Nonparametric Bayesian Statistics: Part II

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

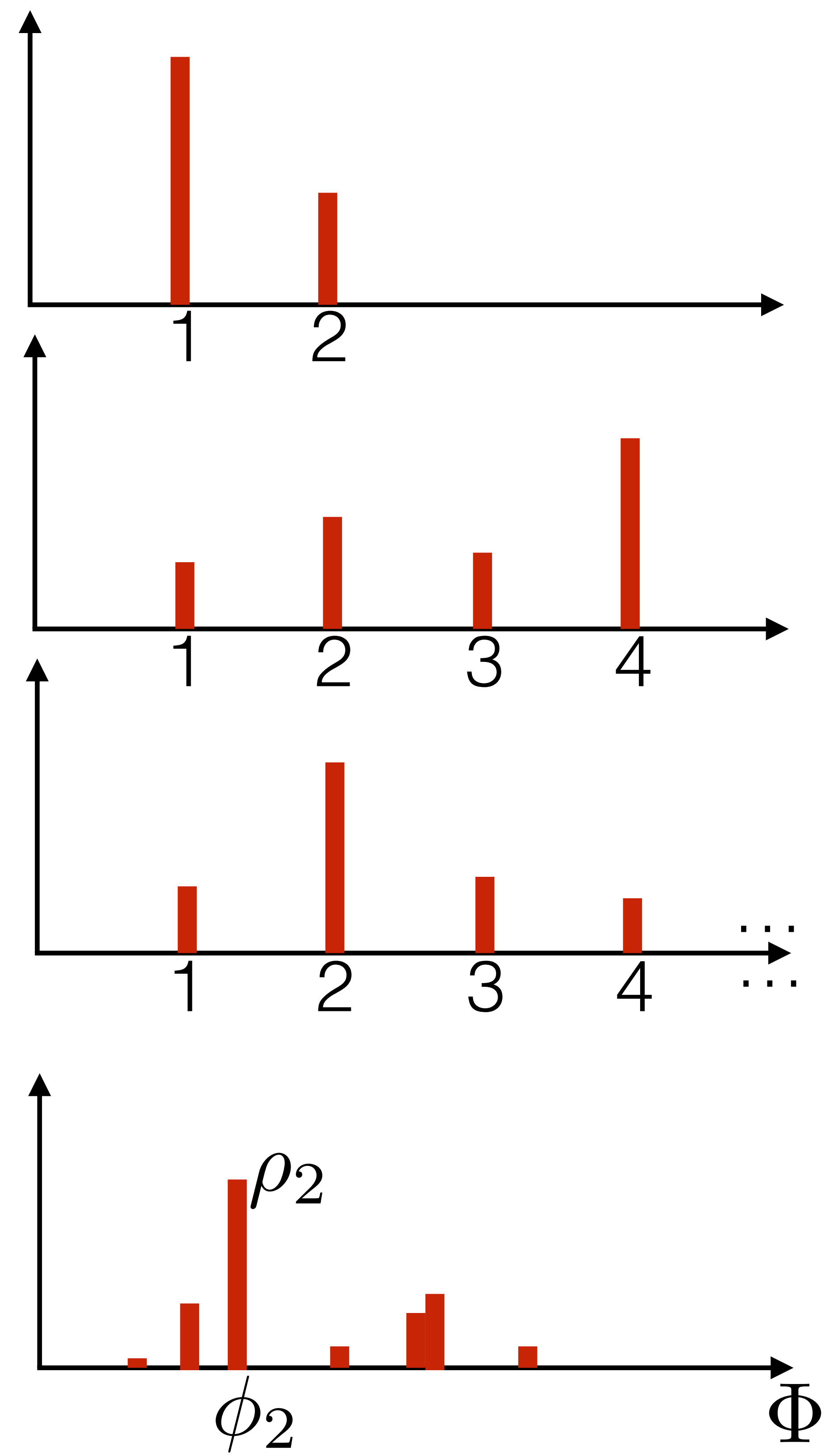
Outline

- Dirichlet process
  - Background for intuition
  - Generative model
  - What does a growing/infinite number of parameters really mean (in Nonparametric Bayes)?
- Chinese restaurant process
- Inference
- Venture further into the wild world of Nonparametric Bayesian statistics

[slides, code: www.tamarabroderick.com/tutorials.html]

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



[Ferguson 1973]

Dirichlet process mixture model

Dirichlet process mixture model

- Gaussian mixture model

Dirichlet process mixture model

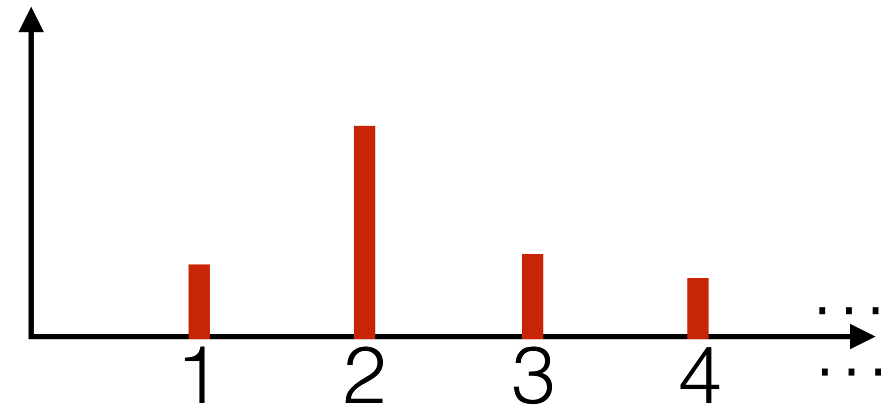
- Gaussian mixture model

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

Dirichlet process mixture model

- Gaussian mixture model

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

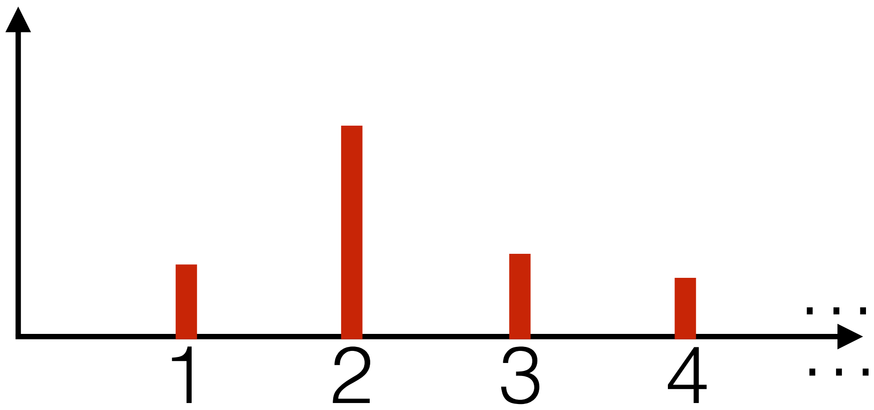


Dirichlet process mixture model

- Gaussian mixture model

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

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

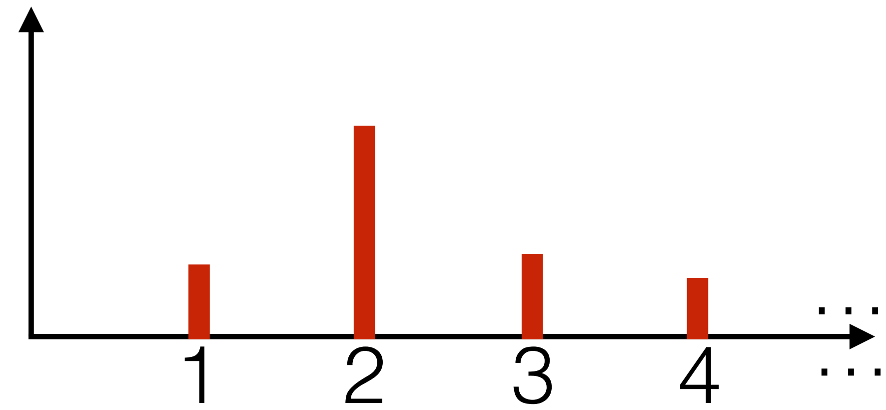


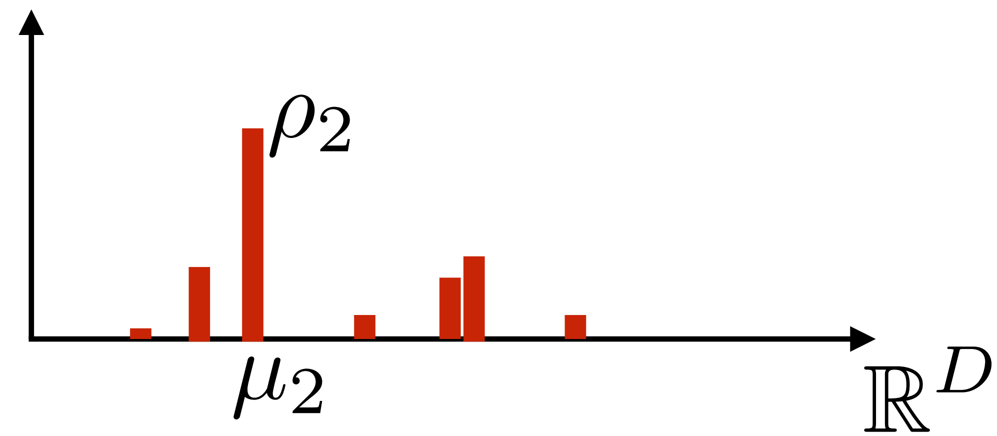
Dirichlet process mixture model

- Gaussian mixture model

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

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





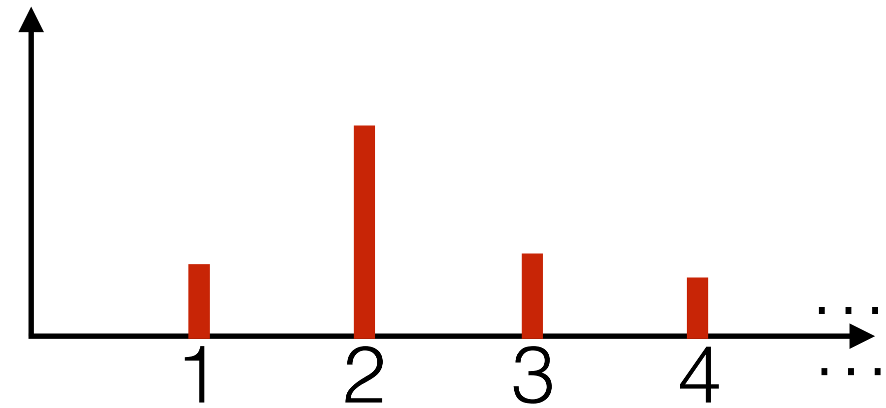
Dirichlet process mixture model

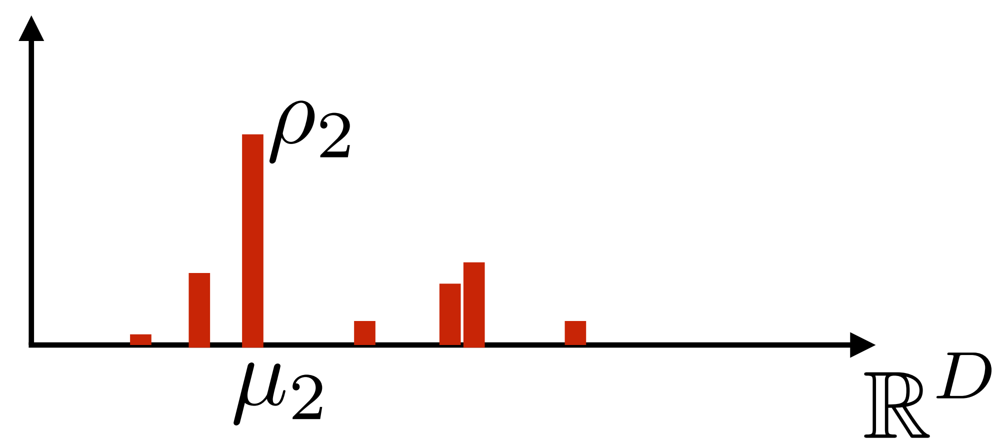
- Gaussian mixture model

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

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

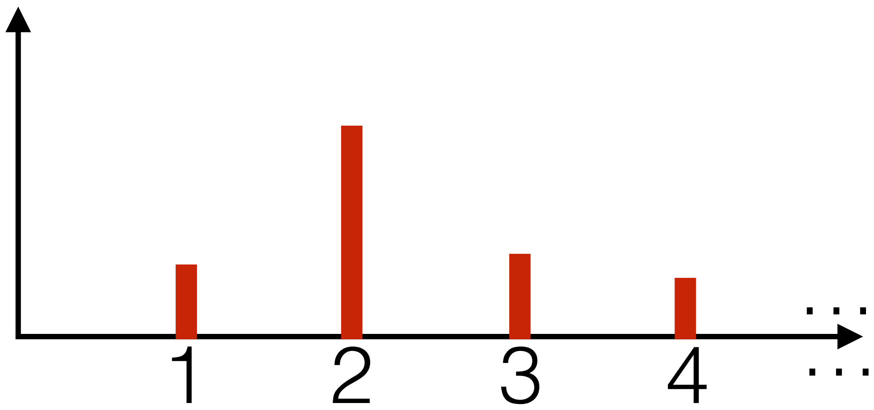
  - i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k}$





Dirichlet process mixture model

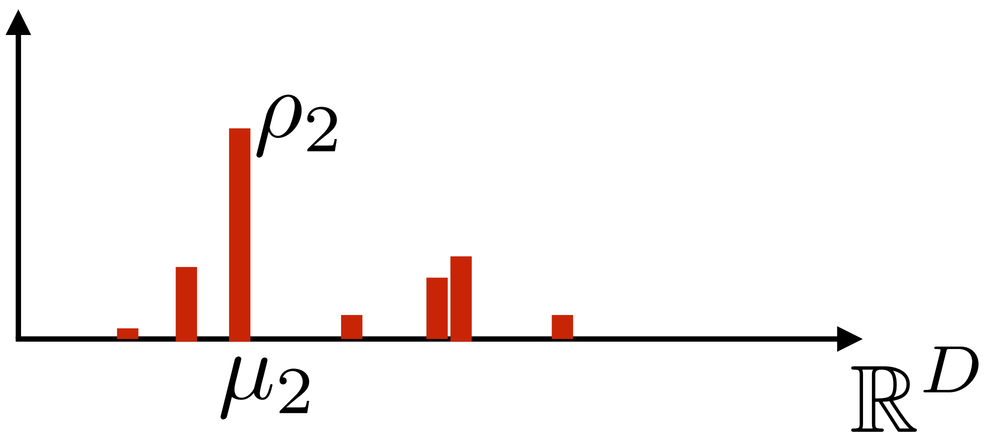
- Gaussian mixture model



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

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

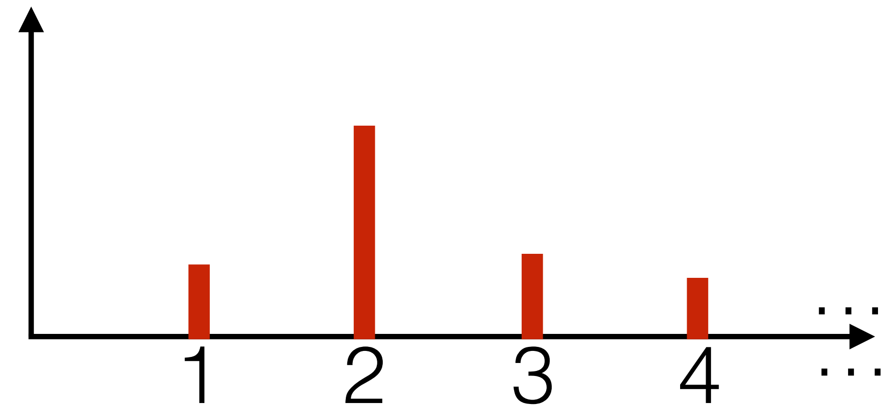
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \stackrel{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



Dirichlet process mixture model

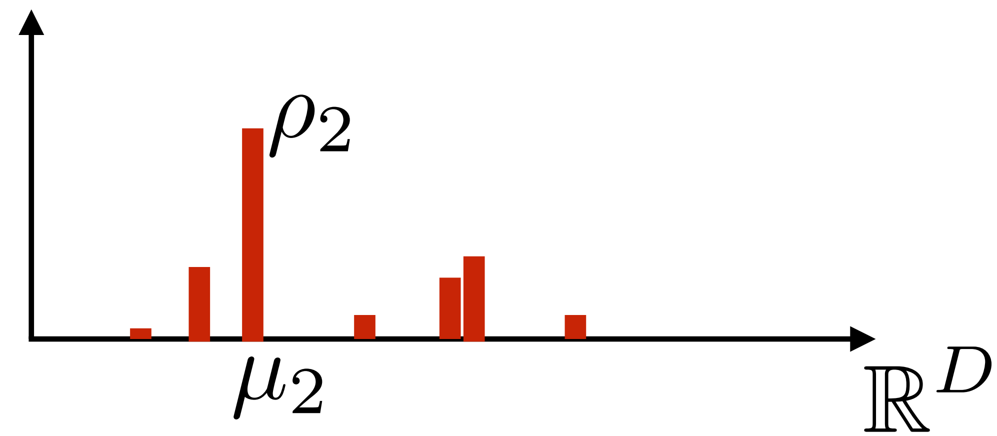
- Gaussian mixture model

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



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

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



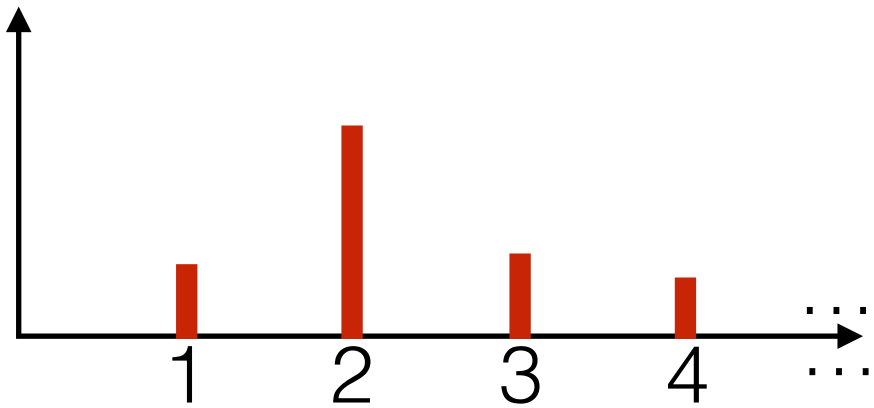
$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

Dirichlet process mixture model

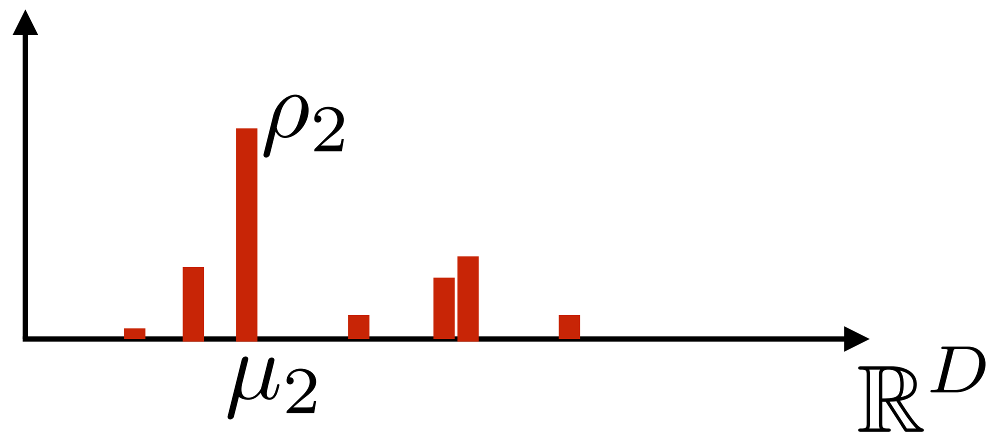
- Gaussian mixture model

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

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



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



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

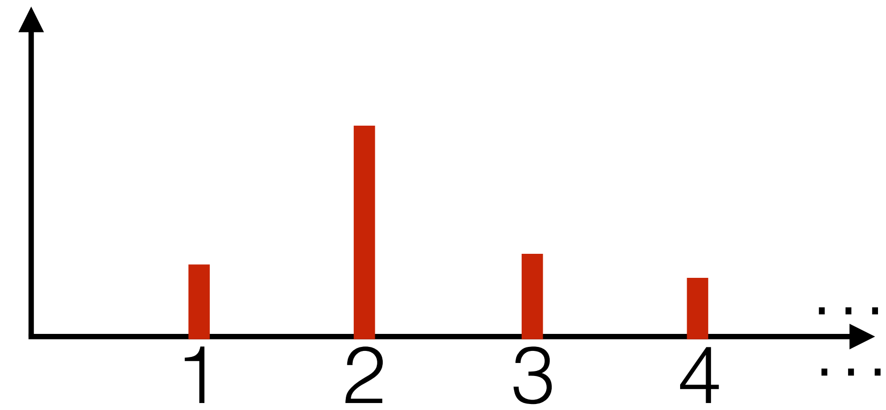
$$\mu_n^* = \mu_{z_n}$$

Dirichlet process mixture model

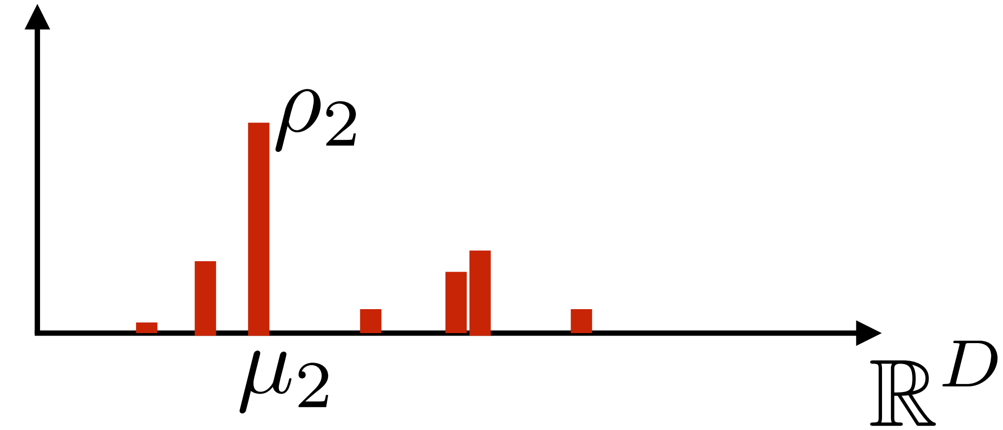
- Gaussian mixture model

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

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

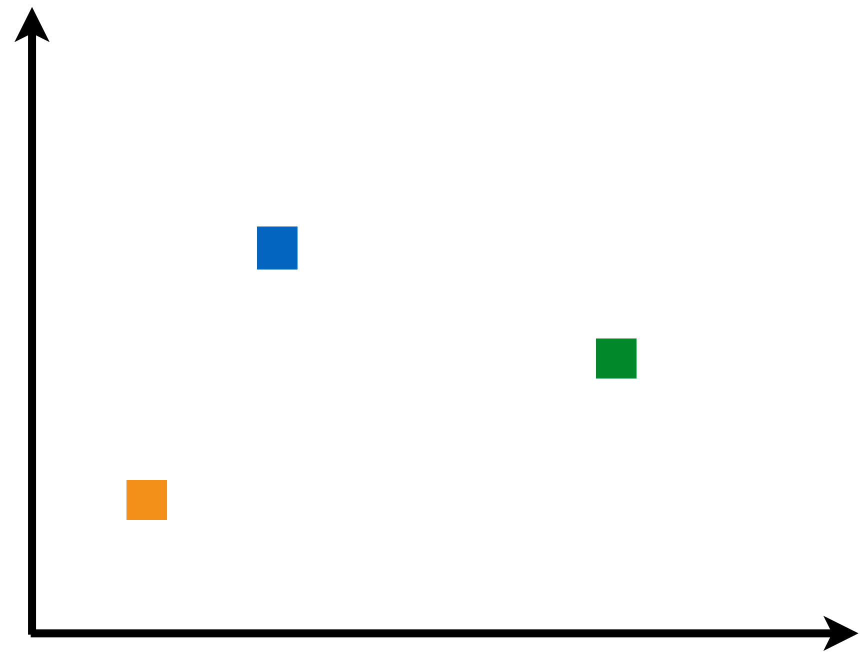


- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

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

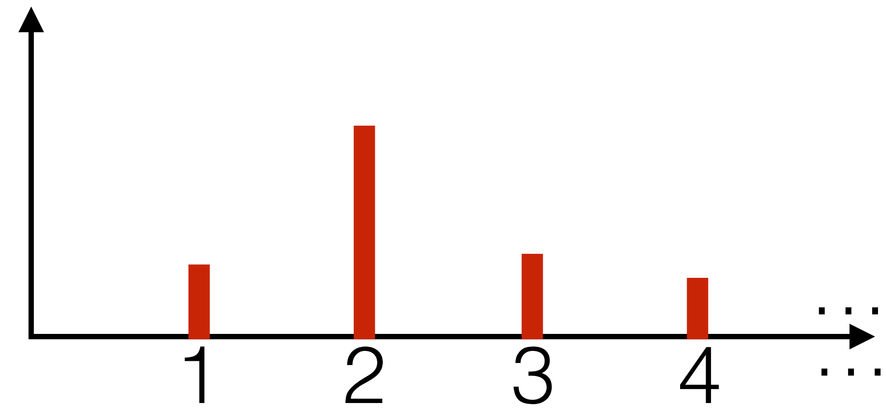


Dirichlet process mixture model

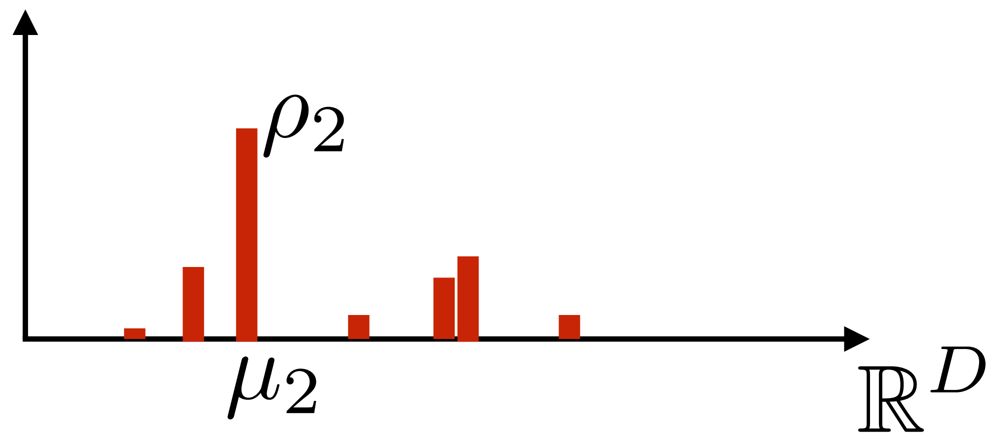
- Gaussian mixture model

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

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



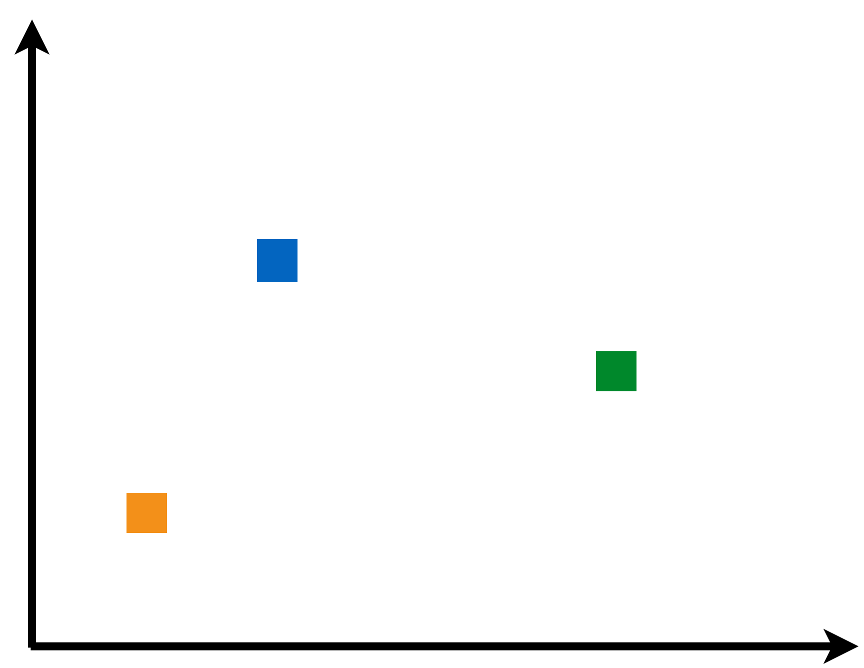
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

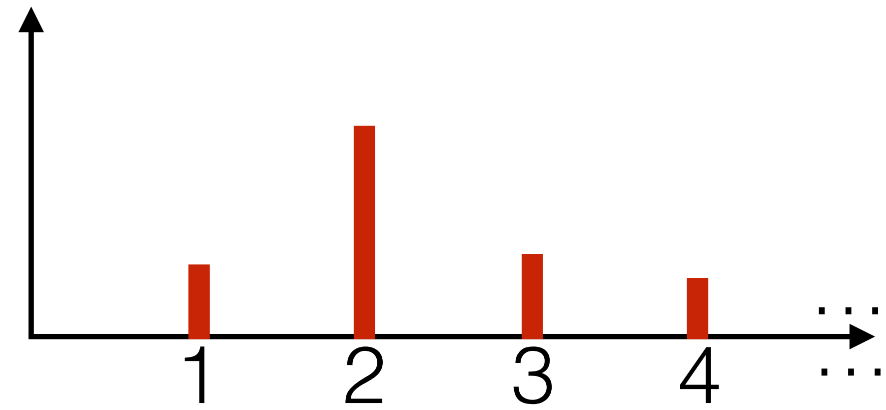
$$\mu_n^* = \mu_{z_n}$$

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



Dirichlet process mixture model

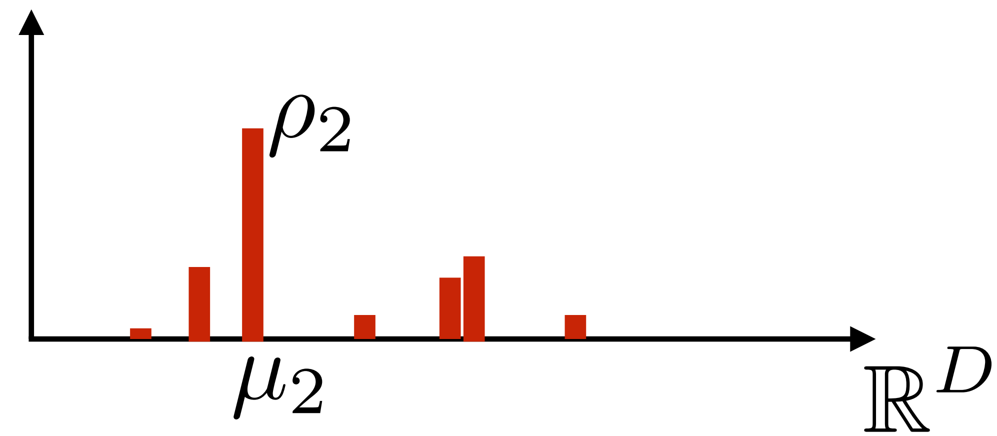
- Gaussian mixture model



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

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

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

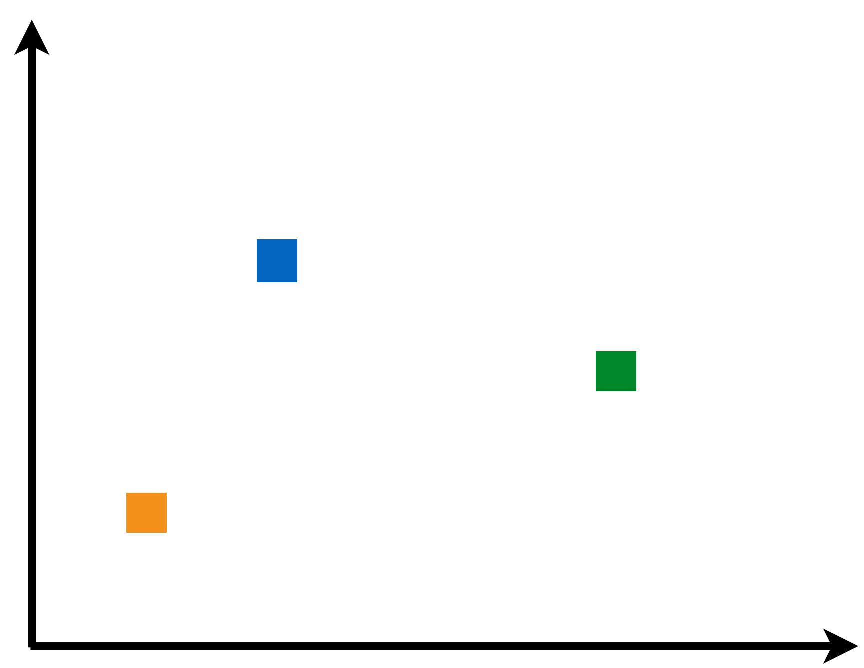


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

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

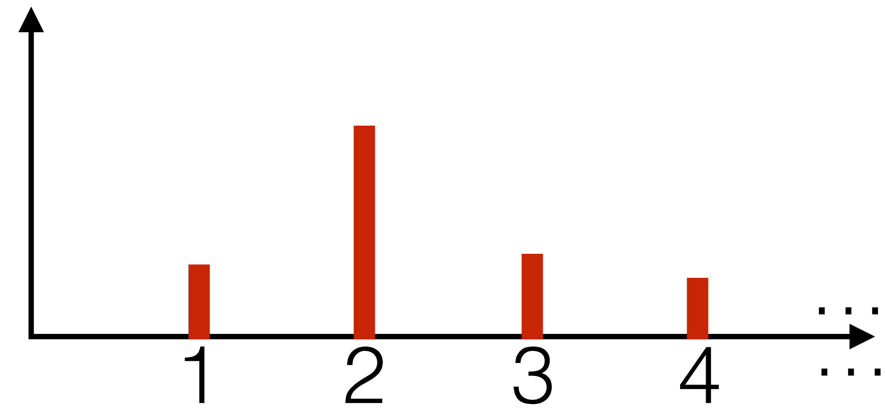
- i.e. $\mu_n^* \overset{iid}{\sim} G$

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



Dirichlet process mixture model

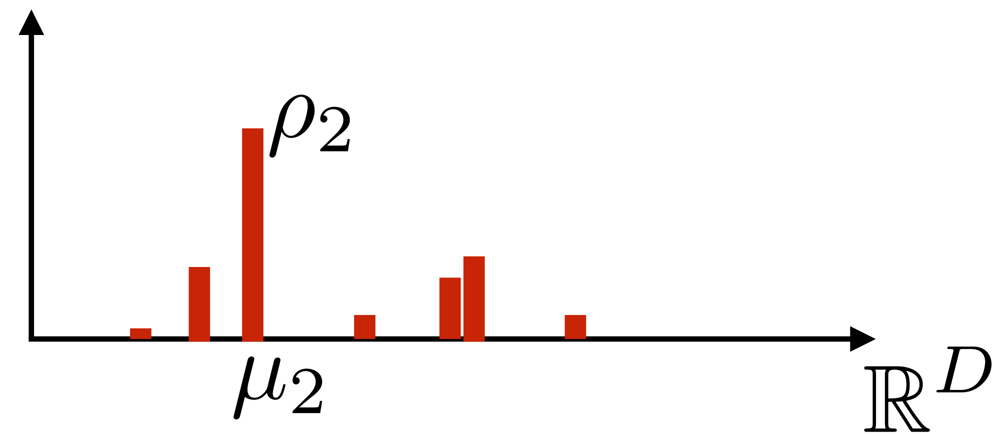
- Gaussian mixture model



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

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

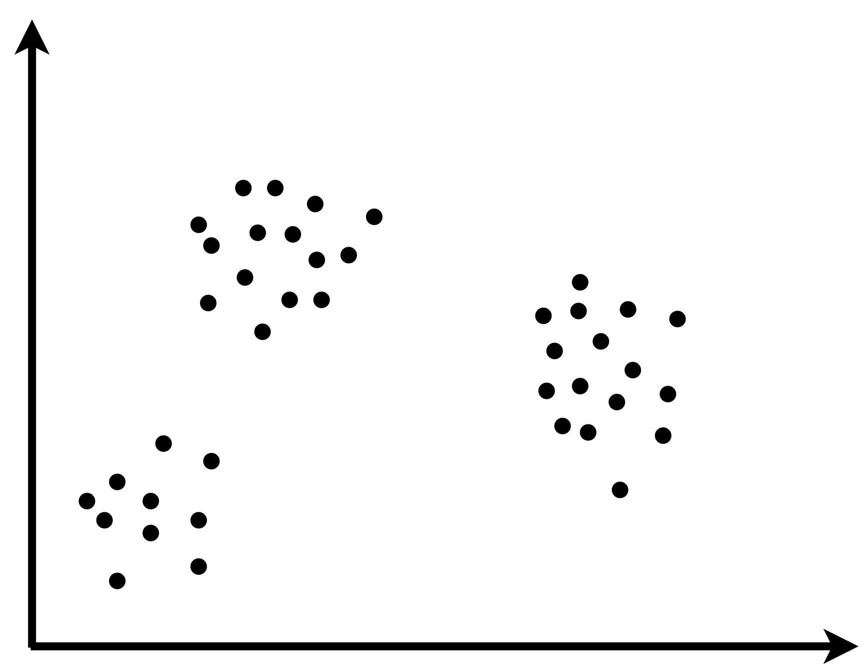
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

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

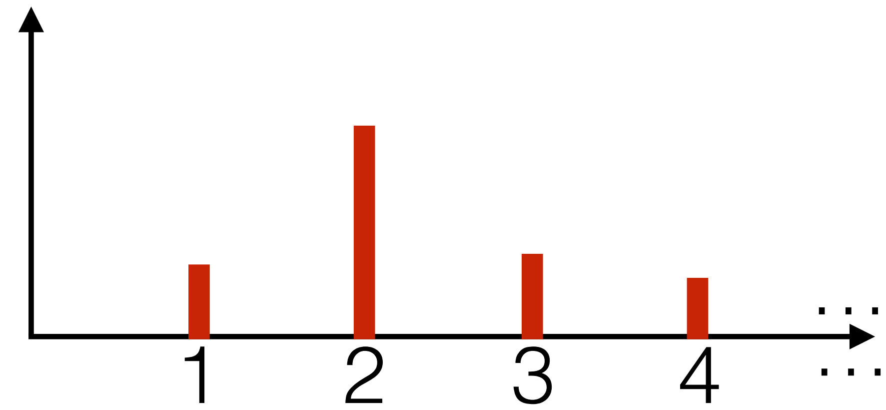
- i.e. $\mu_n^* \overset{iid}{\sim} G$



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

Dirichlet process mixture model

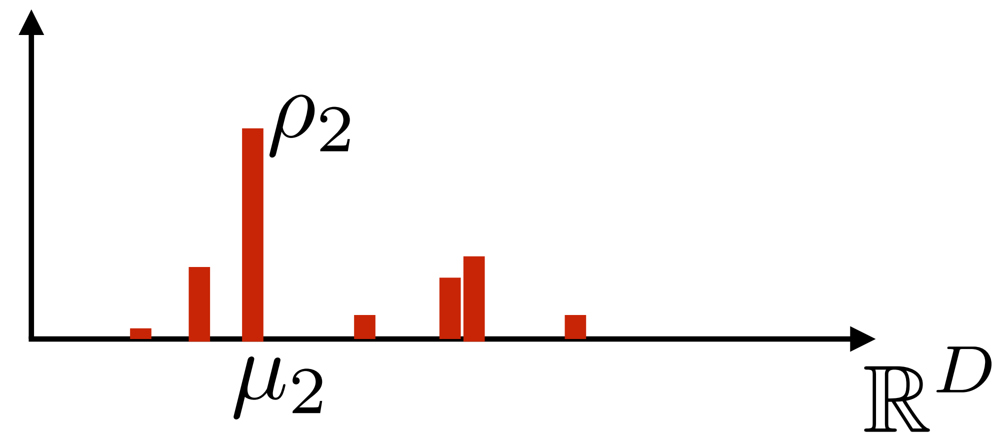
- Gaussian mixture model



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

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

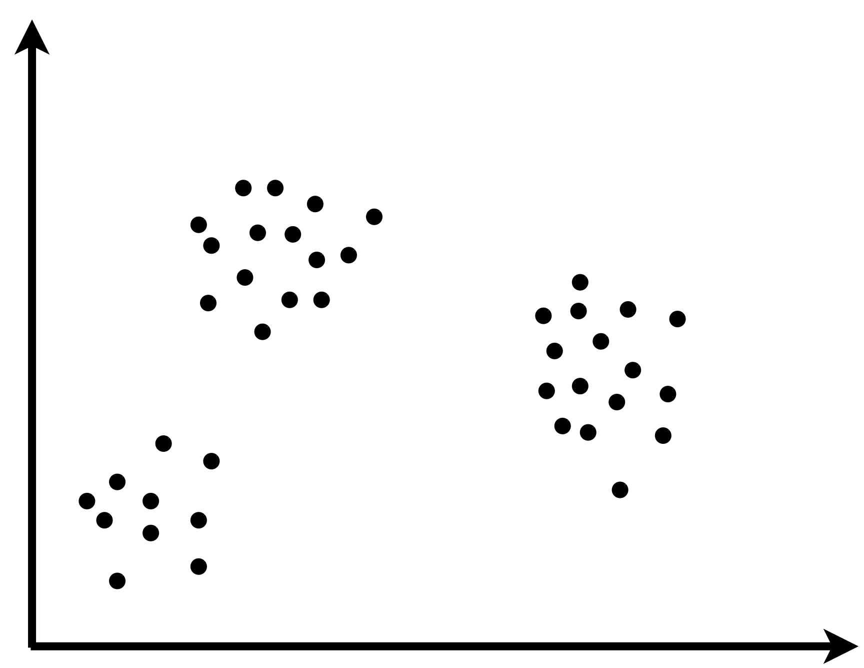
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \overset{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

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

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



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

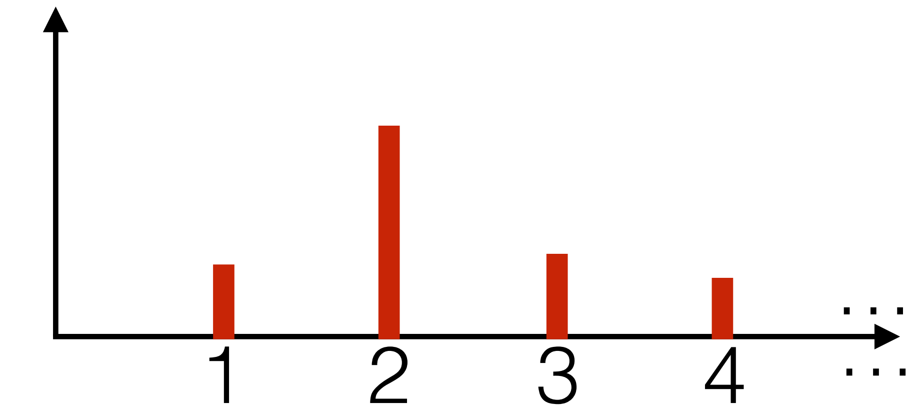
[demo]

Dirichlet process mixture model

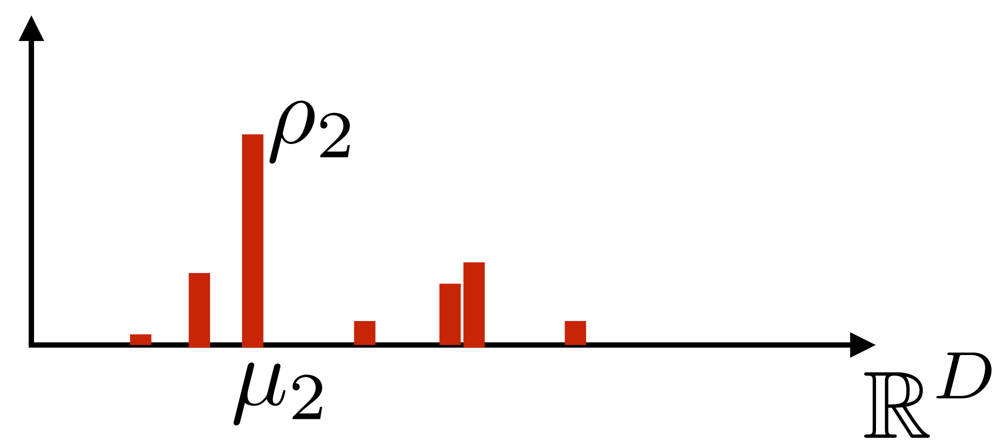
- More generally

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

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



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

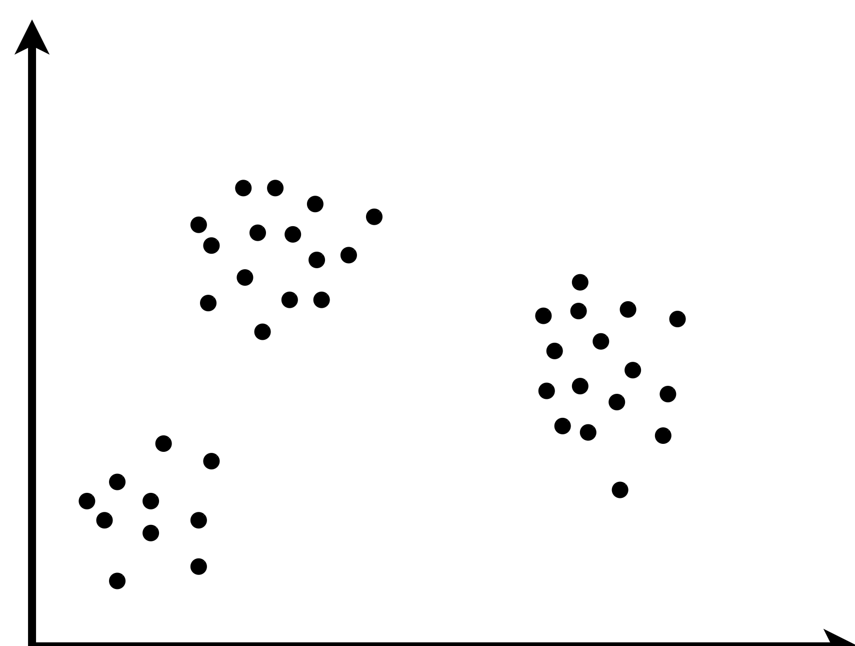


$$z_n \stackrel{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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

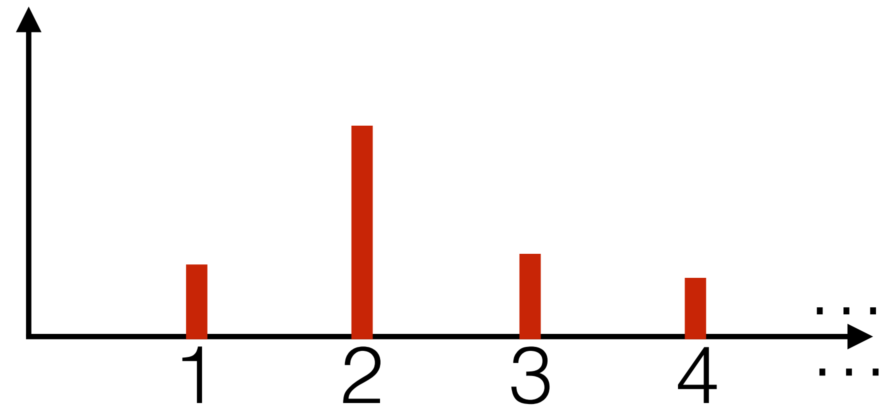


Dirichlet process mixture model

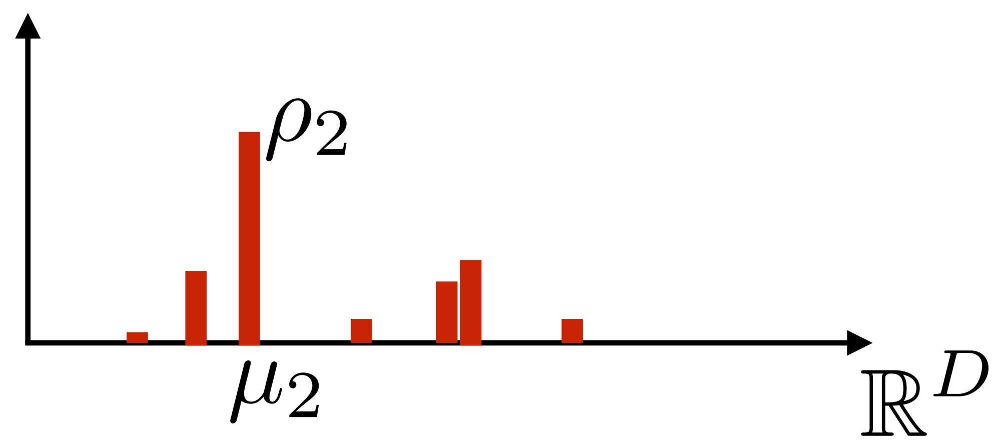
- More generally

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

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



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

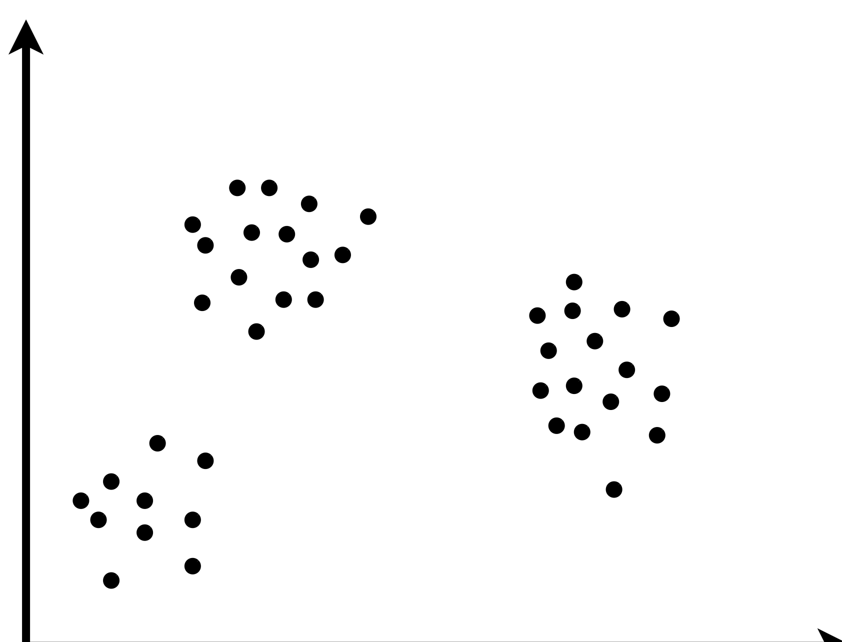


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

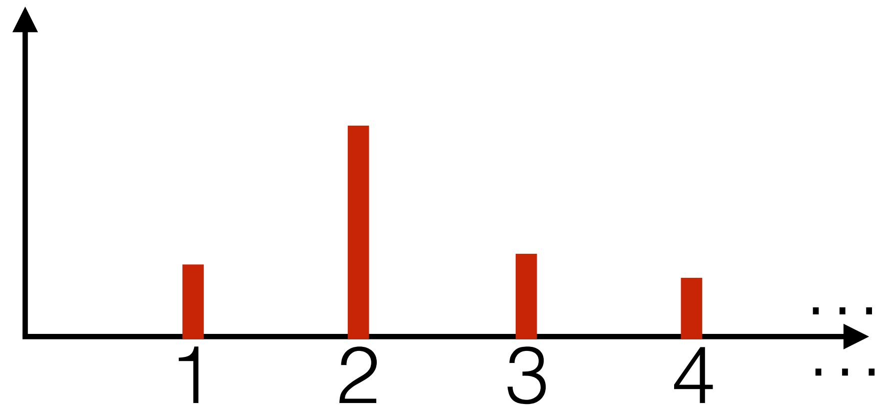
$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$



Dirichlet process mixture model

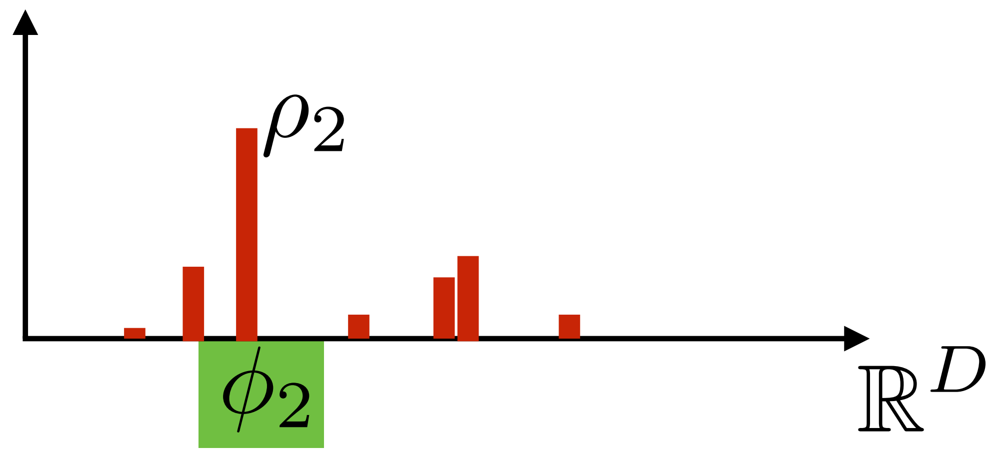
- More generally

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



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

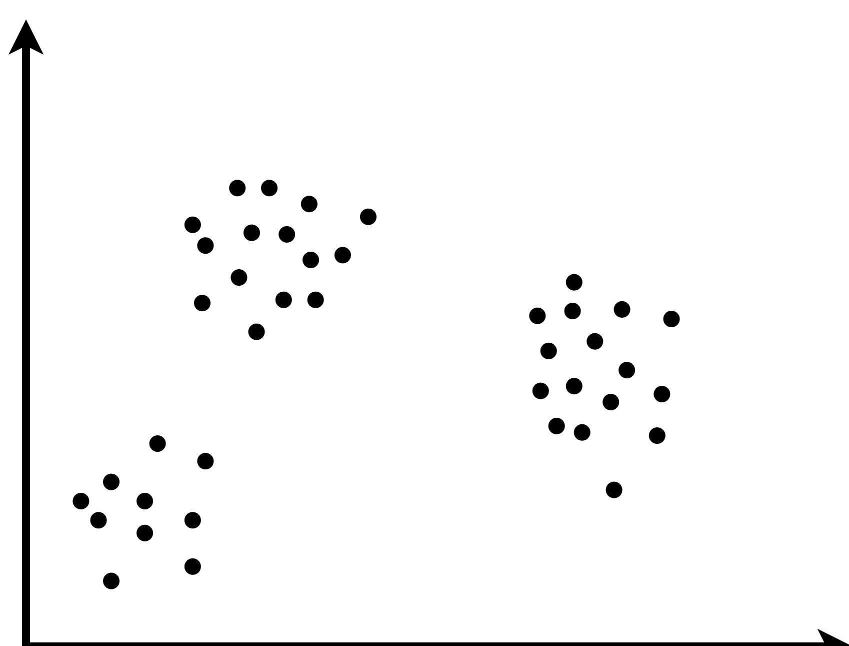
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\mu_k} \stackrel{d}{=} \mathrm{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

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

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



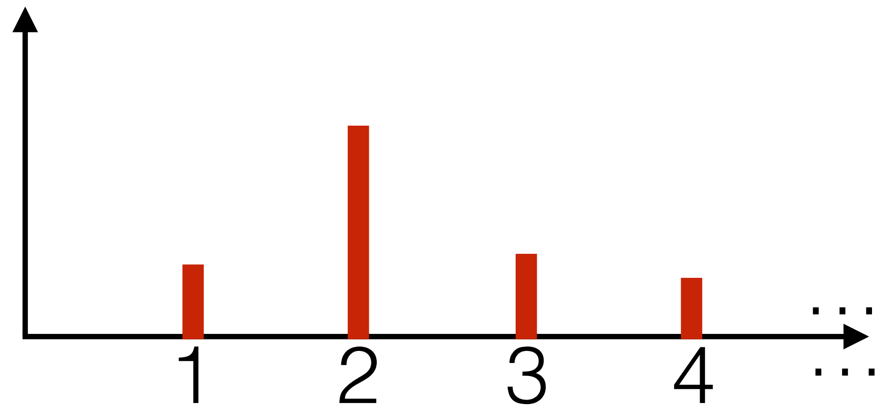
$$x_n \stackrel{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

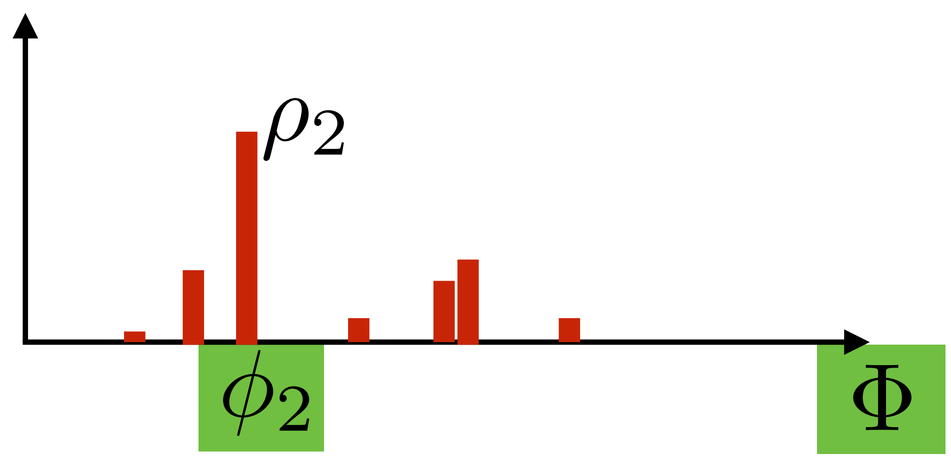
- More generally

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

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



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

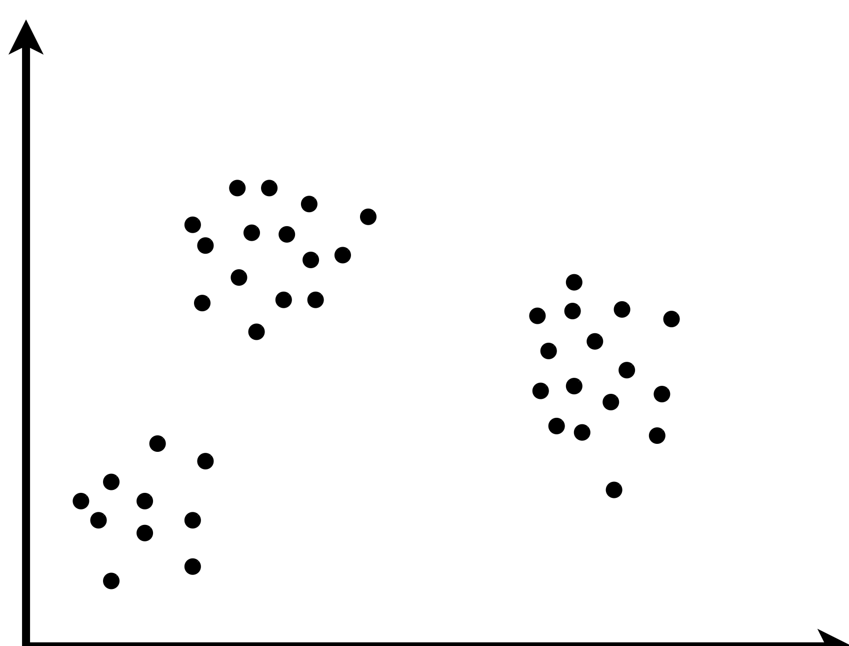


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

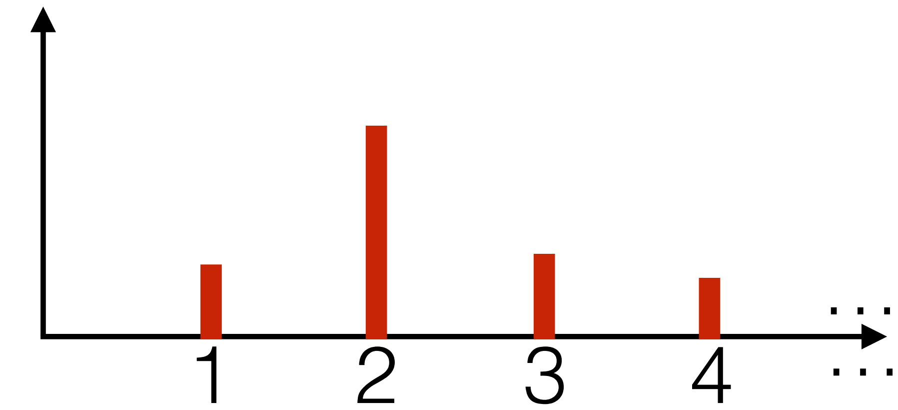
- i.e. $\mu_n^* \overset{iid}{\sim} G$

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



Dirichlet process mixture model

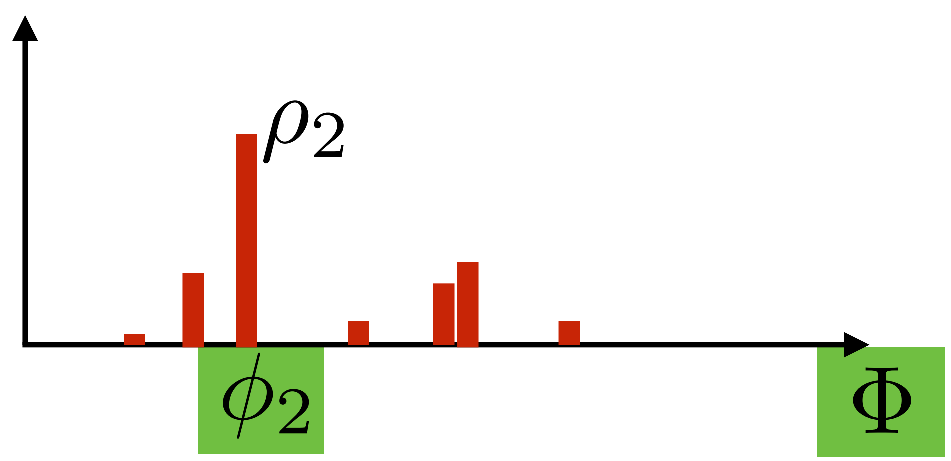
- More generally



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

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

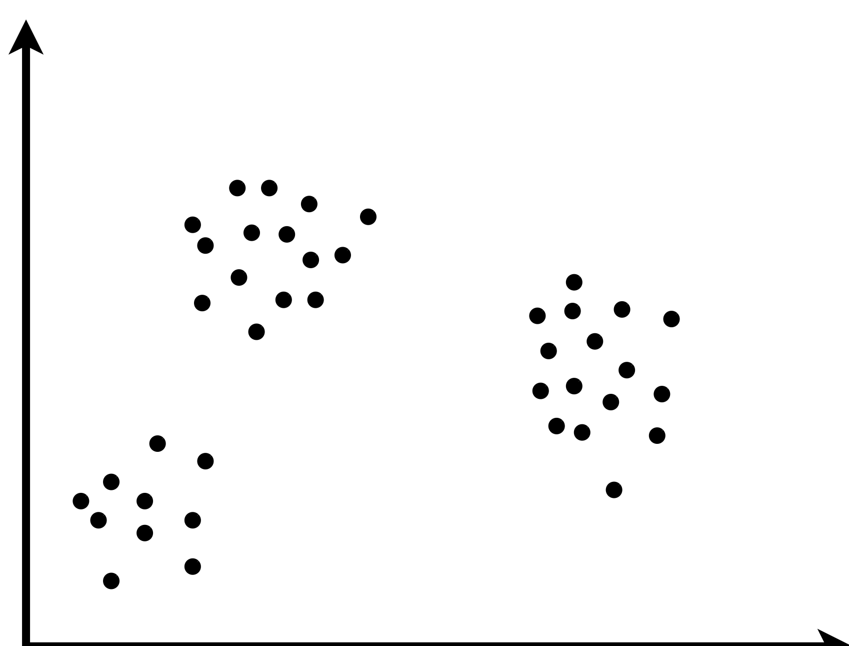
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \text{DP}(\alpha, \mathcal{N}(\mu_0, \Sigma_0))$



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

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

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



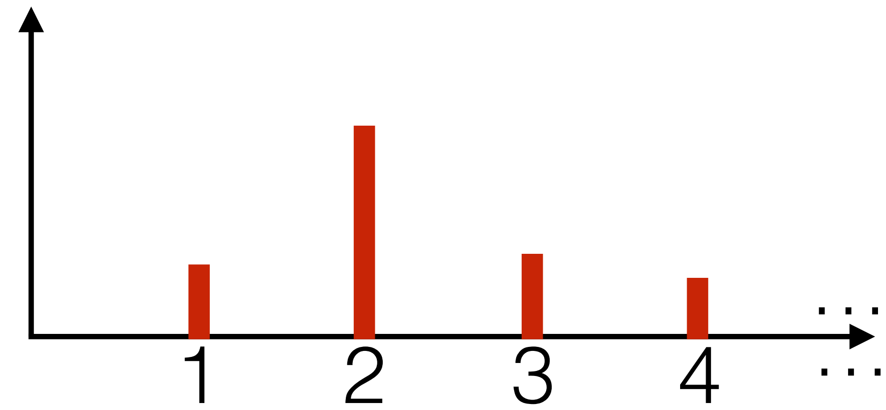
$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_n^*, \Sigma)$$

Dirichlet process mixture model

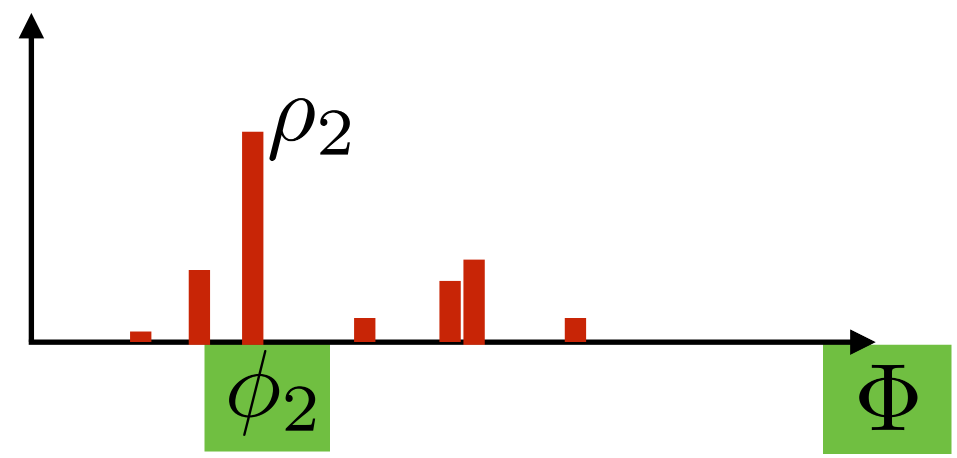
- More generally

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

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



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

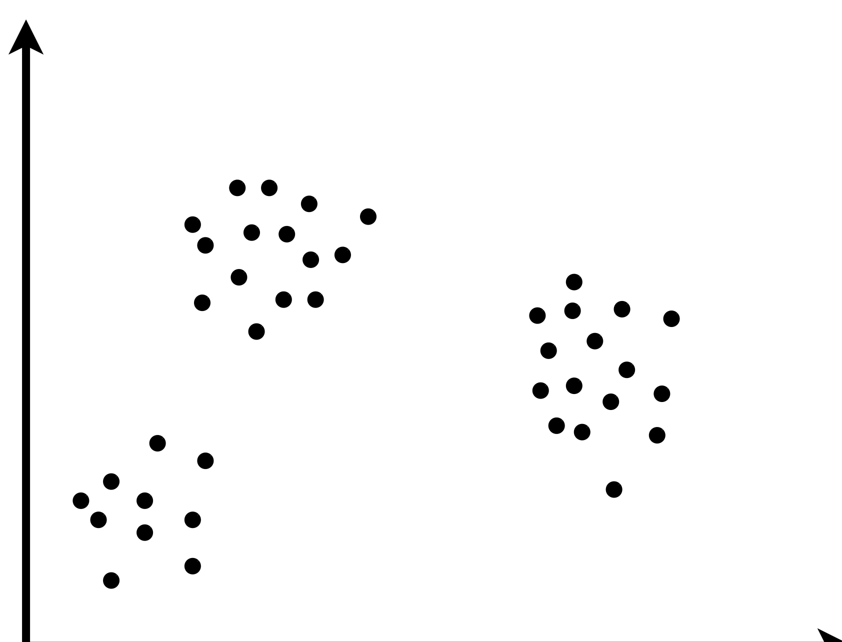


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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



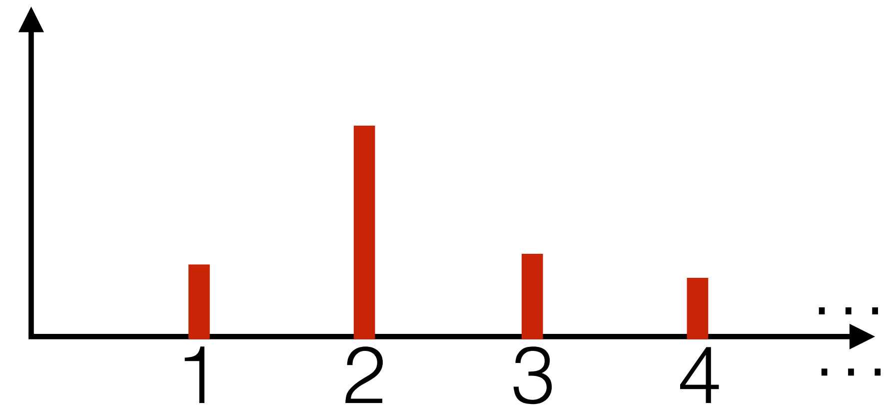
Dirichlet process mixture model

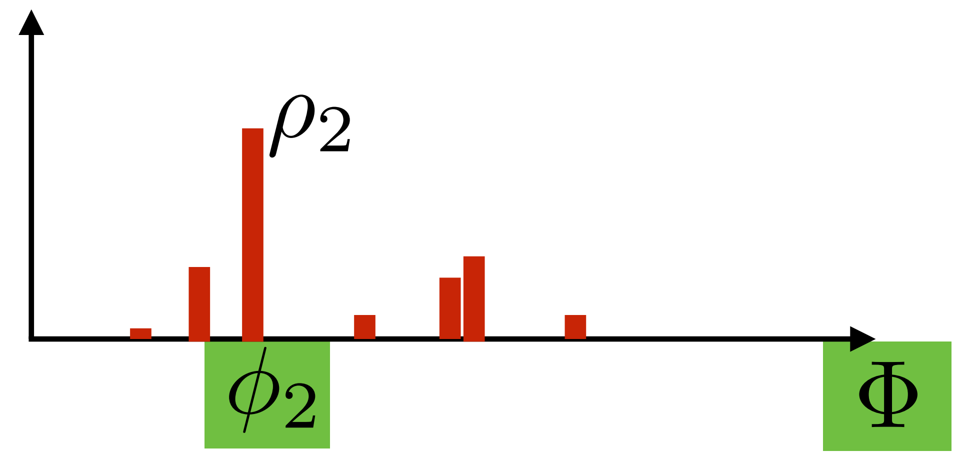
- More generally

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

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

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



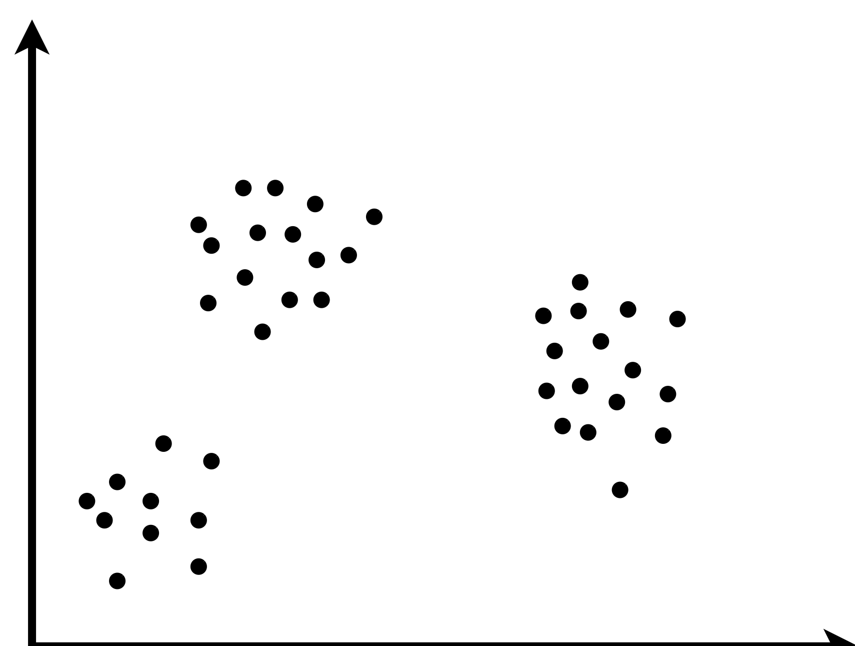


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

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

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

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



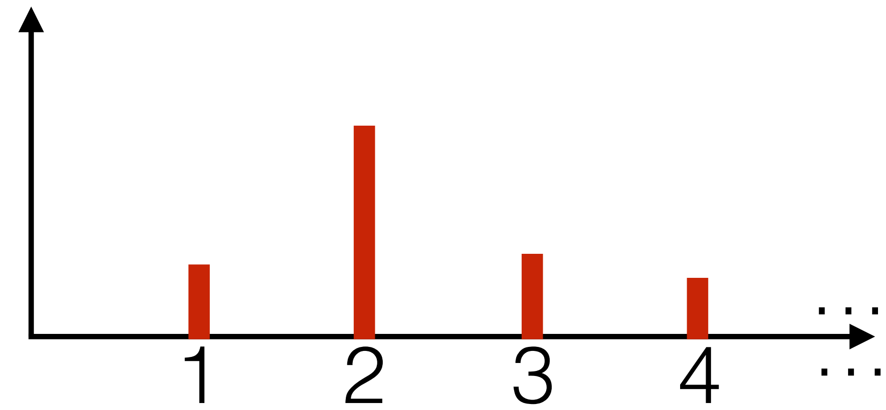
Dirichlet process mixture model

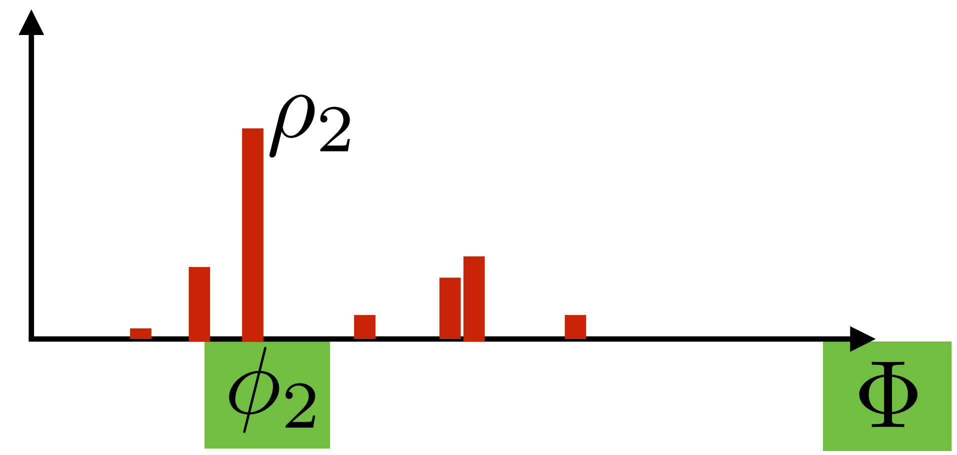
- More generally

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

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

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



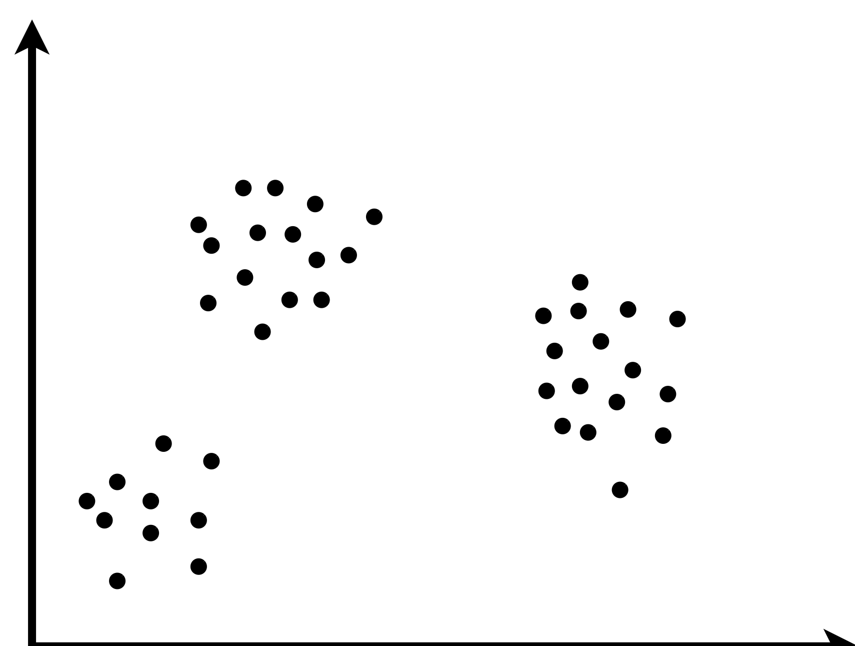


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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

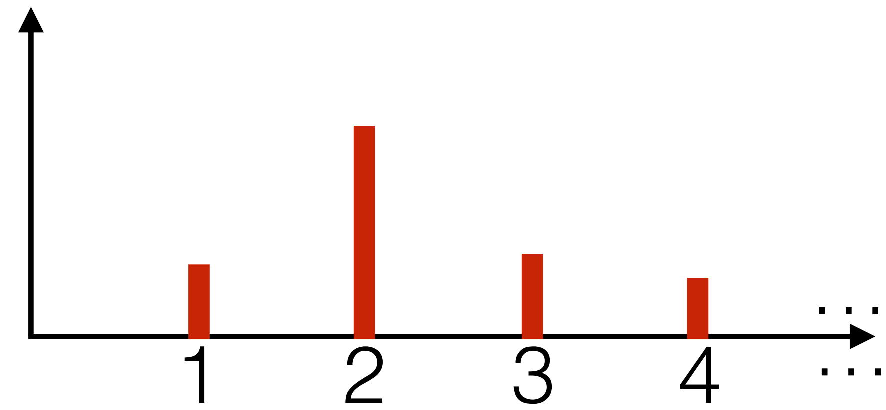


Dirichlet process mixture model

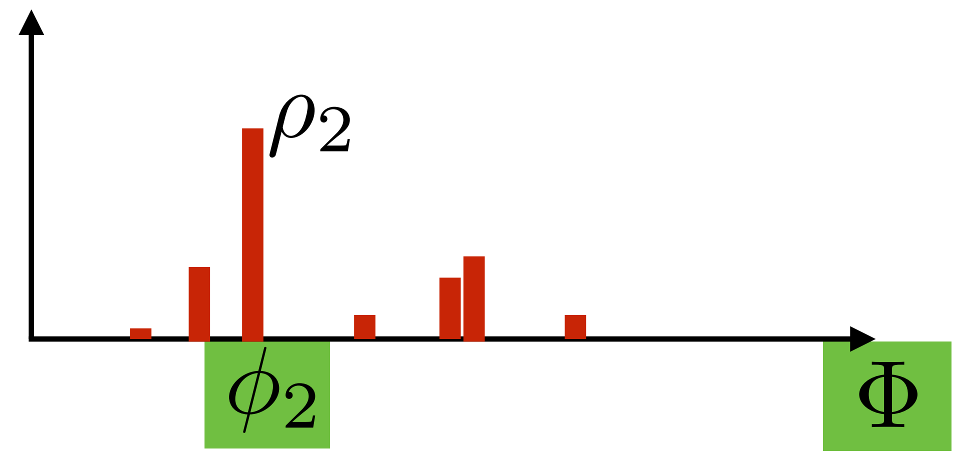
- More generally

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

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



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

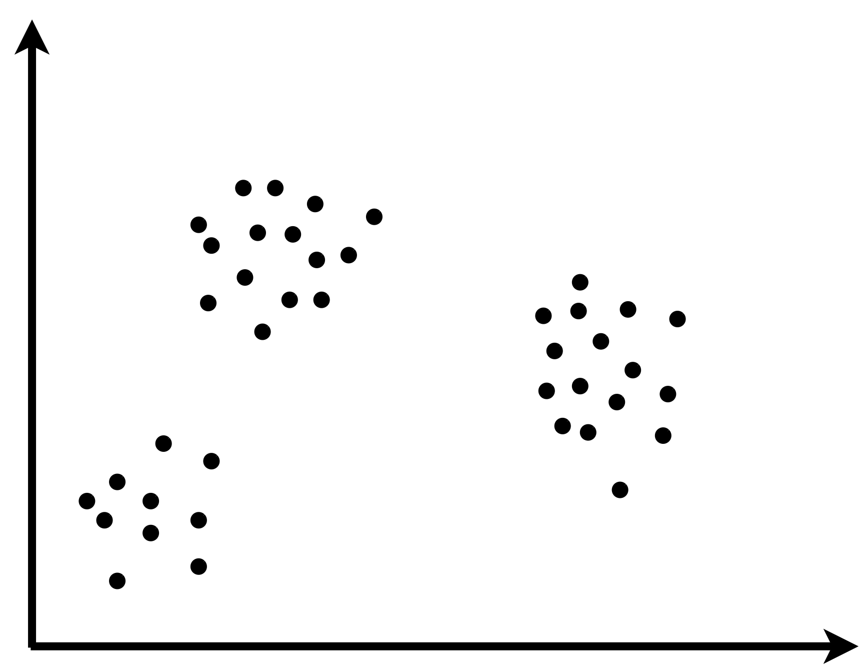


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

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

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

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

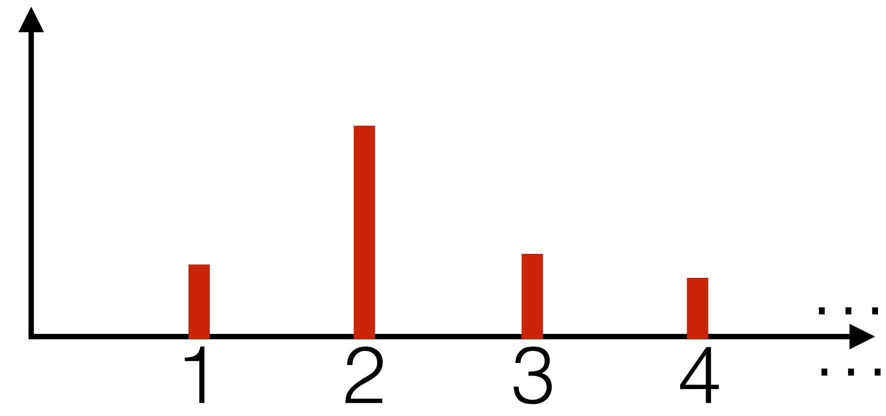


Dirichlet process mixture model

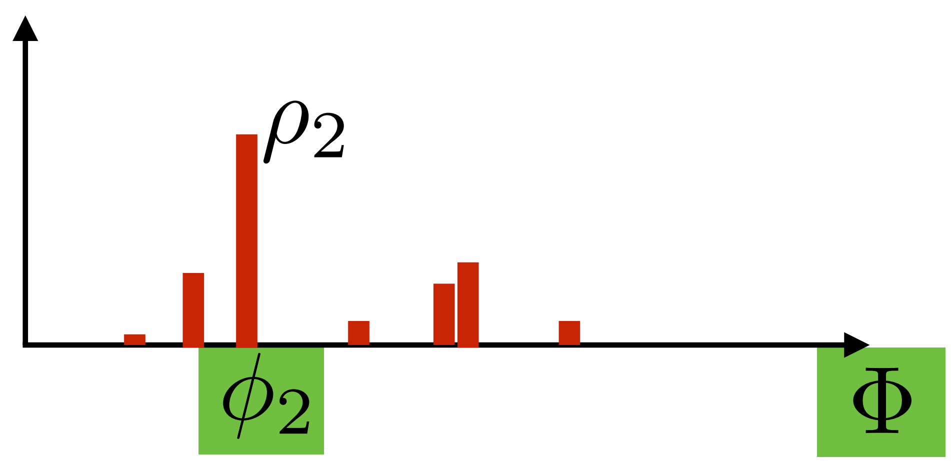
- More generally

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

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



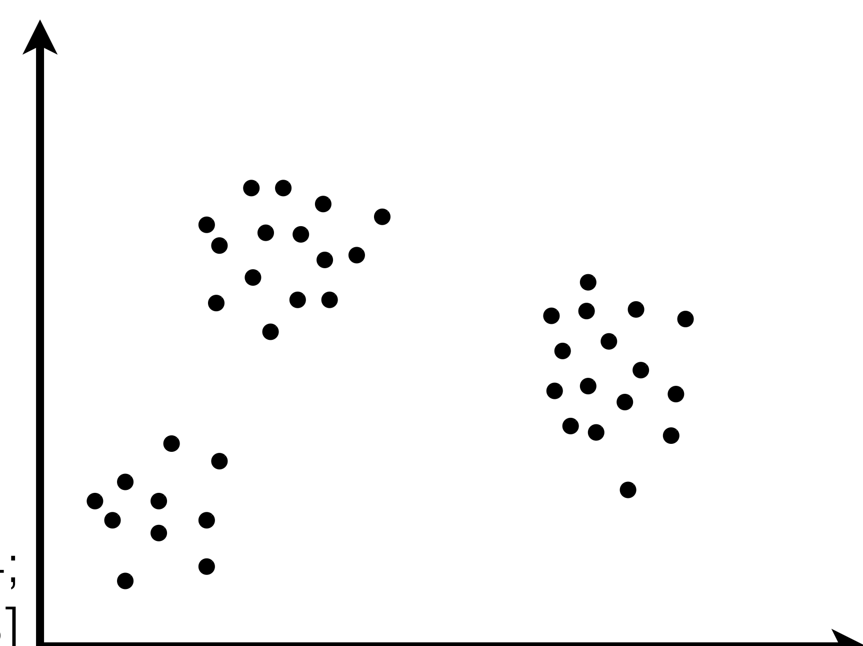
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \mathrm{DP}(\alpha, G_0)$



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

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

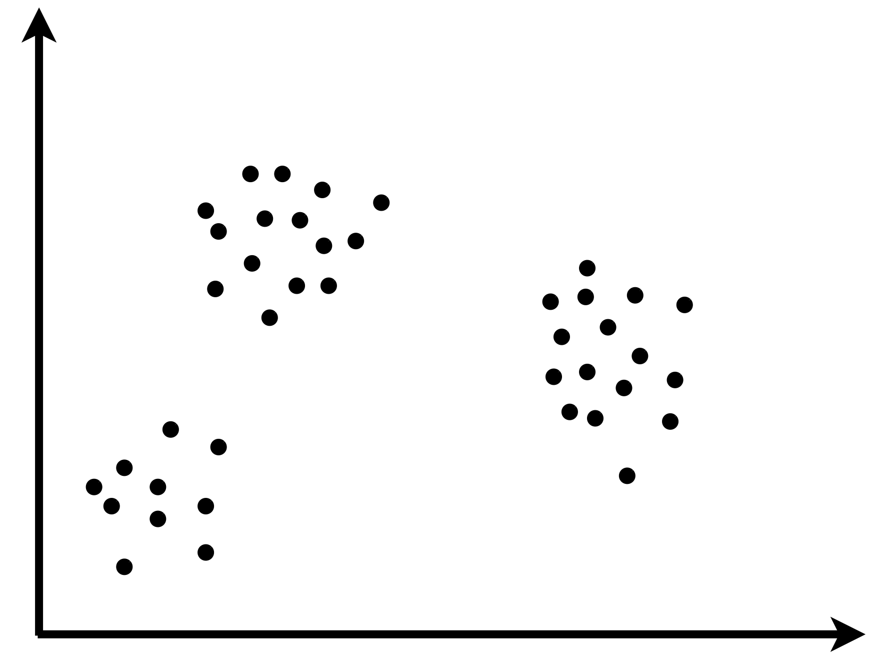
- i.e. $\theta_n \overset{iid}{\sim} G$



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

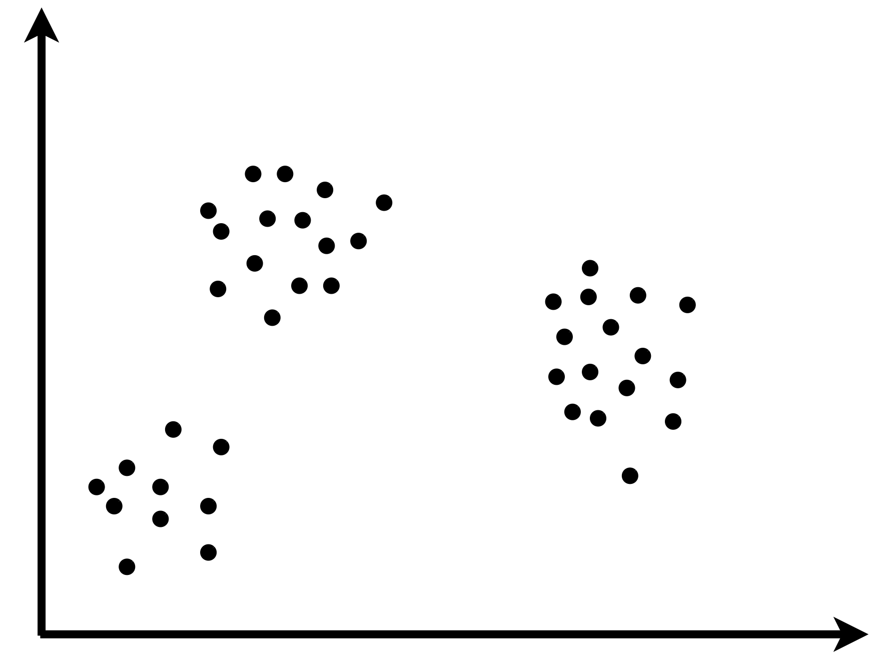
[Antoniak 1974; Ferguson 1983; West, Müller, Escobar 1994; Escobar, West 1995; MacEachern, Müller 1998]

DP or not DP, that is the question



DP or not DP, that is the question

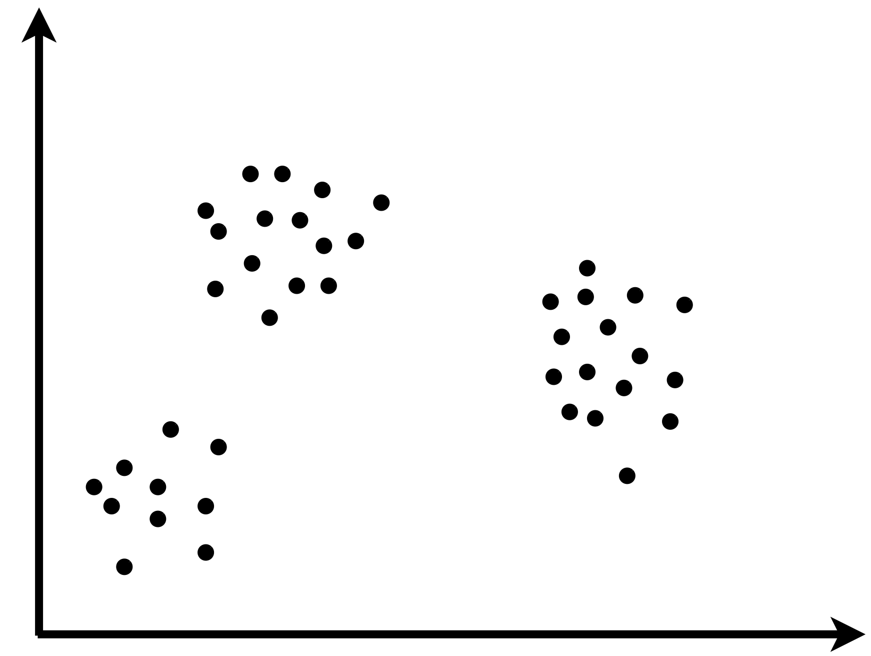
- GEM: 



DP or not DP, that is the question

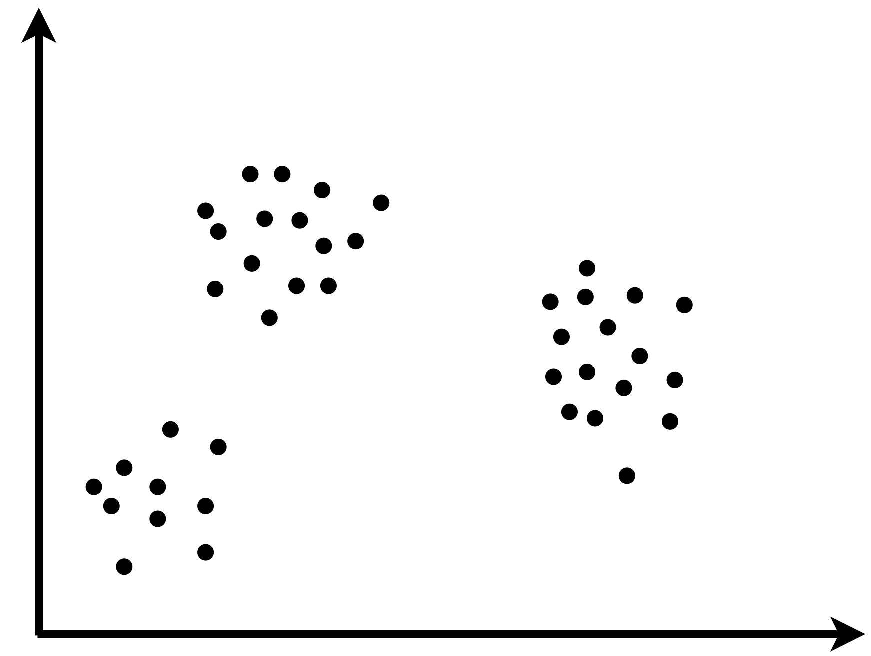
- GEM: 

- Compare to:



DP or not DP, that is the question

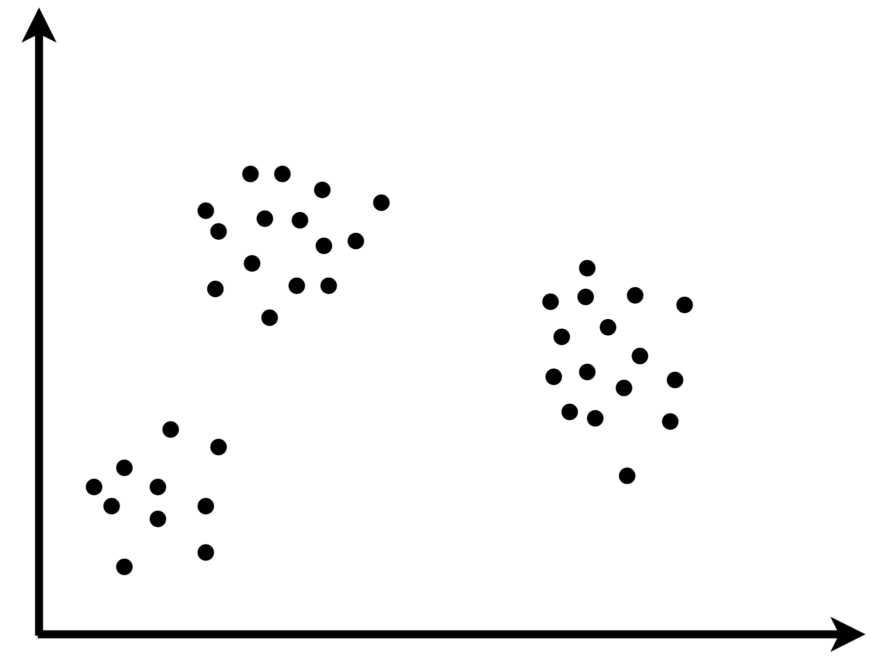
- GEM: 
- Compare to:
  - Finite (small *K*) mixture model



DP or not DP, that is the question

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

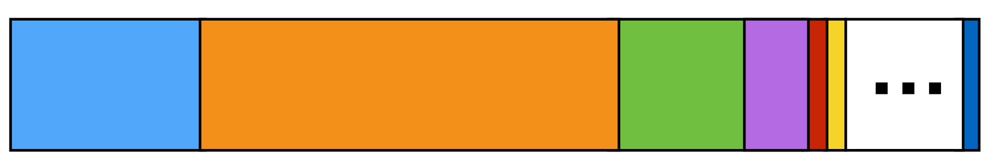
  - Finite (large *K*) mixture model



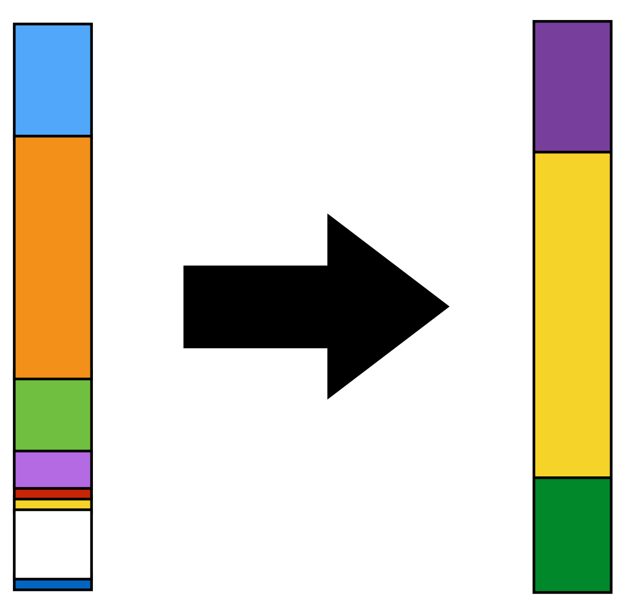
DP or not DP, that is the question

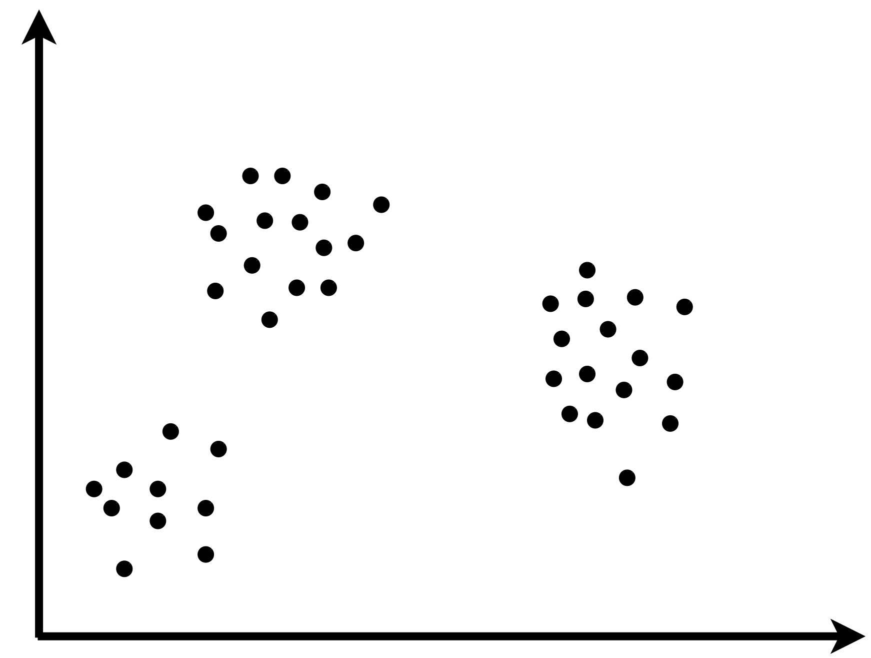
- GEM: 
- Compare to:
  - Finite (small *K*) mixture model

  - Finite (large *K*) mixture model



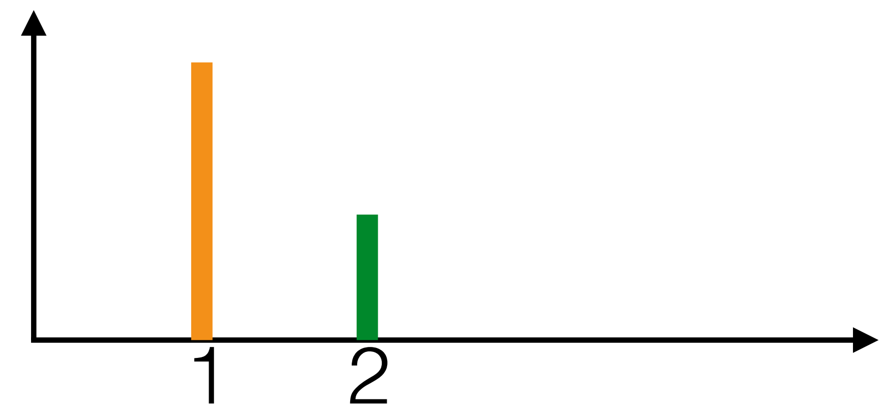
  - Time series





Marginal cluster assignments

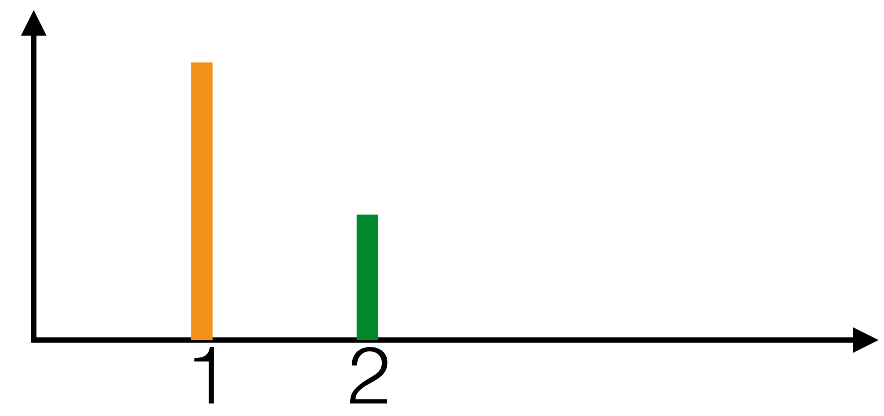
$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$



Marginal cluster assignments

- Integrate out the frequencies

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

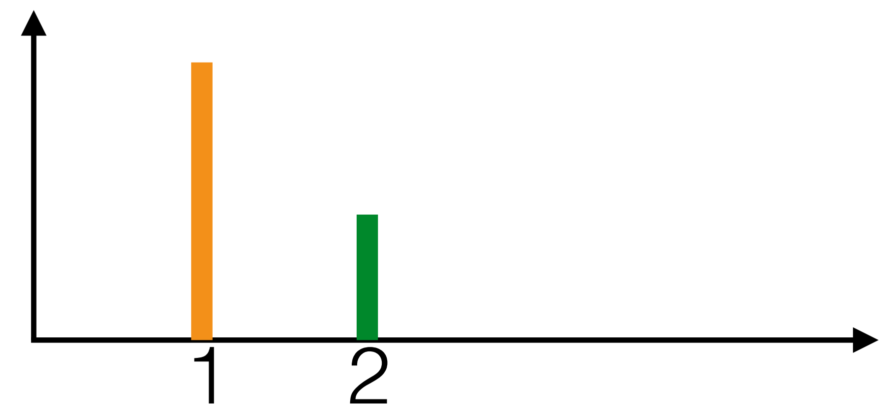


Marginal cluster assignments

- Integrate out the frequencies

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

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

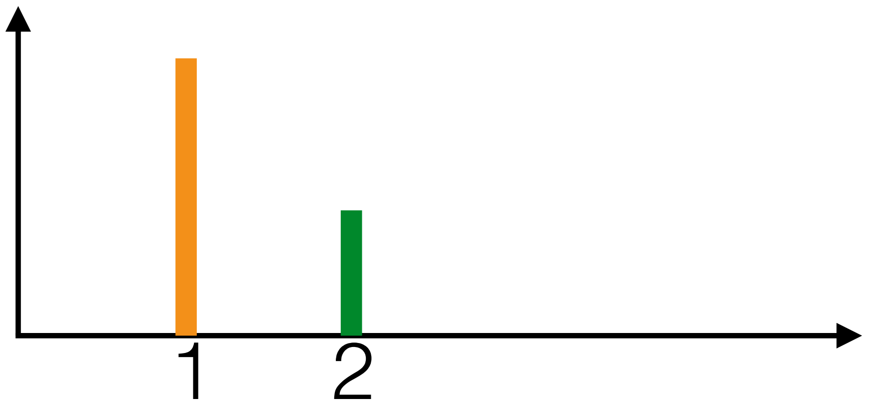


Marginal cluster assignments

- Integrate out the frequencies

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

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

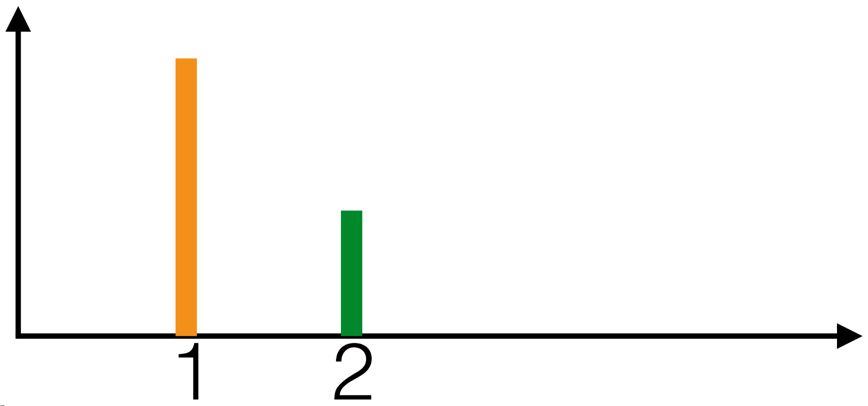


Marginal cluster assignments

- Integrate out the frequencies

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

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

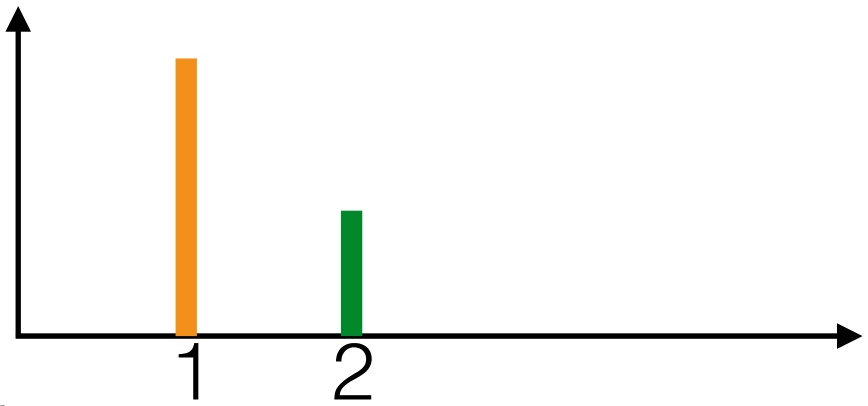


Marginal cluster assignments

- Integrate out the frequencies

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

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

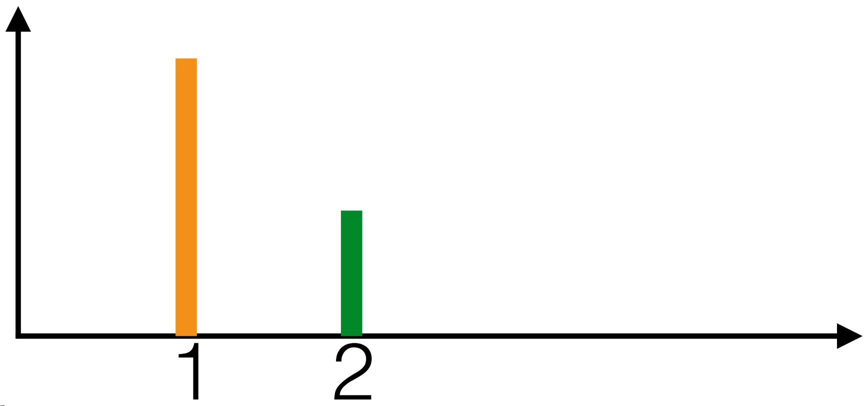


Marginal cluster assignments

- Integrate out the frequencies

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

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



Marginal cluster assignments

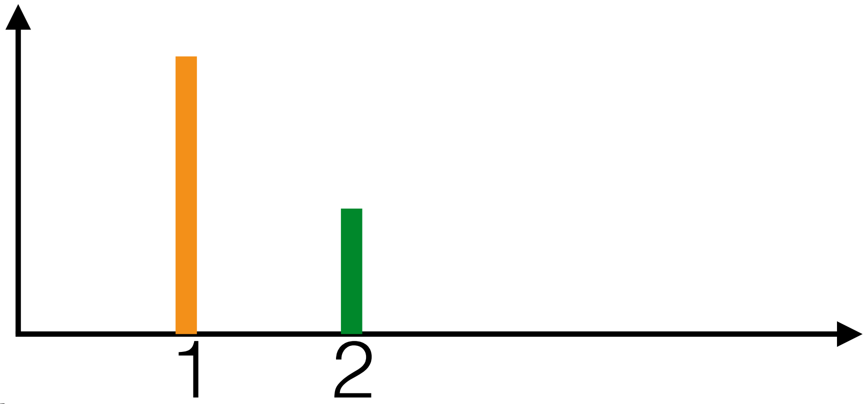
- Integrate out the frequencies

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

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

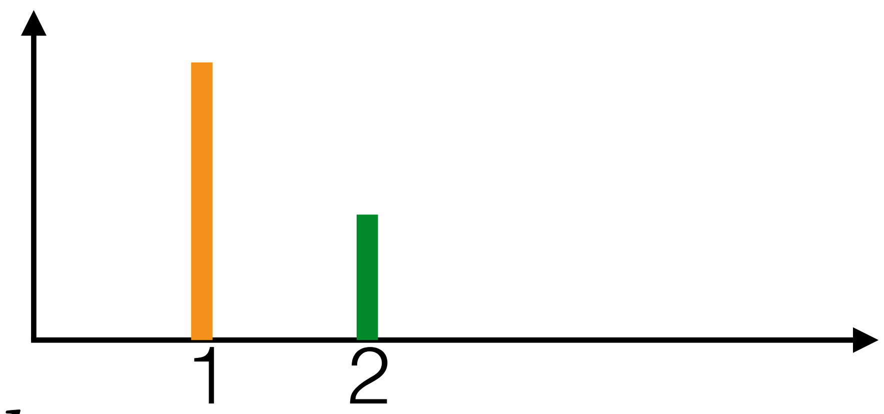
$$= \int p(z_n = 1|\rho_1) p(\rho_1 | z_1, \ldots, z_{n-1}) d\rho_1$$

$$= \int \rho_1$$



Marginal cluster assignments

- Integrate out the frequencies



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

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

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

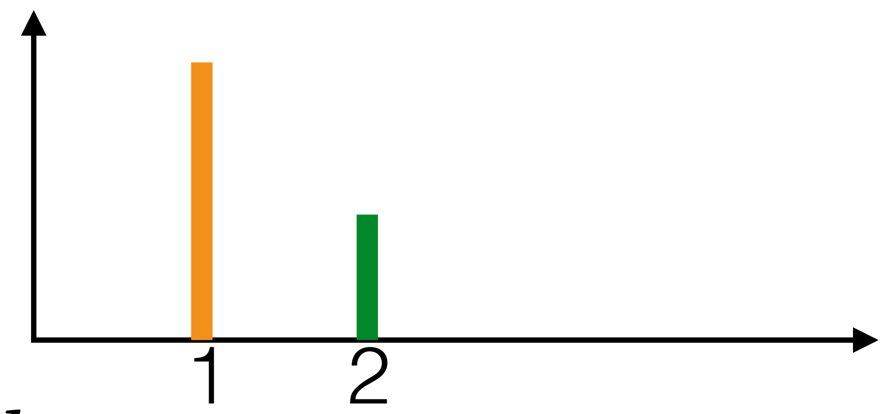
$$= \int \rho_1 \text{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$

Marginal cluster assignments

- Integrate out the frequencies

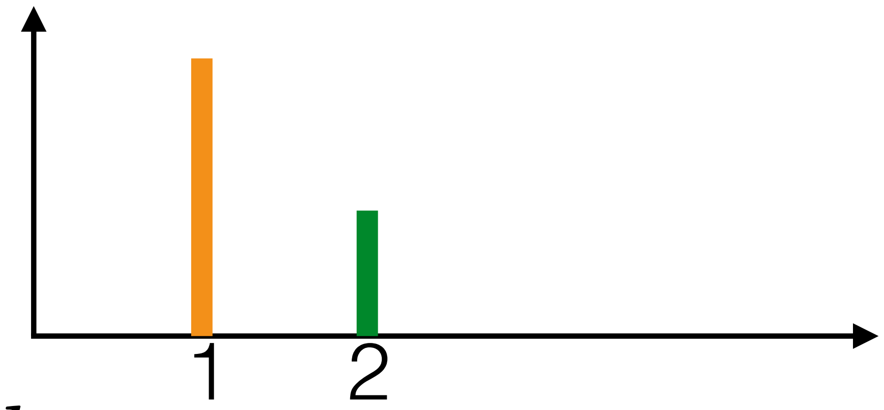
$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \stackrel{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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



Marginal cluster assignments

- Integrate out the frequencies

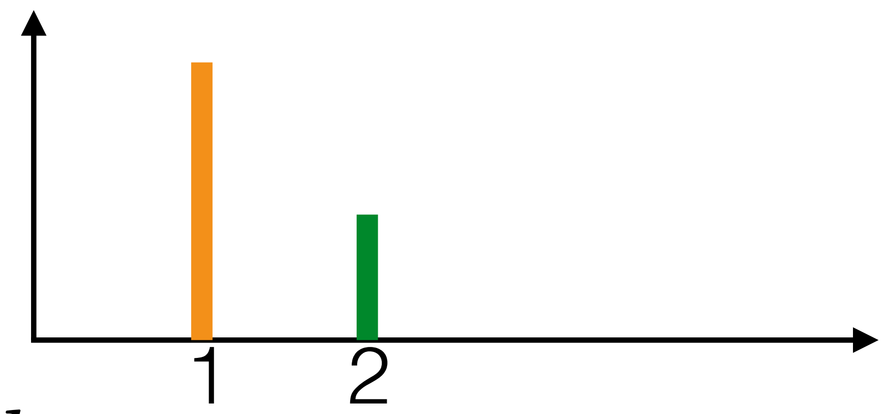


$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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

Marginal cluster assignments

- Integrate out the frequencies



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

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

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

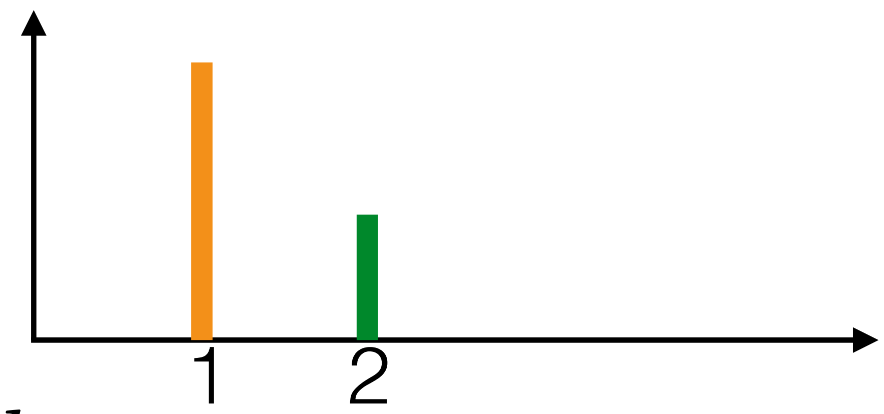
$$= \int \rho_1 \text{Beta}(\rho_1 | a_{1,n}, a_{2,n}) d\rho_1$$

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

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

Marginal cluster assignments

- Integrate out the frequencies

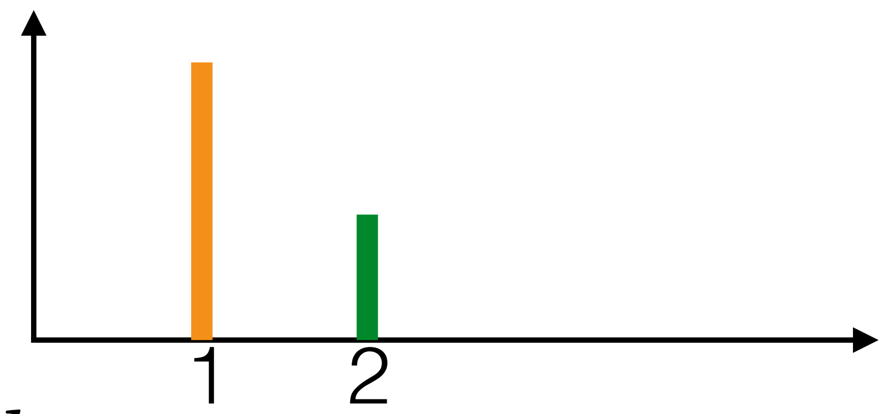


$$\rho_1 \sim \mathrm{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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

Marginal cluster assignments

- Integrate out the frequencies



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

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

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

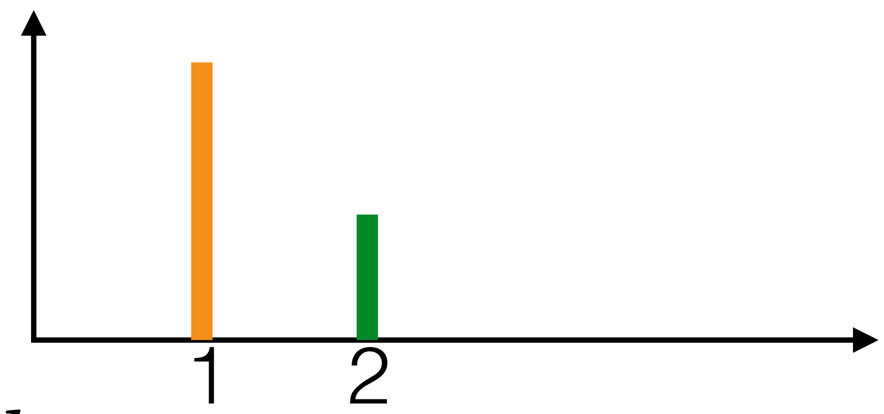
$$= \int \rho_1 \frac{\Gamma(a_{1,n} + a_{2,n})}{\Gamma(a_{1,n})\Gamma(a_{2,n})} \rho_1^{a_{1,n}-1} (1 - \rho_1)^{a_{2,n}-1} d\rho_1$$

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

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

Marginal cluster assignments

- Integrate out the frequencies



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

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

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

Marginal cluster assignments

- Integrate out the frequencies

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

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

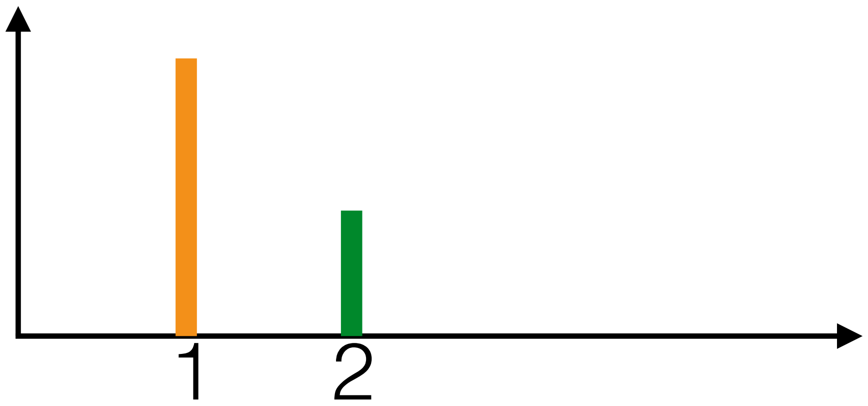
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

- Integrate out the frequencies

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

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

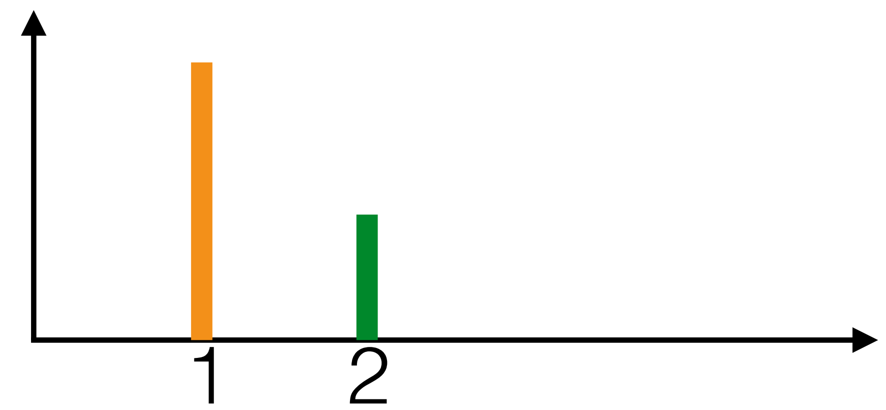


$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

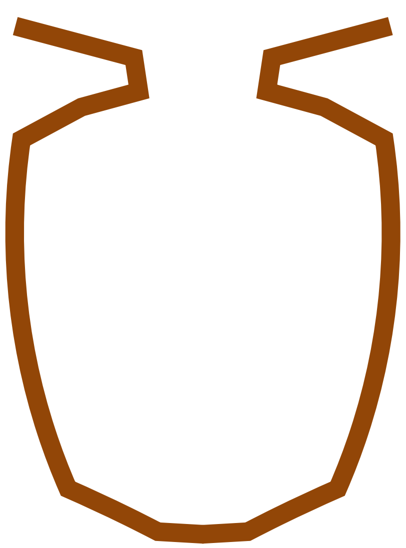
- Integrate out the frequencies

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



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

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

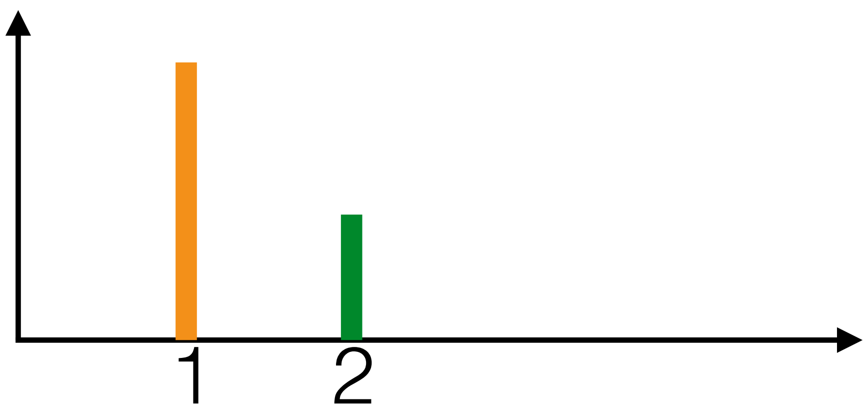


Marginal cluster assignments

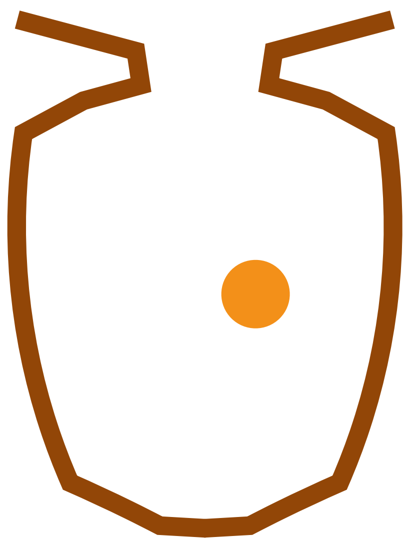
- Integrate out the frequencies

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

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



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

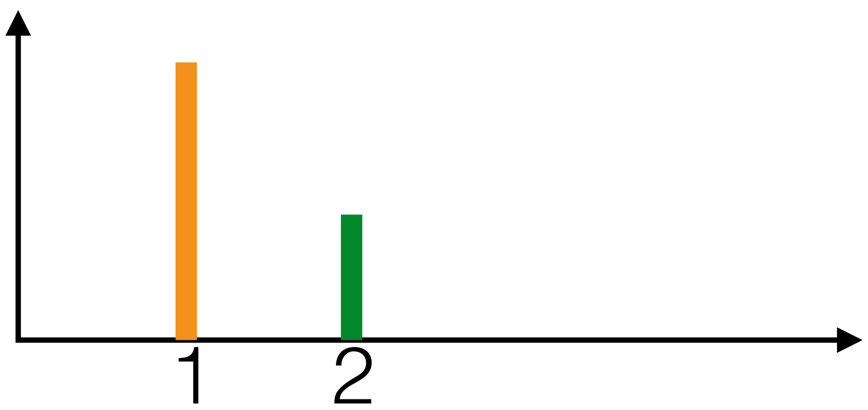


Marginal cluster assignments

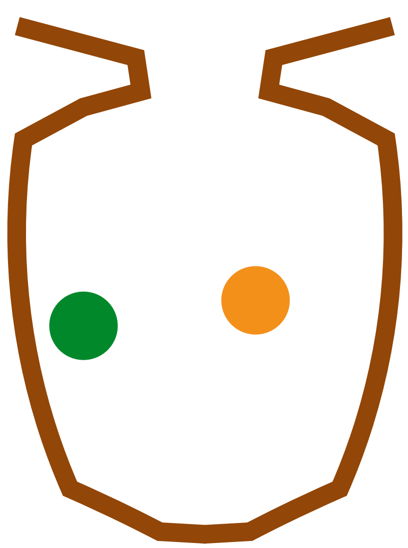
- Integrate out the frequencies

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

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



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

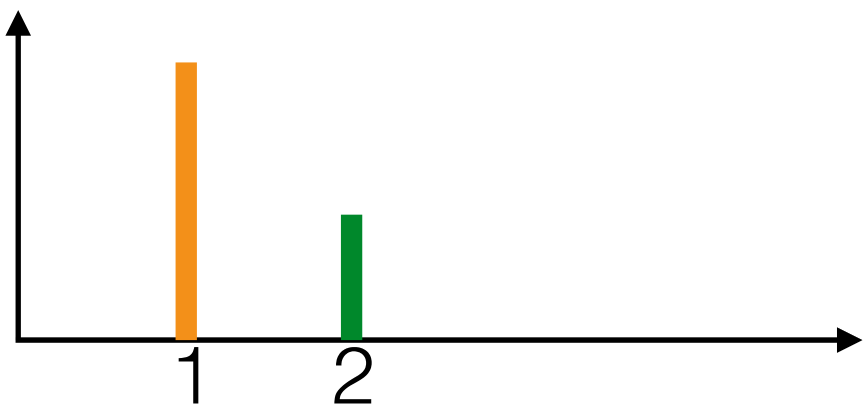


Marginal cluster assignments

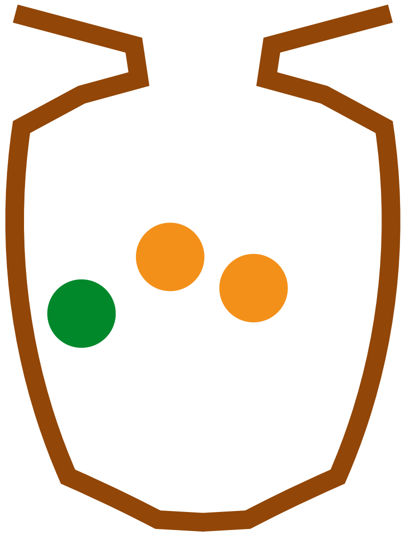
- Integrate out the frequencies

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

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



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

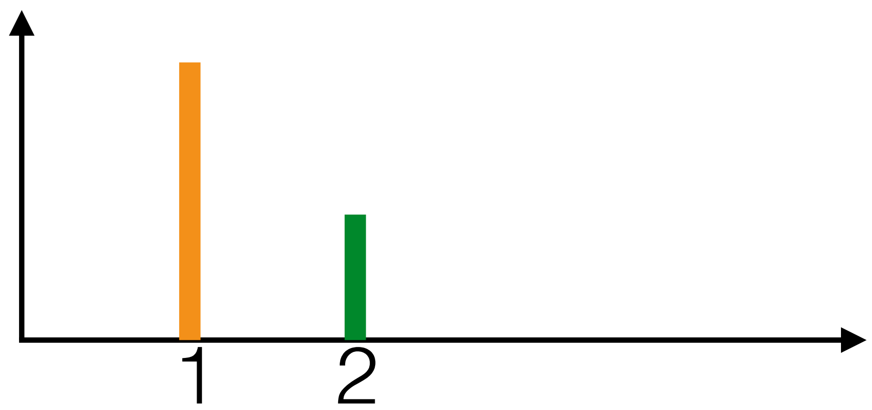


Marginal cluster assignments

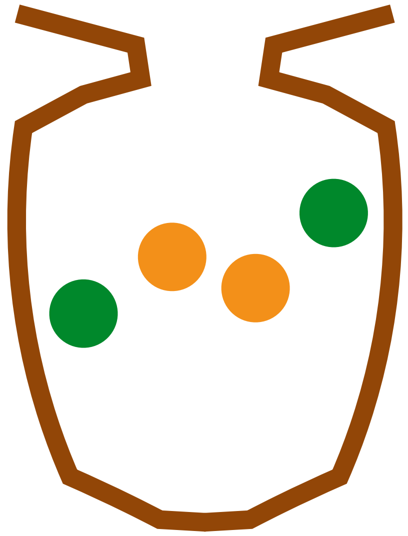
- Integrate out the frequencies

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

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



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

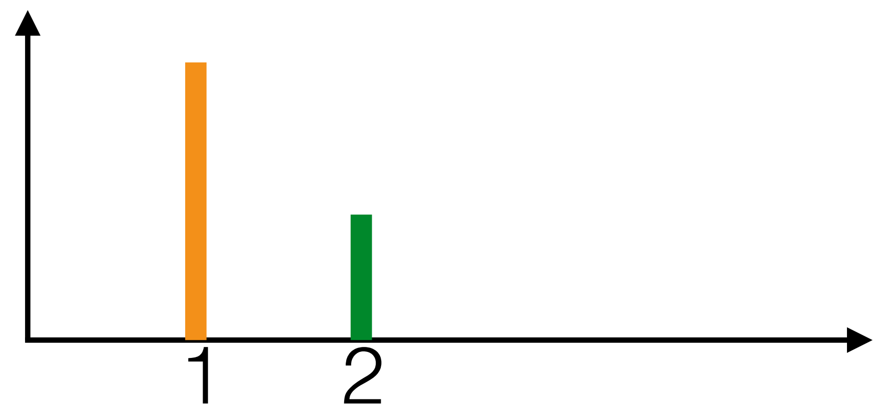


Marginal cluster assignments

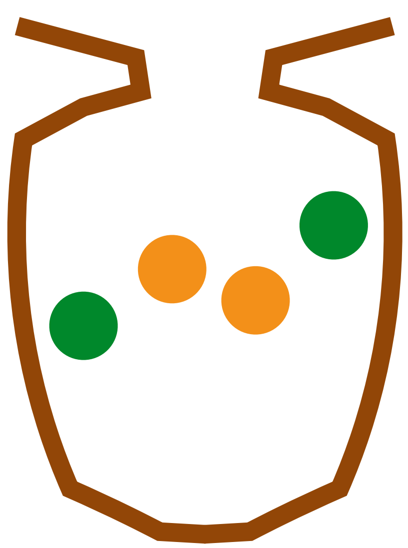
- Integrate out the frequencies

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

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



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

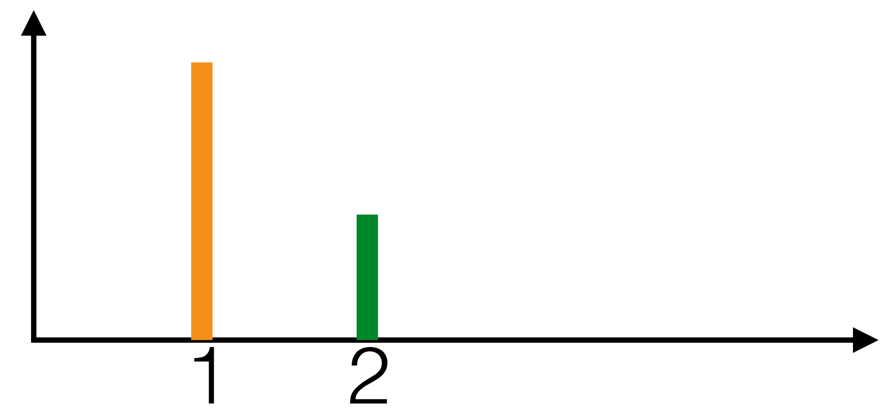


$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}}$$

Marginal cluster assignments

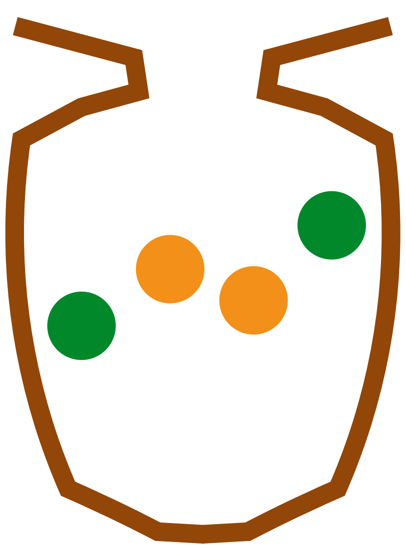
- Integrate out the frequencies

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



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

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



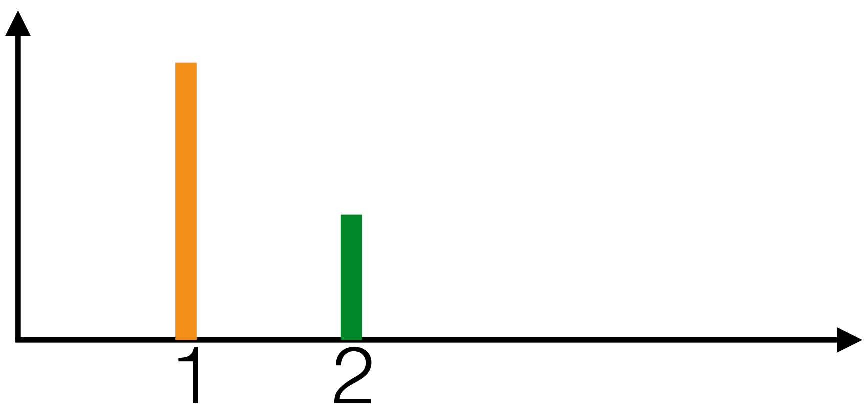
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}}$$

Marginal cluster assignments

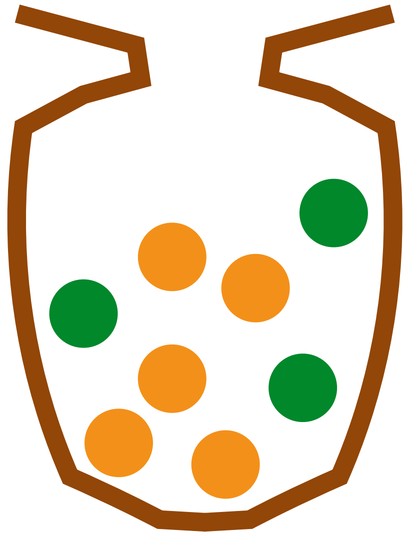
- Integrate out the frequencies

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

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



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

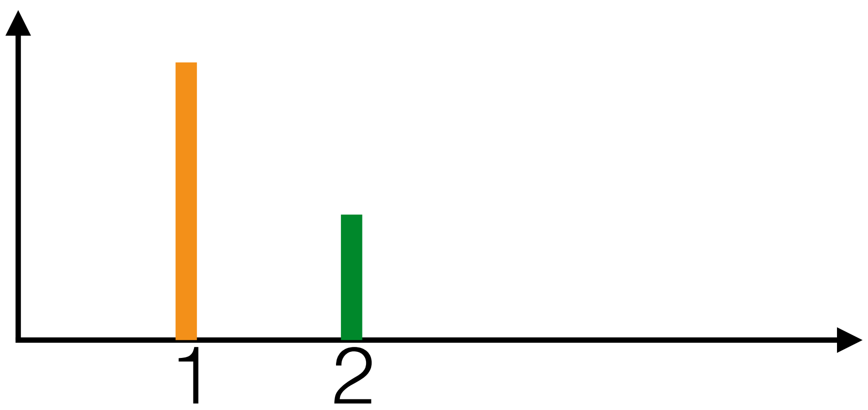


Marginal cluster assignments

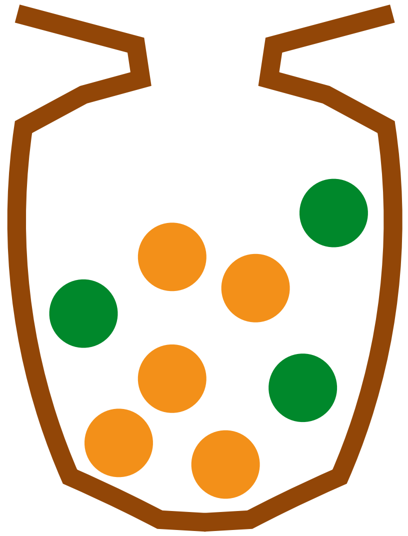
- Integrate out the frequencies

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

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



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

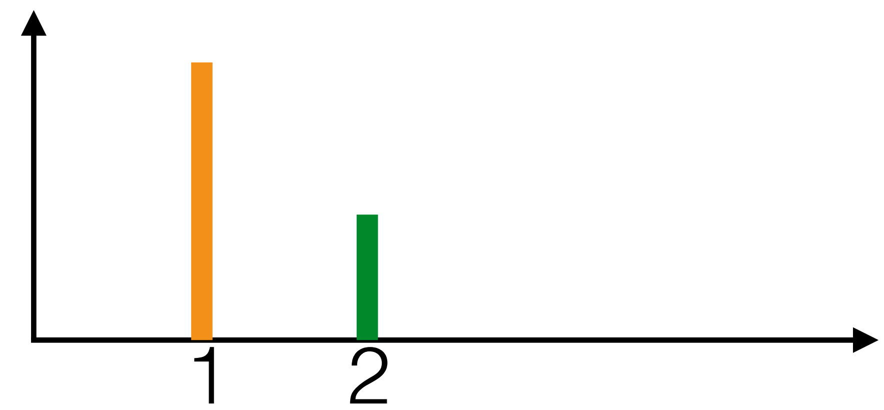


$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}}$$

Marginal cluster assignments

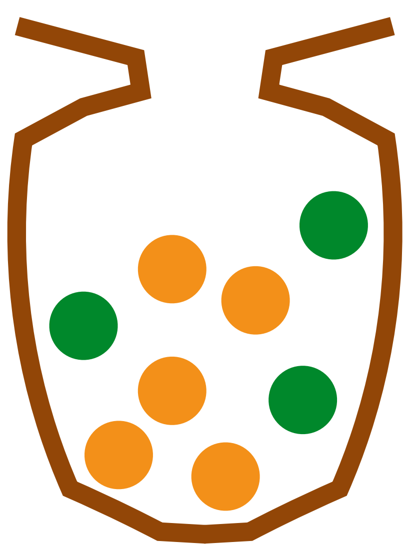
- Integrate out the frequencies

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



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

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



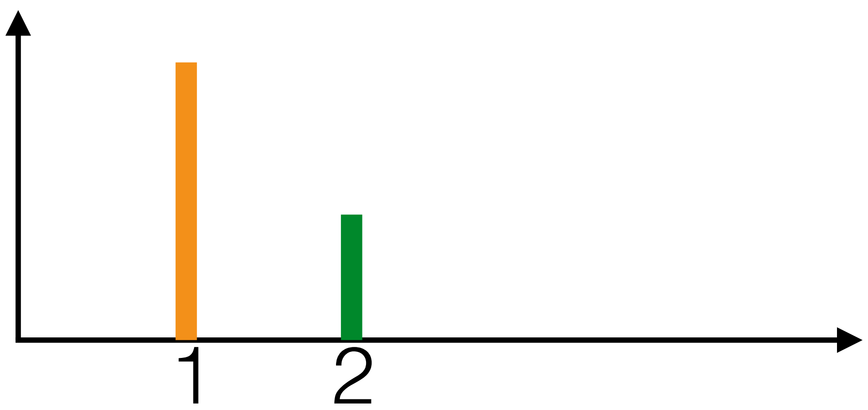
$$\lim_{n \to \infty} \frac{\# \text{ orange}}{\# \text{ total}} = \rho_{\text{orange}}$$

Marginal cluster assignments

- Integrate out the frequencies

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

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



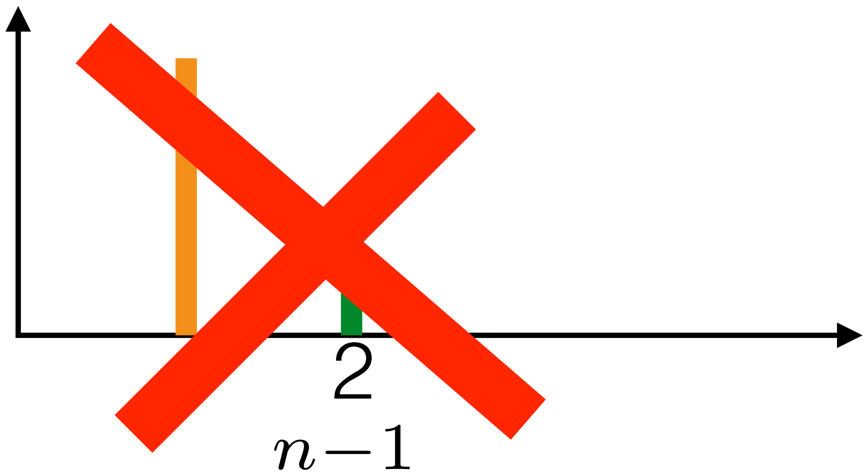
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

- Integrate out the frequencies

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

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

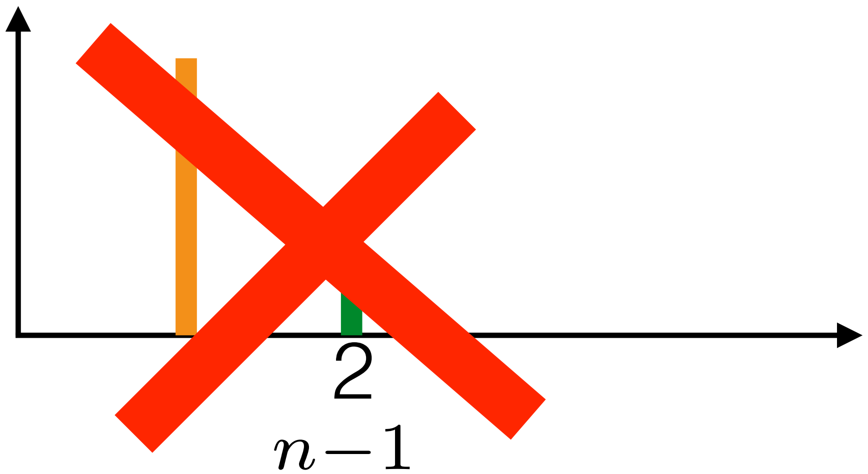


$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\}, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

Marginal cluster assignments

- Integrate out the frequencies

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



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

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

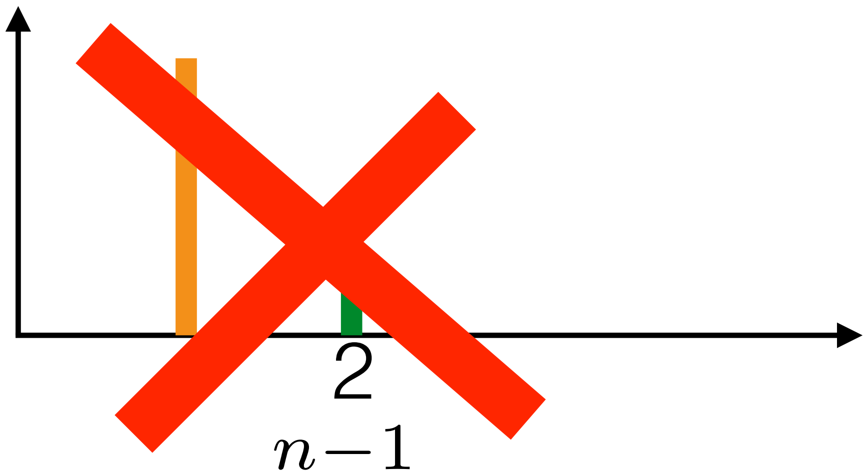
- Pólya urn

Marginal cluster assignments

- Integrate out the frequencies

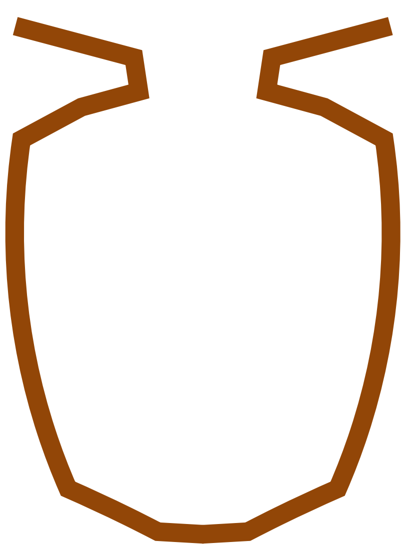
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

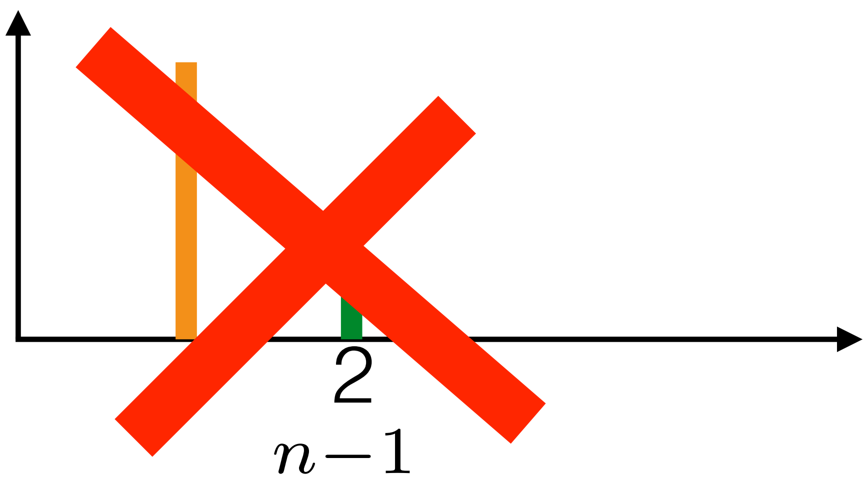
- Pólya urn



Marginal cluster assignments

- Integrate out the frequencies

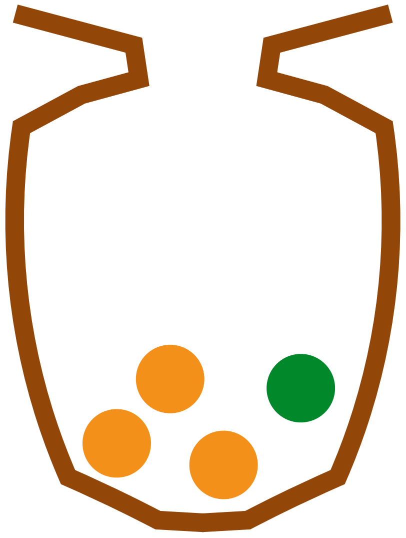
$\rho_1 \sim \text{Beta}(a_1, a_2), z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$



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

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

- Pólya urn

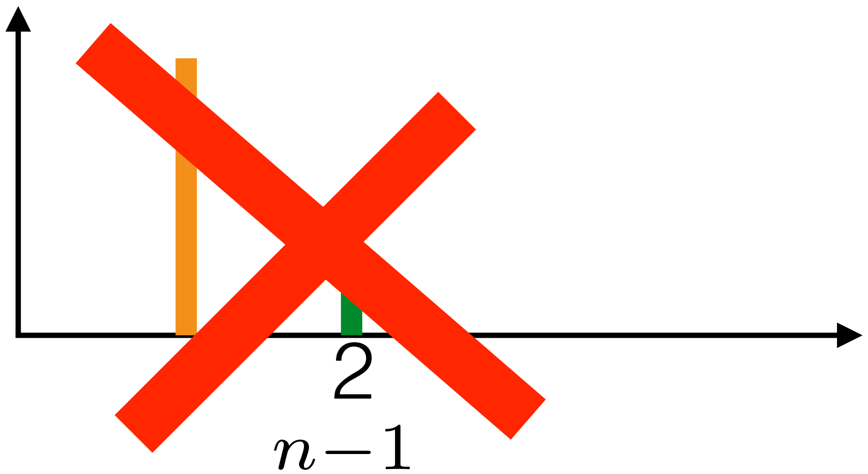


Marginal cluster assignments

- Integrate out the frequencies

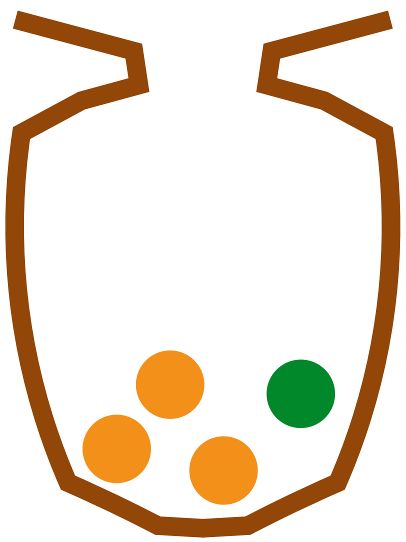
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

- Pólya urn
  - Choose any ball with equal probability

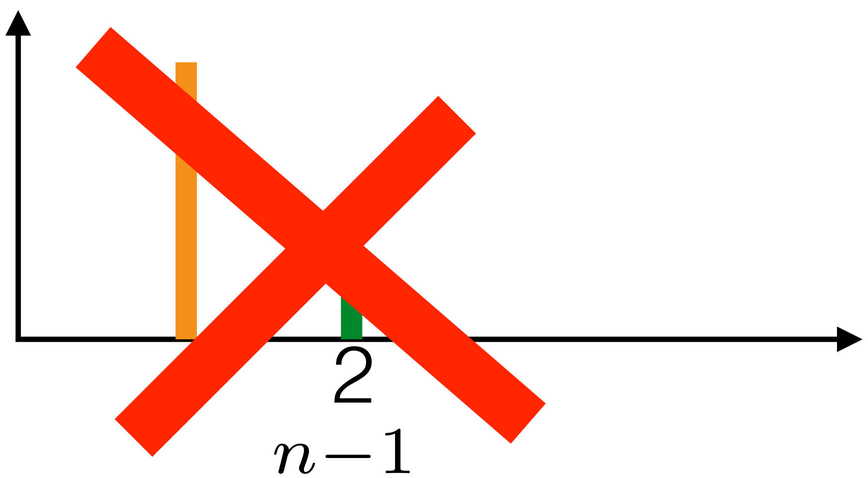


Marginal cluster assignments

- Integrate out the frequencies

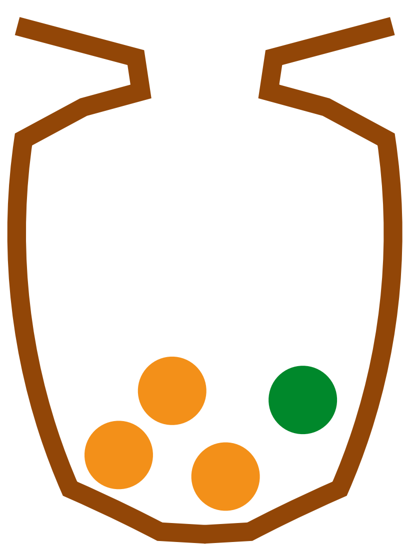
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



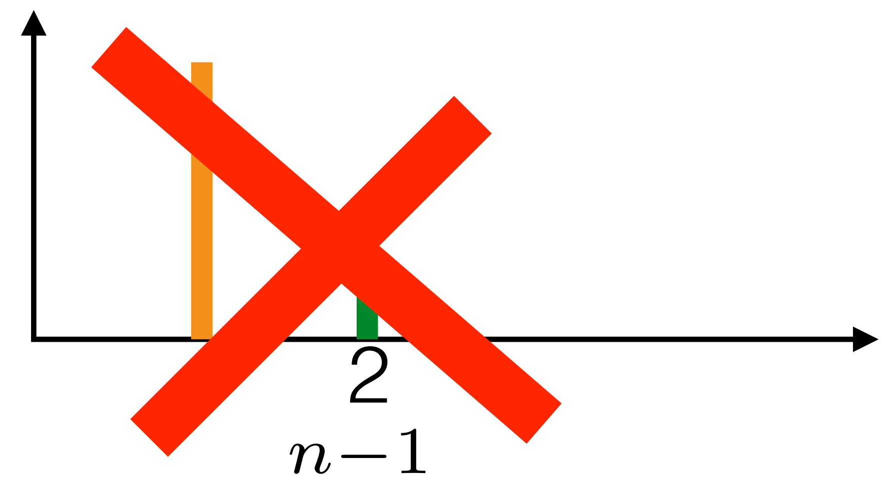
$$a_{1,n} := a_1 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 1\},\, a_{2,n} = a_2 + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = 2\}$$

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



Marginal cluster assignments

- Integrate out the frequencies

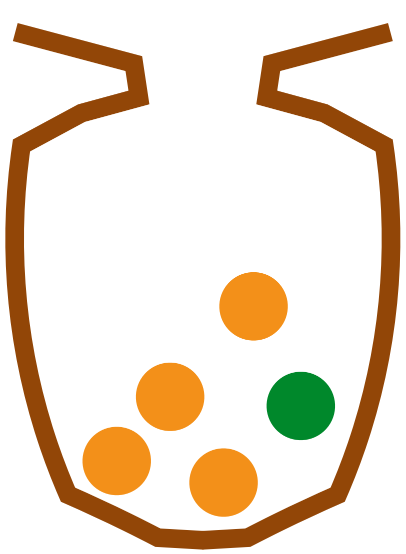


$$\rho_1 \sim \mathrm{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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

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

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

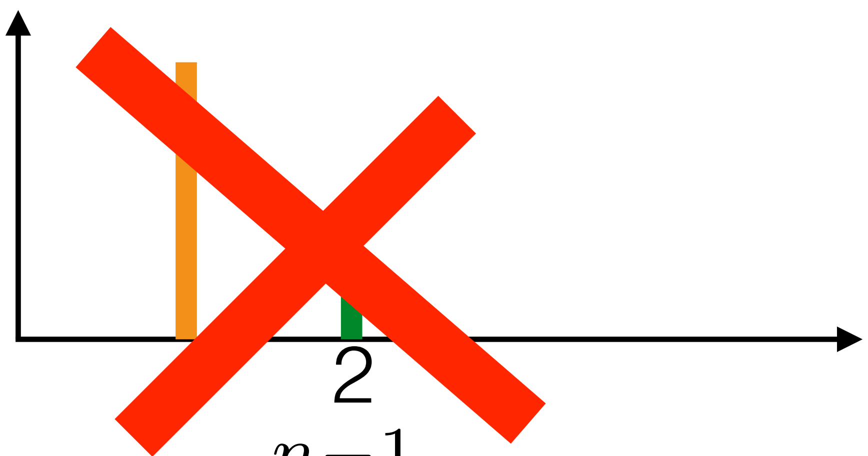


Marginal cluster assignments

- Integrate out the frequencies

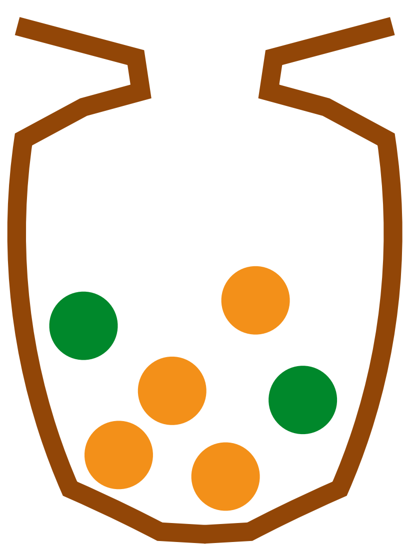
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \stackrel{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

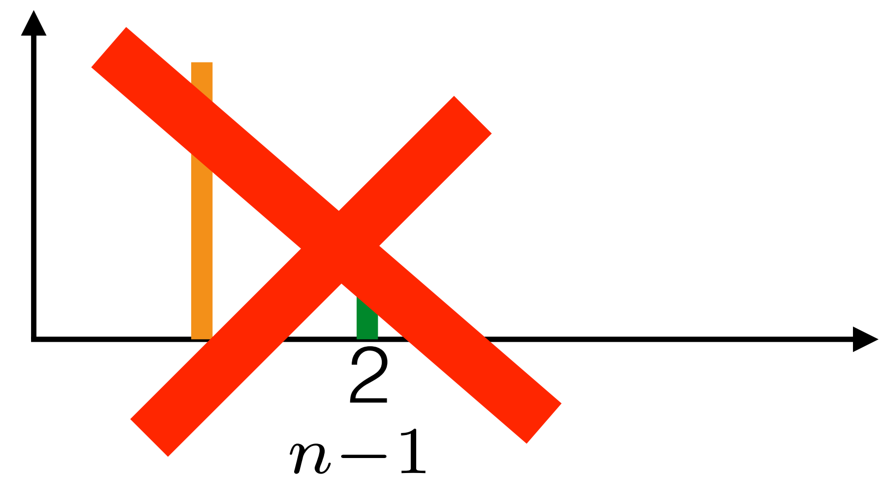
- Pólya urn
  - Choose any ball with equal probability
  - Replace and add ball of same color



Marginal cluster assignments

- Integrate out the frequencies

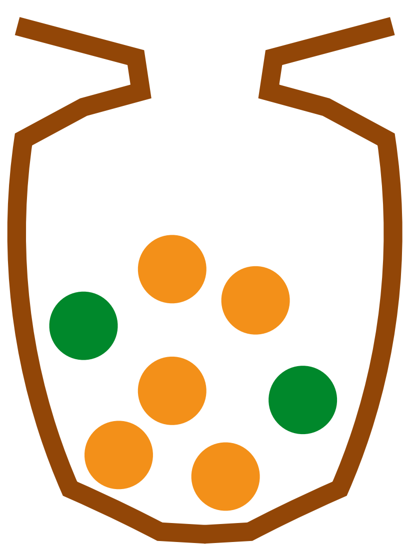
$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$



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

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

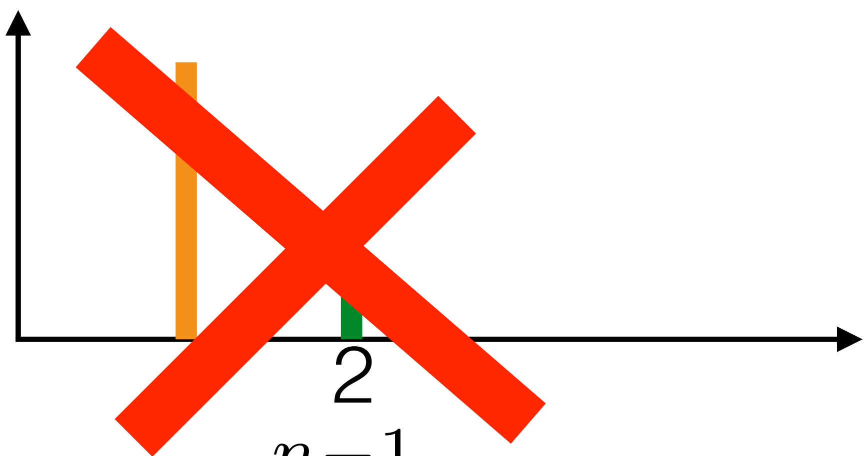
- Pólya urn
    - Choose any ball with equal probability
    - Replace and add ball of same color



Marginal cluster assignments

- Integrate out the frequencies

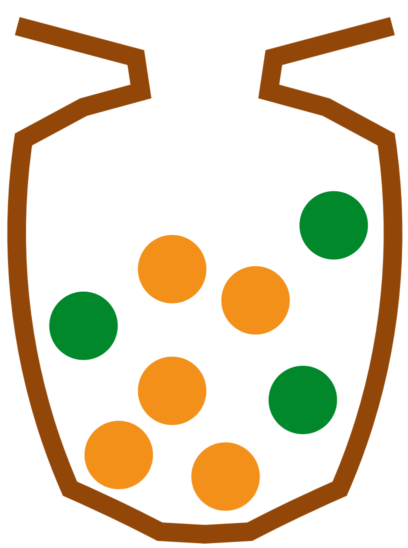
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$



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

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

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



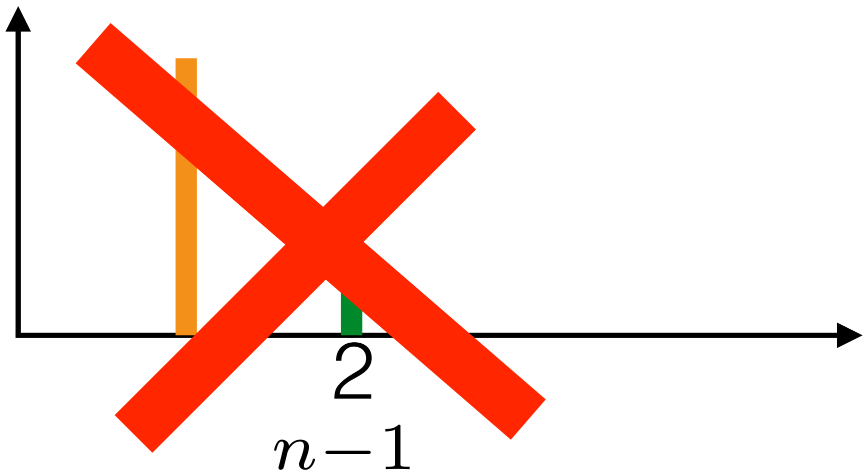
Marginal cluster assignments

- Integrate out the frequencies

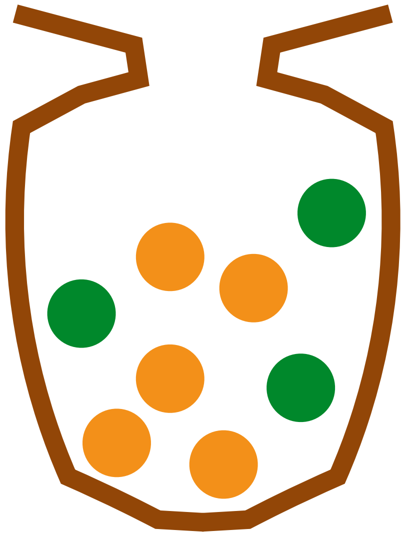
$$\rho_1 \sim \mathrm{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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

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



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



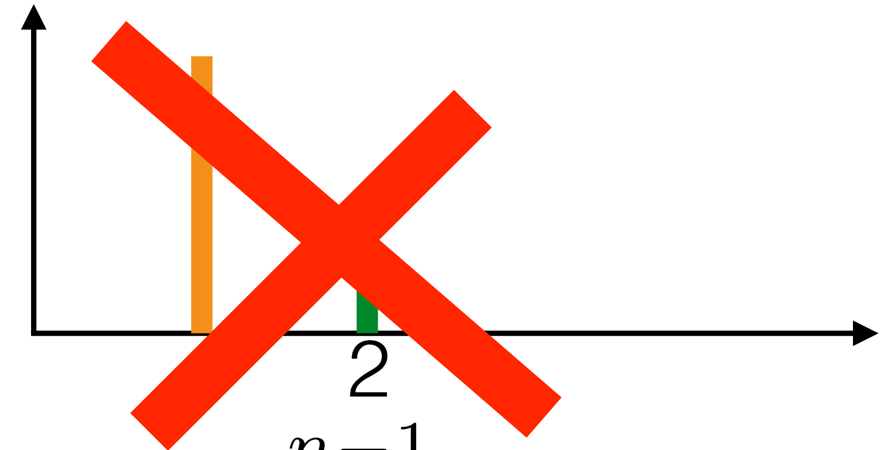
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}}$$

Marginal cluster assignments

- Integrate out the frequencies

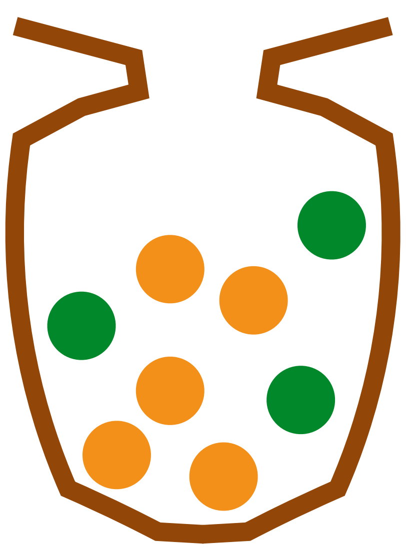
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

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



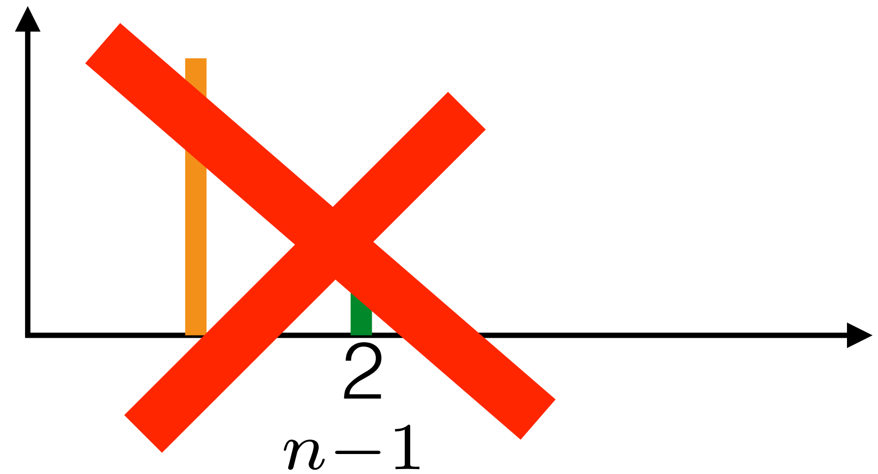
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}}$$

Marginal cluster assignments

- Integrate out the frequencies

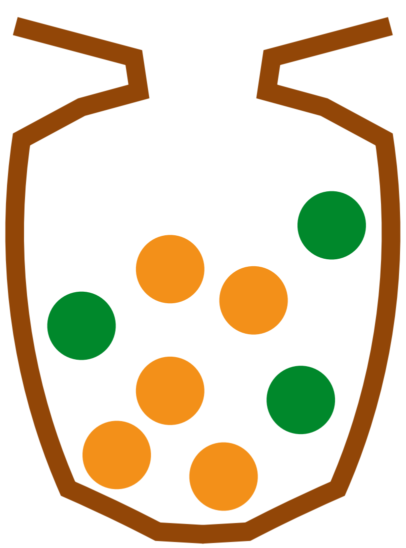
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

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



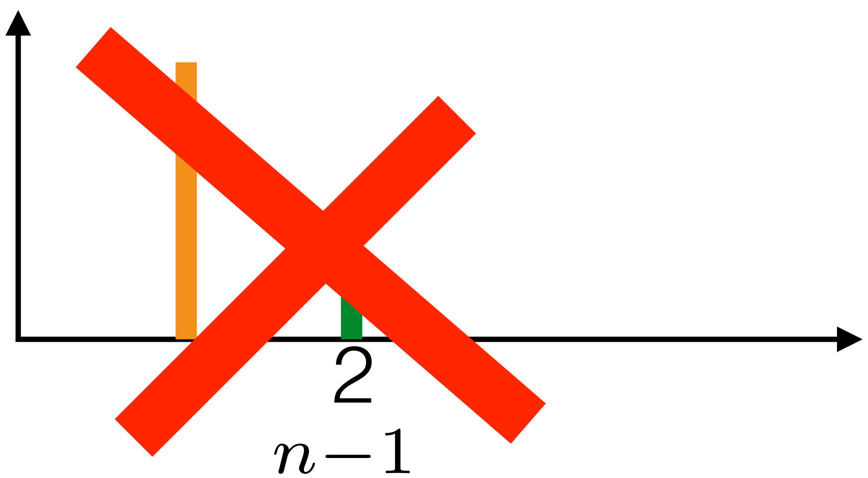
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

Marginal cluster assignments

- Integrate out the frequencies

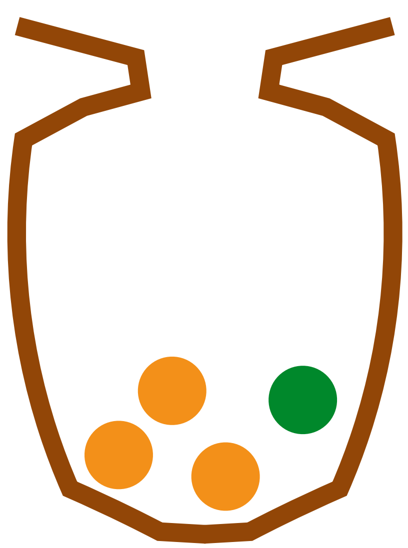
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

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



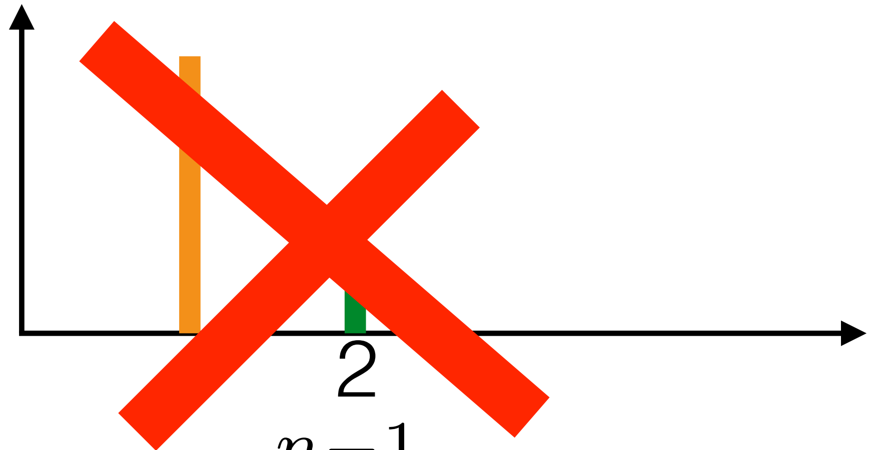
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

Marginal cluster assignments

- Integrate out the frequencies

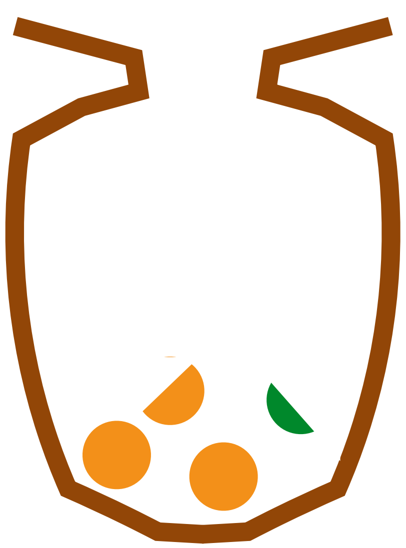
$$\rho_1 \sim \text{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

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



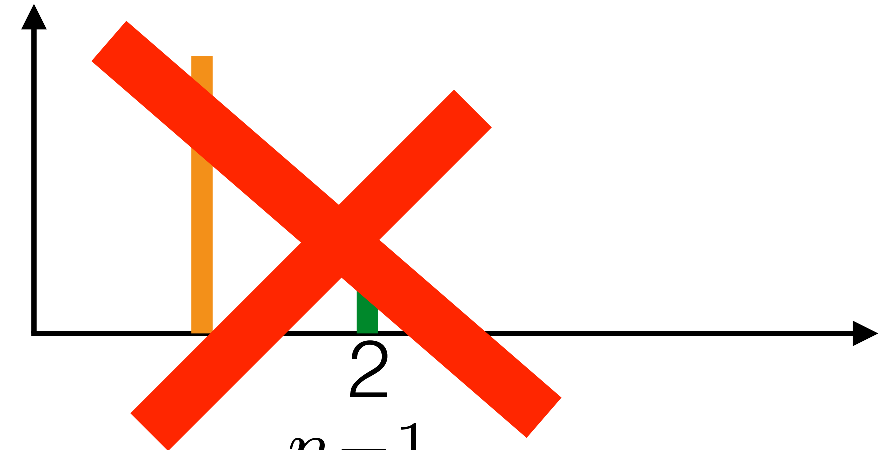
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

Marginal cluster assignments

- Integrate out the frequencies

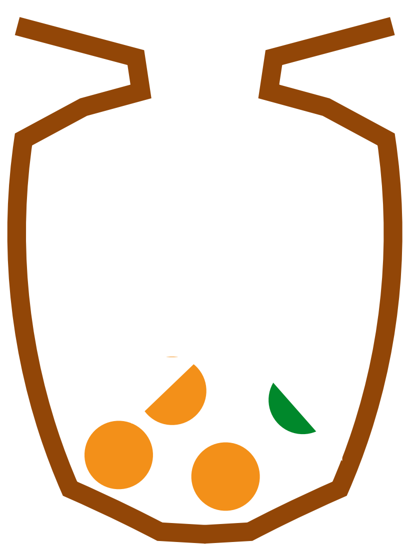
$$\rho_1 \sim \mathrm{Beta}(a_1, a_2),\, z_n \overset{iid}{\sim} \mathrm{Cat}(\rho_1, \rho_2)$$

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



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

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



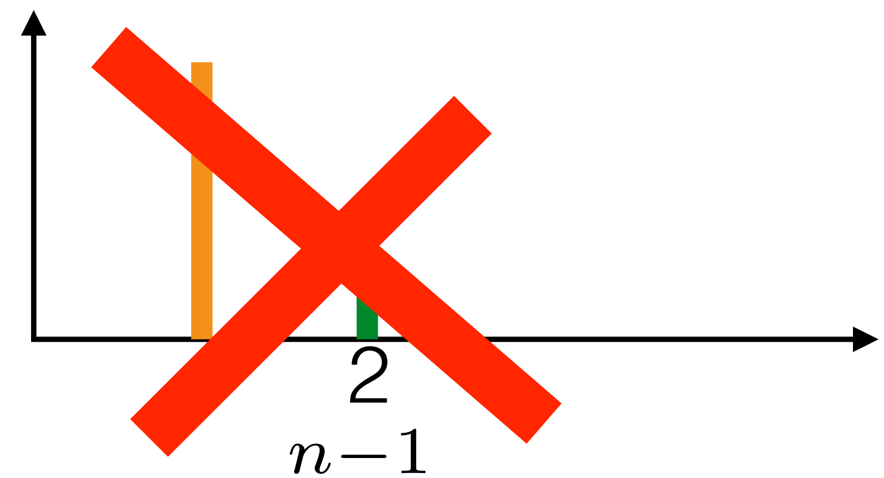
$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \mathrm{Beta}(a_{\text{orange}}, a_{\text{green}})$$

Marginal cluster assignments

- Integrate out the frequencies

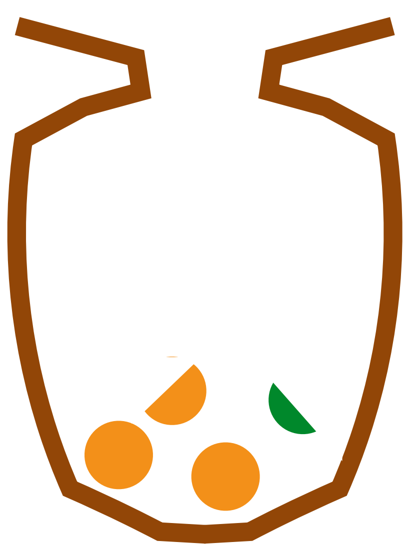
$$\rho_1 \sim \text{Beta}(a_1, a_2),\ z_n \overset{iid}{\sim} \text{Cat}(\rho_1, \rho_2)$$

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



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

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

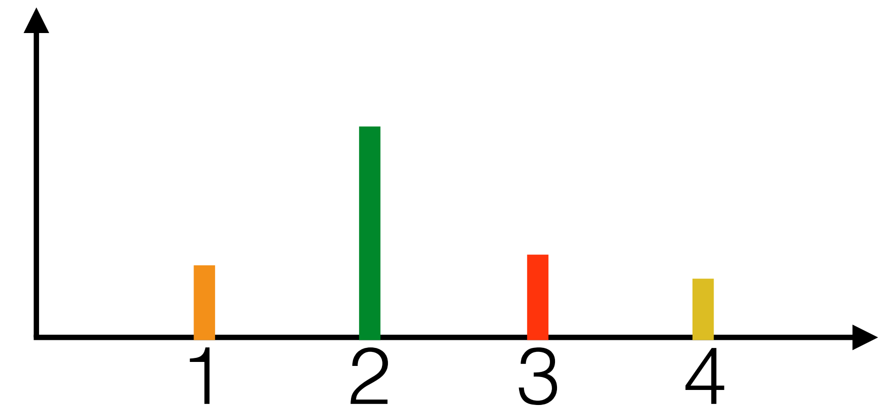


$$\lim_{n \to \infty} \frac{\#\ \text{orange}}{\#\ \text{total}} = \rho_{\text{orange}} \overset{d}{=} \text{Beta}(a_{\text{orange}}, a_{\text{green}})$$

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

Marginal cluster assignments

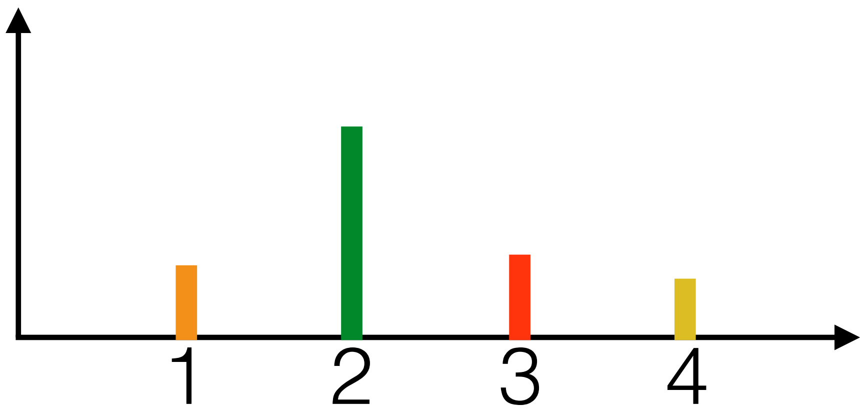
- Integrate out the frequencies



Marginal cluster assignments

- Integrate out the frequencies

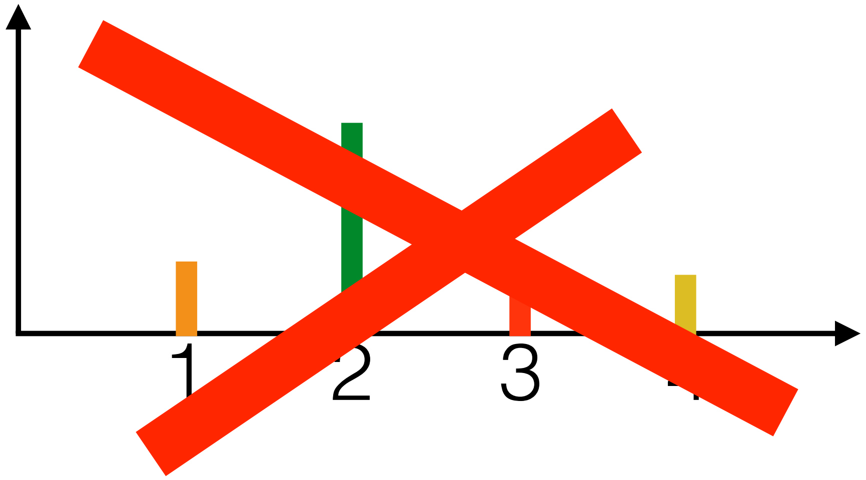
$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \stackrel{iid}{\sim} \text{Cat}(\rho_{1:K})$



Marginal cluster assignments

- Integrate out the frequencies

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

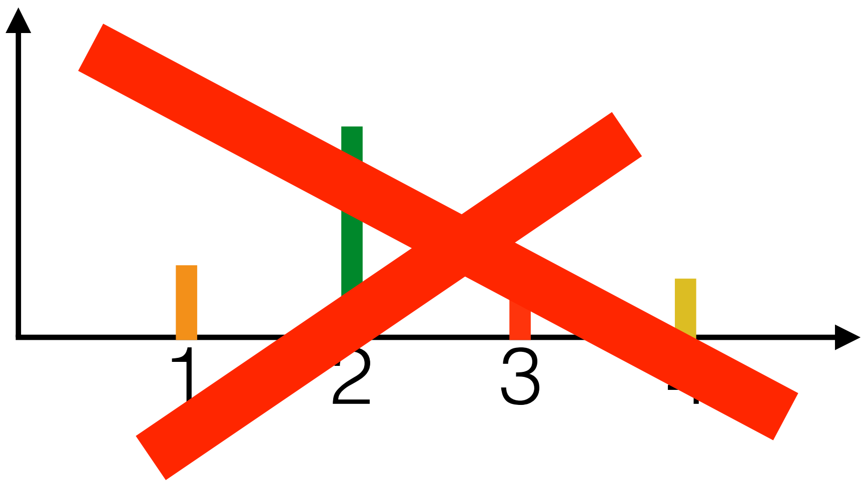


Marginal cluster assignments

- Integrate out the frequencies

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

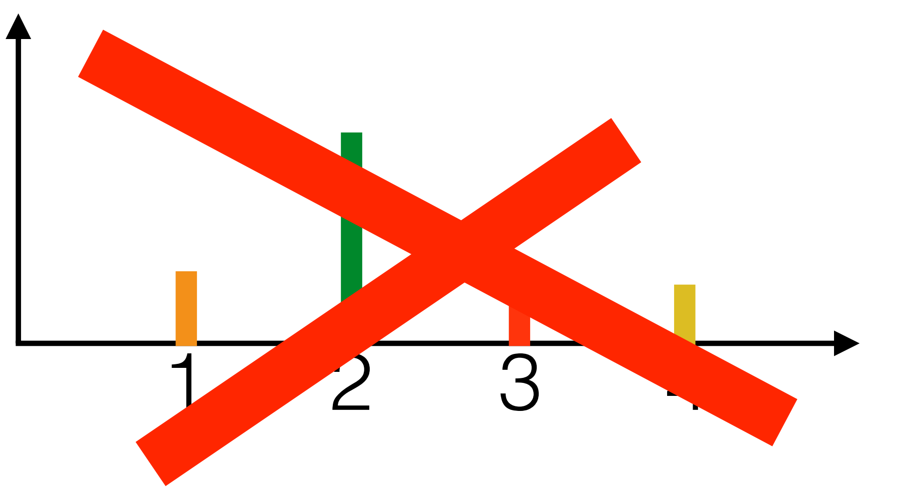
$$p(z_n = k | z_1, \ldots, z_{n-1}) = \frac{a_{k,n}}{\sum_{j=1}^{K} a_{j,n}}$$



Marginal cluster assignments

- Integrate out the frequencies

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



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

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

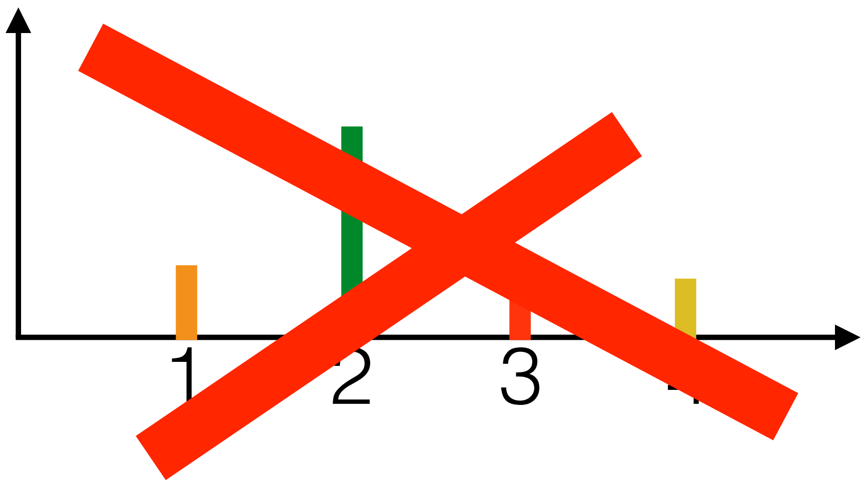
Marginal cluster assignments

- Integrate out the frequencies

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

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

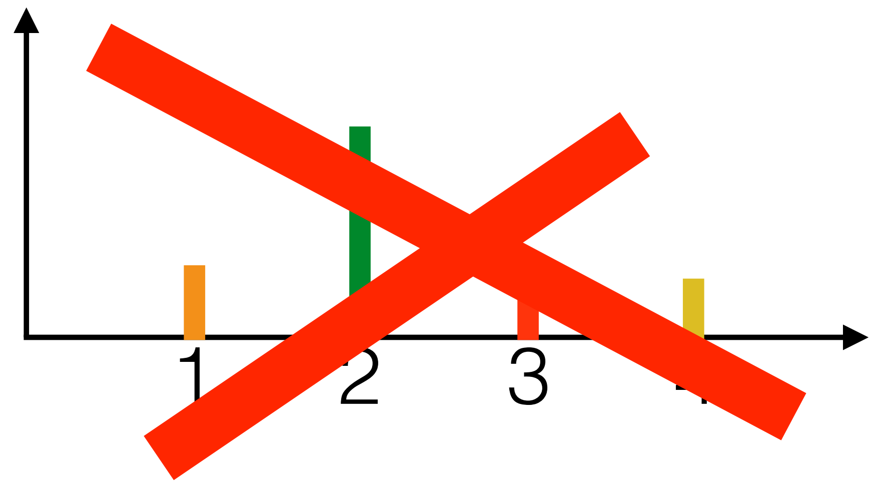
$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$



- multivariate Pólya urn

Marginal cluster assignments

- Integrate out the frequencies



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

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

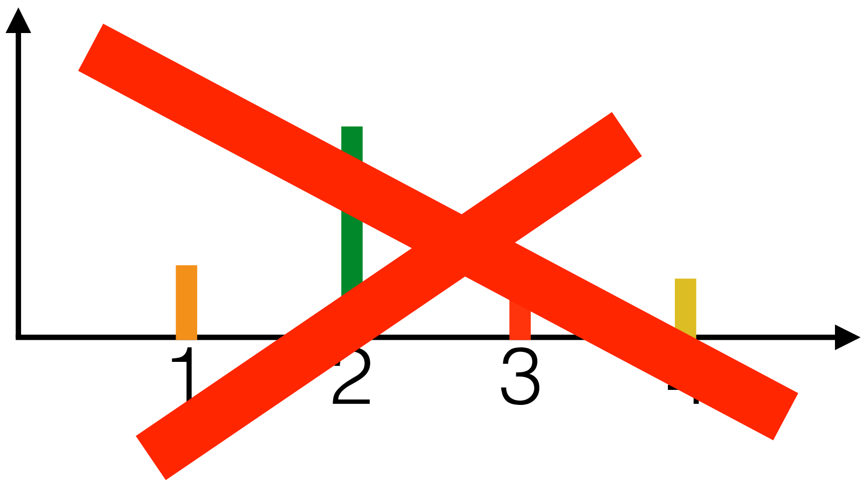
$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

- multivariate Pólya urn

Marginal cluster assignments

- Integrate out the frequencies

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



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

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

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

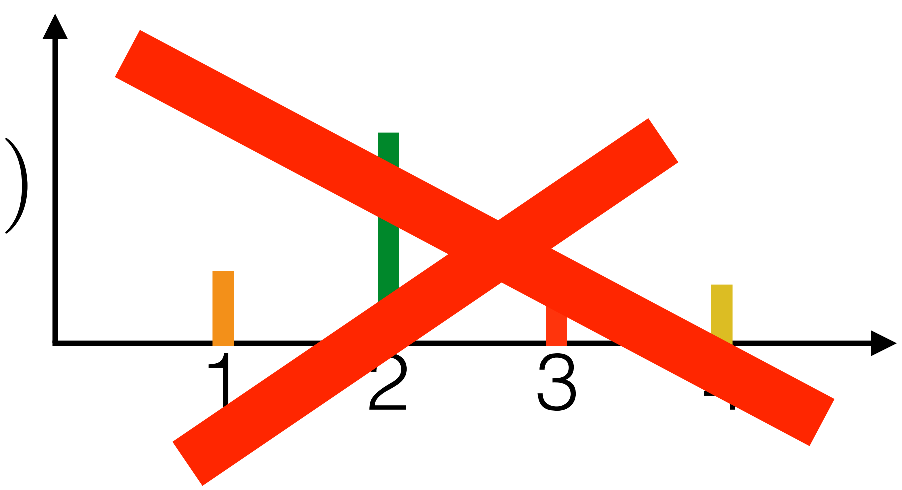
Marginal cluster assignments

- Integrate out the frequencies

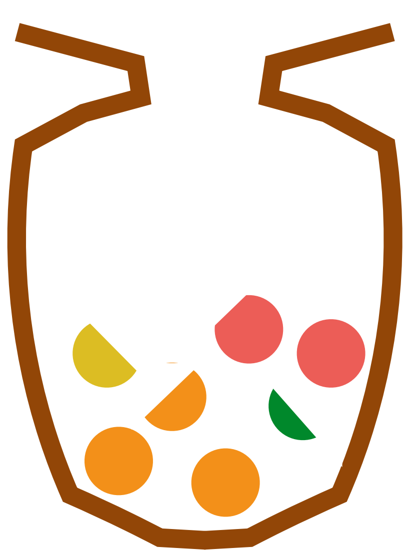
$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K}), z_n \overset{iid}{\sim} \text{Cat}(\rho_{1:K})$$

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

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

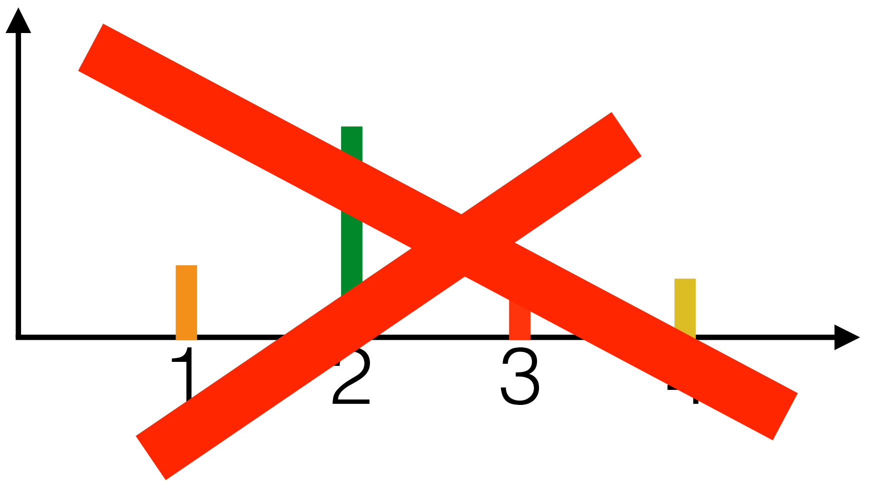


- multivariate Pólya urn
  - Choose any ball with prob proportional to its mass
  - Replace and add ball of same color



Marginal cluster assignments

- Integrate out the frequencies



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

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

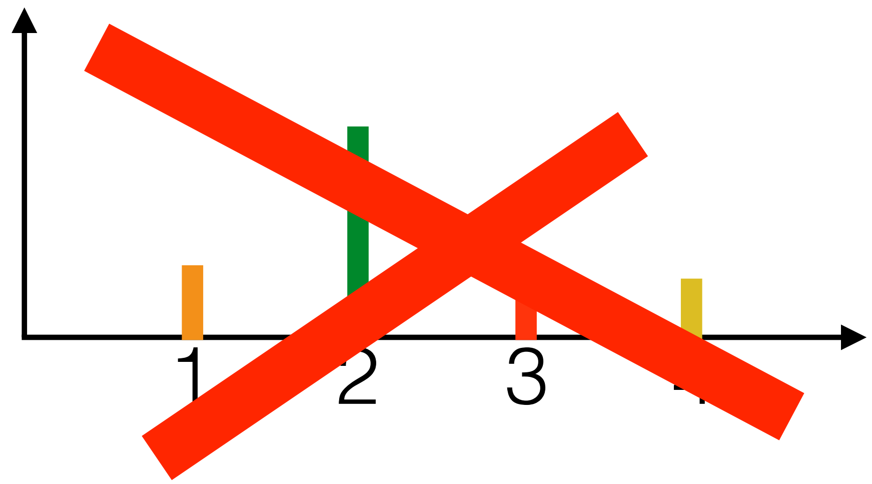
$$a_{k,n} := a_k + \sum_{m=1}^{n-1} \mathbf{1}\{z_m = k\}$$

- 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

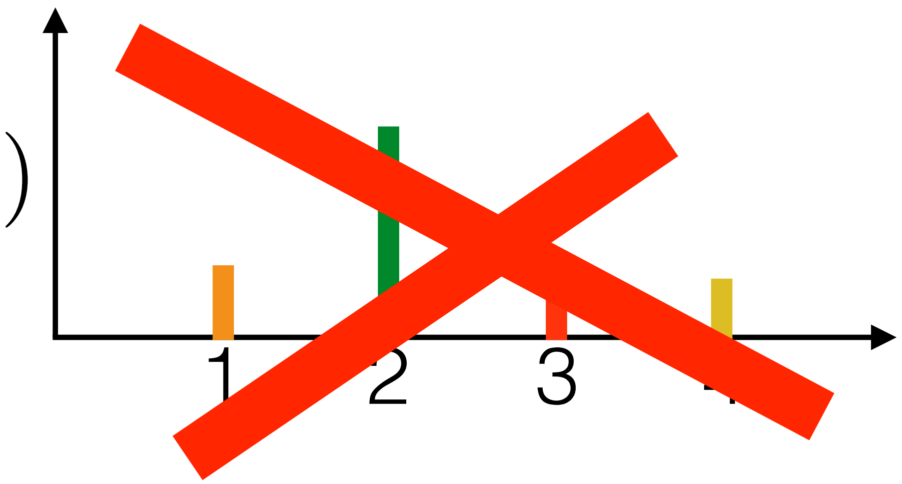
$$\lim_{n \to \infty} \frac{(\#\text{ orange}, \#\text{ green}, \#\text{ red}, \#\text{ yellow})}{\#\text{ total}}$$

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

Marginal cluster assignments

- Integrate out the frequencies

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



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

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

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

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

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

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

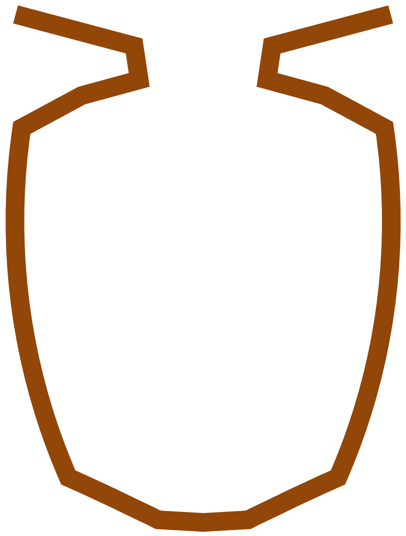
Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

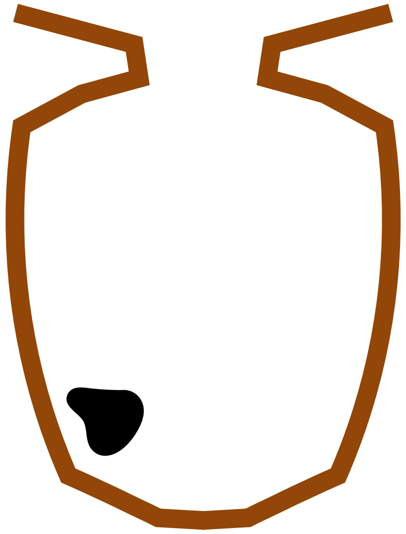
- Hoppe urn / Blackwell-MacQueen urn



[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

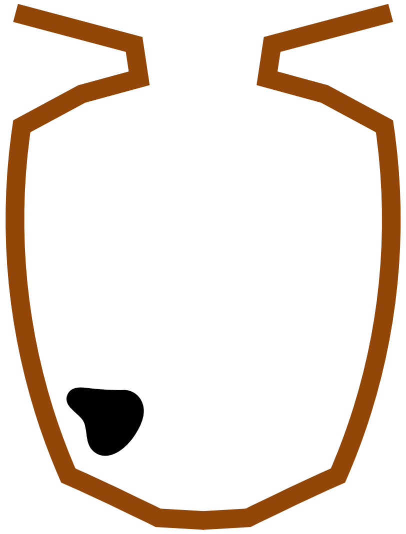
- Hoppe urn / Blackwell-MacQueen urn



[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

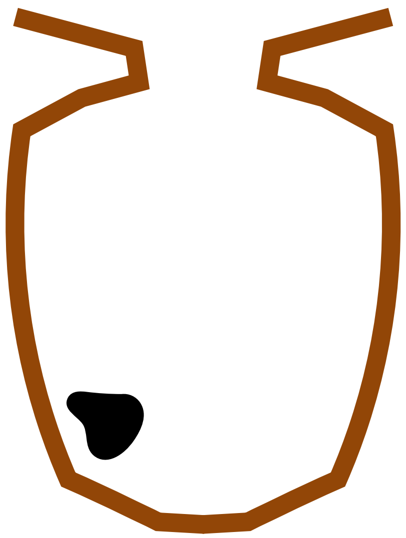


- Choose ball with prob proportional to its mass

[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

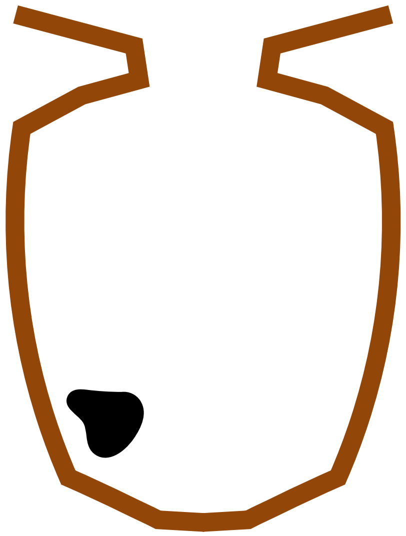


  - Choose ball with prob proportional to its mass
    - If black, replace and add ball of new color

[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn

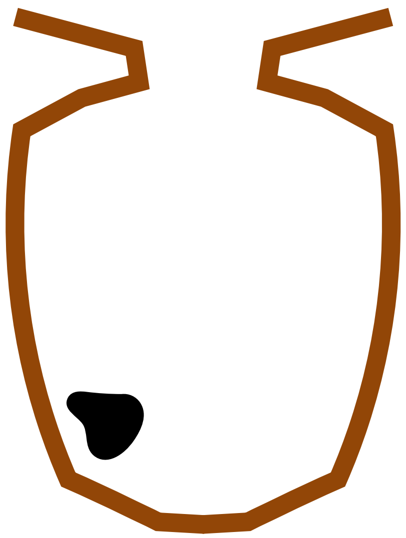


- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color

[Blackwell, MacQueen 1973; Hoppe 1984]

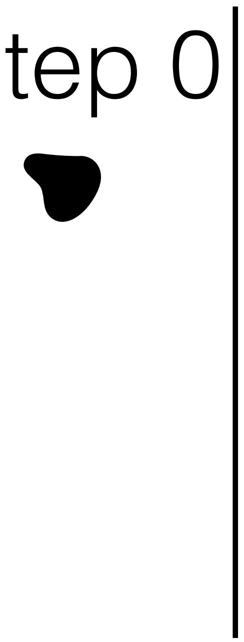
Marginal cluster assignments

- 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

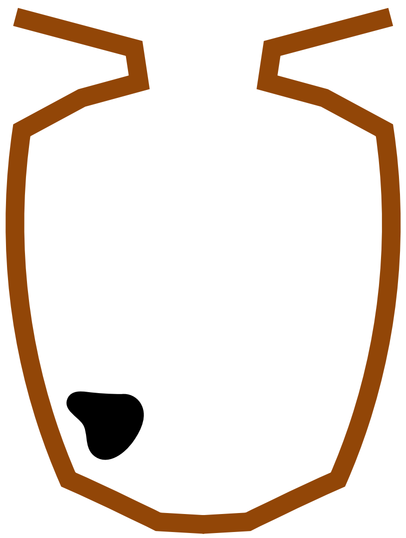
Step 0



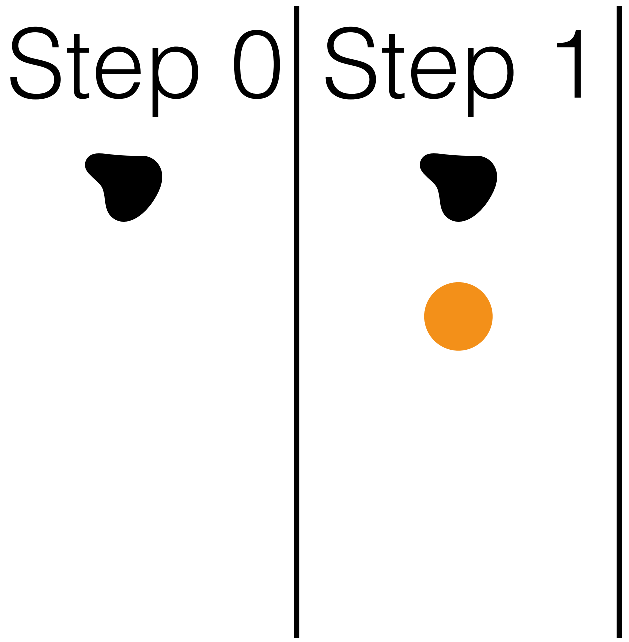
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



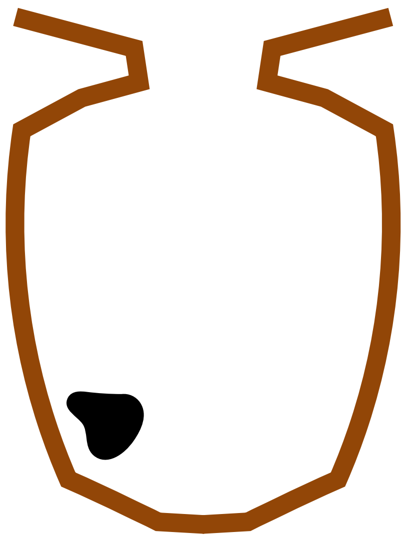
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



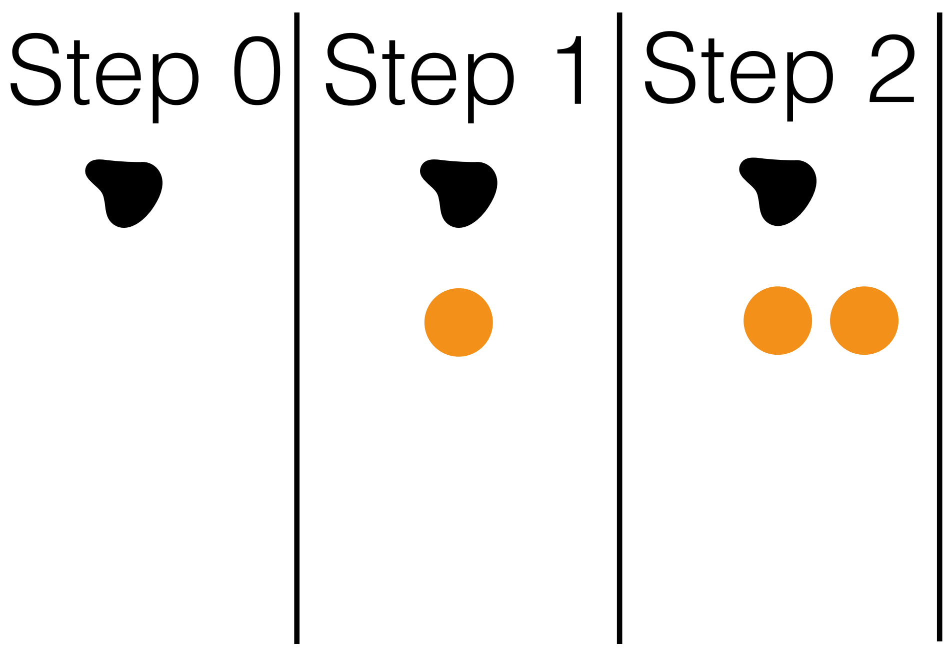
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



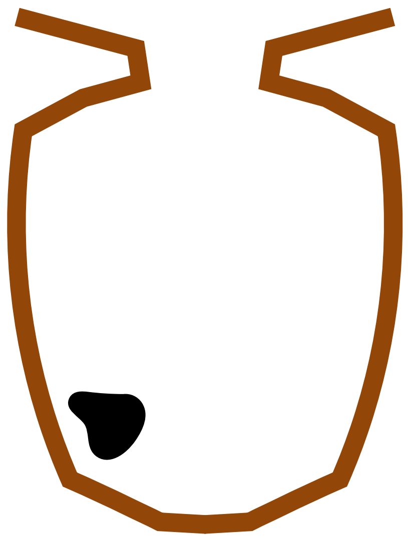
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



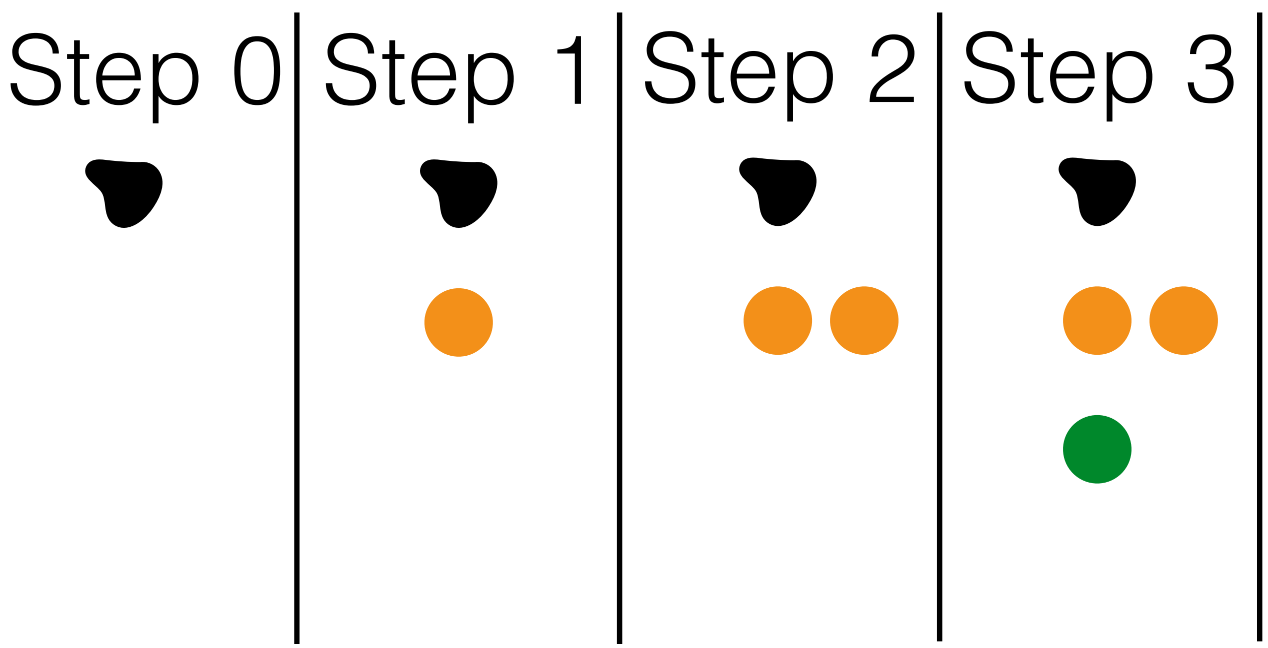
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



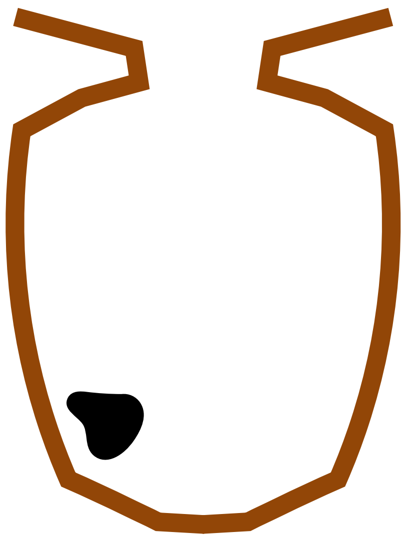
  - Choose ball with prob proportional to its mass
    - If black, replace and add ball of new color
    - Else, replace and add ball of same color



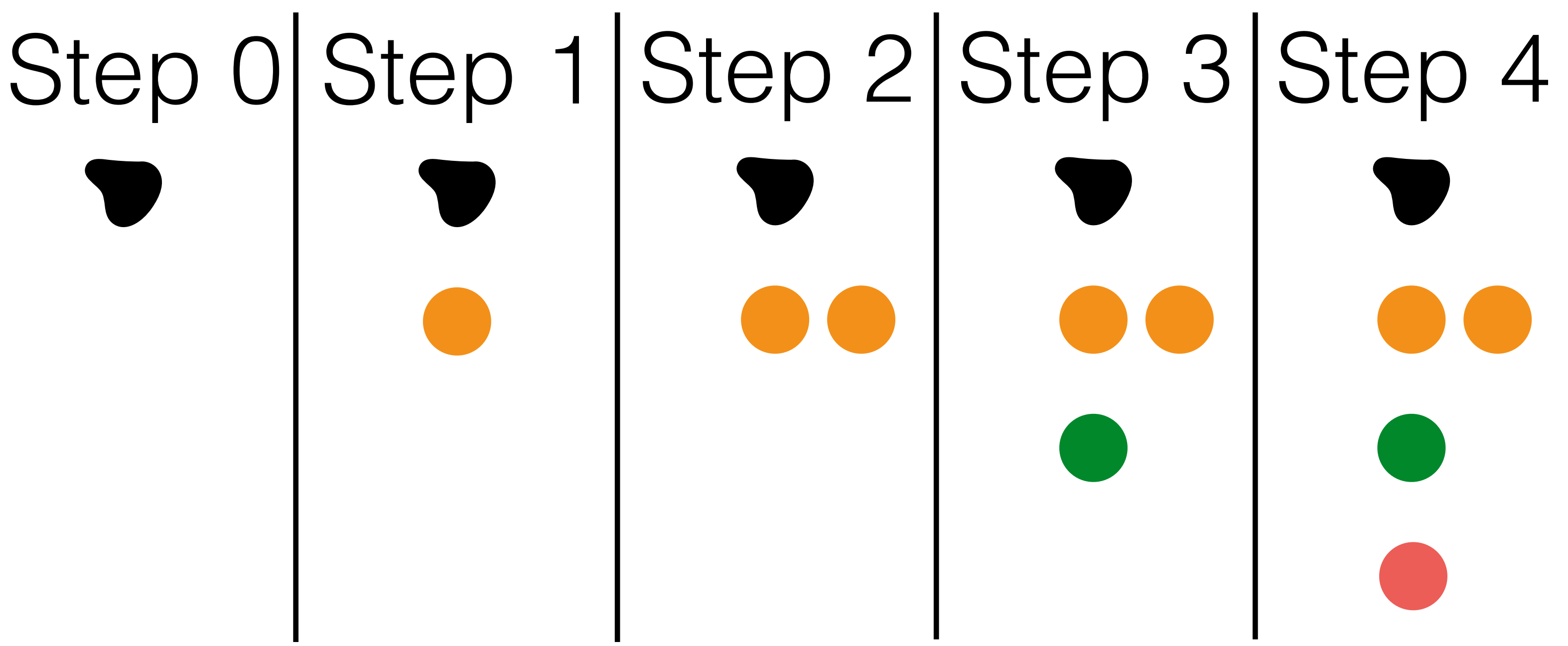
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



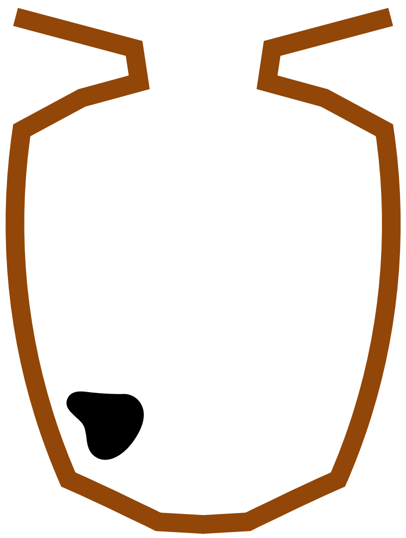
  - Choose ball with prob proportional to its mass
    - If black, replace and add ball of new color
    - Else, replace and add ball of same color



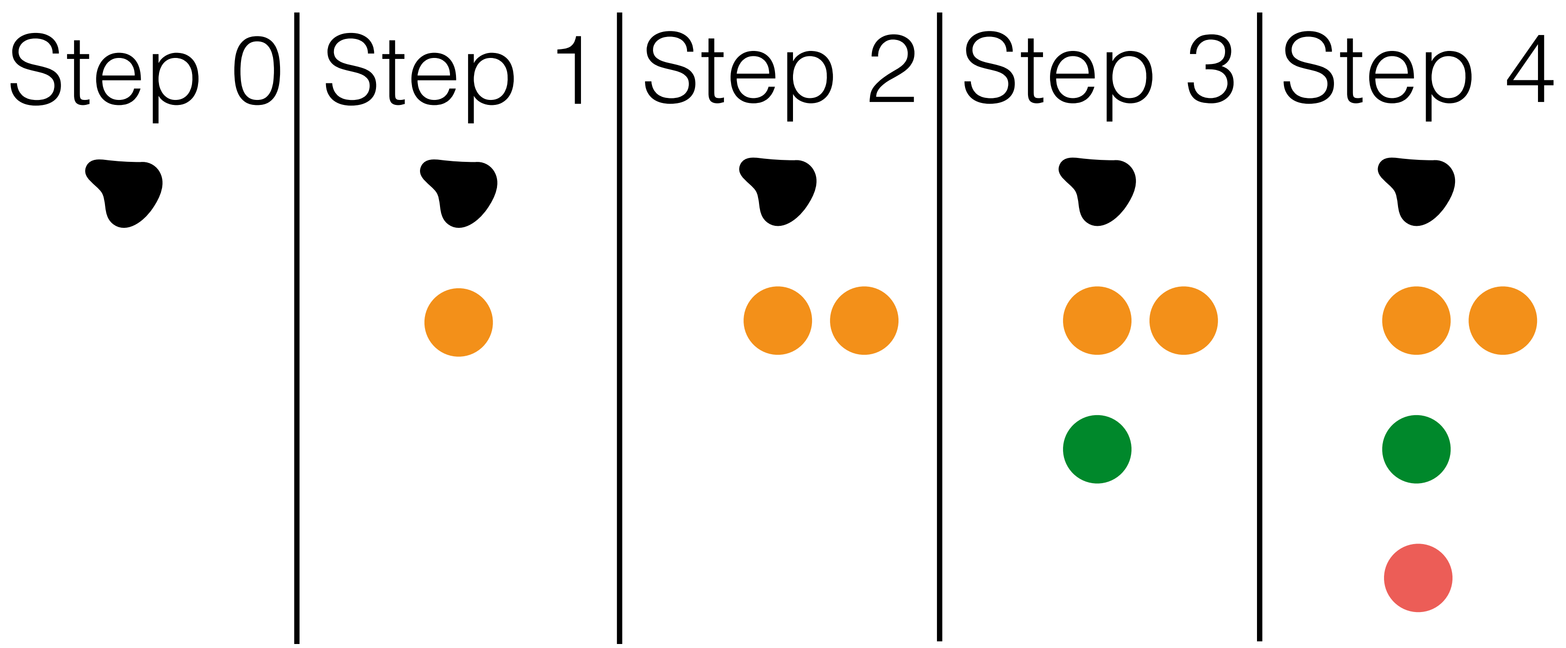
[Blackwell, MacQueen 1973; Hoppe 1984]

Marginal cluster assignments

- Hoppe urn / Blackwell-MacQueen urn



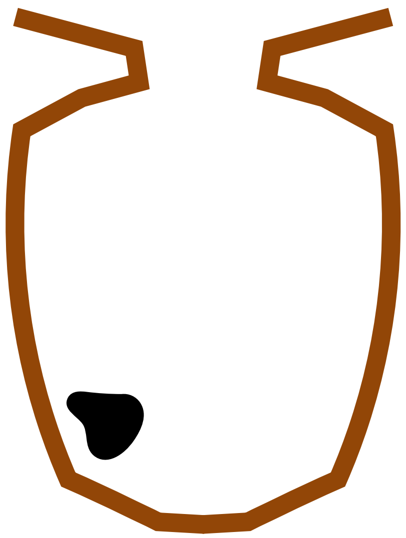
- Choose ball with prob proportional to its mass
  - If black, replace and add ball of new color
  - Else, replace and add ball of same color



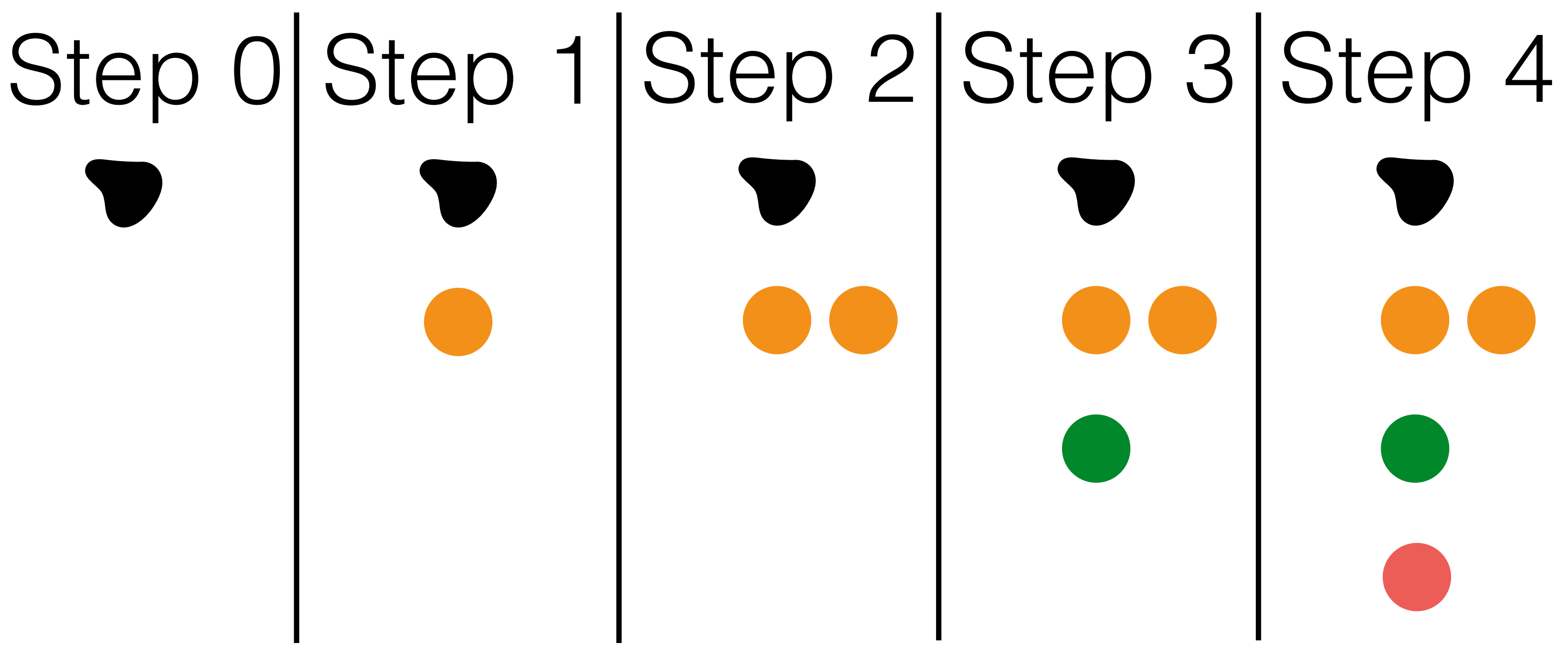
$$(\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$$

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

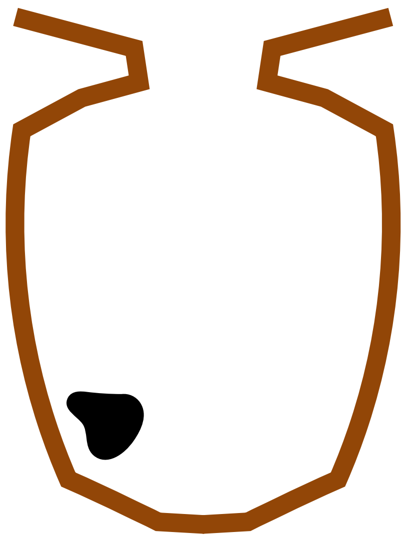


$$(\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$$

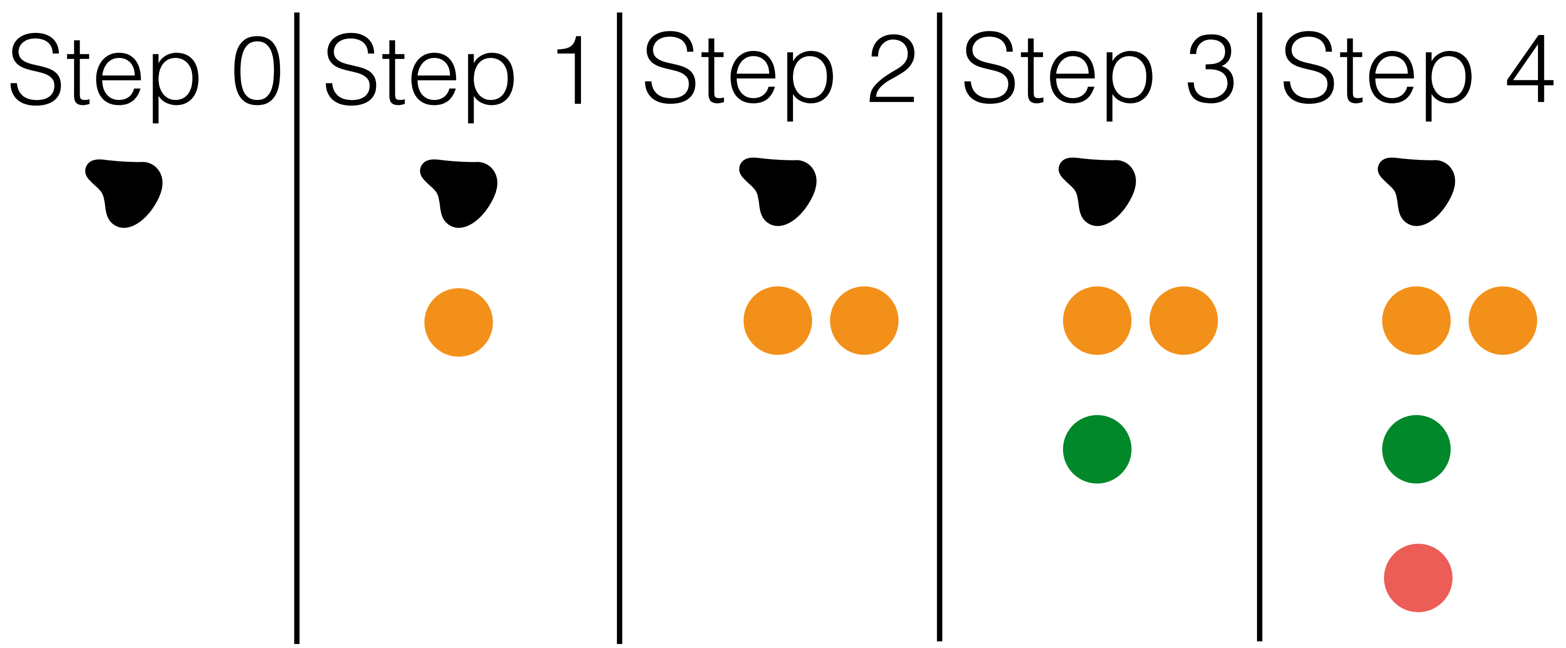
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

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

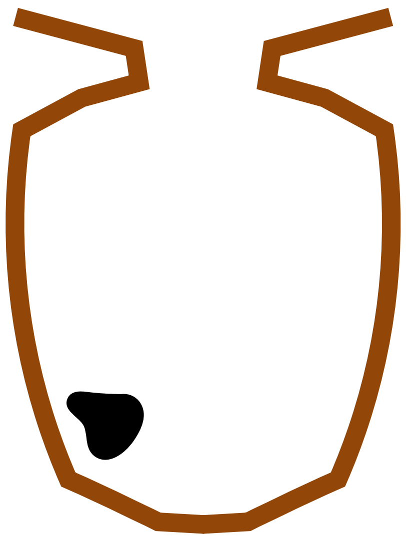


$(\#\text{orange}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

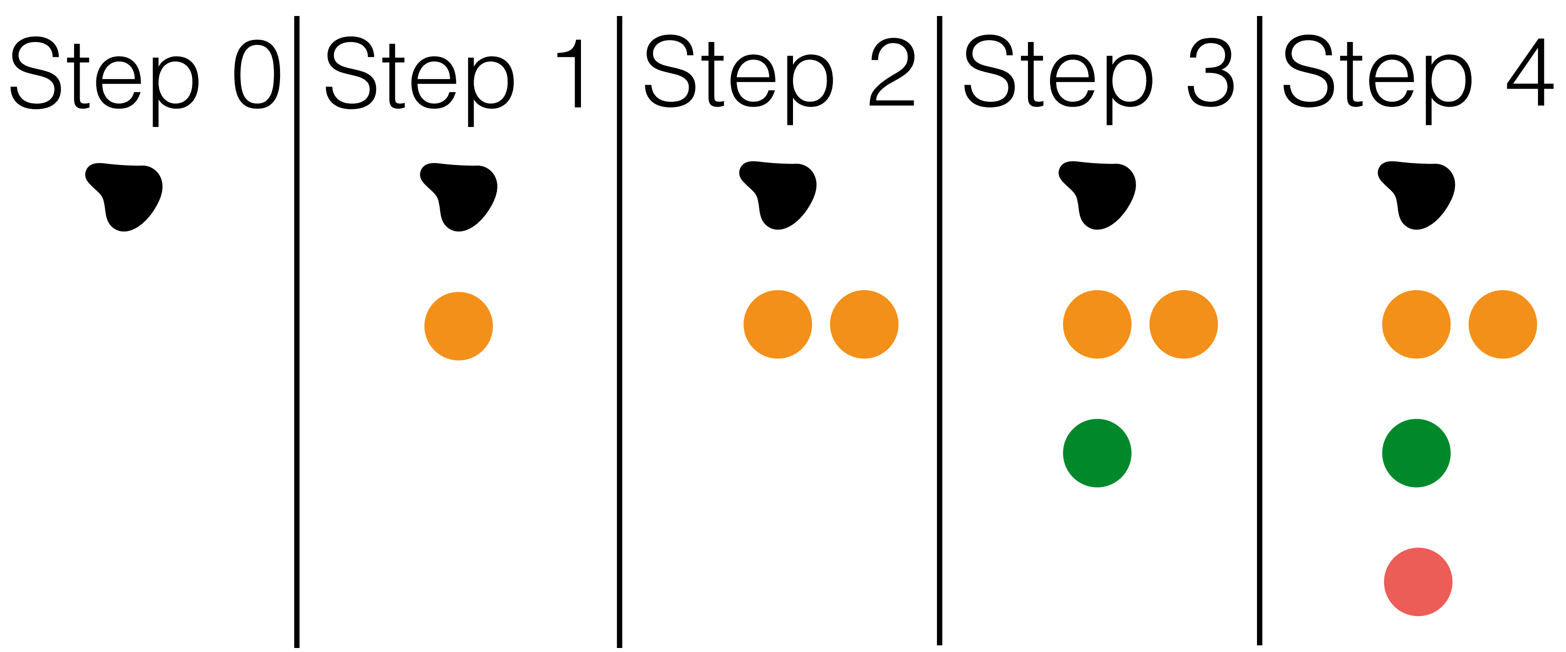
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

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



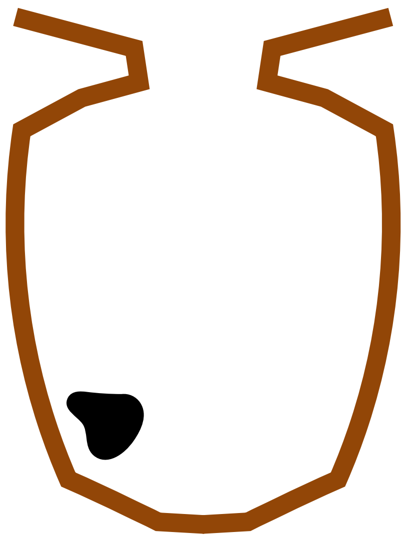
$$V_k \overset{iid}{\sim} \text{Beta}(1, \alpha)$$

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

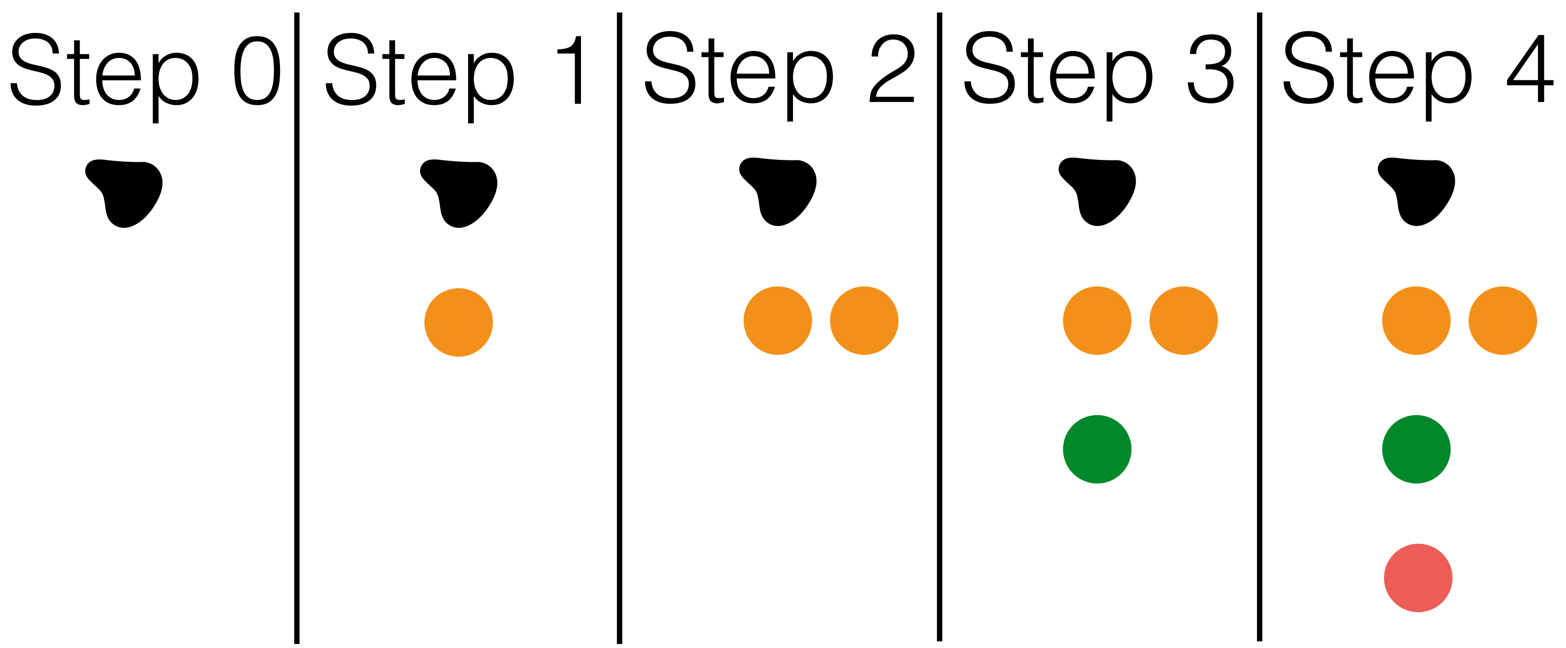
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

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



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

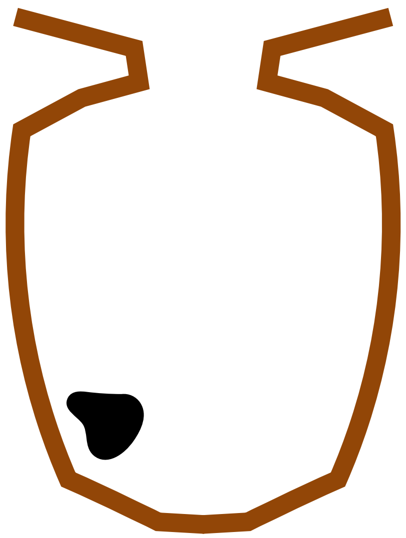
$$\rho_1 = V_1$$

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

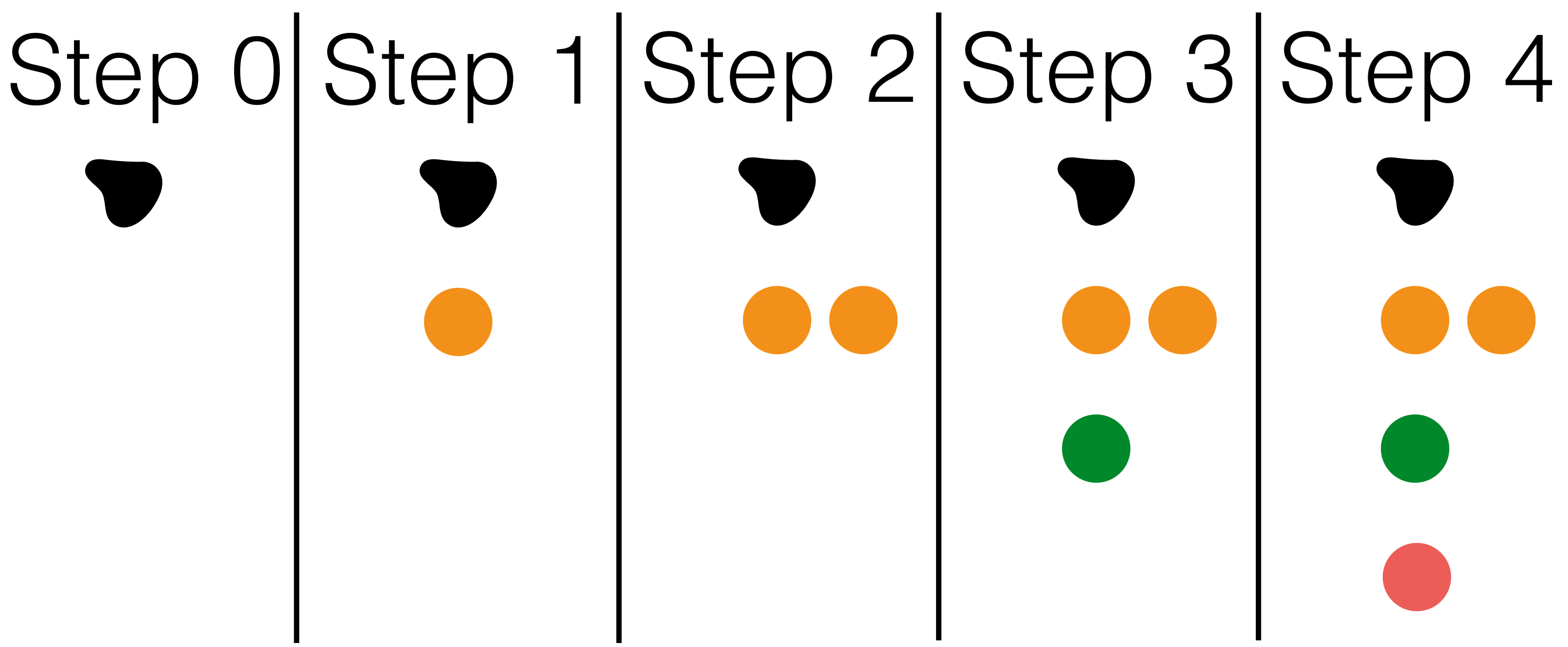
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

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



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

$$\rho_1 = V_1$$

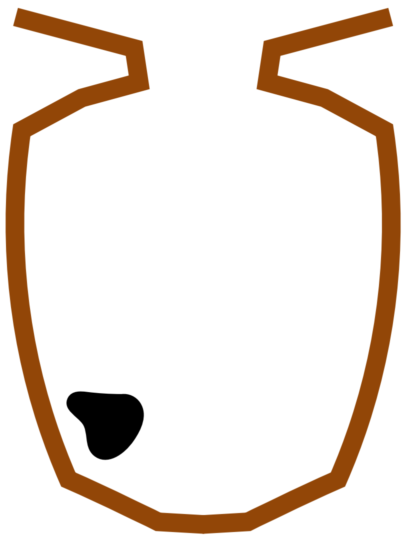
$$\rho_2 = (1 - V_1)V_2$$

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

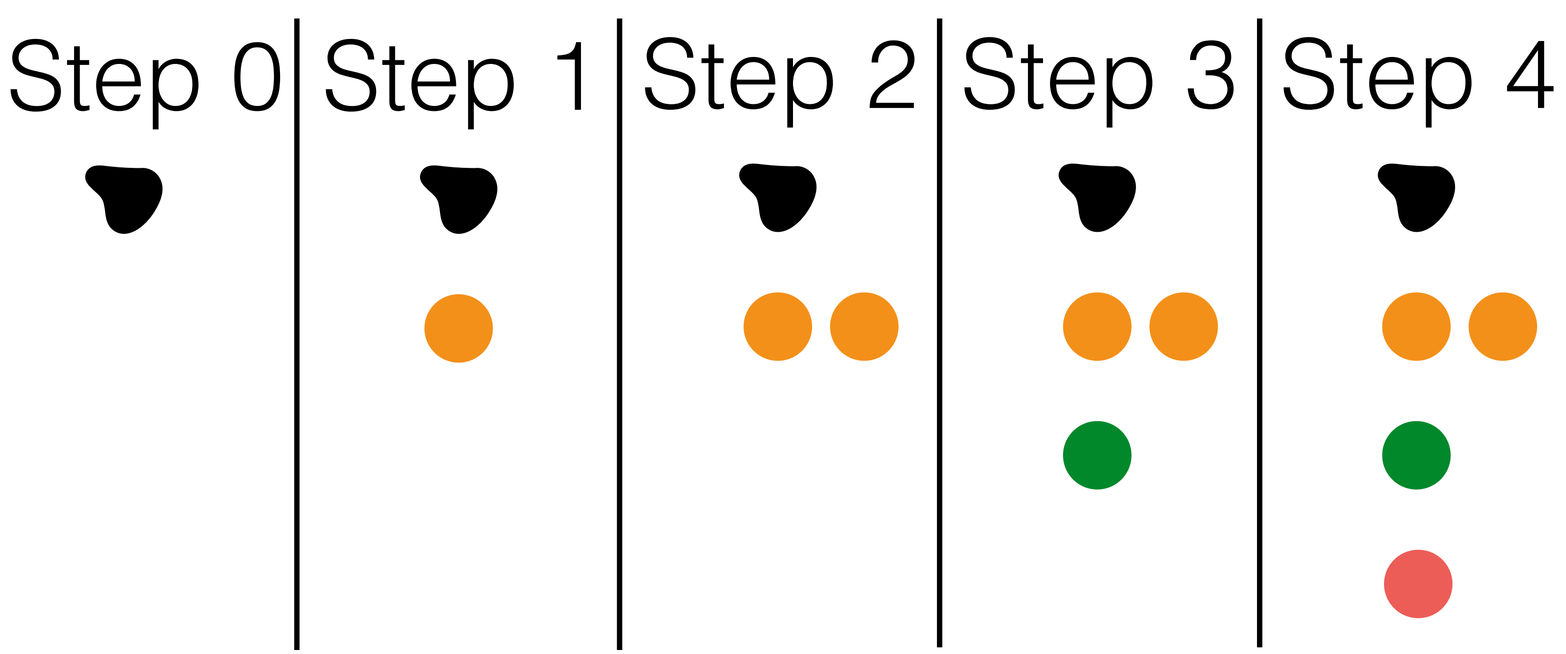
- not orange: $(\#\text{green}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$
- not orange, green: $(\#\text{red}, \#\text{other}) = \text{PolyaUrn}(1, \alpha)$

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



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

$$\rho_1 = V_1$$

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

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

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

- 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

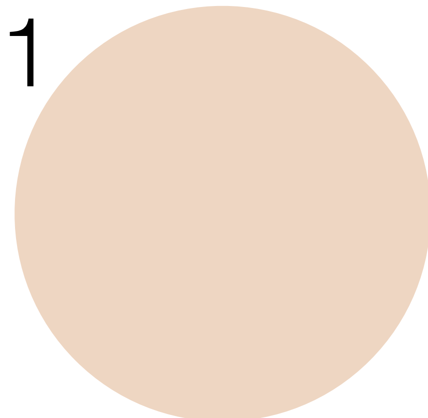
Chinese restaurant process

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

Chinese restaurant process

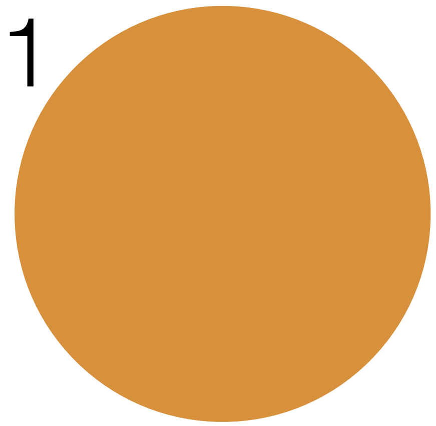
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



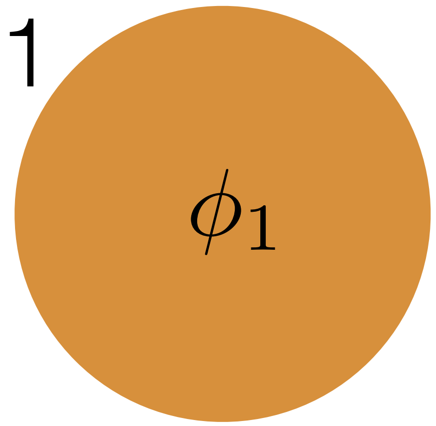
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

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

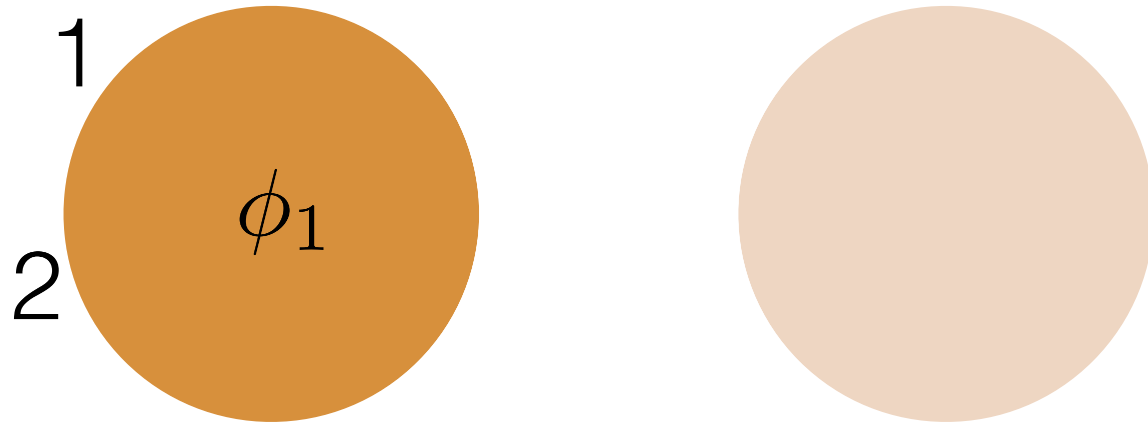


- Same thing we just did
- Each customer walks into the restaurant
    - Sits at existing table with prob proportional to # people there
    - Forms new table with prob proportional to α

Chinese restaurant process

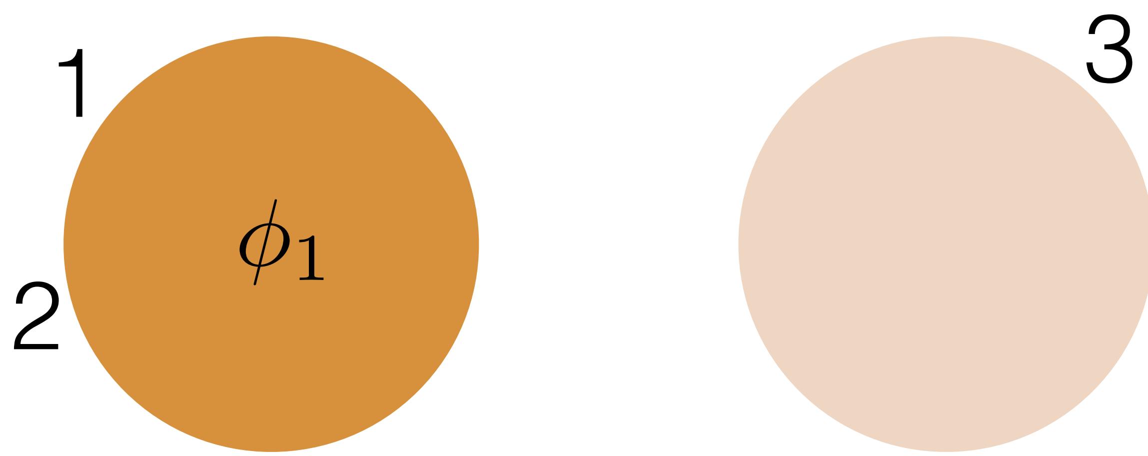
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

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

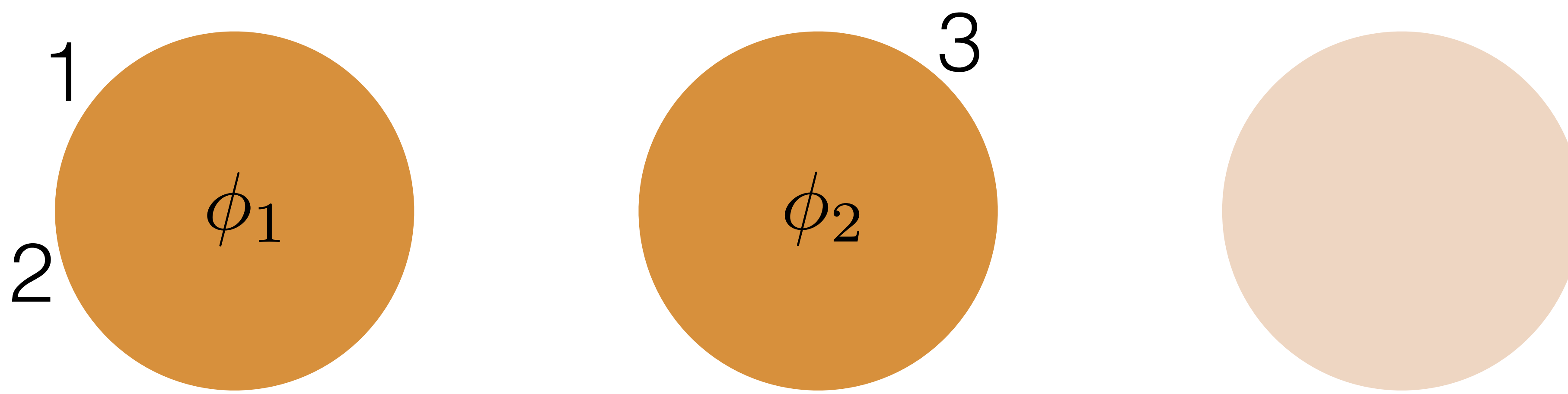
Chinese restaurant process

- 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

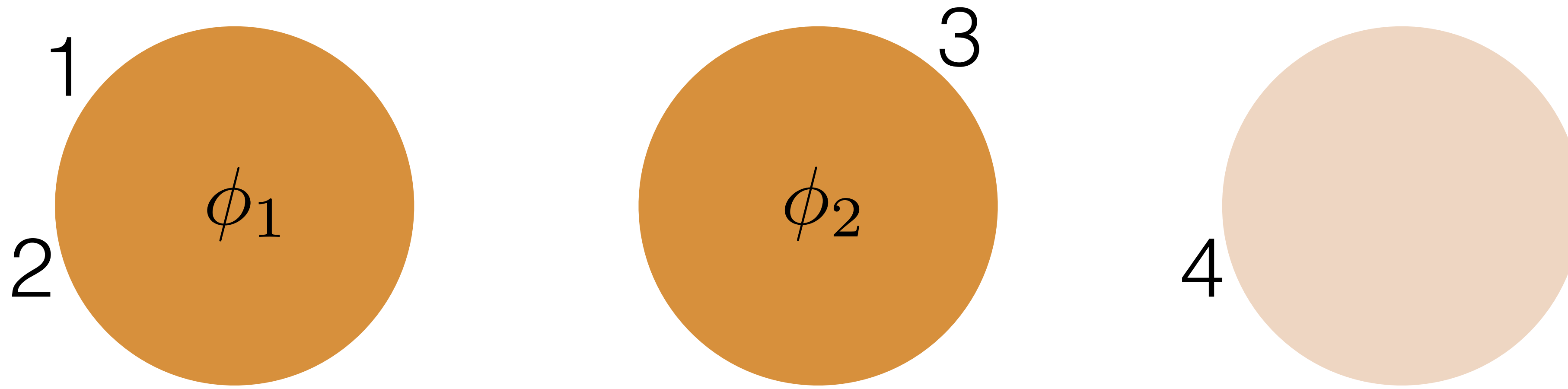
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



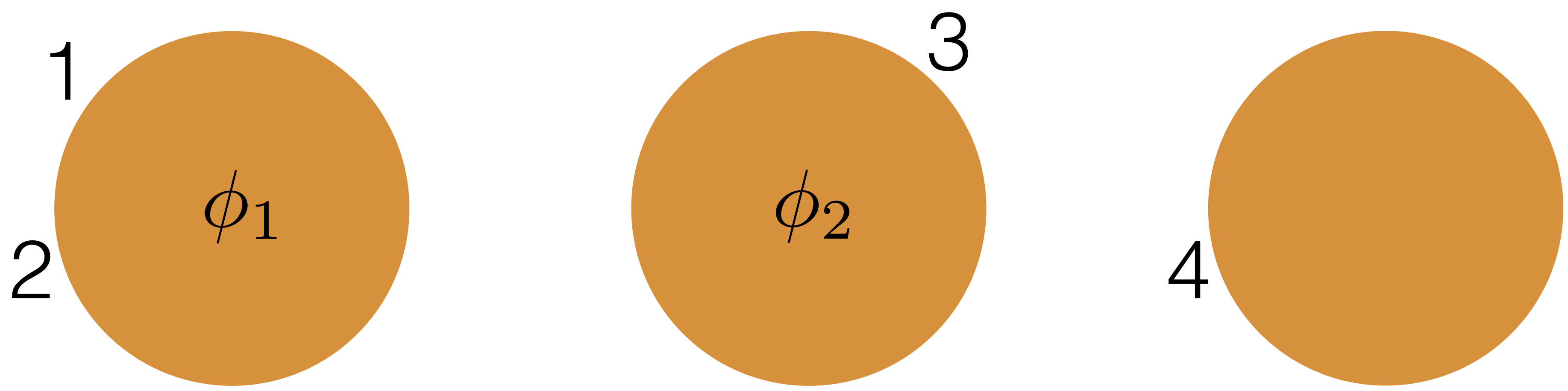
- Same thing we just did
- Each customer walks into the restaurant
    - Sits at existing table with prob proportional to # people there
    - Forms new table with prob proportional to α

Chinese restaurant process



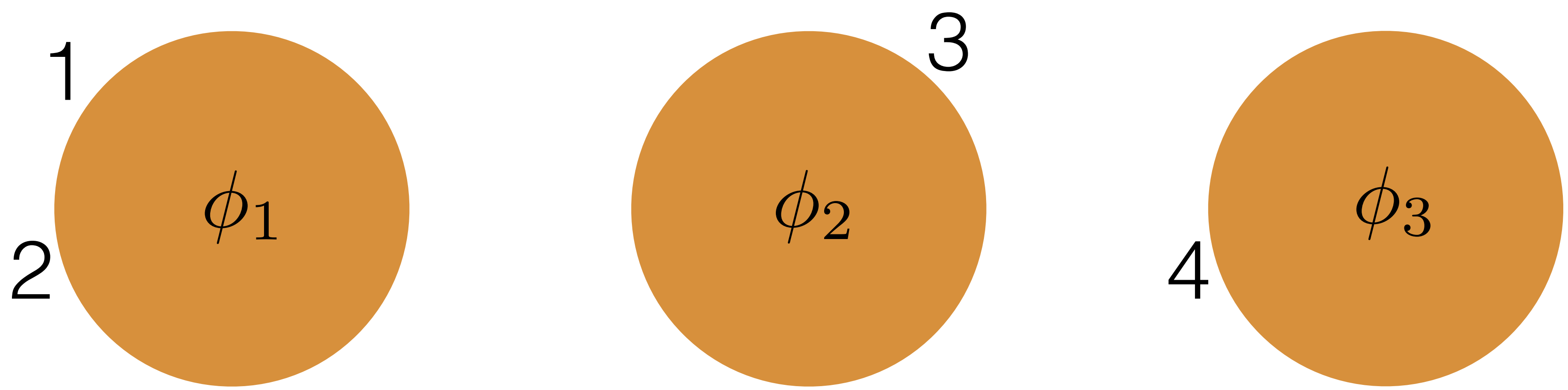
- Same thing we just did
- Each customer walks into the restaurant
    - Sits at existing table with prob proportional to # people there
    - Forms new table with prob proportional to α

Chinese restaurant process



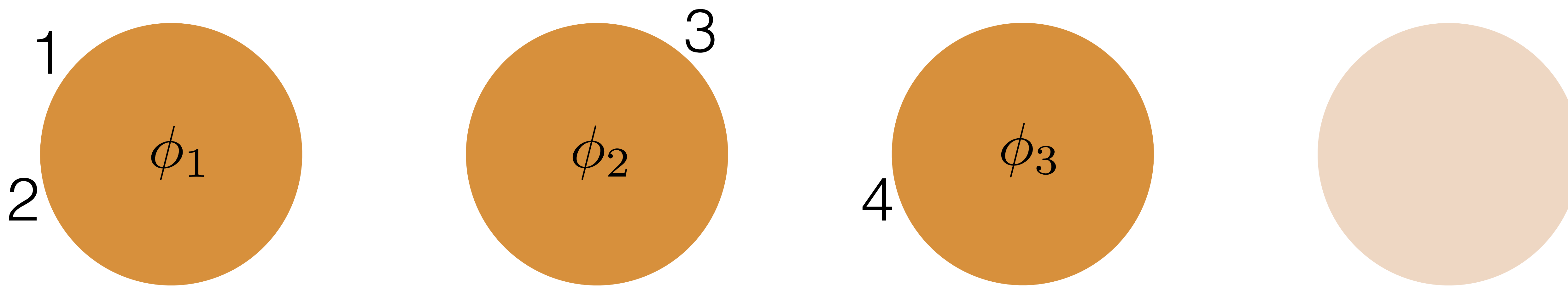
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



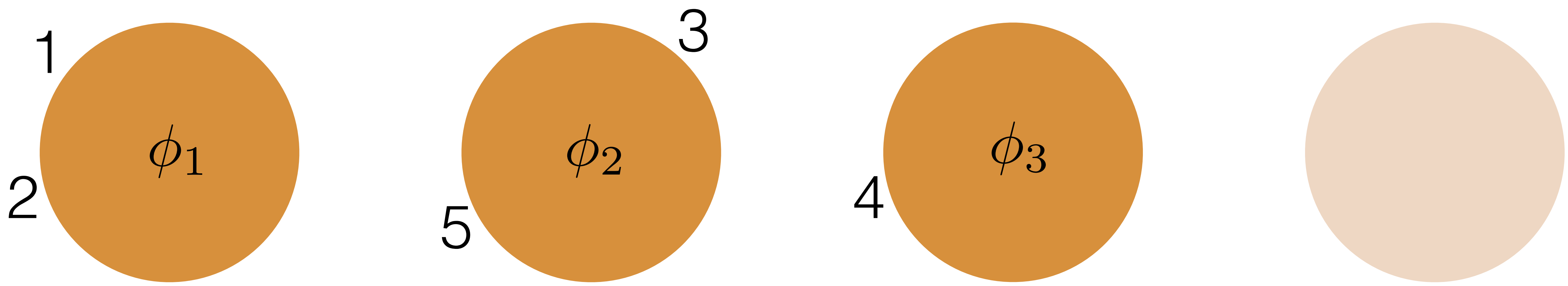
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



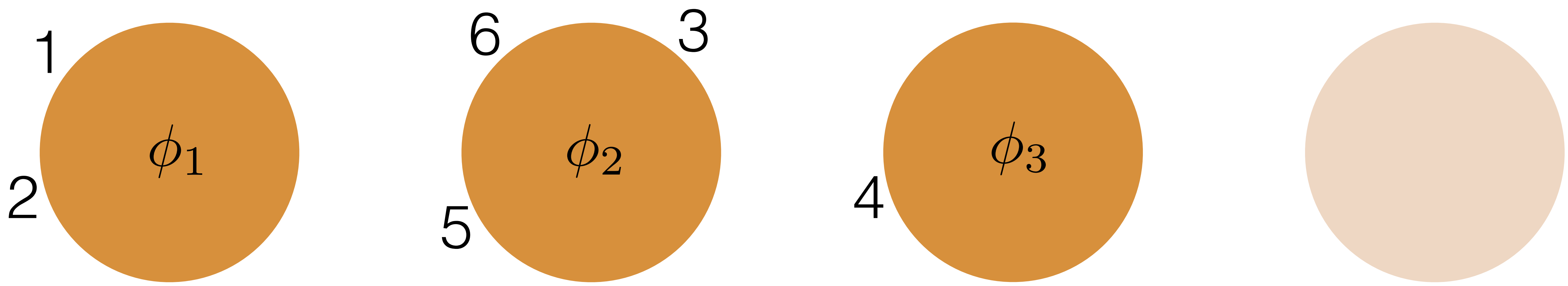
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



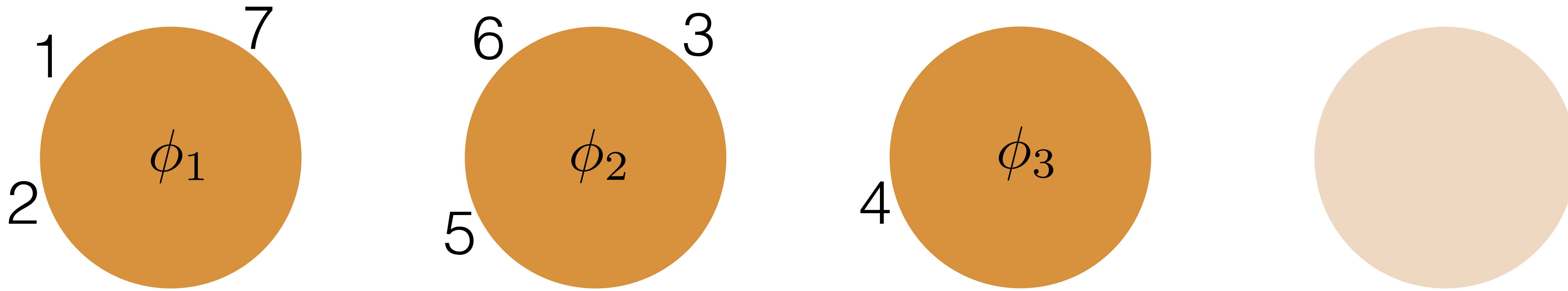
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

Chinese restaurant process



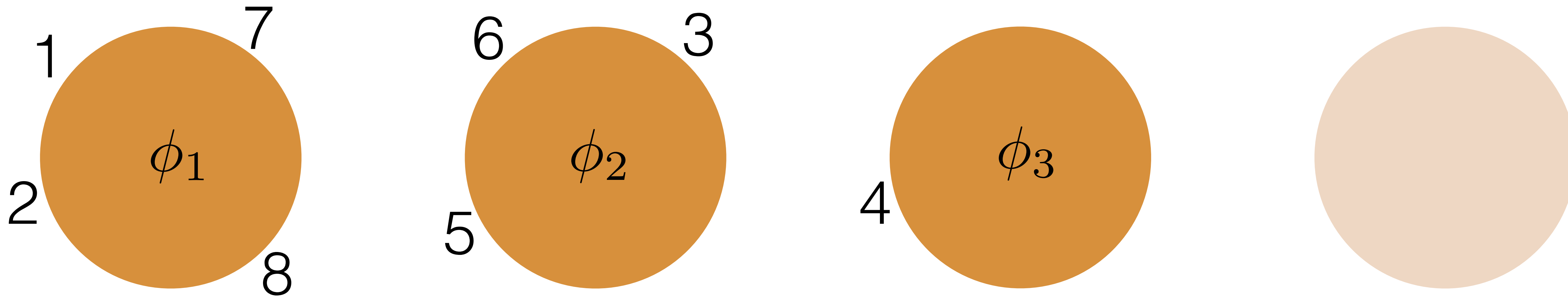
- Same thing we just did
- Each customer walks into the restaurant
    - Sits at existing table with prob proportional to # people there
    - Forms new table with prob proportional to α

Chinese restaurant process



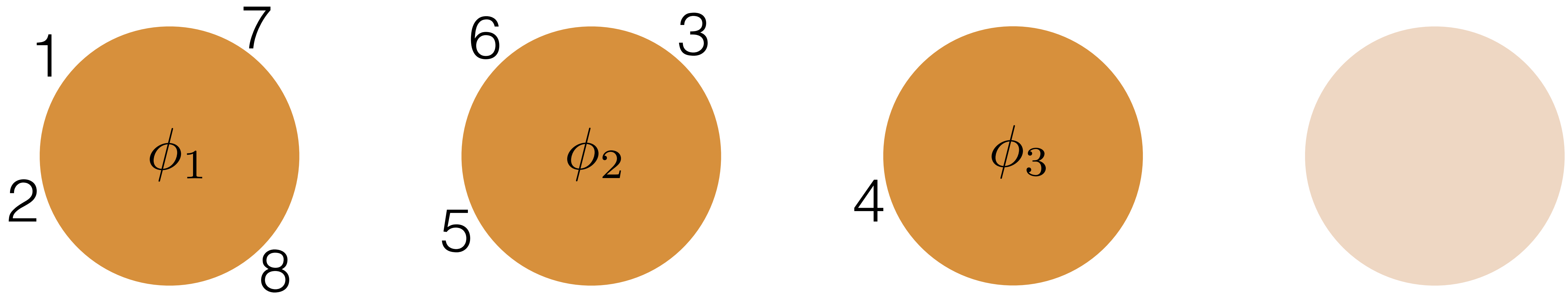
- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α

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

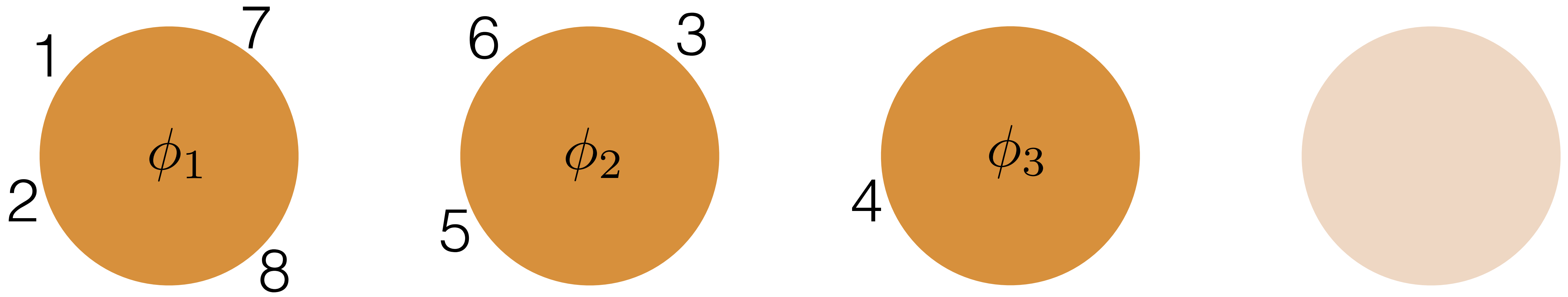
Chinese restaurant process



- 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

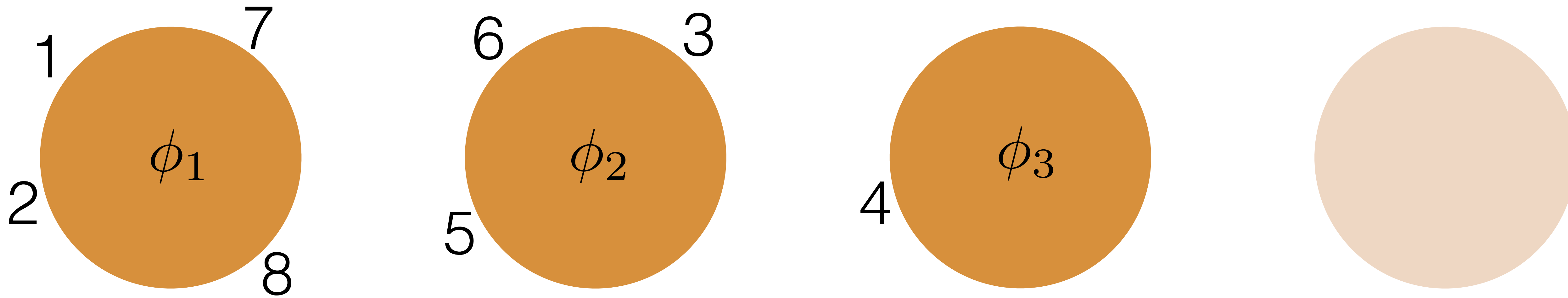
[Aldous 1983]

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

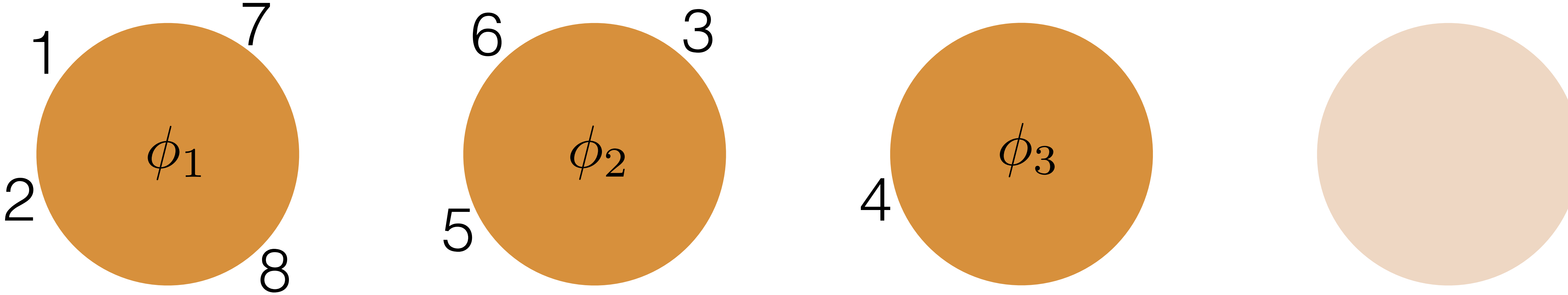


- Same thing we just did
- Each customer walks into the restaurant
  - Sits at existing table with prob proportional to # people there
  - Forms new table with prob proportional to α
- Marginal for the Categorical likelihood with GEM prior

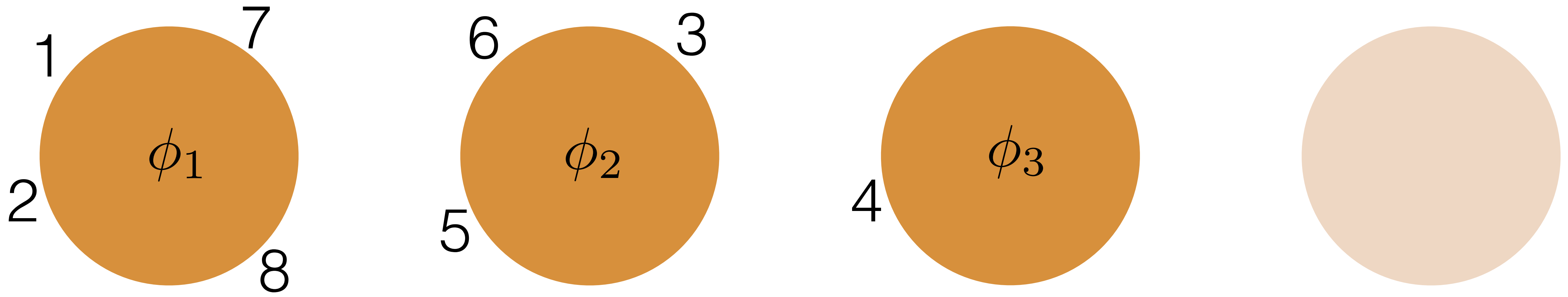
So far: Dirichlet process, Chinese restaurant process

- Infinity of parameters, growing number of parameters

Exercises

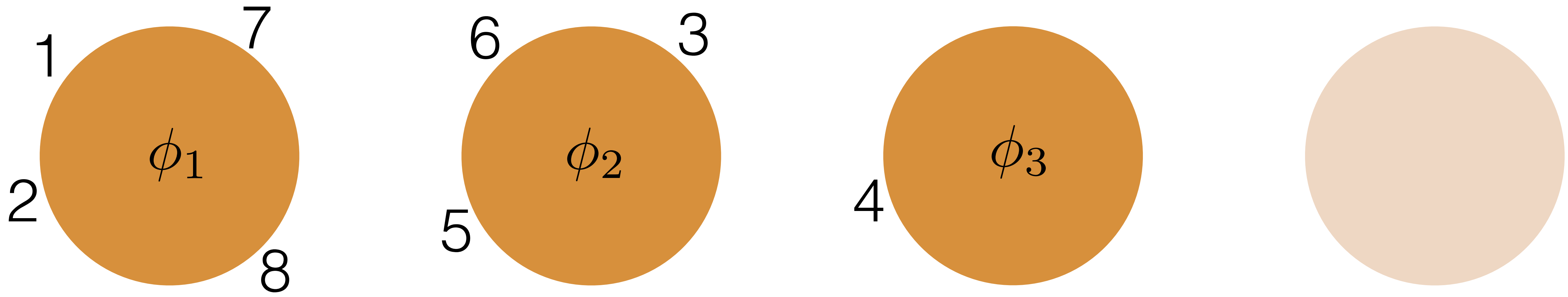


Exercises



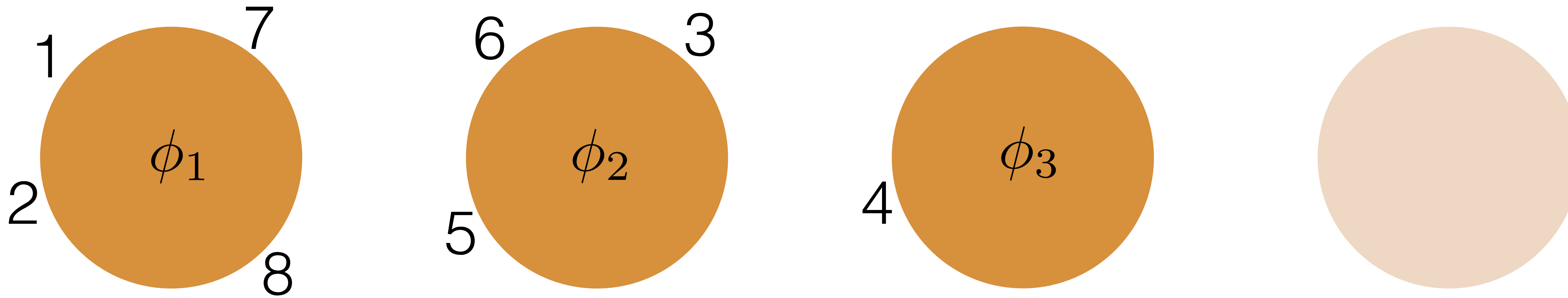
- Review Gibbs sampling

Exercises



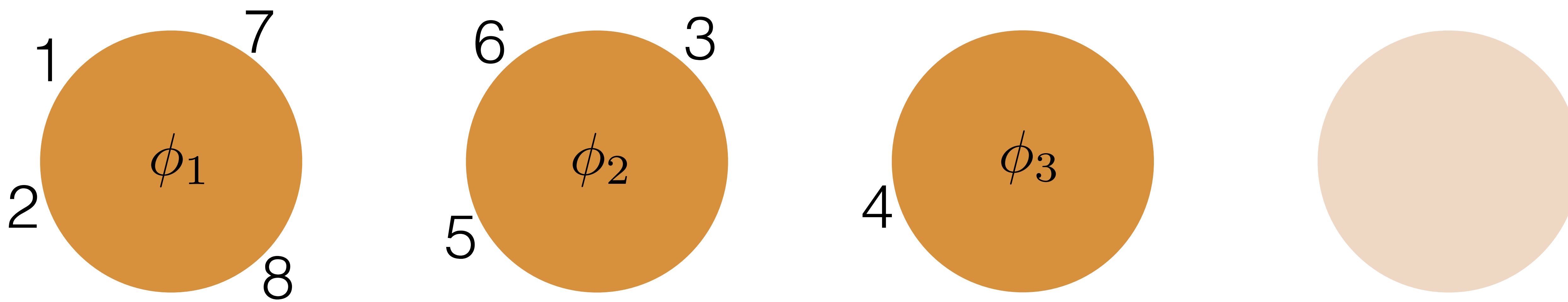
- Review Gibbs sampling
- What are the advantages and disadvantages of the DP and CRP representations?

Exercises



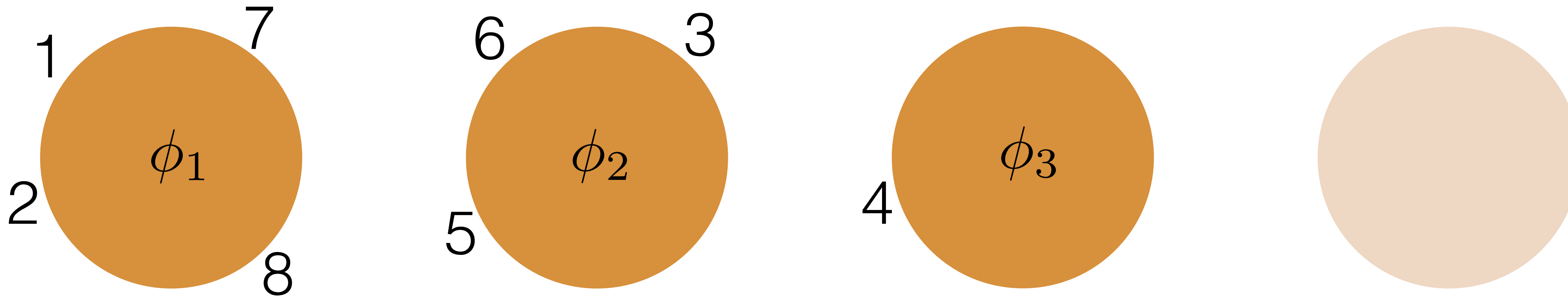
- Review Gibbs sampling
- What are the advantages and disadvantages of the DP and CRP representations?
- What is the expected number of clusters generated by a CRP(α) after N data points?

Exercises



- Review Gibbs sampling
- What are the advantages and disadvantages of the DP and CRP representations?
- 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



- Review Gibbs sampling
- What are the advantages and disadvantages of the DP and CRP representations?
- 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?
- Code a CRP sampler. Examine the empirical distribution of the number of clusters after *N* customers.

References

A full reference list is provided at the end of the “Part III” slides.