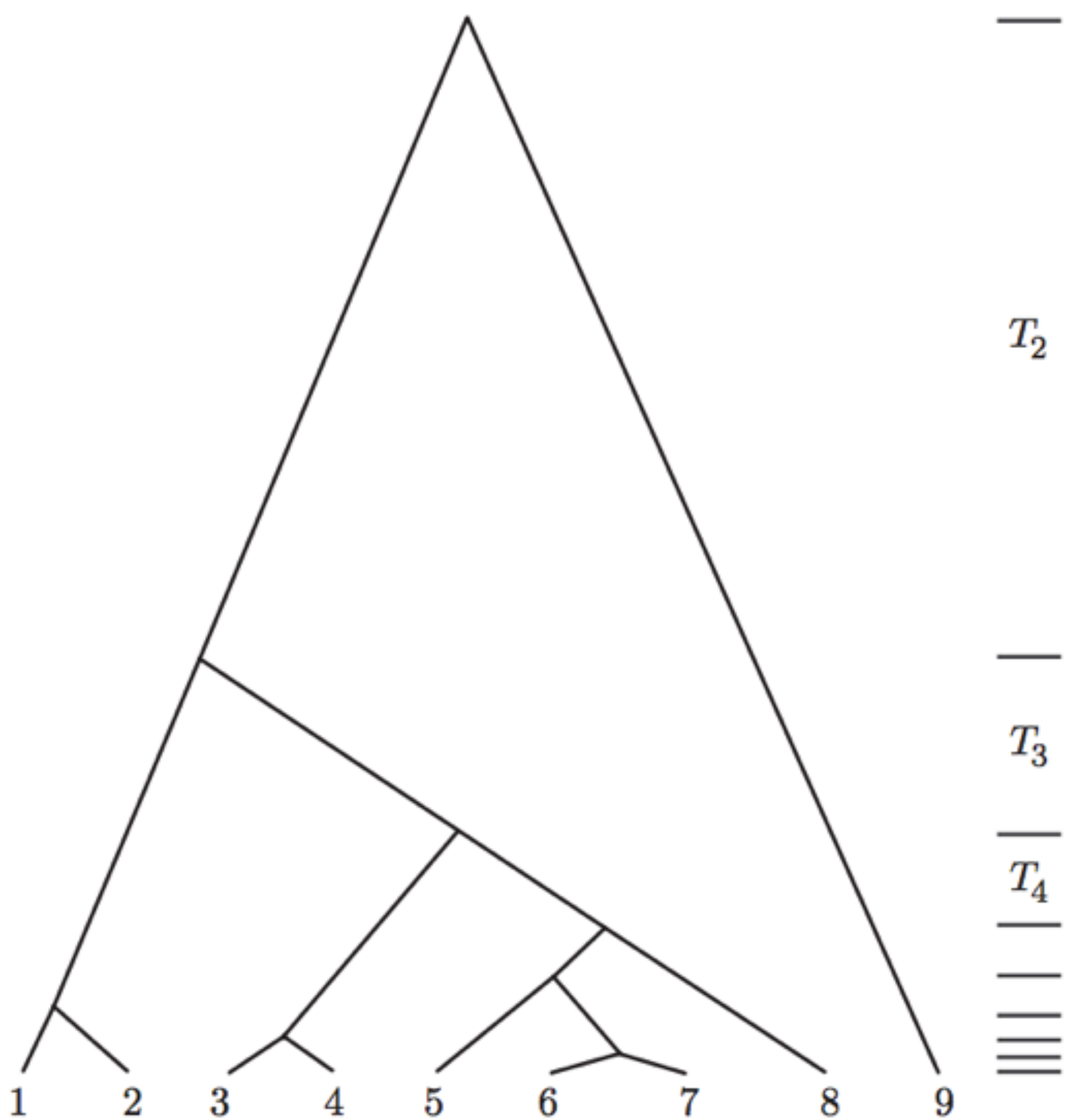


[Wakeley 2008]

- Kingman coalescent

[Kingman 1982

Genealogy, trees, beyond trees

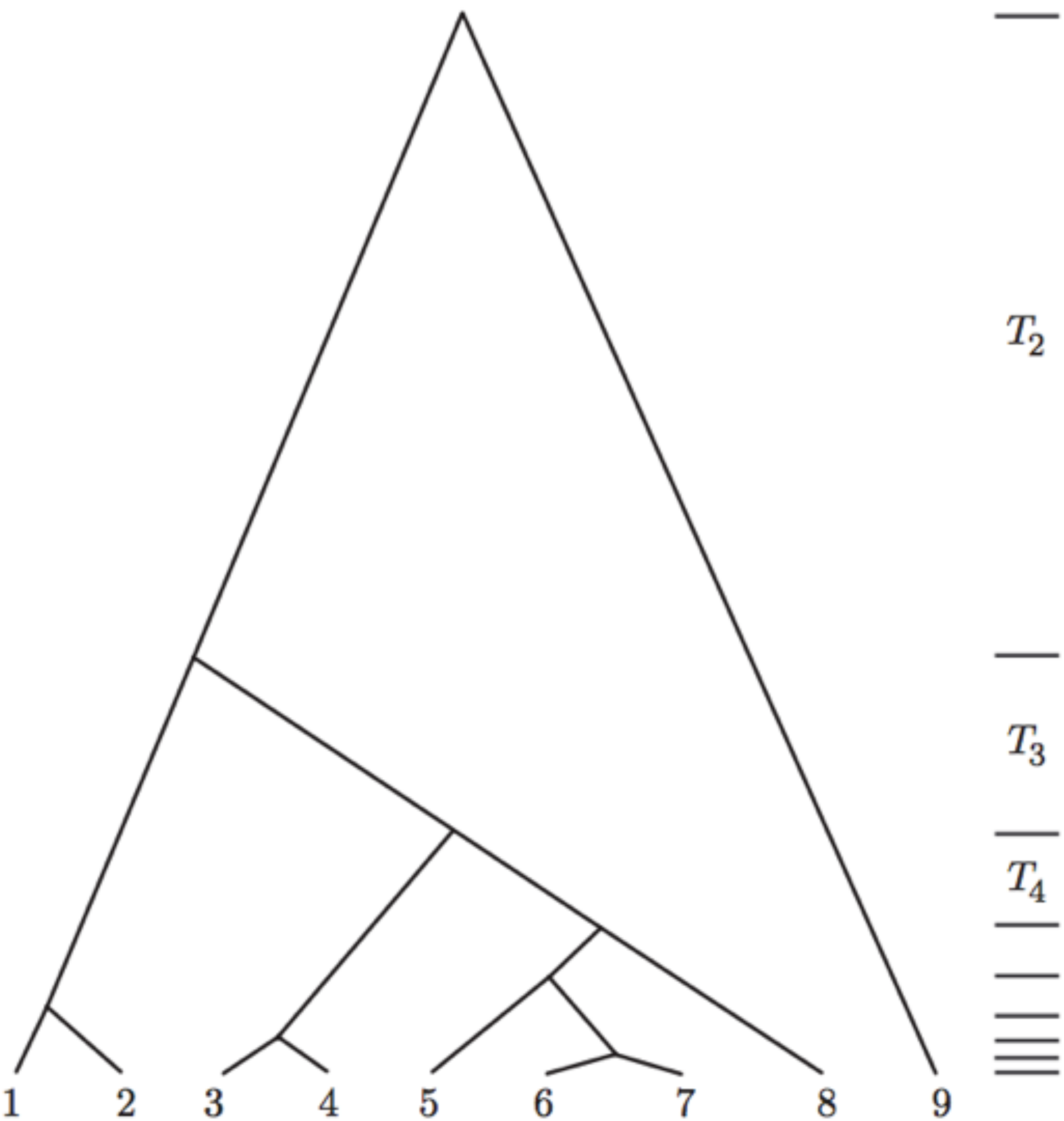


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation

[Kingman 1982

Genealogy, trees, beyond trees

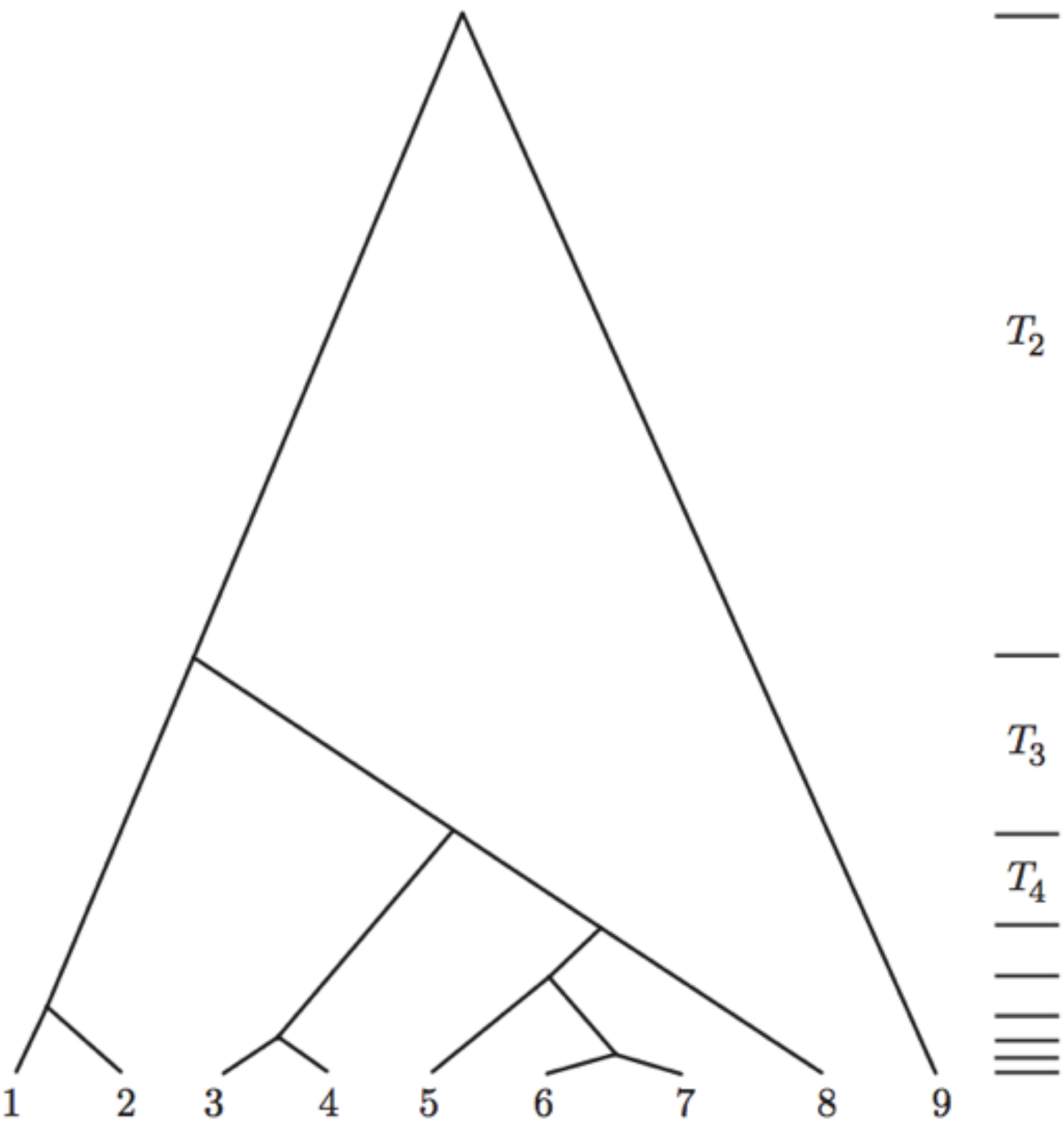


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation

[Kingman 1982, Bertoin 2006, Teh et al 2011

Genealogy, trees, beyond trees

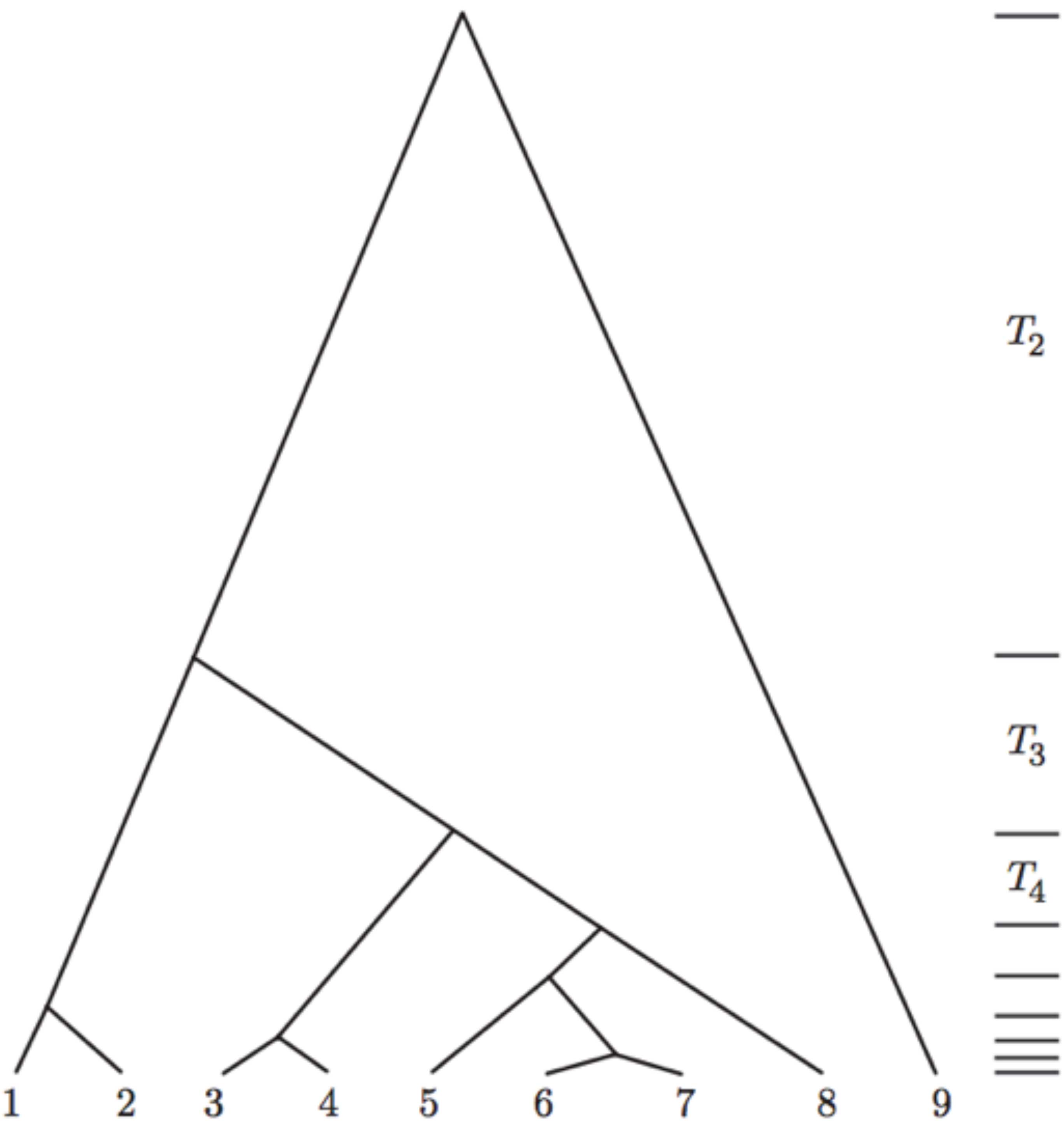


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

[Kingman 1982, Bertoin 2006, Teh et al 2011

Genealogy, trees, beyond trees

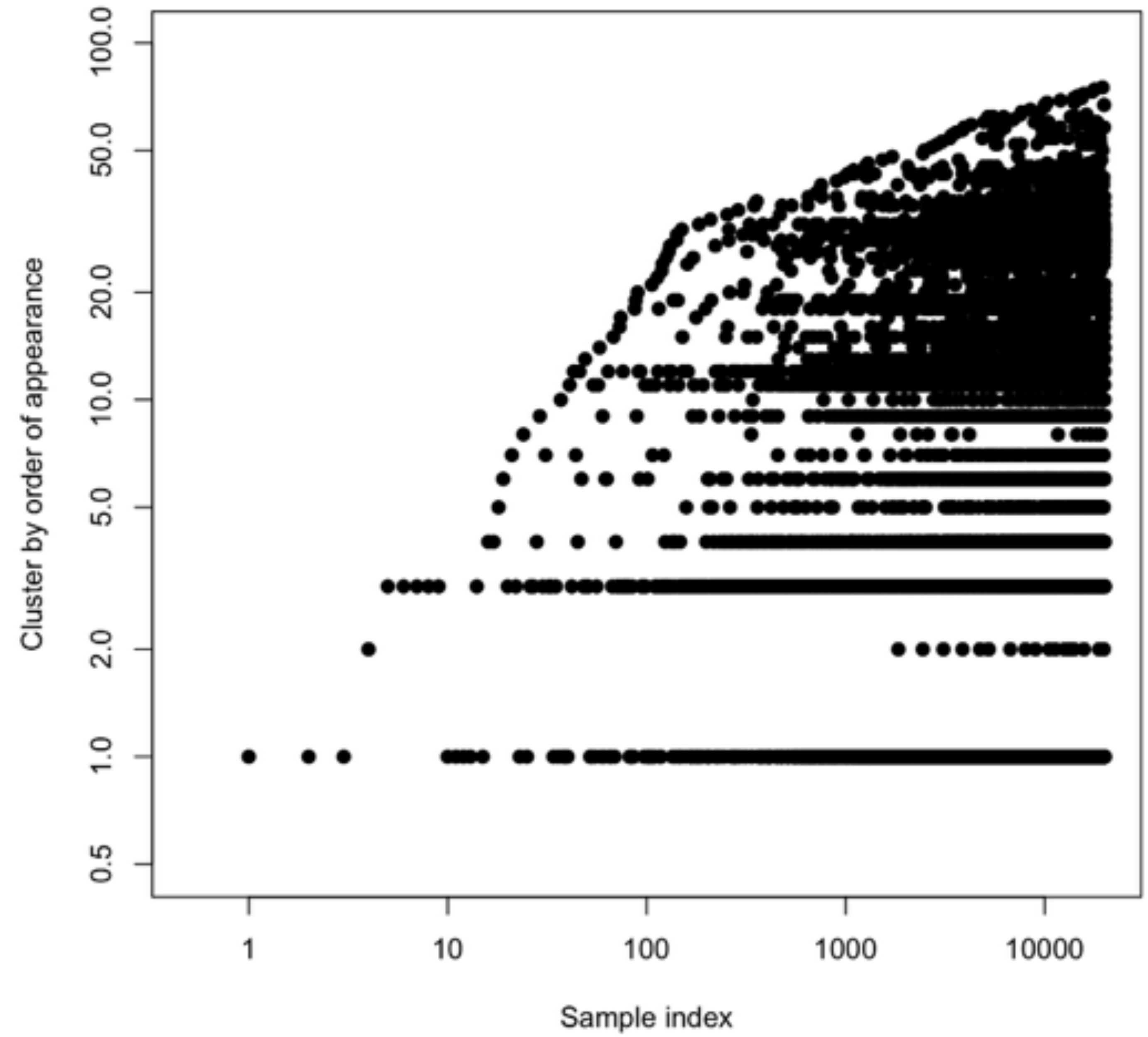


[Wakeley 2008]

- Kingman coalescent
- Fragmentation
- Coagulation
- Dirichlet diffusion tree

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

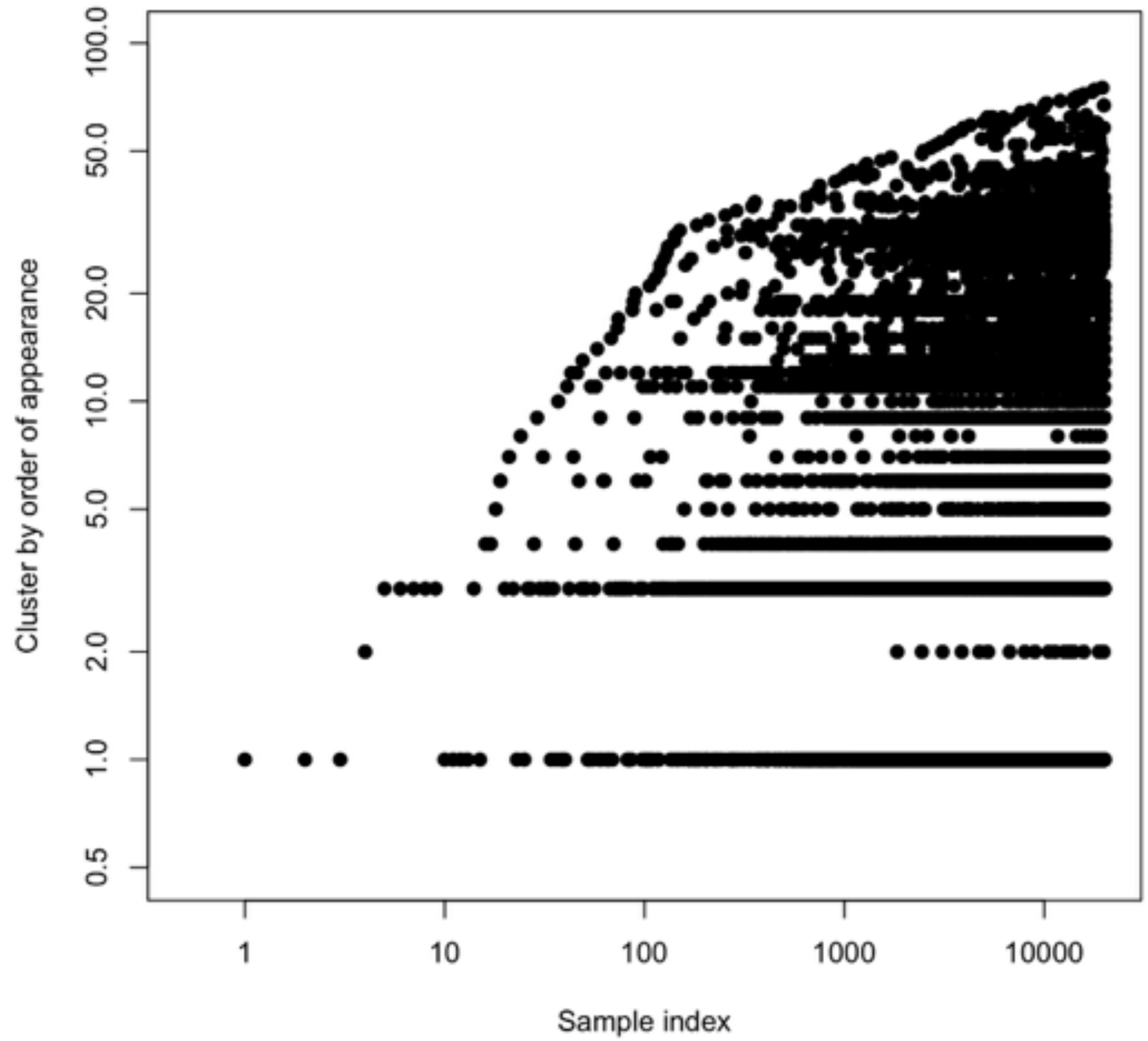
Power laws



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

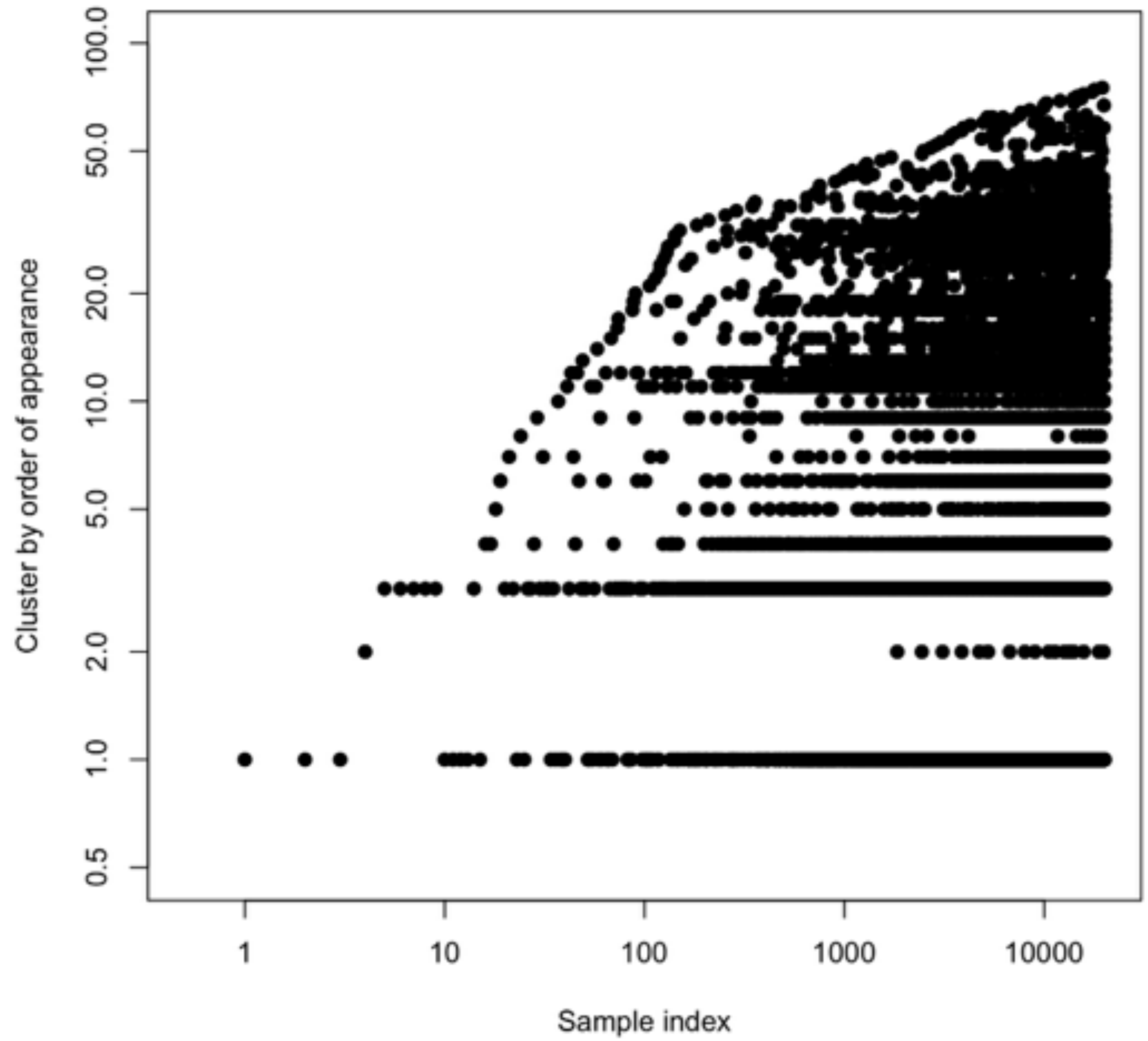
Power laws

- K_N := # clusters occupied by N data points



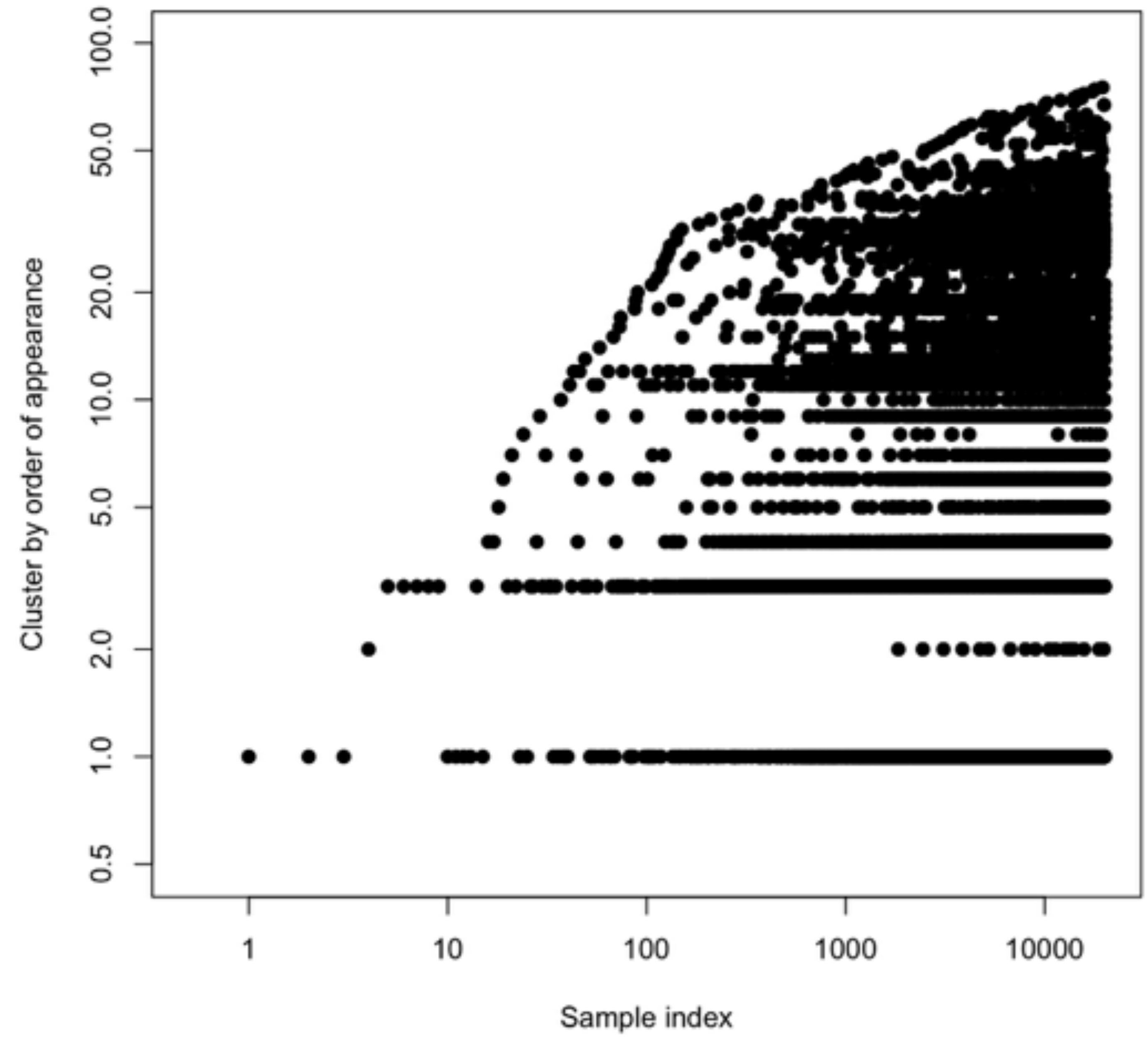
Power laws

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



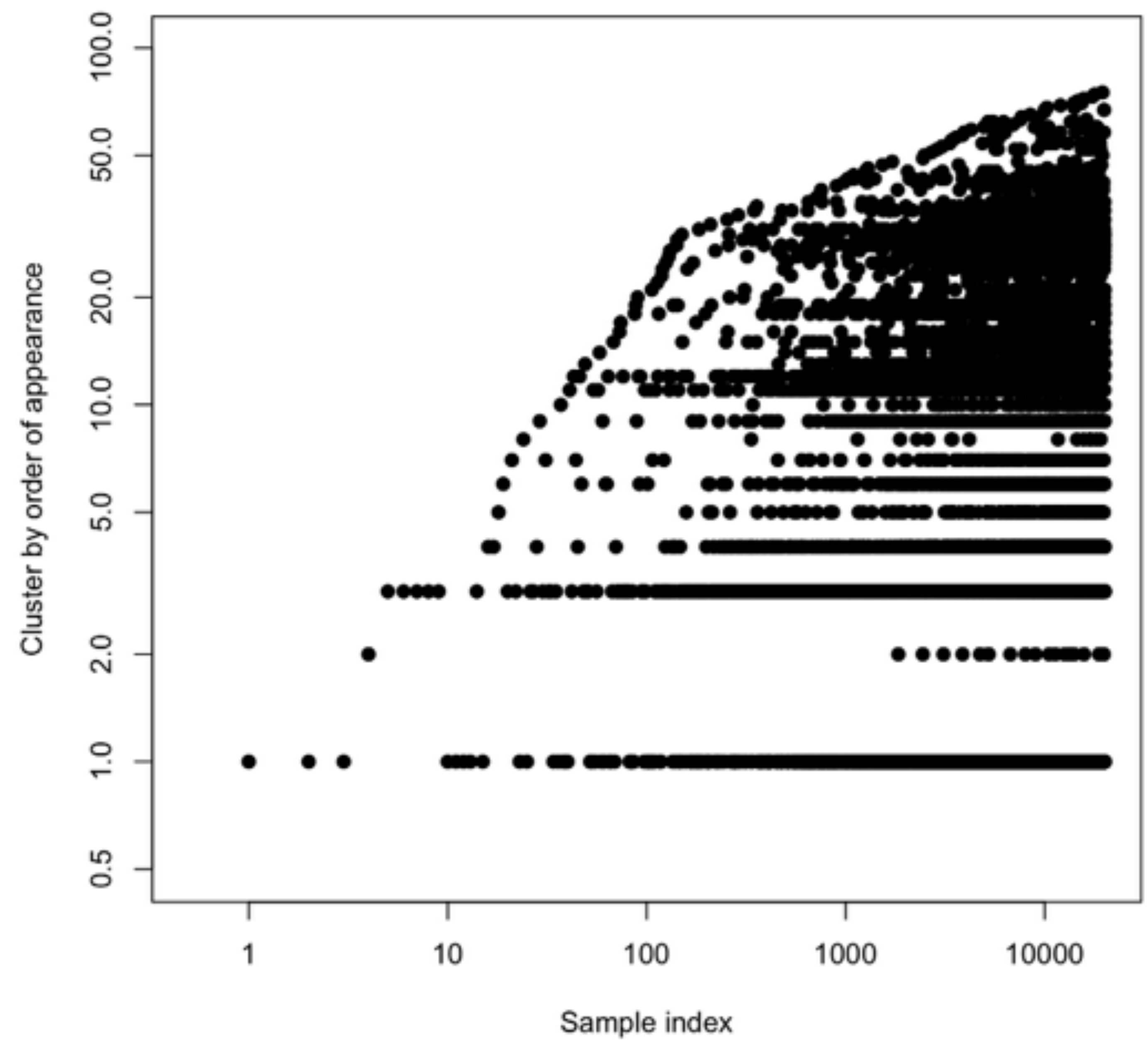
Power laws

- 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

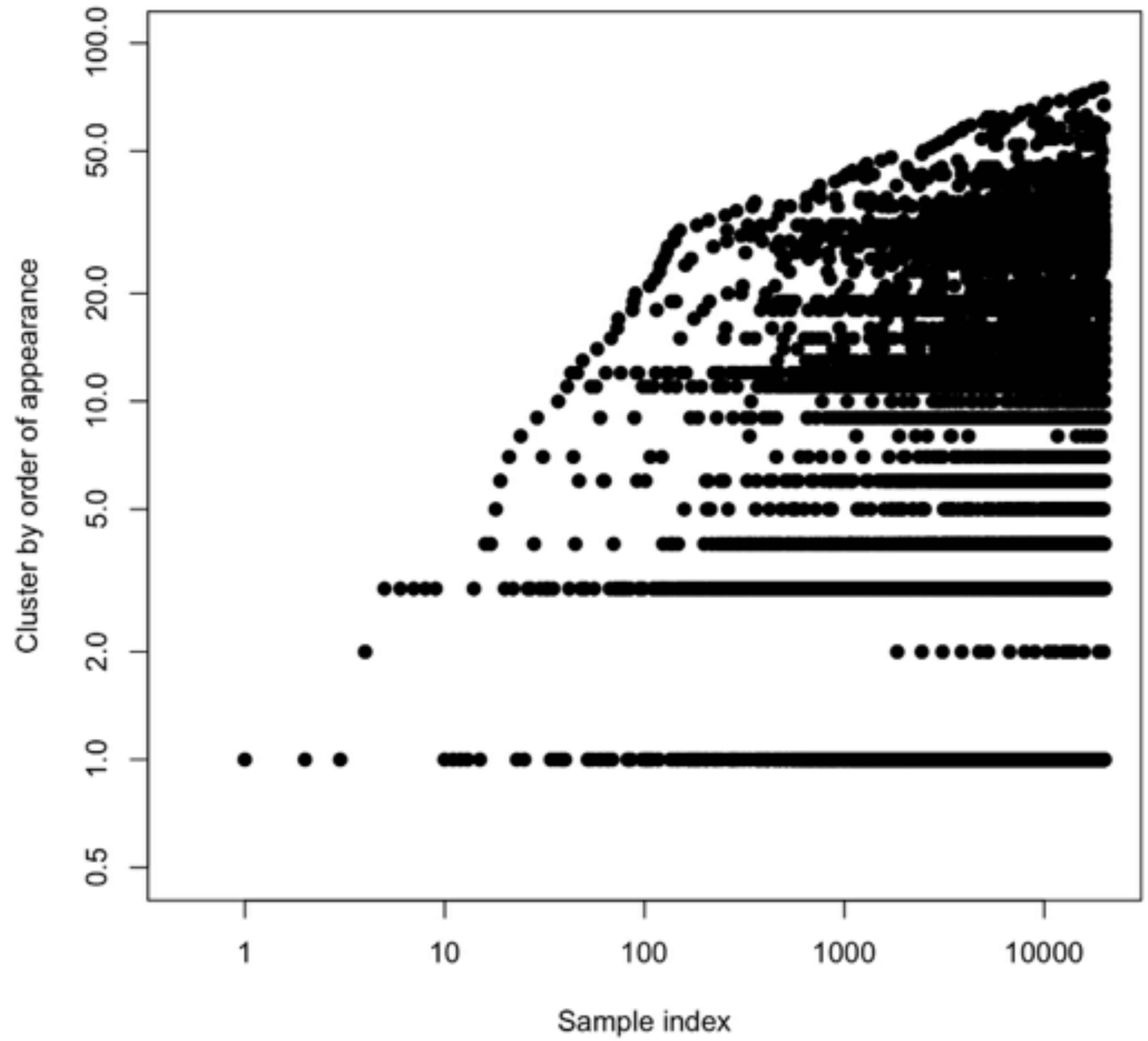
- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc



[Gnedin, et al 2007

Power laws

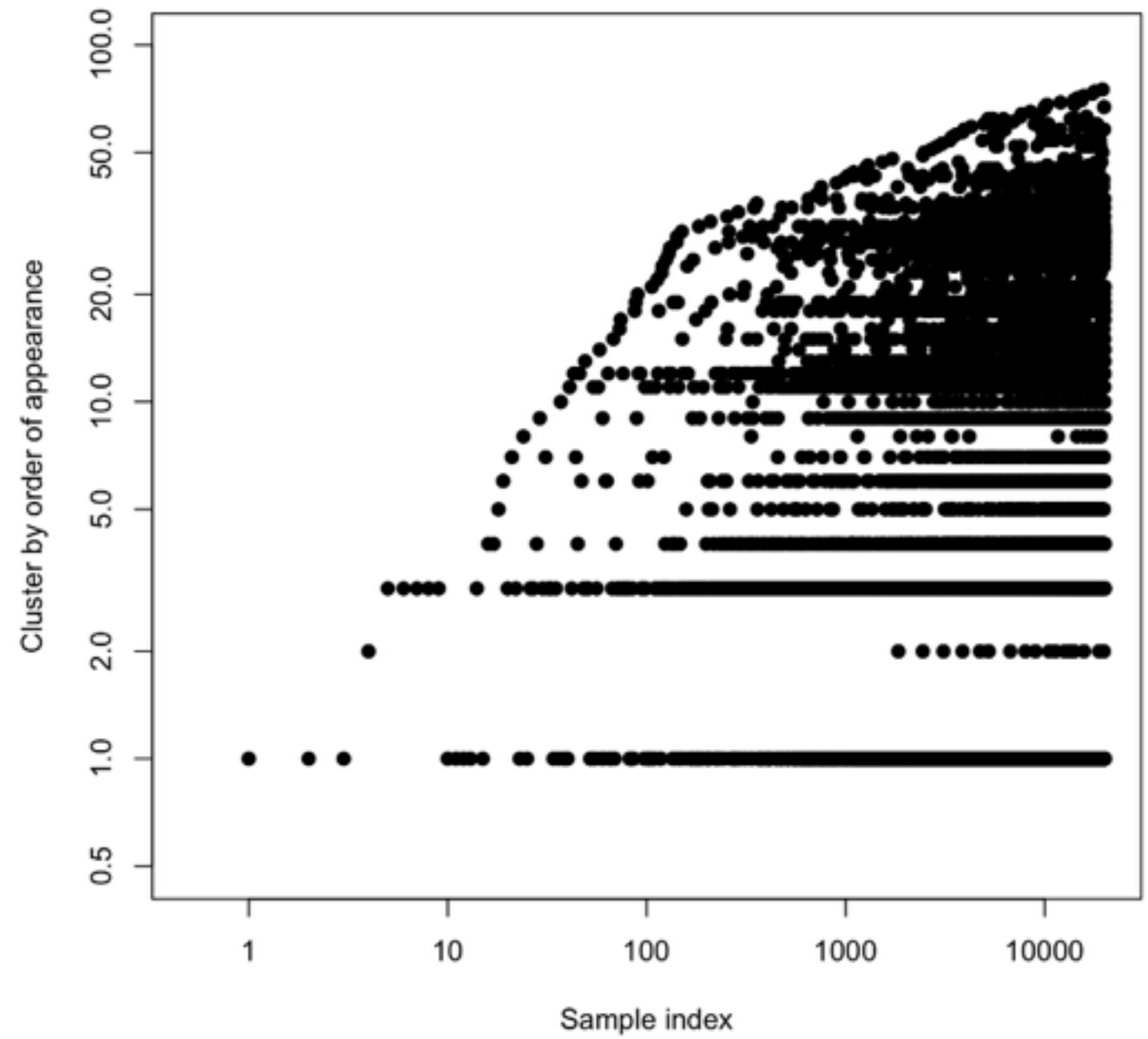
- K_N := # clusters occupied by N data points
- CRP: $K_N \sim \alpha \log N \text{ w.p. } 1$
  - vs. Heaps' law, Herdan's law, etc
- Pitman-Yor process:



[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:

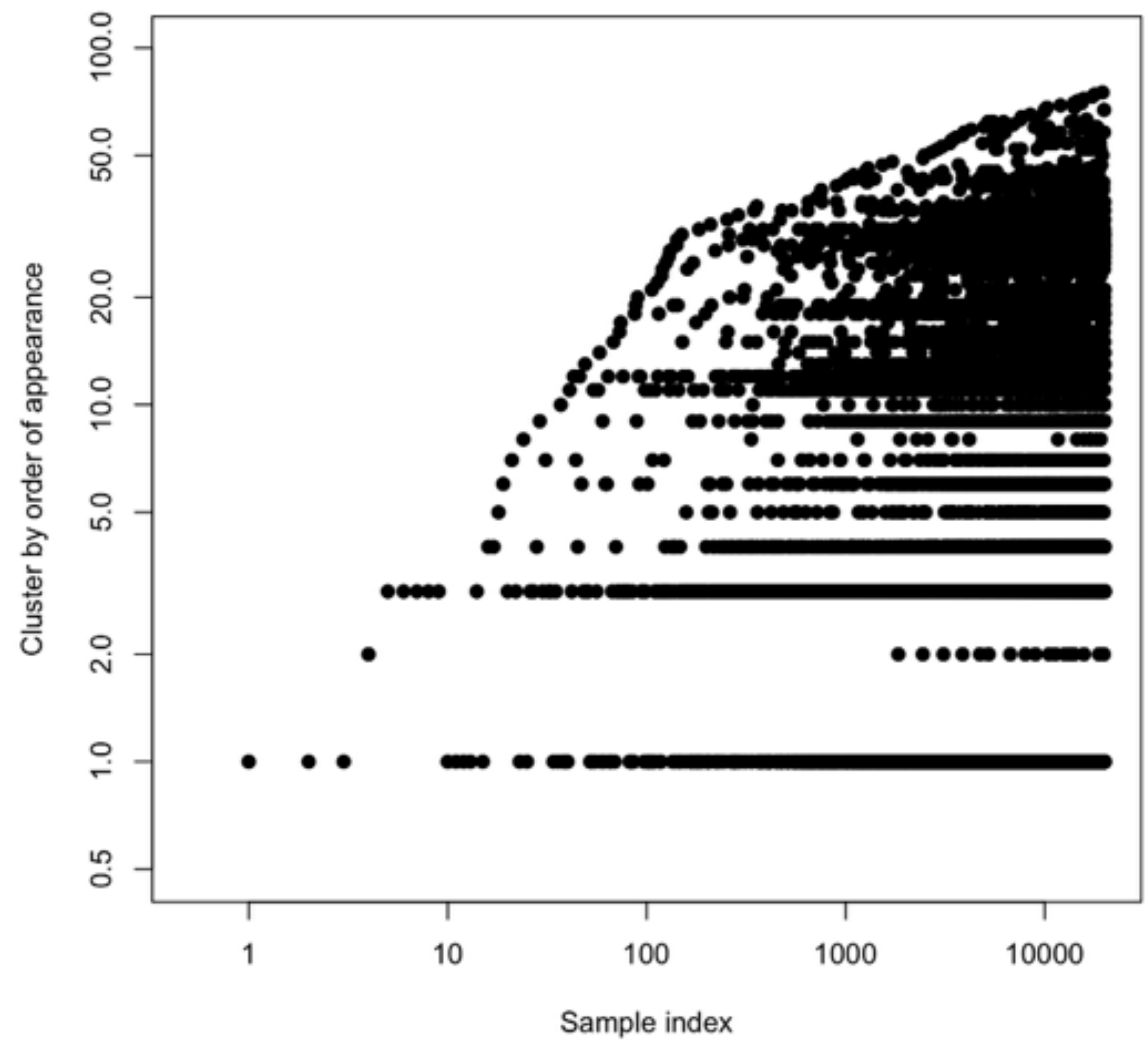


[Gnedin, et al 2007, Pitman, Yor 1997

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:

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

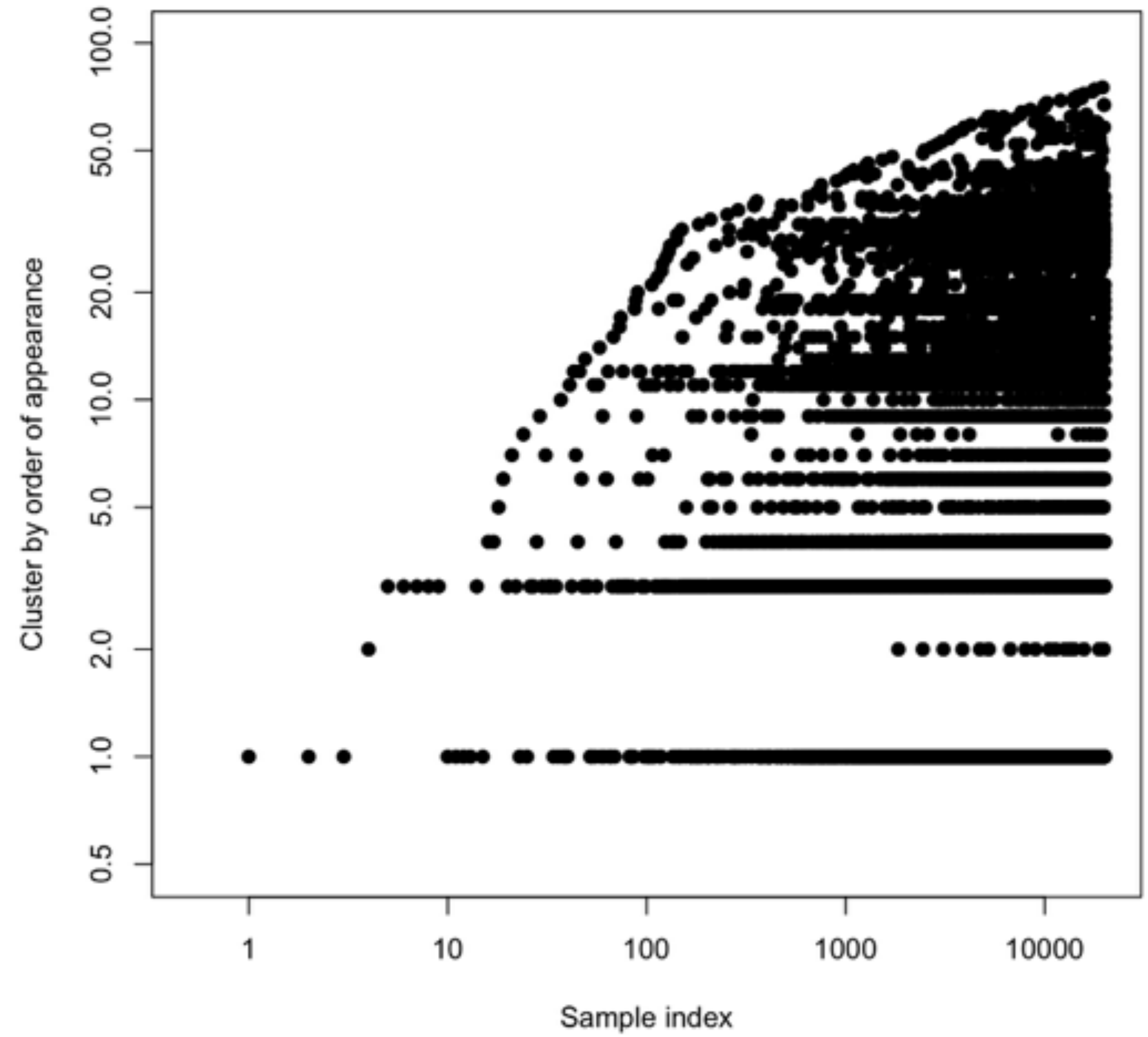


[Gnedin, et al 2007, Pitman, Yor 1997

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:

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



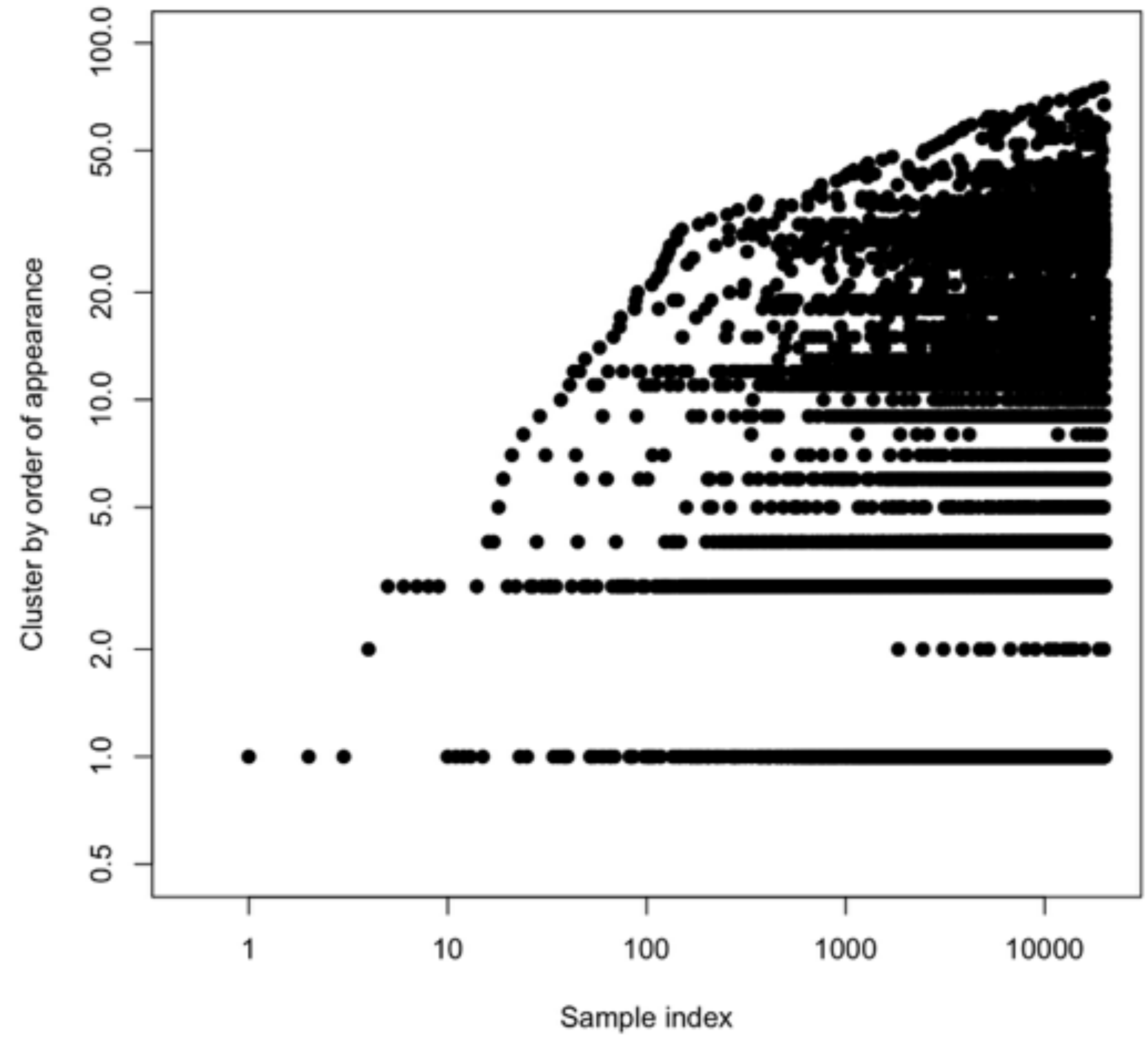
[Gnedin, et al 2007, Pitman, Yor 1997, Goldwater et al 2005, Teh 2006

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:

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

  - related to Zipf's law (ranked frequencies)



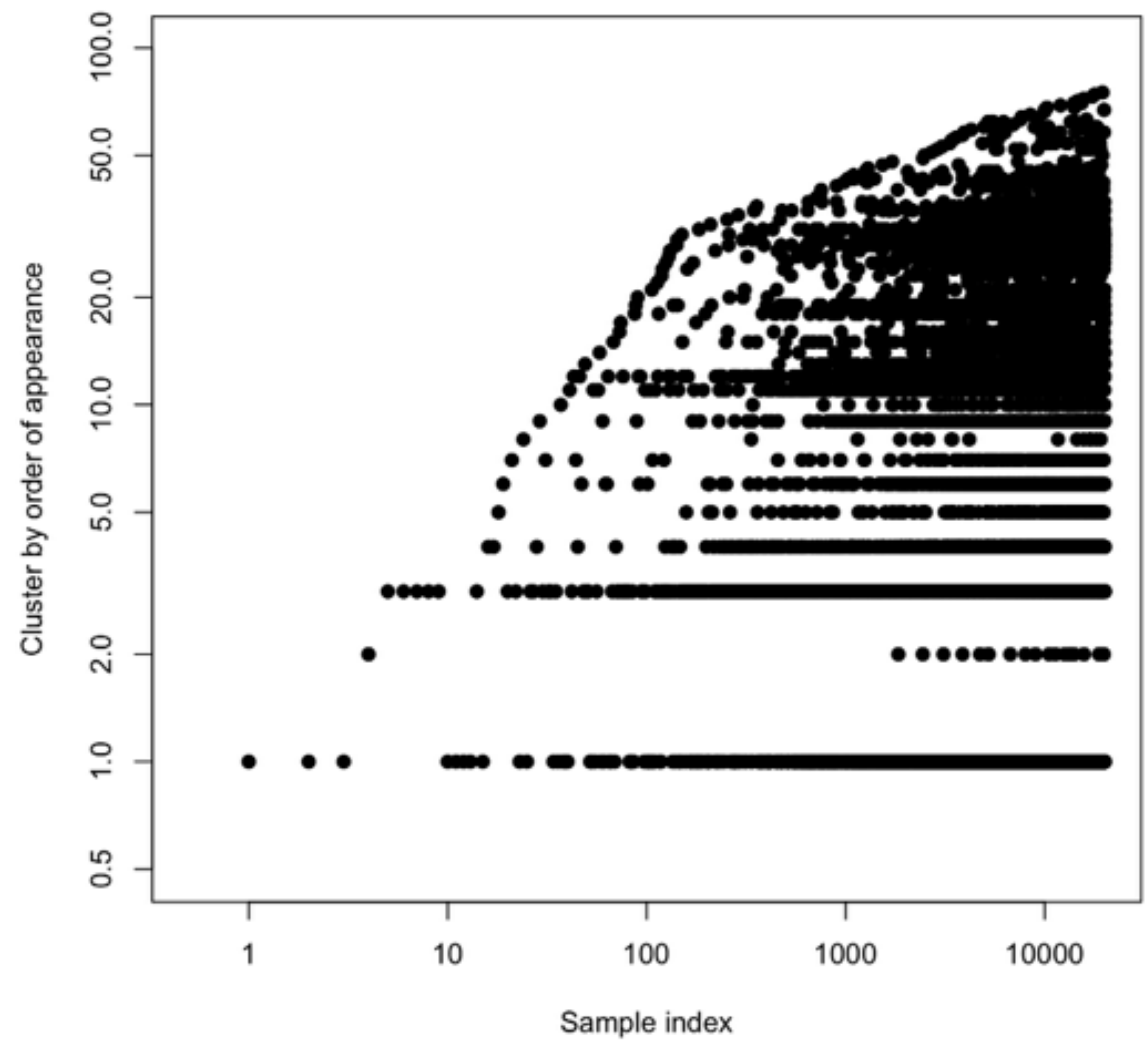
[Gnedin, et al 2007, Pitman, Yor 1997, Goldwater et al 2005, Teh 2006

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:

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

  - related to Zipf's law (ranked frequencies)



[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

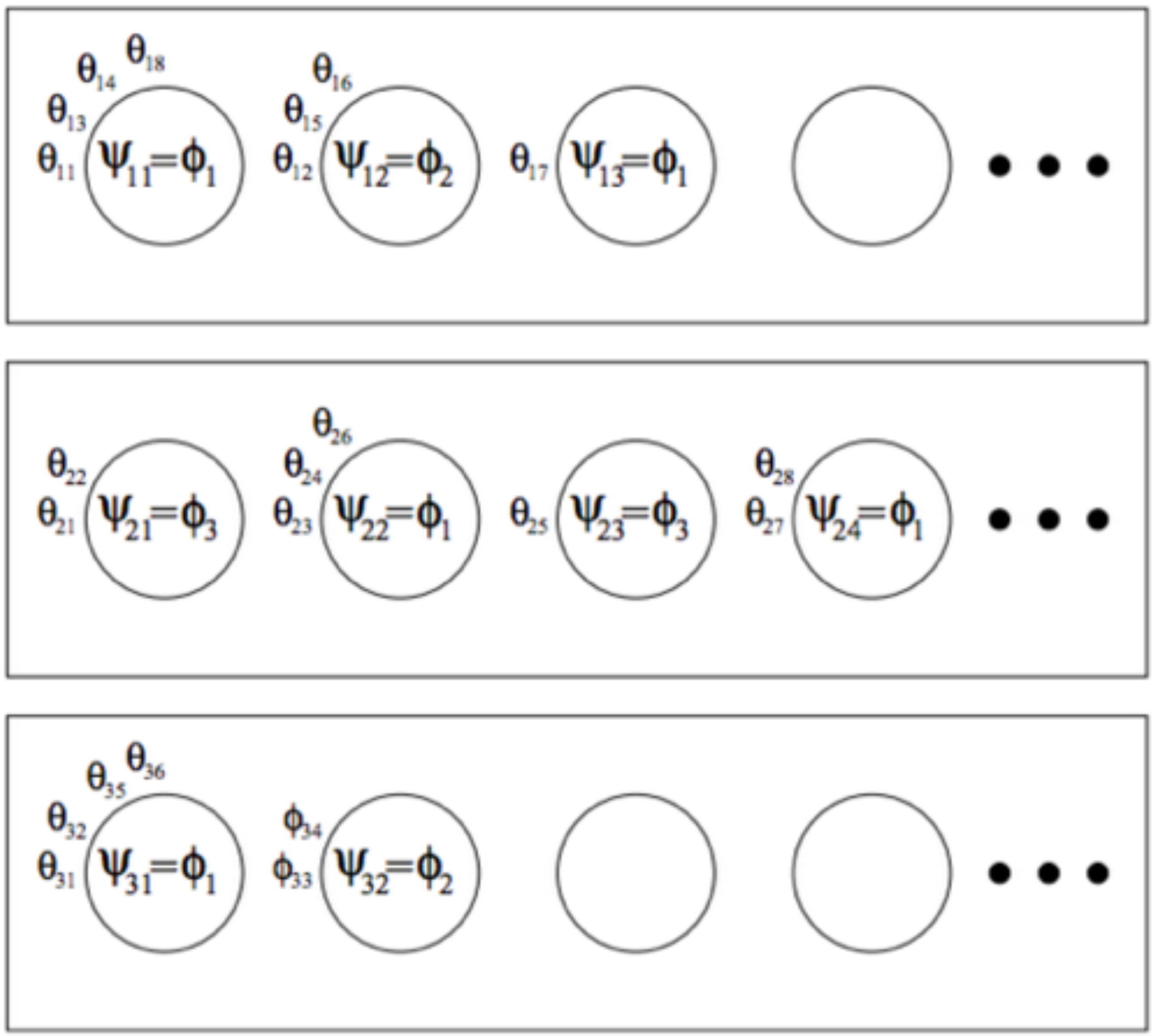
[Teh et al 2006, Rodríguez et al 2008

Hierarchies

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008

Hierarchies

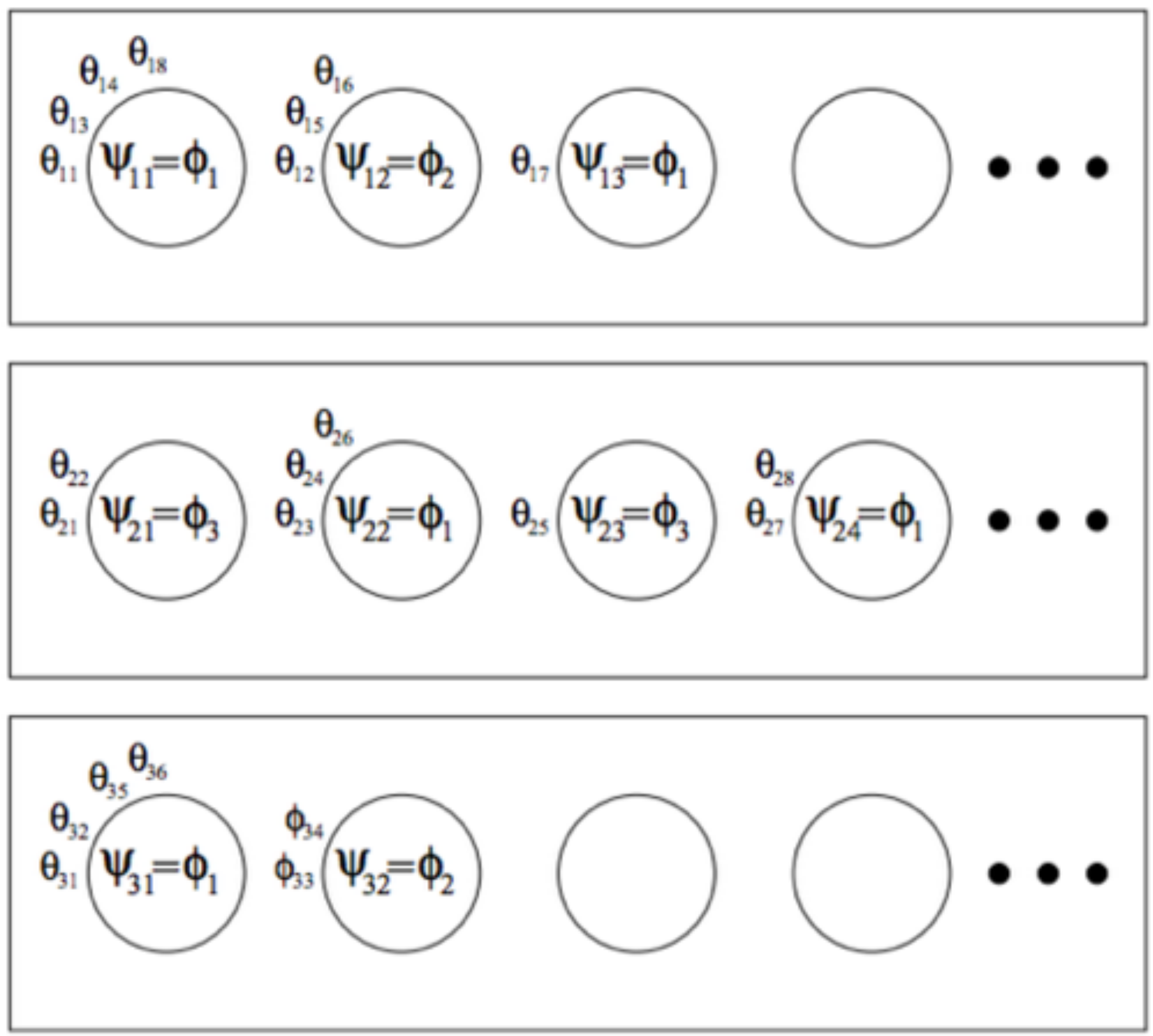


[Teh et al 2006]

- Hierarchical Dirichlet process
- Chinese restaurant franchise

[Teh et al 2006, Rodríguez et al 2008

Hierarchies

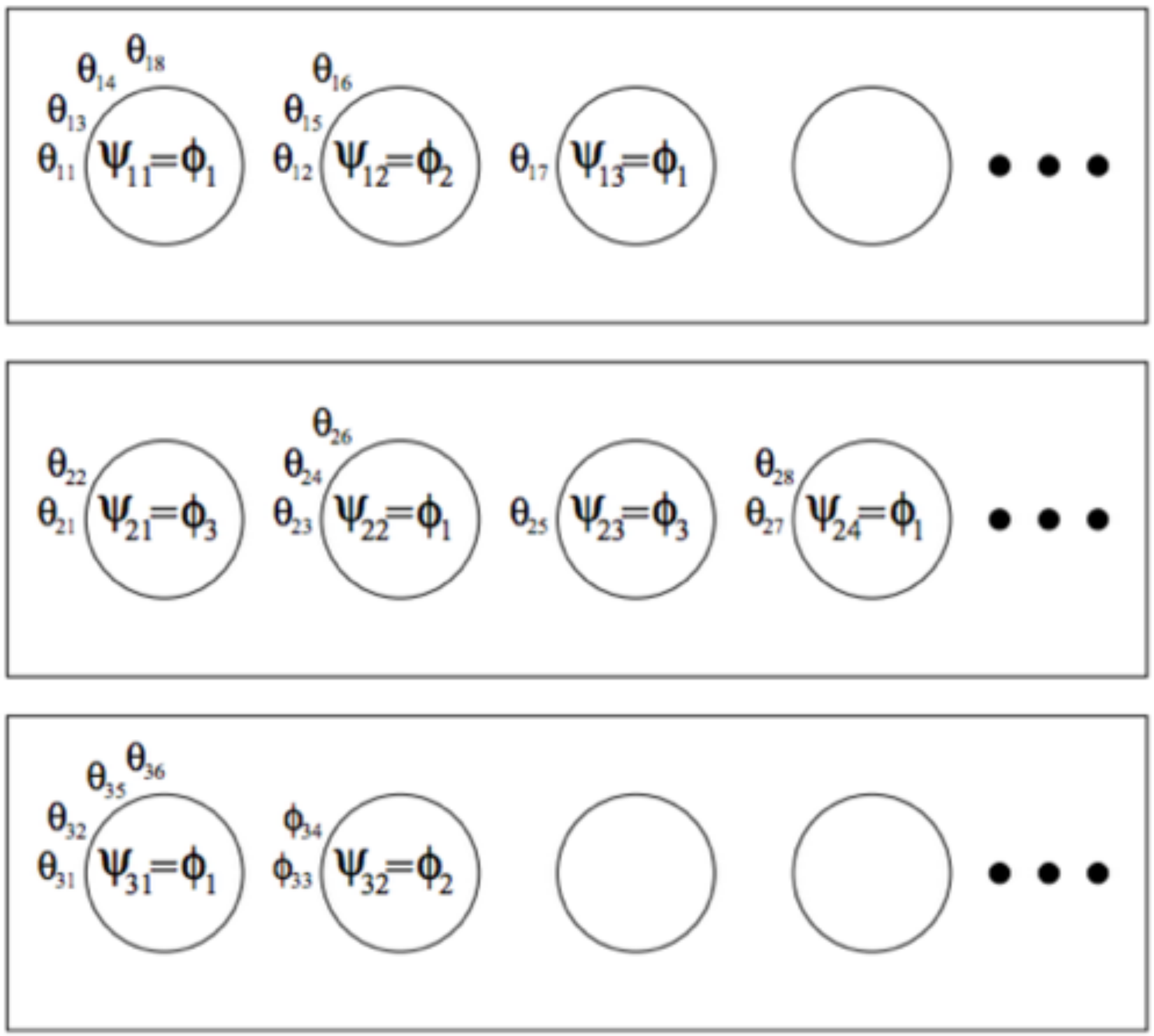


[Teh et al 2006]

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

[Teh et al 2006, Rodríguez et al 2008

Hierarchies



[Teh et al 2006]

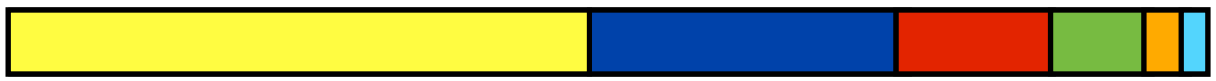
- Hierarchical Dirichlet process
- Chinese restaurant franchise
- Hierarchical beta process

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

De Finetti mixing measures

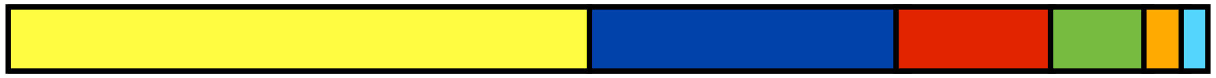
De Finetti mixing measures

- Clustering: Kingman paintbox



De Finetti mixing measures

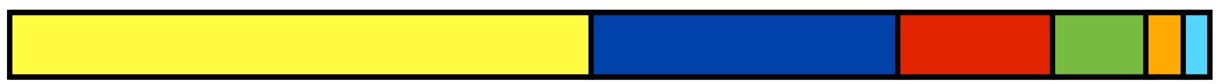
- Clustering: Kingman paintbox



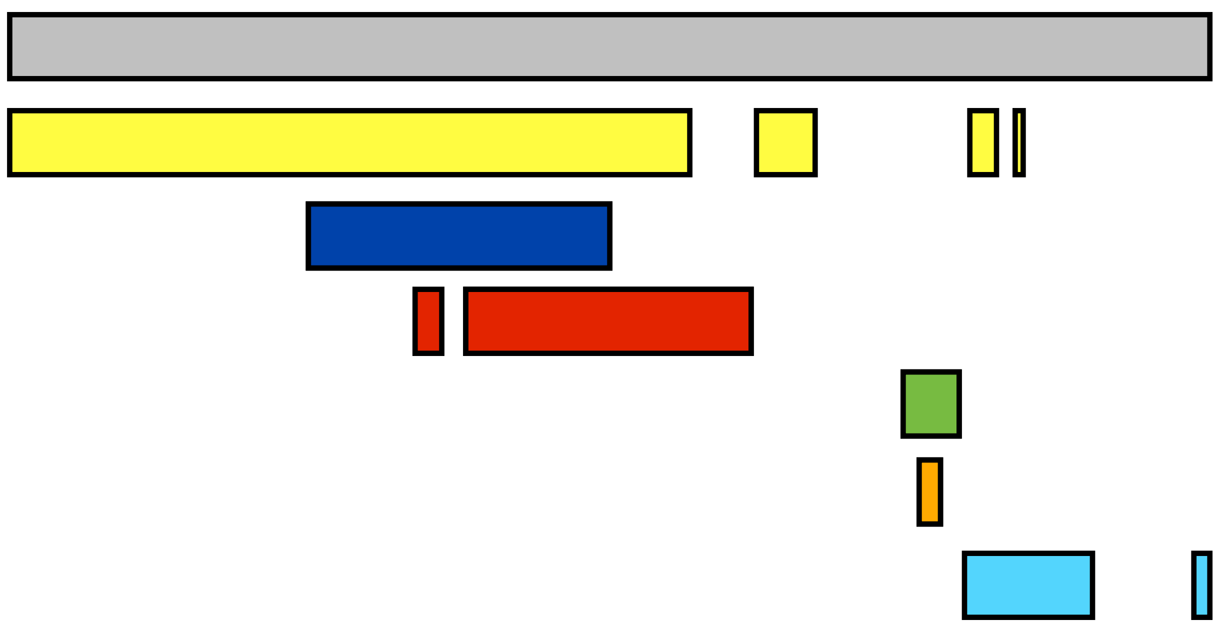
[Kingman 1978

De Finetti mixing measures

- Clustering: Kingman paintbox



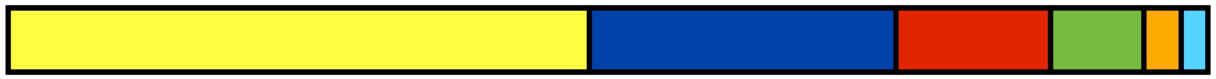
- Feature allocation: Feature paintbox



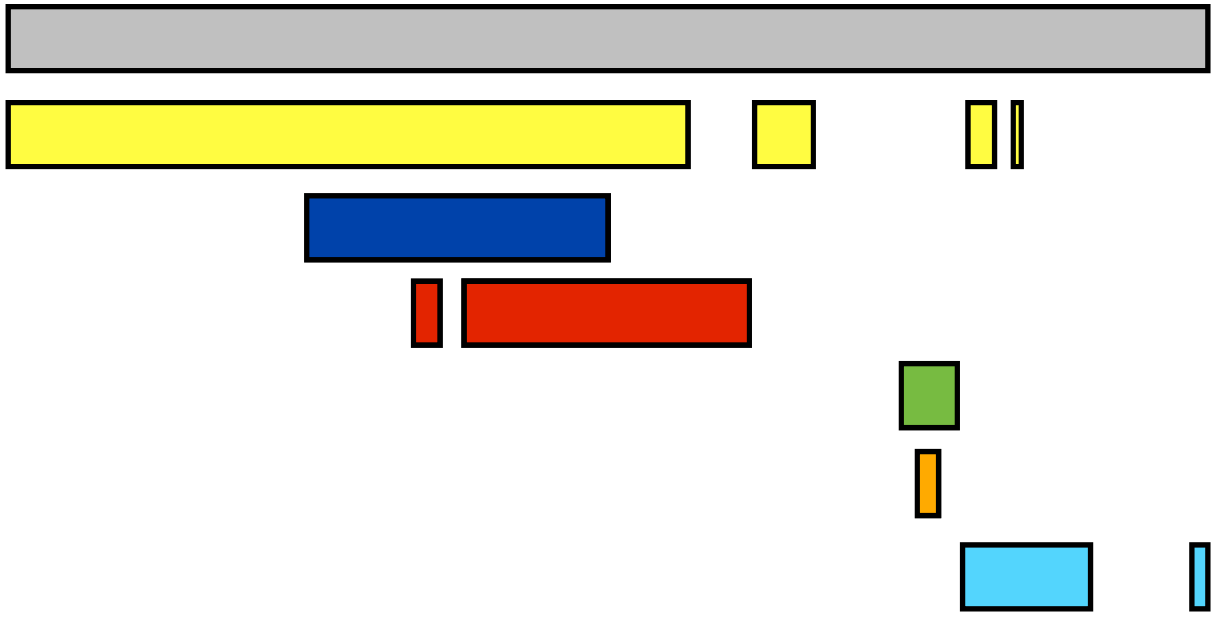
[Kingman 1978

De Finetti mixing measures

- Clustering: Kingman paintbox



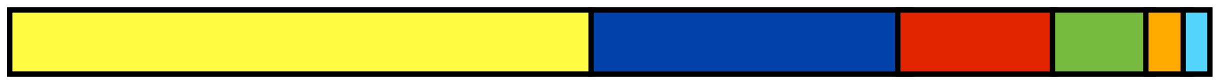
- Feature allocation: Feature paintbox



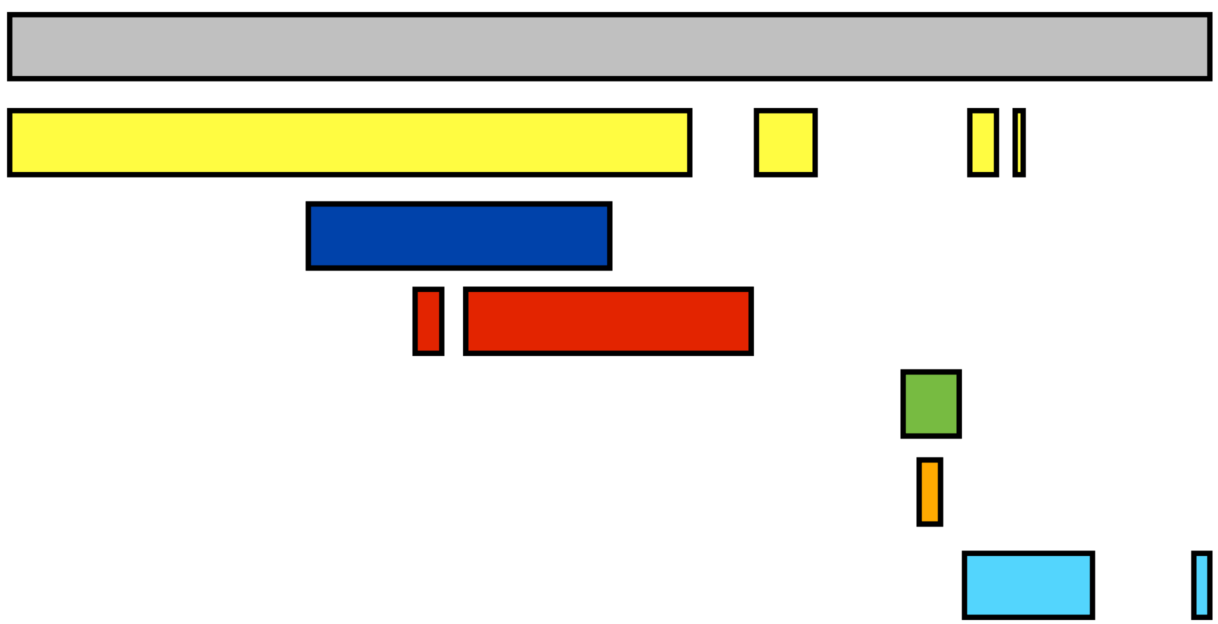
[Kingman 1978, Broderick, Pitman, Jordan 2013]

De Finetti mixing measures

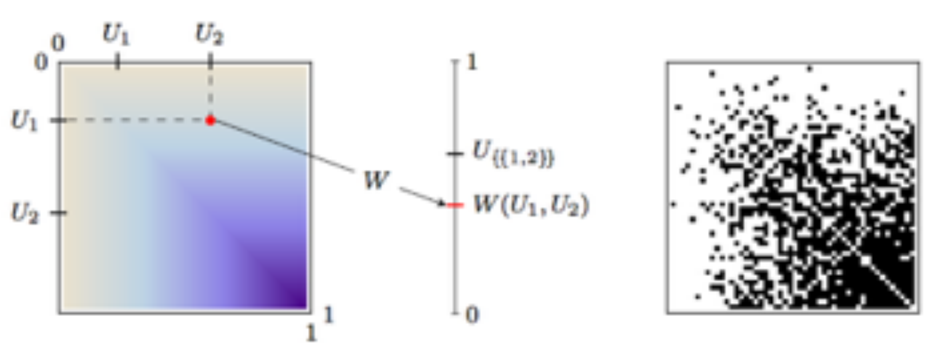
- Clustering: Kingman paintbox



- Feature allocation: Feature paintbox



- Graphs/networks: Aldous-Hoover theorem

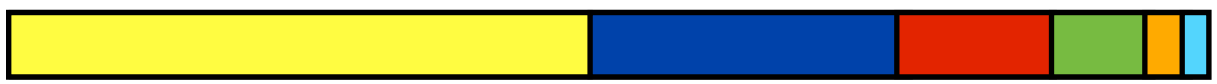


[Lloyd 2012]

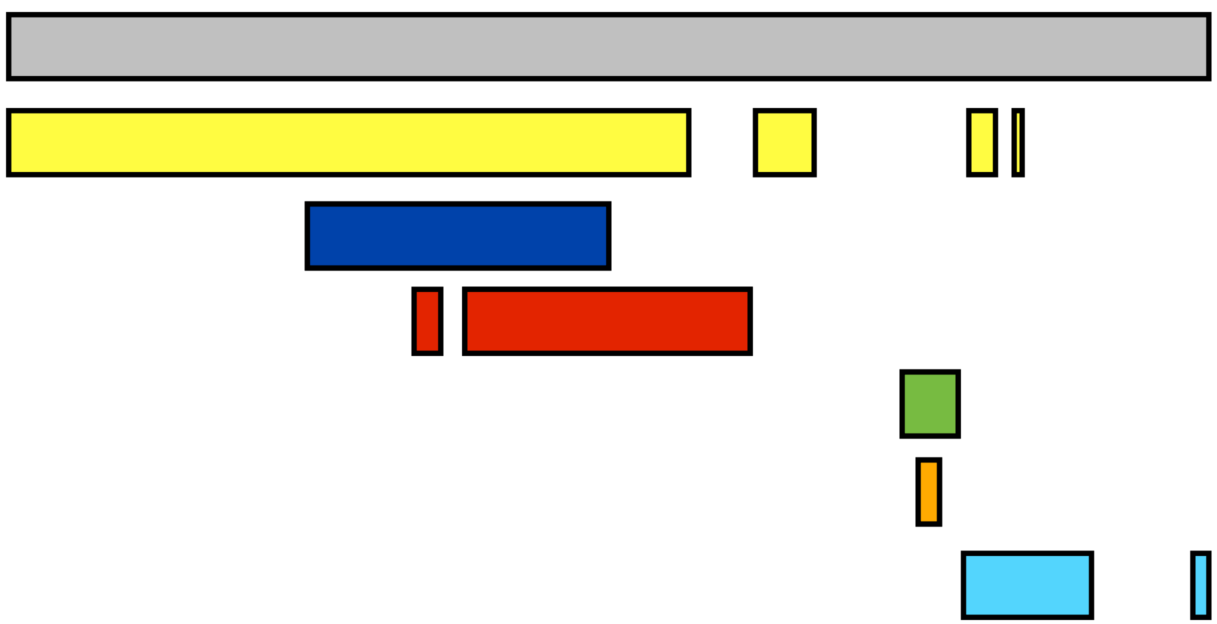
[Kingman 1978, Broderick, Pitman, Jordan 2013,

De Finetti mixing measures

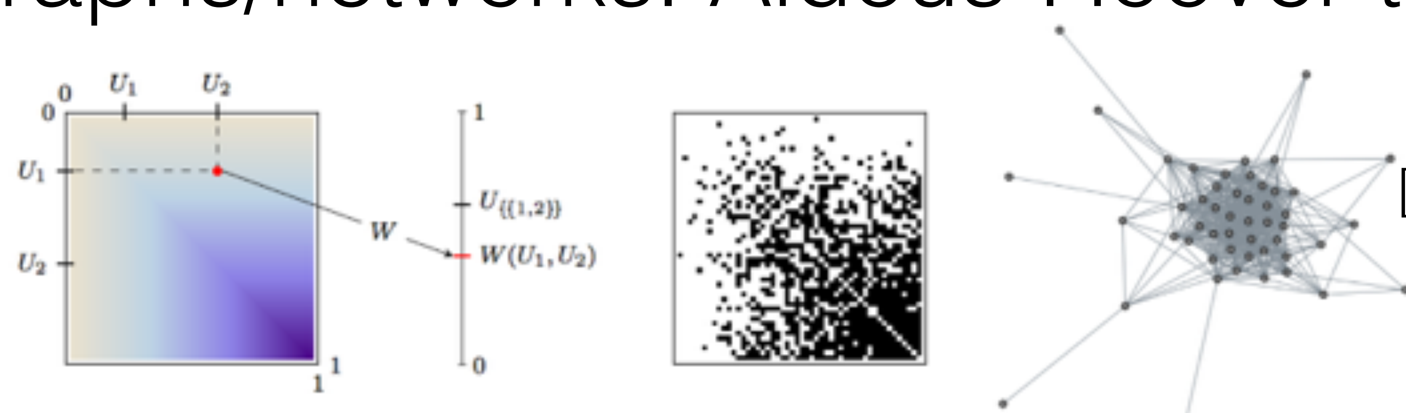
- Clustering: Kingman paintbox



- Feature allocation: Feature paintbox



- Graphs/networks: Aldous-Hoover theorem



[Lloyd 2012]

[Kingman 1978, Broderick, Pitman, Jordan 2013, Aldous 1983, Hoover 1979, Orbanz, Roy 2015]

Conjugacy & Poisson point processes

Orbanz 2009, Orbanz 2010

Conjugacy & Poisson point processes

- Beta process, Bernoulli process (Indian buffet)

Orbanz 2009, Orbanz 2010

Conjugacy & Poisson point processes

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

Orbanz 2009, Orbanz 2010

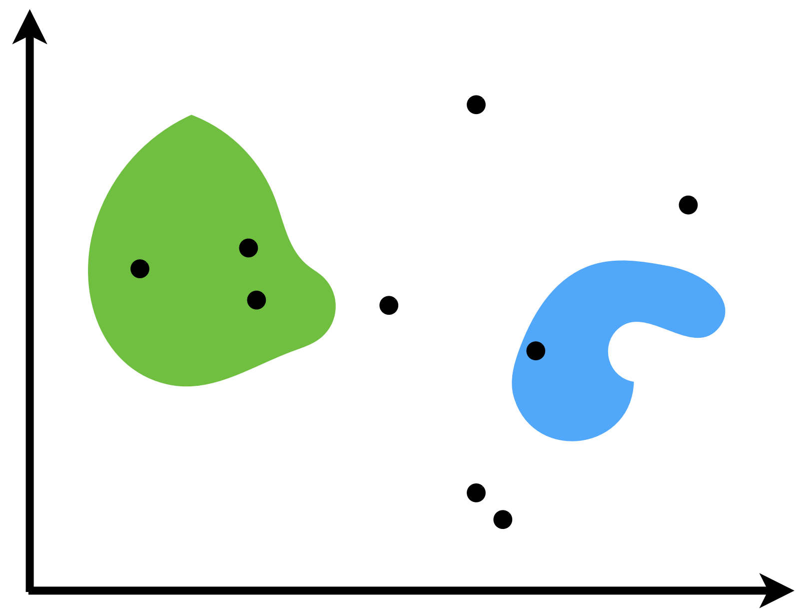
Conjugacy & Poisson point processes

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

Orbanz 2009, Orbanz 2010

Conjugacy & Poisson point processes

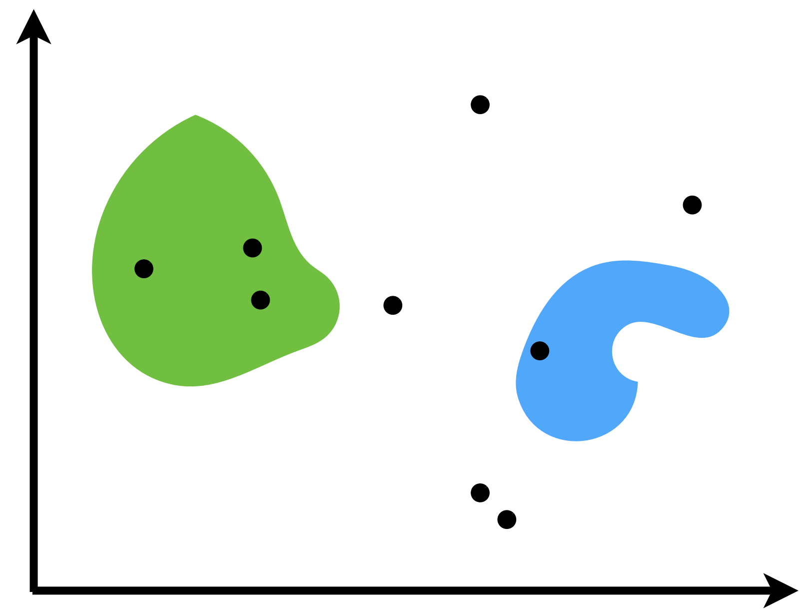
- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process



Orbanz 2009, Orbanz 2010

Conjugacy & Poisson point processes

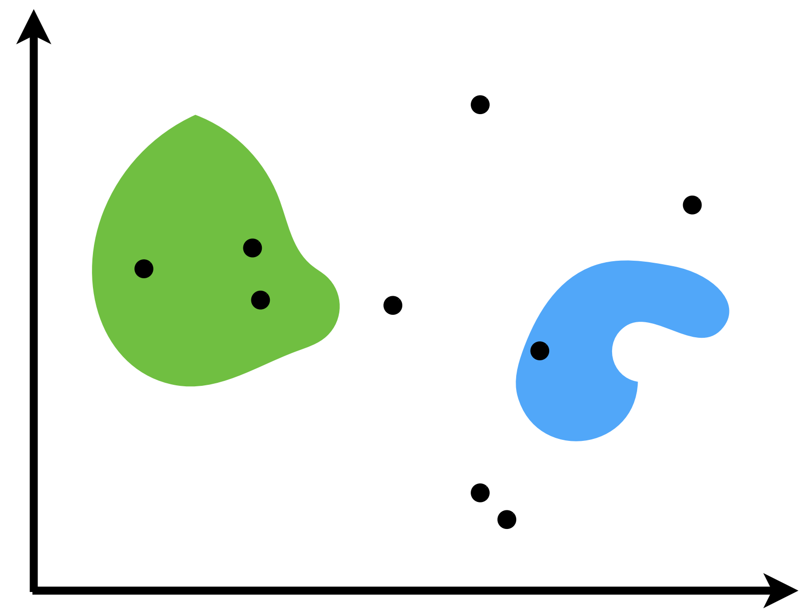
- Beta process, Bernoulli process (Indian buffet)
- Gamma process, Poisson likelihood process (DP, CRP)
- Beta process, negative binomial process



[Kingman 1992, Orbanz 2009, Orbanz 2010

Conjugacy & Poisson point processes

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

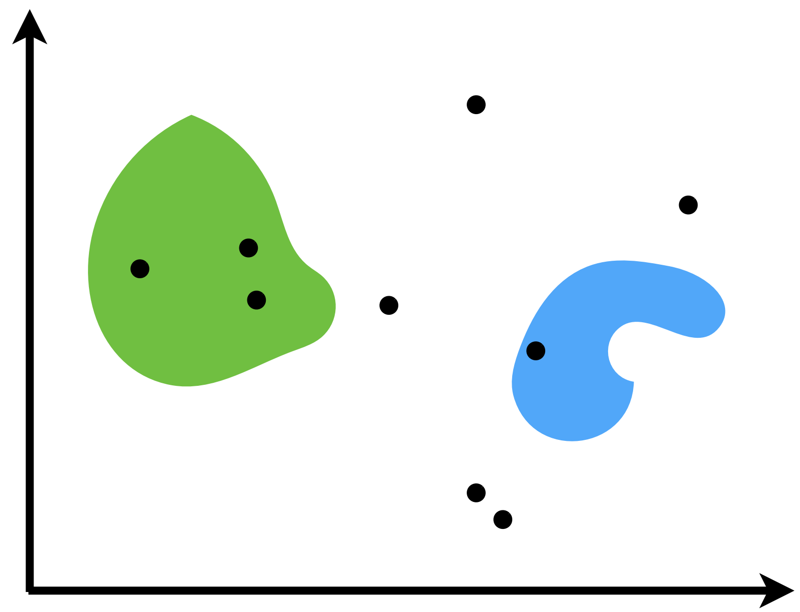


- Posteriors, conjugacy, and exponential families for completely random measures

[Kingman 1992, Orbanz 2009, Orbanz 2010

Conjugacy & Poisson point processes

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



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

[Kingman 1992, Orbanz 2009, Orbanz 2010, 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

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

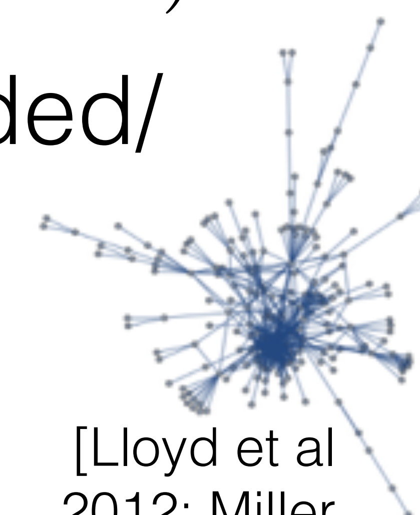
- Not parametric (i.e. not finite parameter, unbounded/ growing/infinite number of parameters)

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)

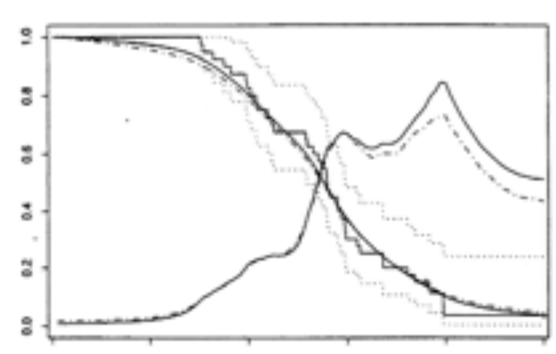


[Lloyd et al 2012; Miller et al, 2010]

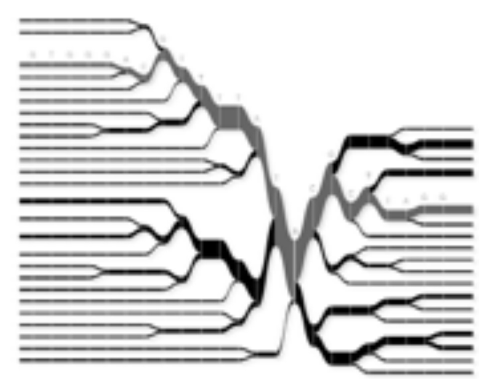
[wikipedia.org]

[Ed Bowlby, NOAA]

[Fox, et al 2014]

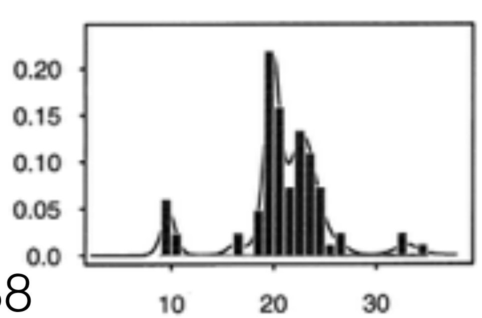


[Arjas, Gasbarra 1994]



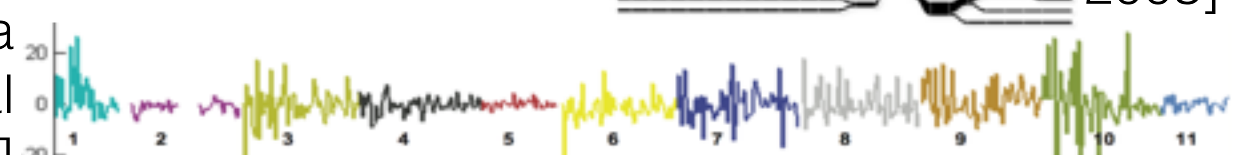
[Ewens, 1972; Hartl, Clark 2003]

[Sudderth, Jordan 2009]



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

[Saria et al 2010]



References (page 1 of 5)

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.

DM Blei and MI Jordan. Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, 2006.

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

T Campbell*, J Huggins*, and T Broderick. Truncated random measures. ArXiv:1603.00861, 2016.

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

References (page 2 of 5)

MD Escobar and M West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 1995. W Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 1972.

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

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

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

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.

RJ Giordano, T Broderick, and MI Jordan. Linear response methods for accurate covariance estimates from mean field variational Bayes. *NIPS*, 2015.

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.

References (page 3 of 5)

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.

M Kalli, JE Griffin, and SG Walker. Slice sampling mixture models. *Statistics and Computing*, 2011.

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.

References (page 4 of 5)

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.

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.

References (page 4 of 5)

Time Magazine. Retrieved from: http://time.com/4359750/peacock-spiders-discovered-photos/

J Wakeley. *Coalescent Theory: An Introduction*, Chapter 3, 2008.

SG Walker. Sampling the Dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation*, 2007.

M West, P Müller, and MD Escobar. Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation. *Aspects of Uncertainty*, 1994.