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

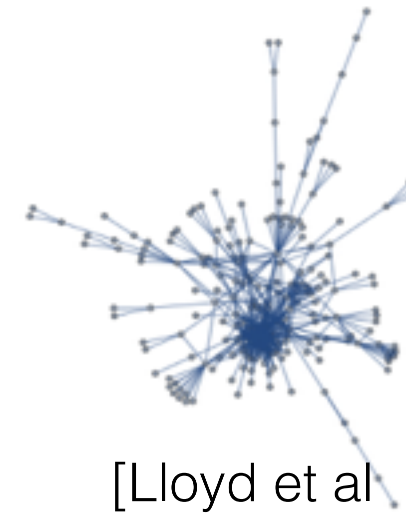
Tamara Broderick
ITT Career Development Assistant Professor
Electrical Engineering & Computer Science
MIT

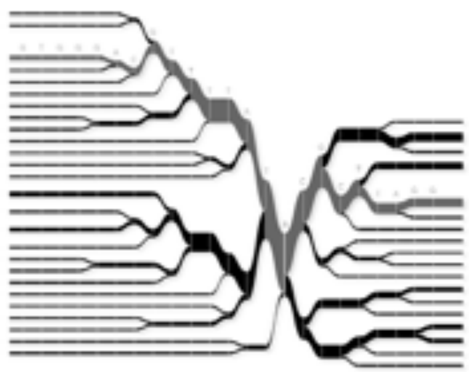
Applications

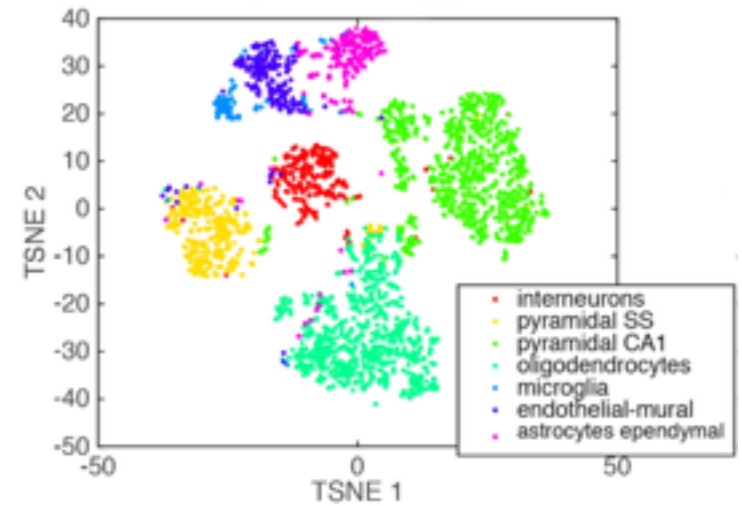
[wikipedia.org]


[Ed Bowlby, NOAA]


[Fox et al 2014]

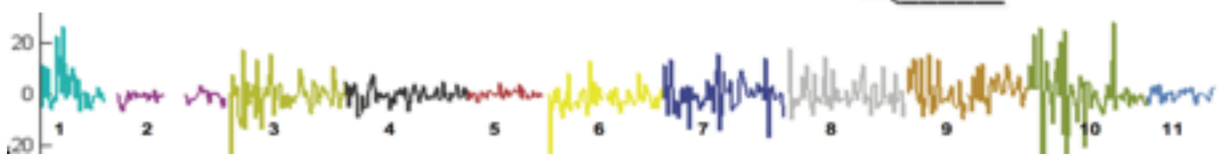

[Lloyd et al 2012; Miller et al 2010]


[Ewens 1972; Hartl, Clark 2003]

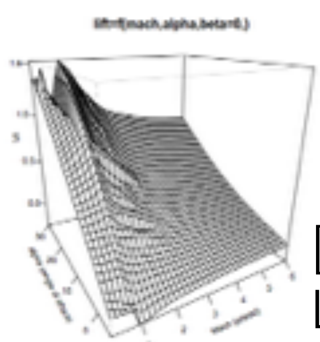


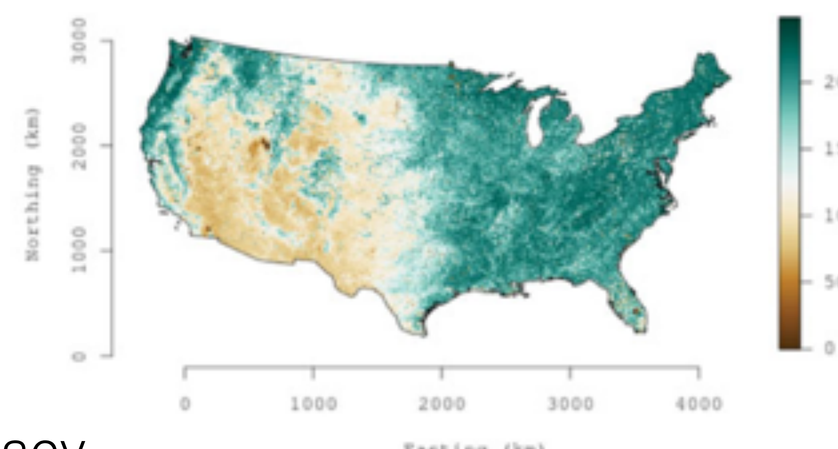
[Prabhakaran, Azizi, Carr, Pe'er 2016]


[Sudderth, Jordan 2009]

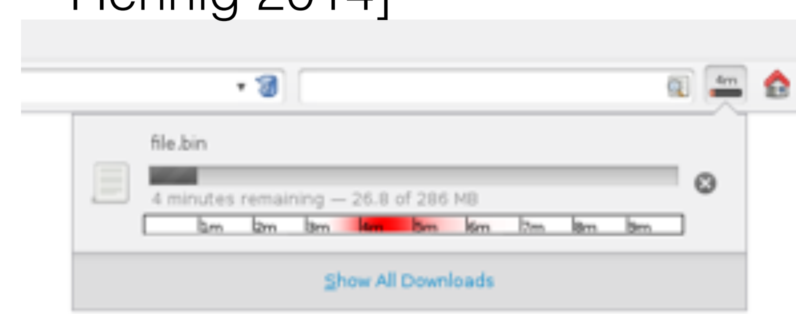
[Saria et al 2010]



[US CDC PHIL; Futoma, Hariharan, Heller 2017]


[Gramacy, Lee 2009]



[Datta, Banerjee, Finley, Gelfand 2016]

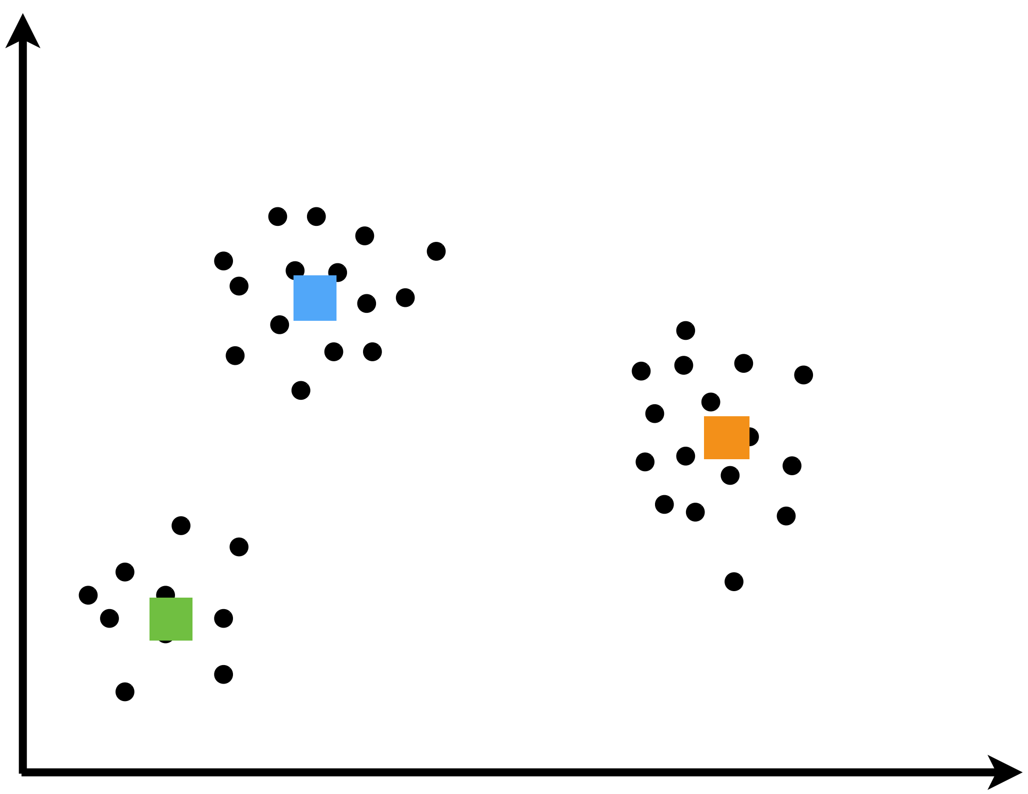
[Kiefel, Schuler, Hennig 2014]



[Chati, Balakrishnan 2017]

[Deisenroth, Fox, Rasmussen 2015]


Generative model

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



- Finite Gaussian mixture model (*K* clusters)

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

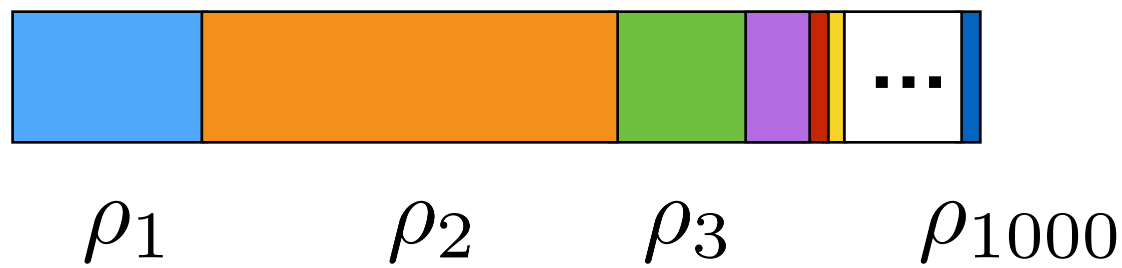
$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_{1:K})$$

$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$

ρ_1 ρ_2 ρ_3

What if $K > N$?

- e.g. species sampling, topic modeling, groups on a social network, etc.



- Components: number of latent groups
- Clusters: number of components represented in the data
- [demo 1, demo 2]
- Number of clusters for N data points is random
- Number of clusters grows with N

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \text{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \mathrm{Dirichlet}(a_2, \ldots, a_K)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \tfrac{(\rho_2,\ldots,\rho_K)}{1-\rho_1} \stackrel{d}{=} \mathrm{Dirichlet}(a_2,\ldots,a_K)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \tfrac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \mathrm{Dirichlet}(a_2, \ldots, a_K)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp\!\!\!\perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \mathrm{Dirichlet}(a_2, \ldots, a_K)$$

- "Stick breaking"

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp\!\!\!\perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \mathrm{Dirichlet}(a_2, \ldots, a_K)$$

- "Stick breaking"

$$V_1 \sim \mathrm{Beta}(a_1, a_2 + a_3 + a_4)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \text{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \tfrac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)$$

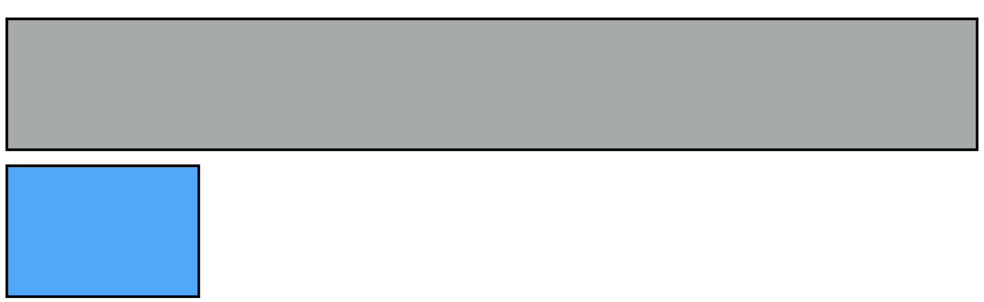
- "Stick breaking"

$V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4)$ $\quad\quad$ $\rho_1 = V_1$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \text{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \tfrac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)$$



- "Stick breaking"

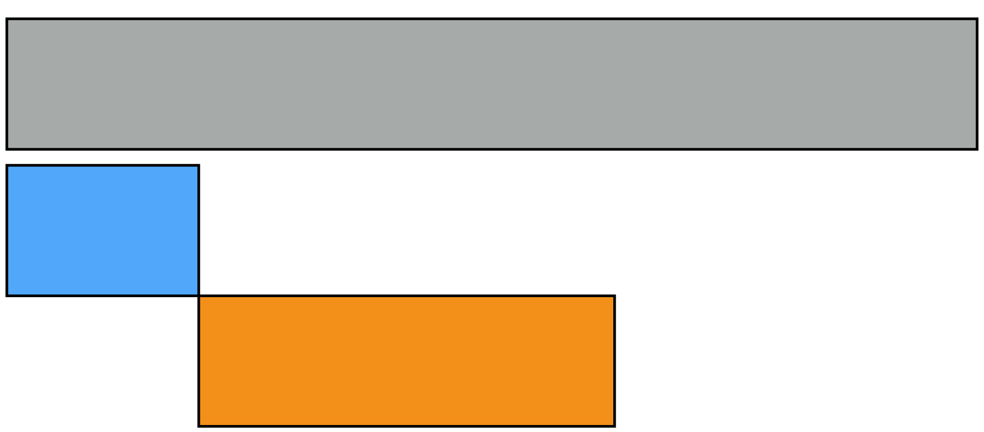
$$V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \qquad \rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, a_3 + a_4)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
    - Observation: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \text{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \frac{(\rho_2, \ldots, \rho_K)}{1 - \rho_1} \stackrel{d}{=} \text{Dirichlet}(a_2, \ldots, a_K)$$



- "Stick breaking"

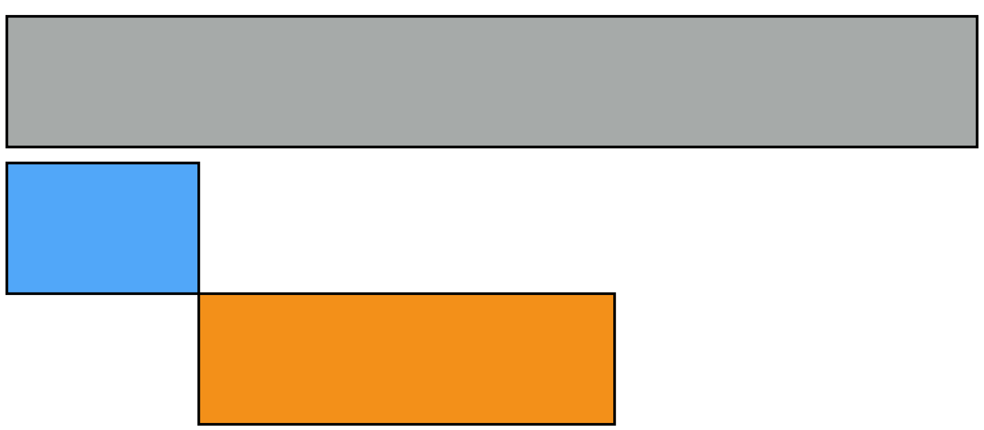
$$V_1 \sim \text{Beta}(a_1, a_2 + a_3 + a_4) \qquad \rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, a_3 + a_4) \qquad \rho_2 = (1 - V_1) V_2$$

Choosing $K = \infty$

- Here, difficult to choose finite *K* in advance (contrast with small *K*): don't know *K*, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \overset{d}{=} \mathrm{Beta}(a_1, \sum_{k=1}^{K} a_k - a_1) \perp\!\!\!\perp \tfrac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \overset{d}{=} \mathrm{Dirichlet}(a_2, \ldots, a_K)$$



- "Stick breaking"

$$V_1 \sim \mathrm{Beta}(a_1, a_2 + a_3 + a_4) \qquad \rho_1 = V_1$$

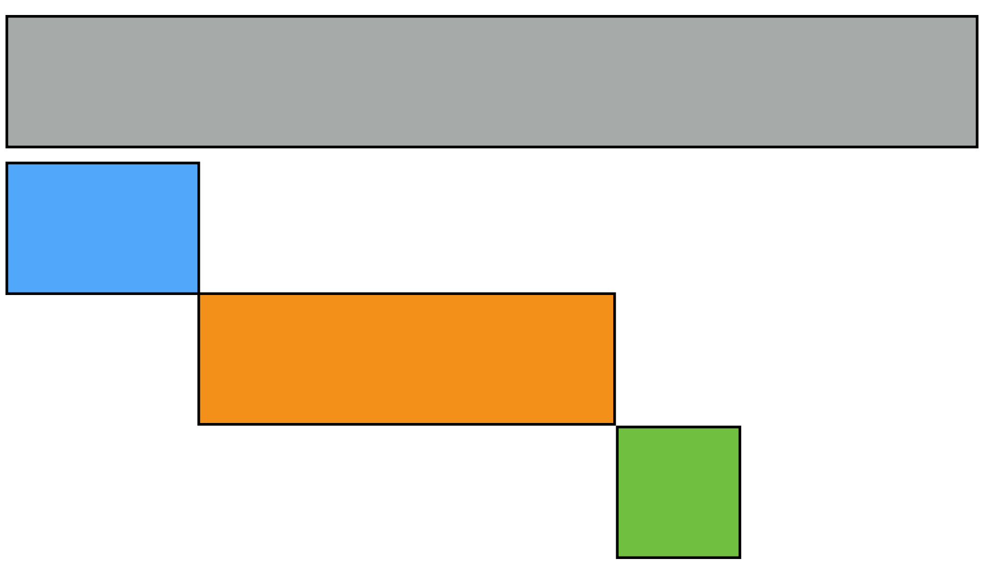
$$V_2 \sim \mathrm{Beta}(a_2, a_3 + a_4) \qquad \rho_2 = (1 - V_1)V_2$$

$$V_3 \sim \mathrm{Beta}(a_3, a_4)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \mathrm{Dirichlet}(a_2, \ldots, a_K)$$



- "Stick breaking"

$$V_1 \sim \mathrm{Beta}(a_1, a_2 + a_3 + a_4) \qquad \rho_1 = V_1$$

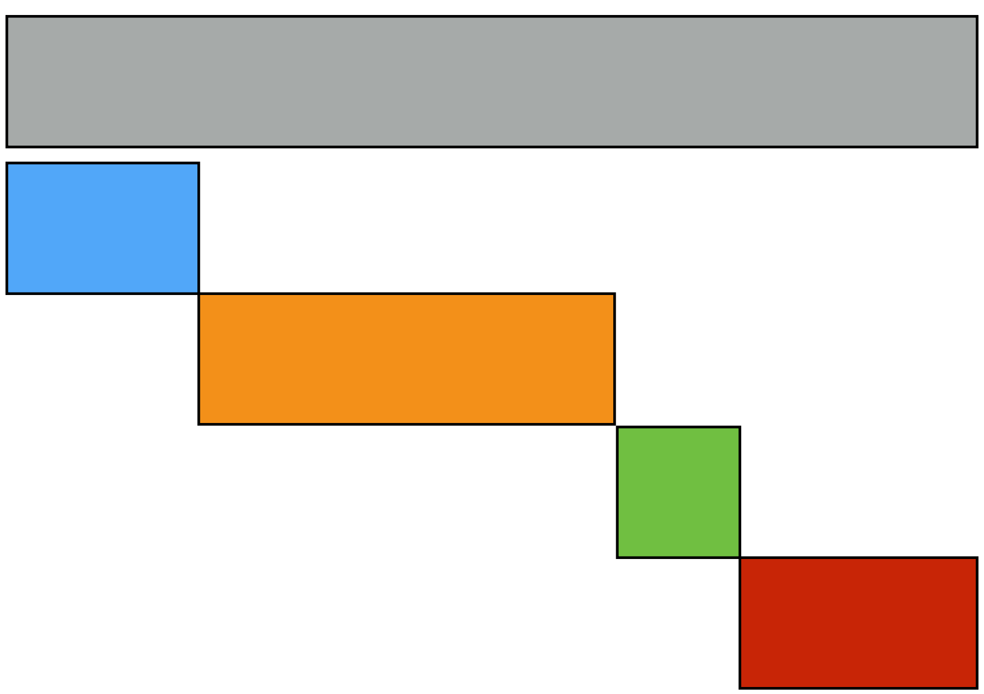
$$V_2 \sim \mathrm{Beta}(a_2, a_3 + a_4) \qquad \rho_2 = (1 - V_1)V_2$$

$$V_3 \sim \mathrm{Beta}(a_3, a_4) \qquad \rho_3 = (1 - V_1)(1 - V_2)V_3$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - Observation: $\rho_{1:K} \sim \mathrm{Dirichlet}(a_{1:K})$

$$\Leftrightarrow \rho_1 \stackrel{d}{=} \mathrm{Beta}\left(a_1, \sum_{k=1}^{K} a_k - a_1\right) \perp\!\!\!\perp \frac{(\rho_2, \ldots, \rho_K)}{1-\rho_1} \stackrel{d}{=} \mathrm{Dirichlet}(a_2, \ldots, a_K)$$



- "Stick breaking"

$$V_1 \sim \mathrm{Beta}(a_1, a_2 + a_3 + a_4) \qquad \rho_1 = V_1$$

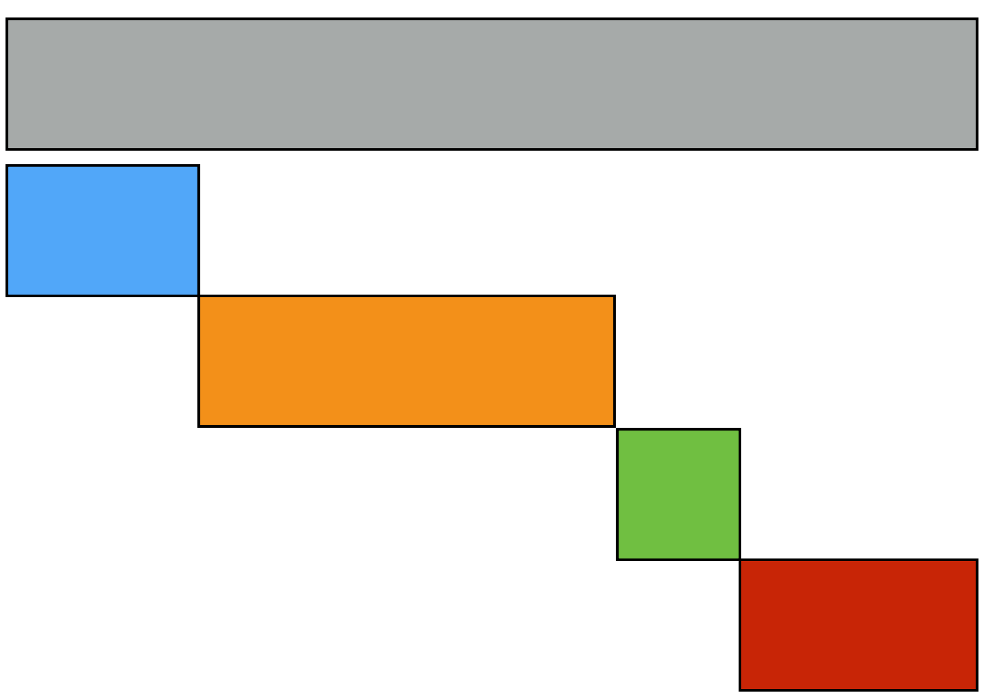
$$V_2 \sim \mathrm{Beta}(a_2, a_3 + a_4) \qquad \rho_2 = (1 - V_1)V_2$$

$$V_3 \sim \mathrm{Beta}(a_3, a_4) \qquad \rho_3 = (1 - V_1)(1 - V_2)V_3$$

$$\rho_4 = 1 - \sum_{k=1}^{3} \rho_k$$

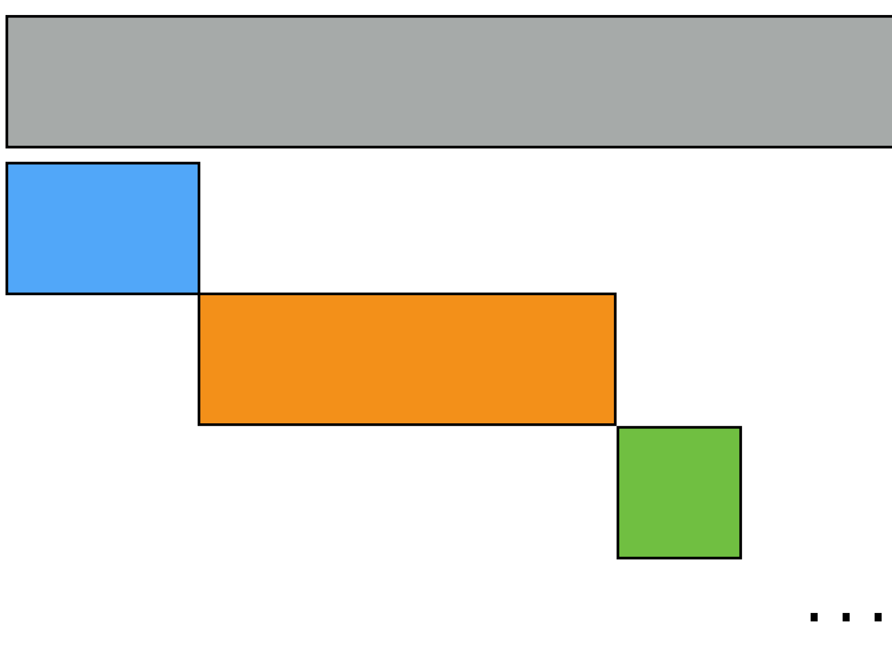
Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?

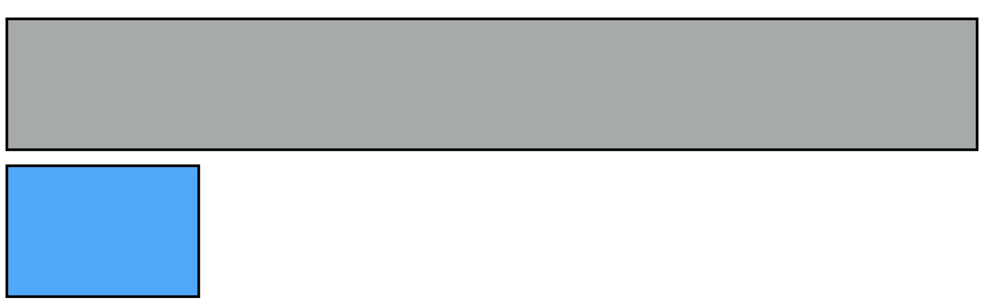
Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?

$$V_1 \sim \mathrm{Beta}(a_1, b_1)$$

Choosing $K = \infty$

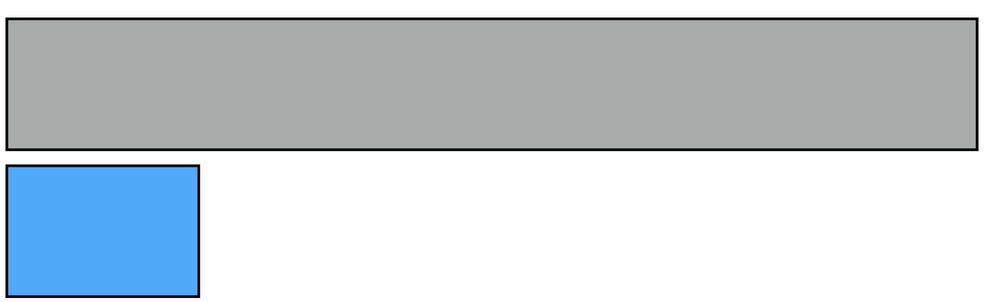
- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



$$V_1 \sim \mathrm{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$

Choosing $K = \infty$

- Here, difficult to choose finite *K* in advance (contrast with small *K*): don't know *K*, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



$$V_1 \sim \mathrm{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$
$$V_2 \sim \mathrm{Beta}(a_2, b_2)$$

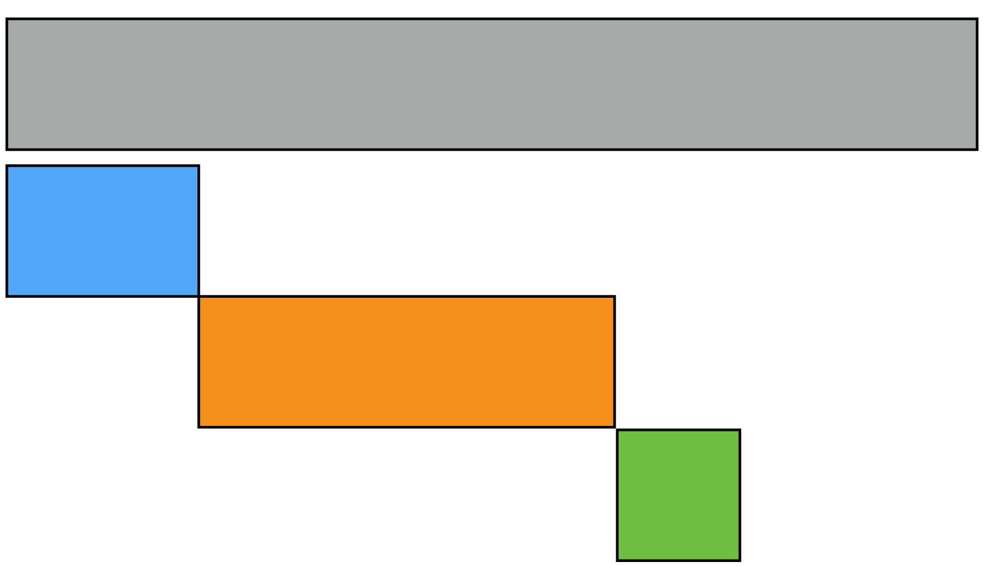
Choosing $K = \infty$

- Here, difficult to choose finite *K* in advance (contrast with small *K*): don't know *K*, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?

$$V_1 \sim \text{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$
$$V_2 \sim \text{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?

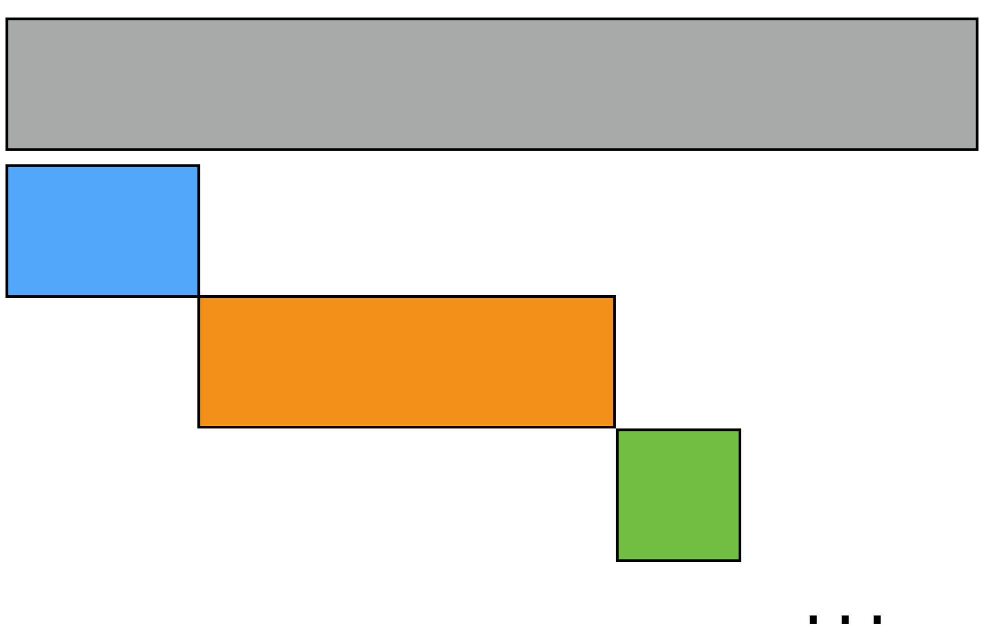


$$V_1 \sim \text{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

Choosing $K = \infty$

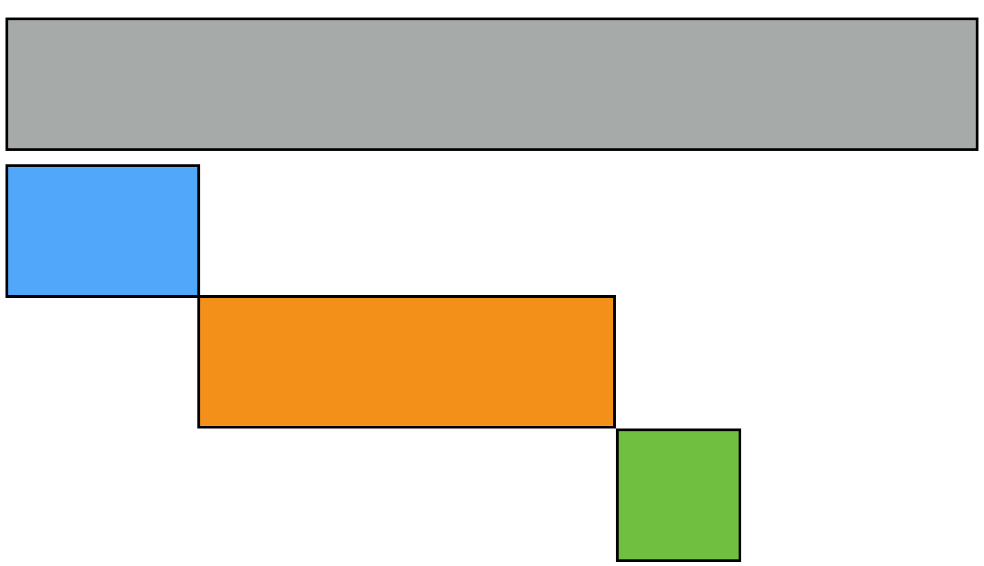
- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



$$V_1 \sim \text{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$
$$V_2 \sim \text{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



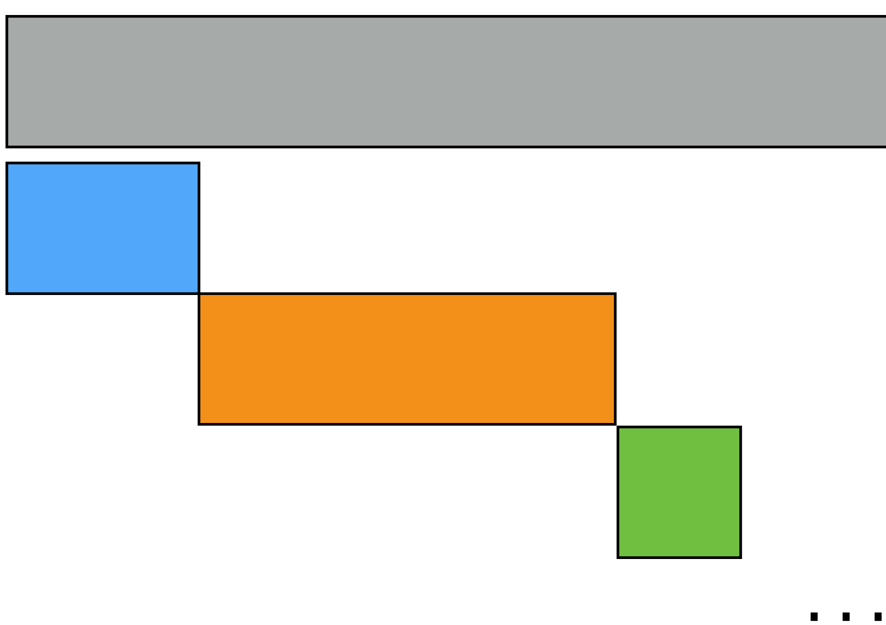
$$V_1 \sim \text{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

$$\cdots \quad V_k \sim \text{Beta}(a_k, b_k)$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



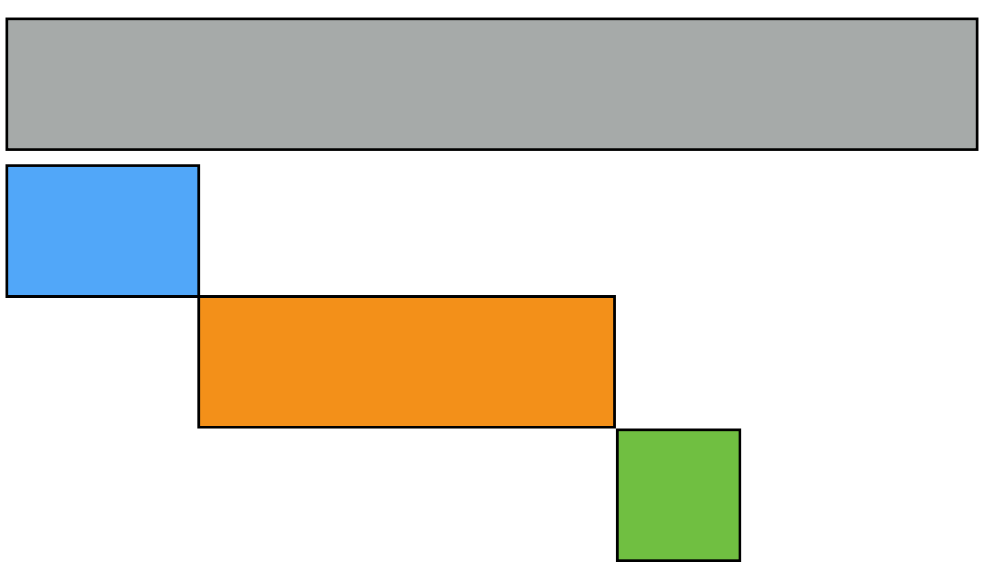
$$V_1 \sim \mathrm{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$

$$V_2 \sim \mathrm{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

$$V_k \sim \mathrm{Beta}(a_k, b_k) \qquad \rho_k = \left[\prod_{j=1}^{k-1}(1 - V_j)\right] V_k$$

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



$$V_1 \sim \mathrm{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$

$$V_2 \sim \mathrm{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

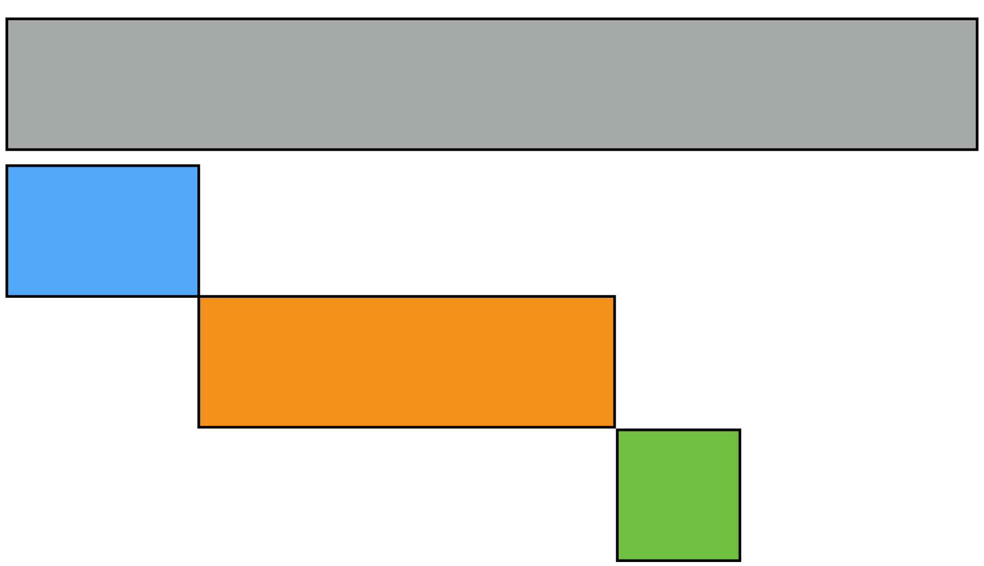
$$\cdots$$

$$V_k \sim \mathrm{Beta}(a_k, b_k) \qquad \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j) \right] V_k$$

[Ishwaran, James 2001]

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - **Dirichlet process stick-breaking**: $a_k = 1, b_k = \alpha > 0$



$\cdots$

$$V_1 \sim \mathrm{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$

$$V_2 \sim \mathrm{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

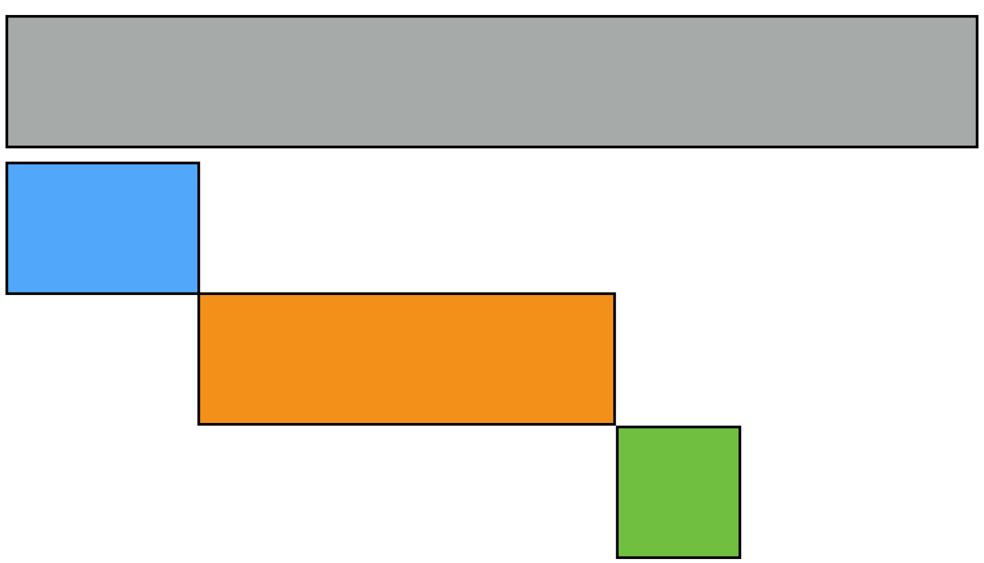
$$V_k \sim \mathrm{Beta}(a_k, b_k) \qquad \rho_k = \left[\prod_{j=1}^{k-1}(1 - V_j)\right] V_k$$

[Ishwaran, James 2001]

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - **Dirichlet process stick-breaking**: $a_k = 1, b_k = \alpha > 0$
  - Griffiths-Engen-McCloskey (**GEM**) distribution:

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



$$V_1 \sim \mathrm{Beta}(a_1, b_1) \qquad \rho_1 = V_1$$

$$V_2 \sim \mathrm{Beta}(a_2, b_2) \qquad \rho_2 = (1 - V_1)V_2$$

$\cdots$

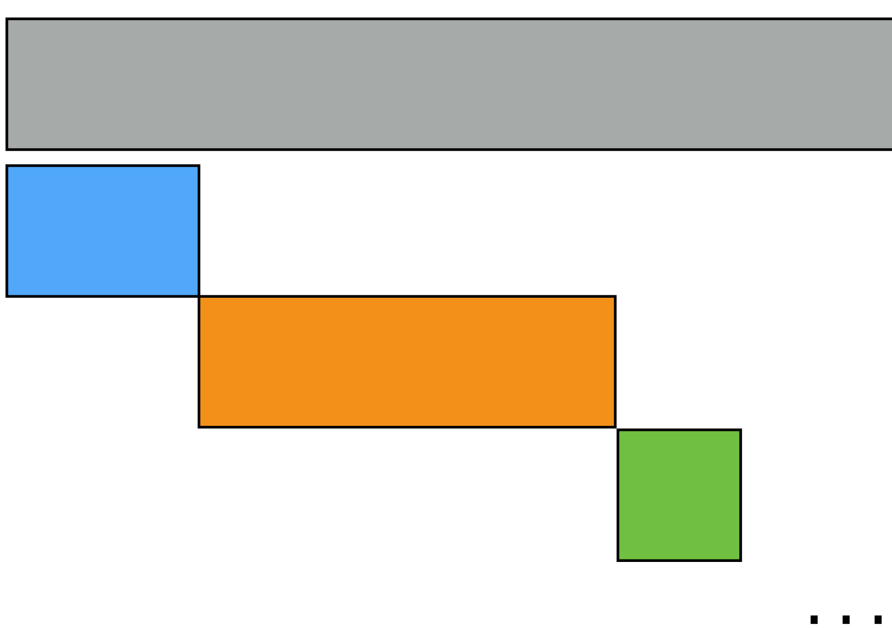
$$V_k \sim \mathrm{Beta}(a_k, b_k) \qquad \rho_k = \left[\prod_{j=1}^{k-1}(1 - V_j)\right] V_k$$

 [McCloskey 1965; Engen 1975; Patil and Taillie 1977; Ewens 1987; Sethuraman 1994; Ishwaran, James 2001]

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - **Dirichlet process stick-breaking**: $a_k = 1, b_k = \alpha > 0$
  - Griffiths-Engen-McCloskey (**GEM**) distribution:

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



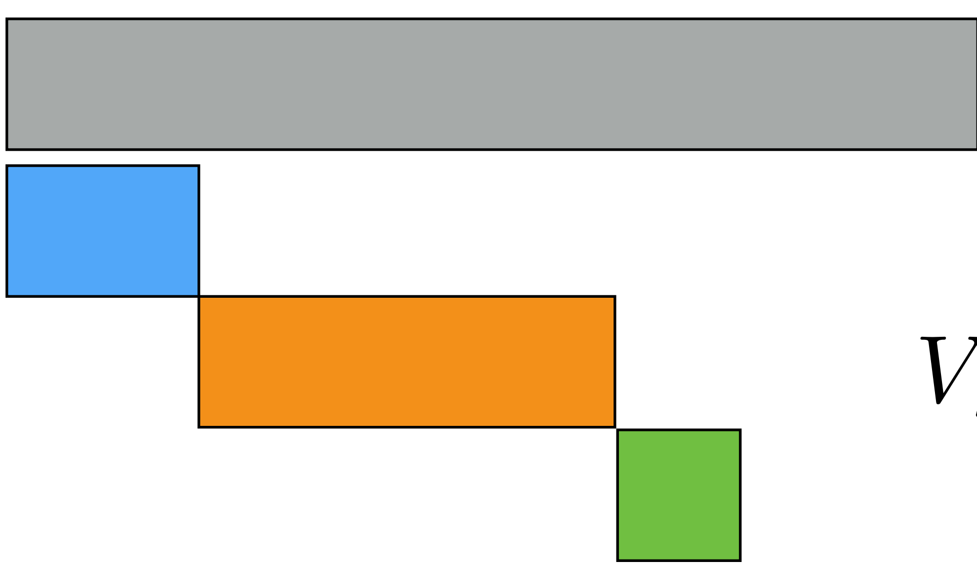
$$V_k \overset{iid}{\sim} \text{Beta}(1, \alpha) \quad \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j)\right] V_k$$

[McCloskey 1965; Engen 1975; Patil and Taillie 1977; Ewens 1987; Sethuraman 1994; Ishwaran, James 2001]

Choosing $K = \infty$

- Here, difficult to choose finite K in advance (contrast with small K): don't know K, difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?
  - **Dirichlet process stick-breaking**: $a_k = 1, b_k = \alpha > 0$
  - Griffiths-Engen-McCloskey (**GEM**) distribution:

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



$\cdots$

$$V_k \overset{iid}{\sim} \mathrm{Beta}(1, \alpha) \quad \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j) \right] V_k$$

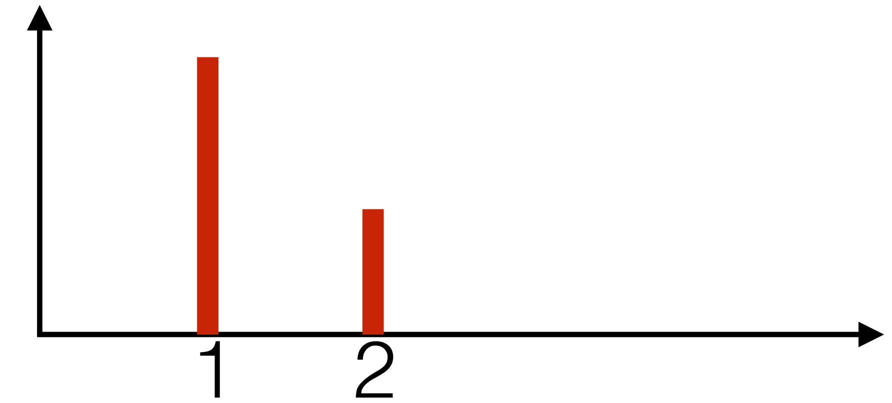
[demo]

[McCloskey 1965; Engen 1975; Patil and Taillie 1977; Ewens 1987; Sethuraman 1994; Ishwaran, James 2001]

Distributions

Distributions

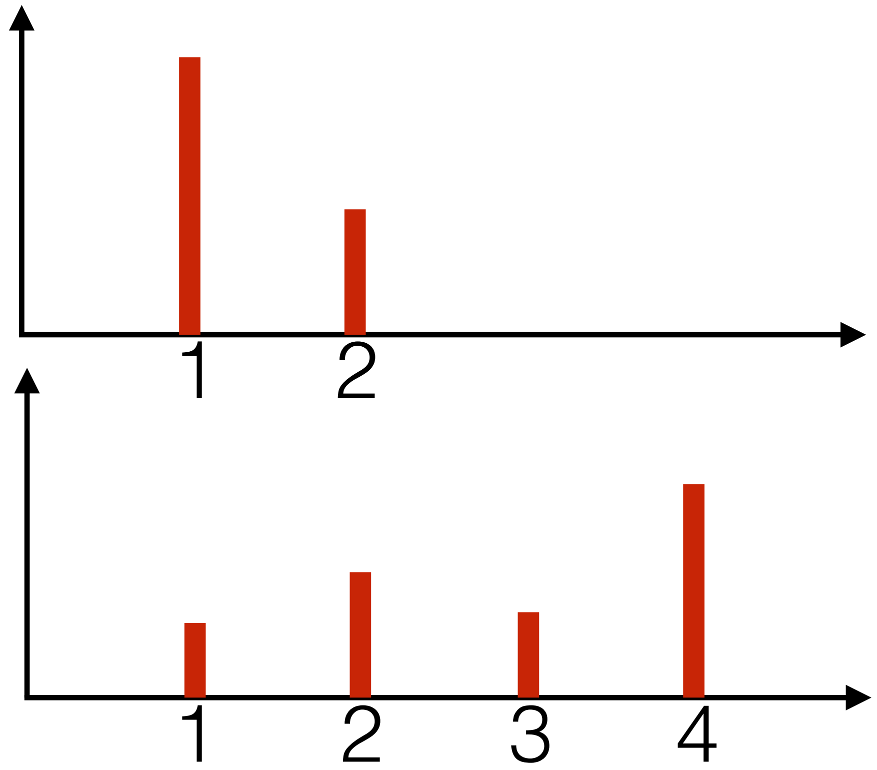
- Beta → random distribution over $1, 2$



Distributions

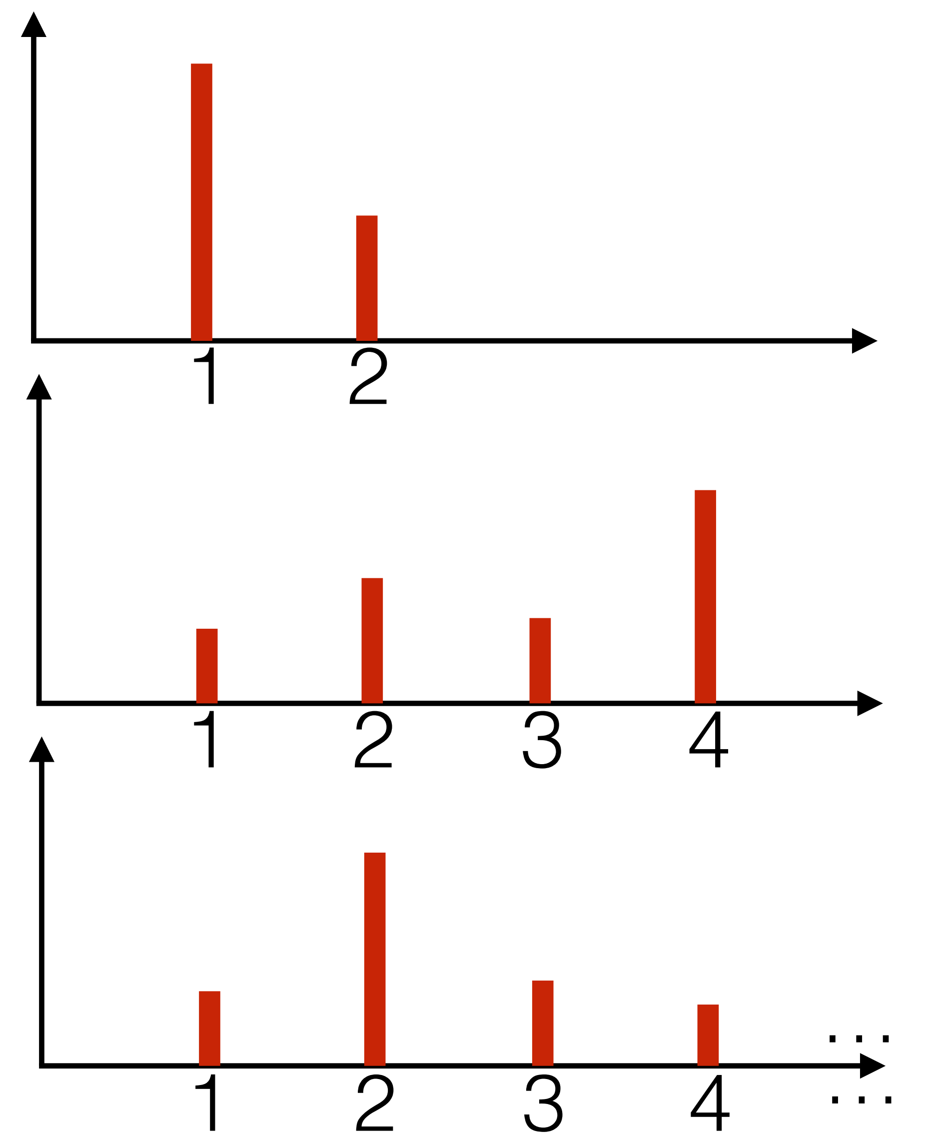
- Beta → random distribution over $1, 2$

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



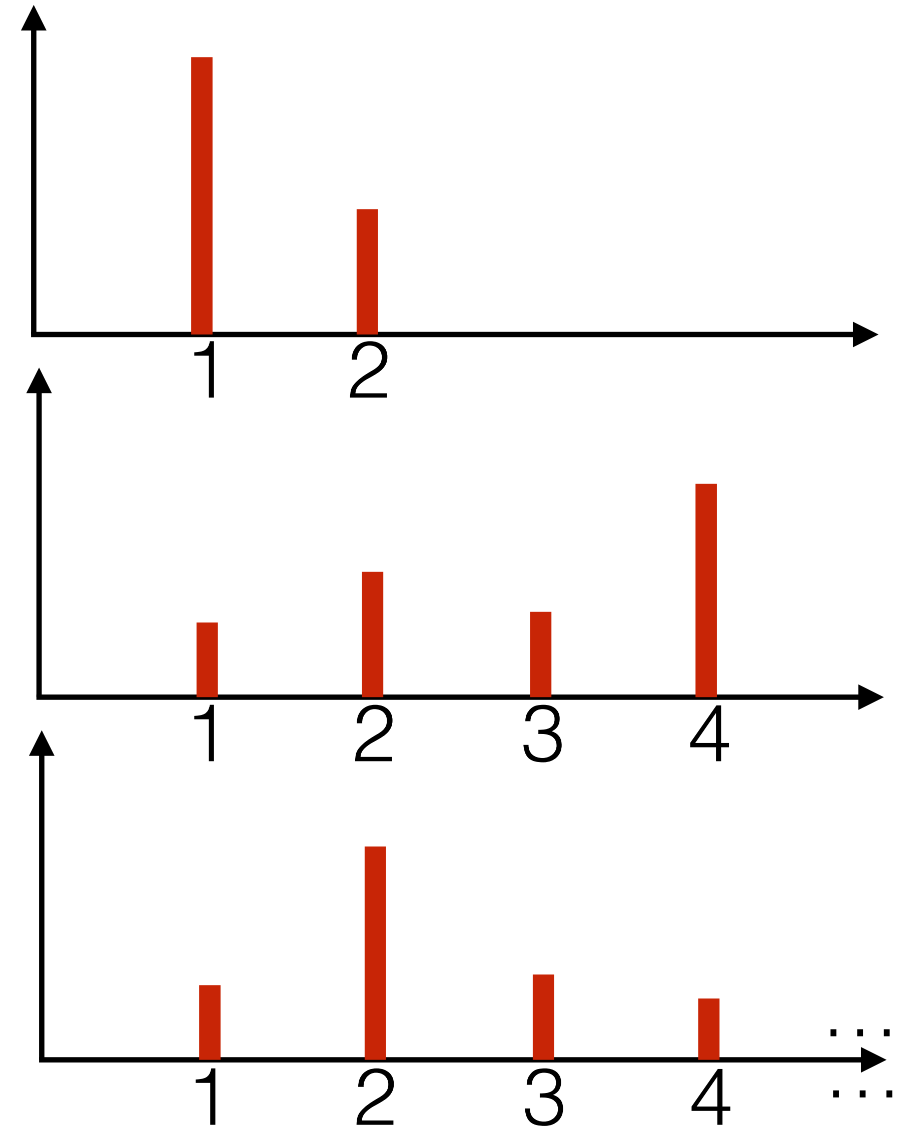
Distributions

- 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 $1, 2, \ldots$



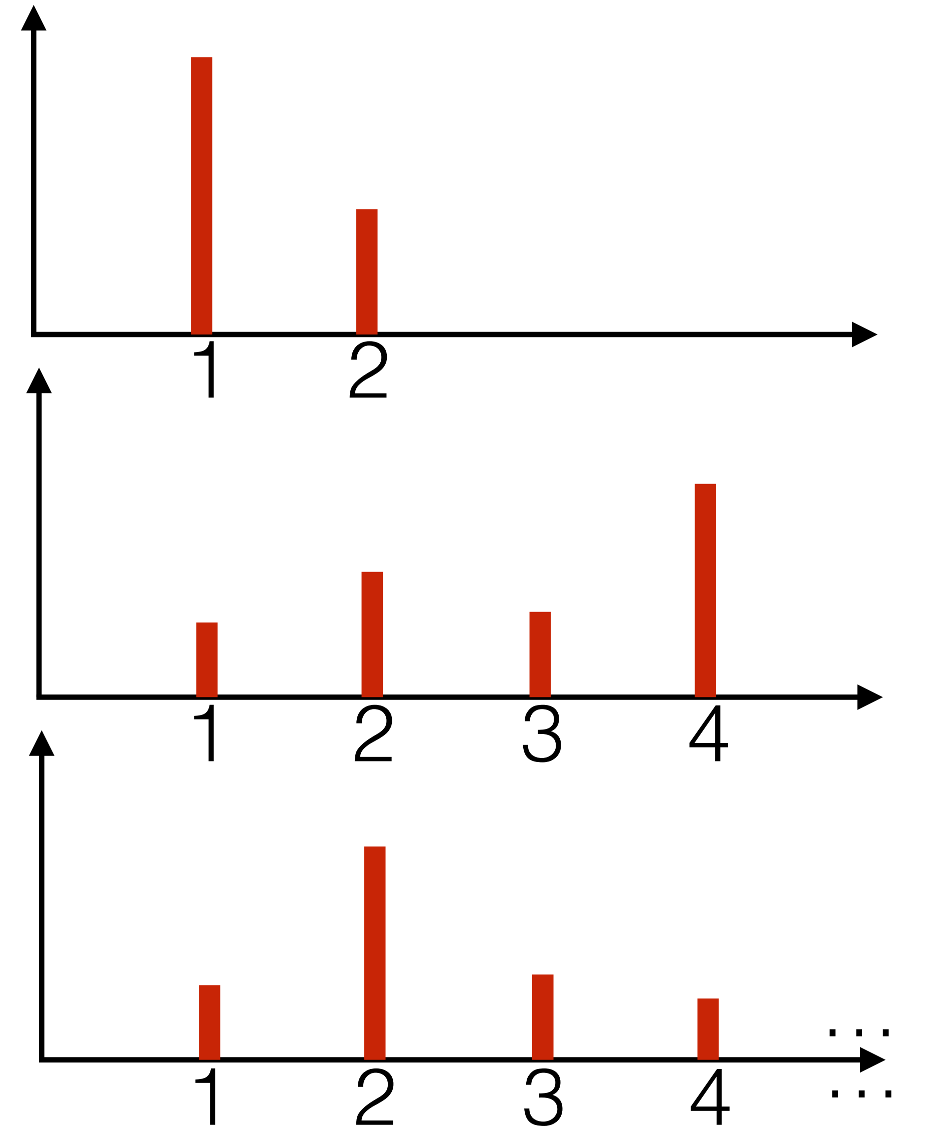
- Infinity of parameters: components
- Growing number of parameters: clusters

Distributions

- Beta → random distribution over $1, 2$

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

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



Dirichlet process mixture model

Dirichlet process mixture model

- Gaussian mixture model

Dirichlet process mixture model

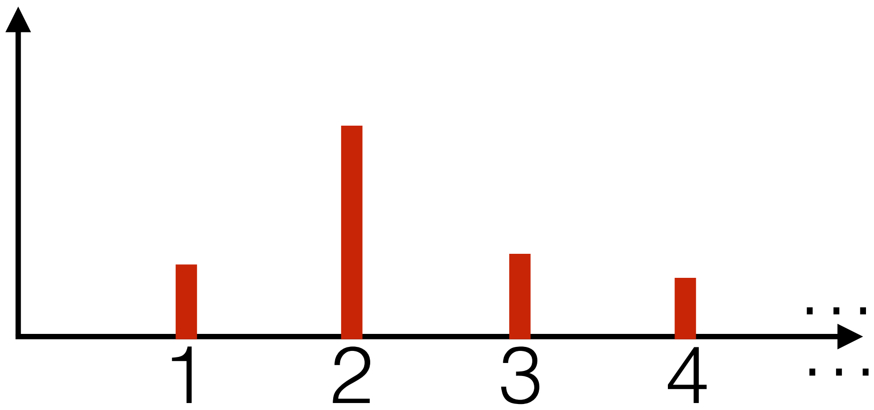
- Gaussian mixture model

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

Dirichlet process mixture model

- Gaussian mixture model

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

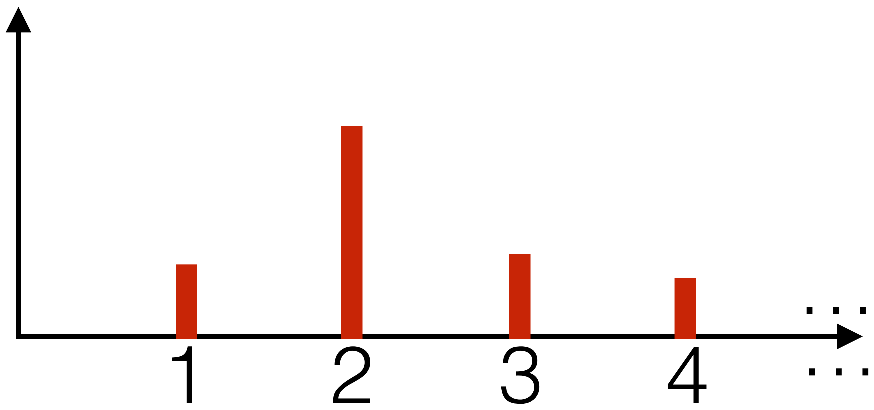


Dirichlet process mixture model

- Gaussian mixture model

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

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

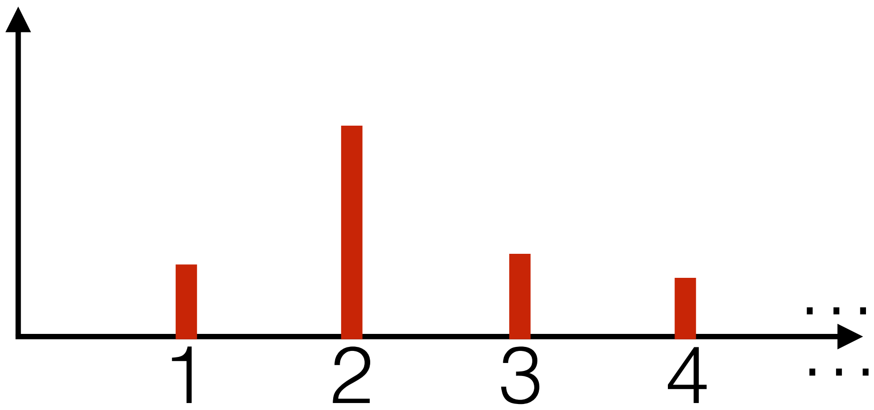


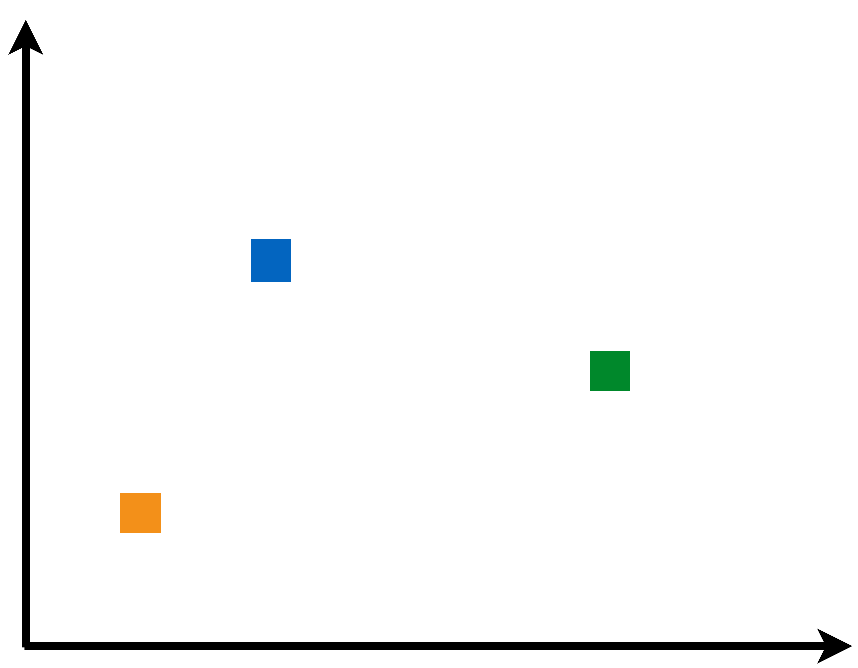
Dirichlet process mixture model

- Gaussian mixture model

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

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



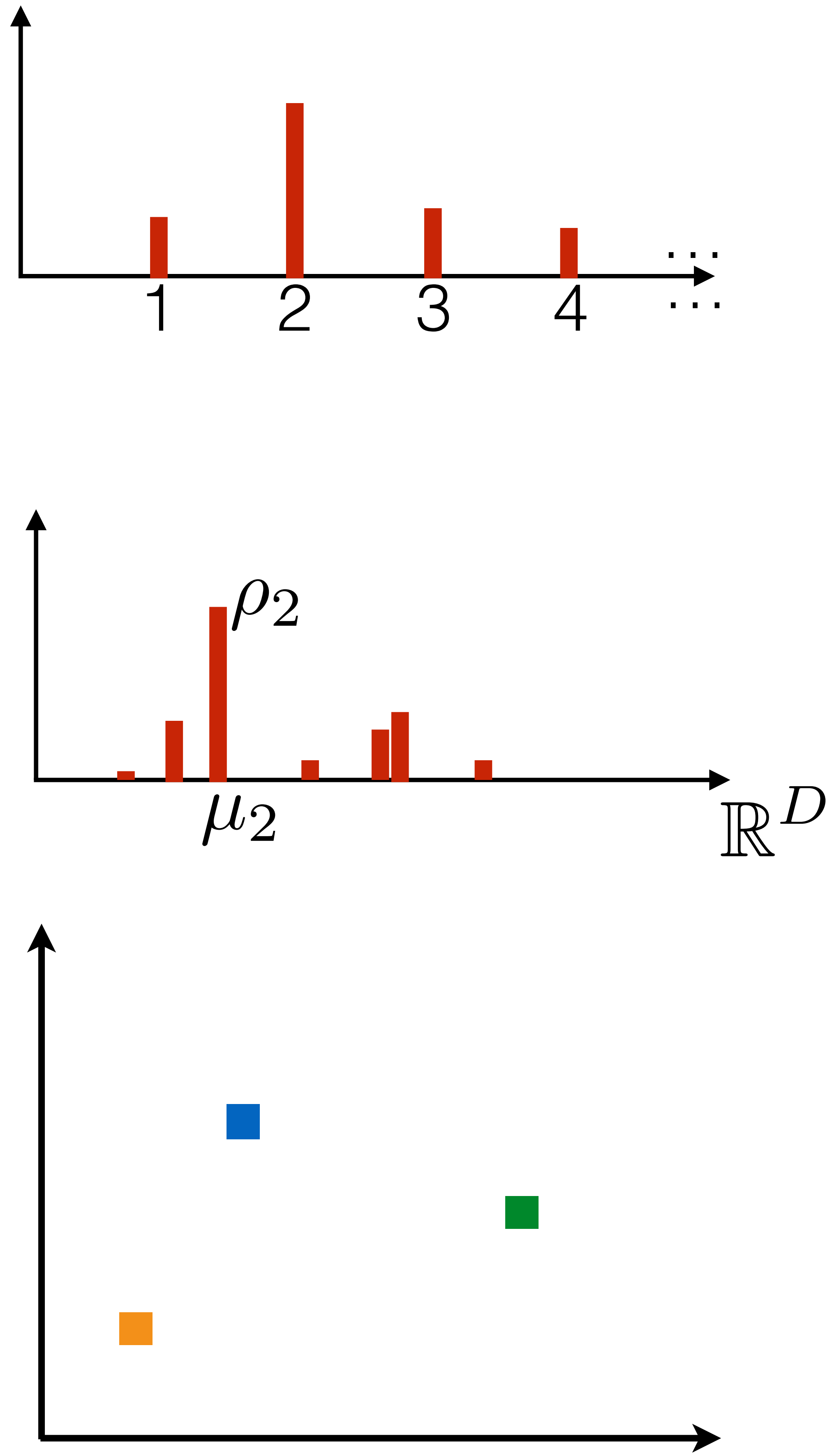


Dirichlet process mixture model

- Gaussian mixture model

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

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



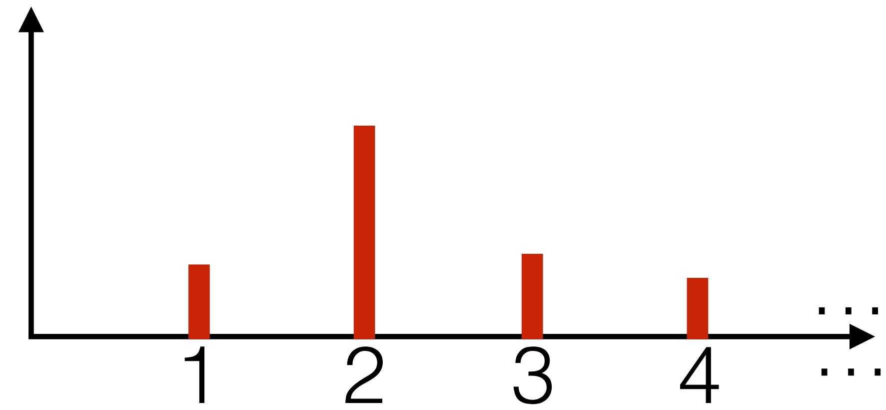
Dirichlet process mixture model

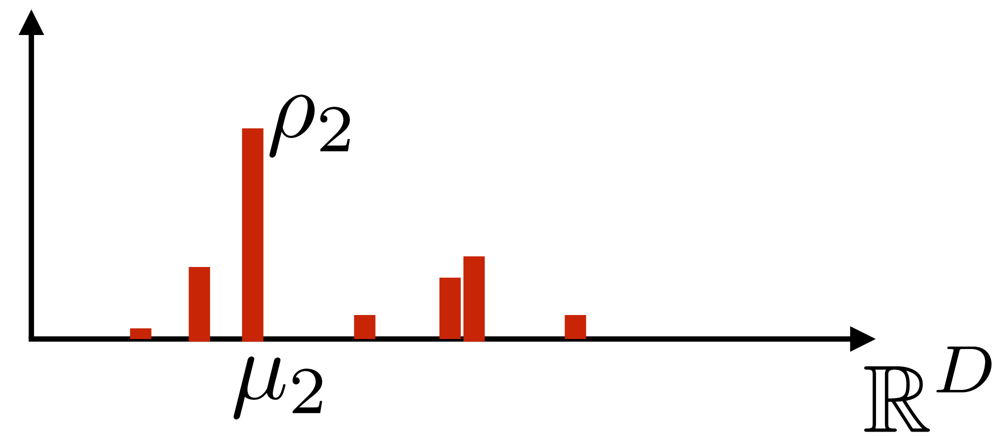
- Gaussian mixture model

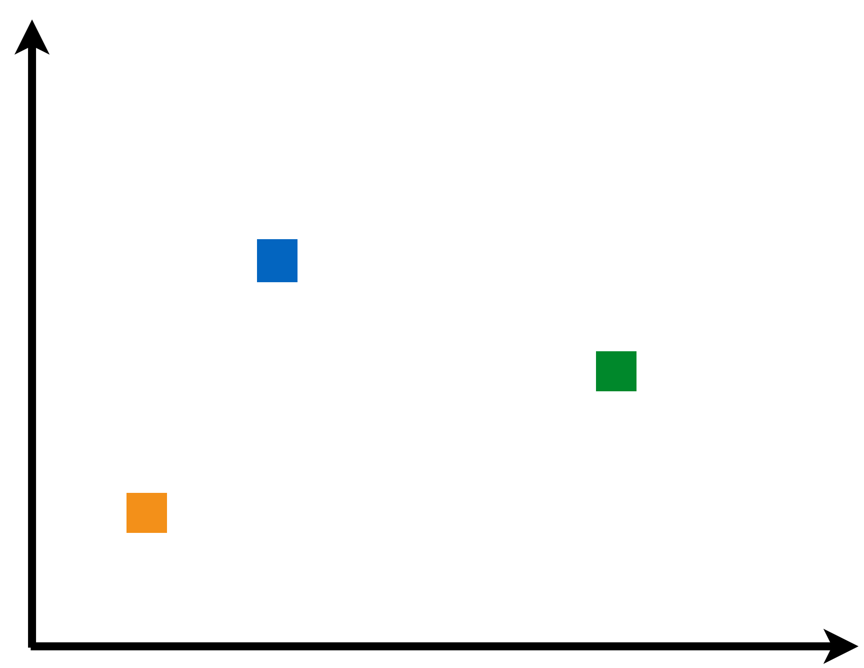
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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

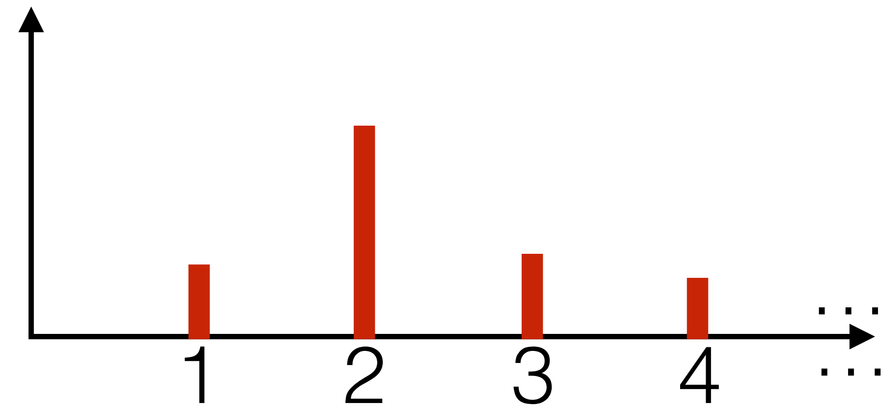






Dirichlet process mixture model

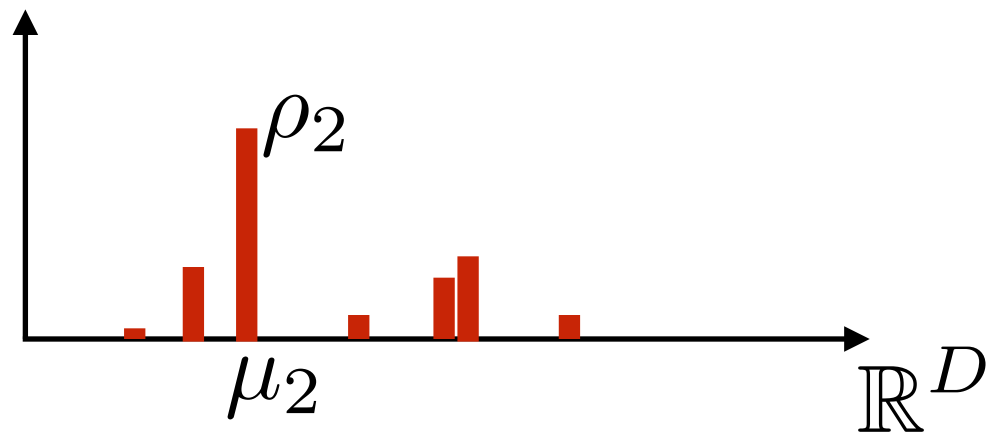
- Gaussian mixture model

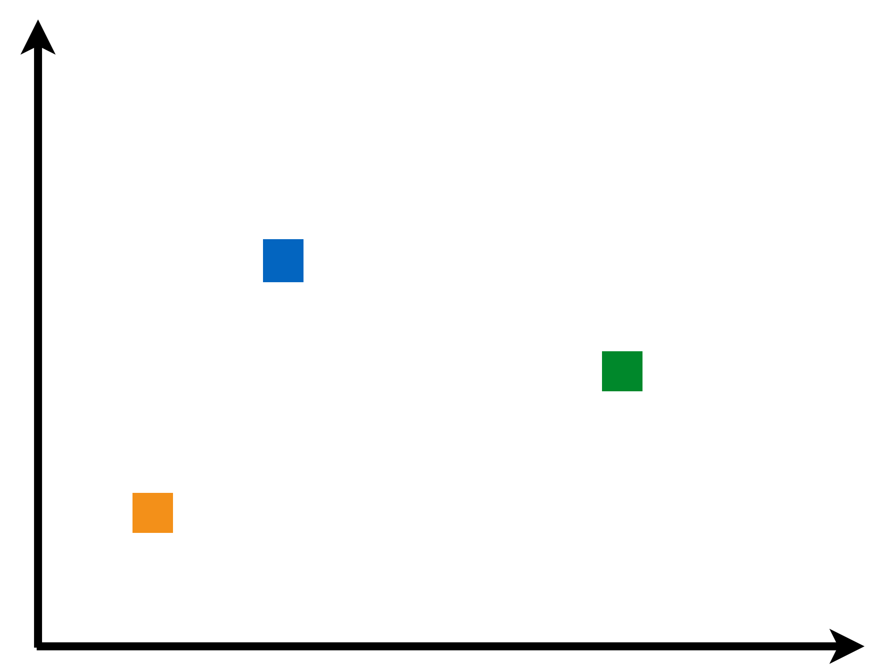


$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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





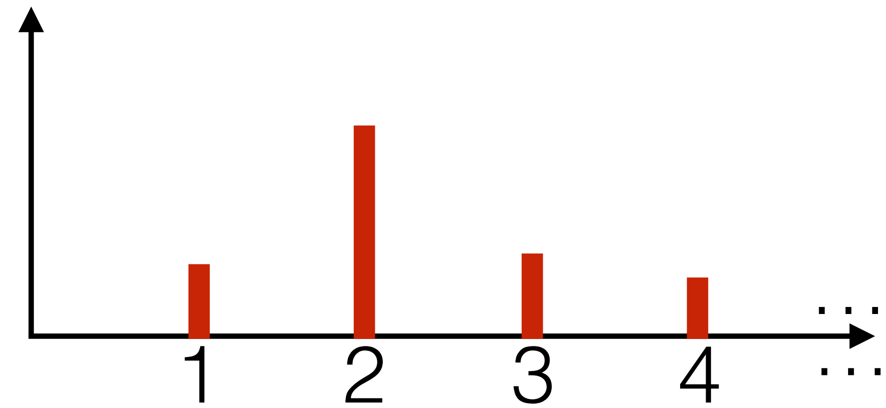
Dirichlet process mixture model

- Gaussian mixture model

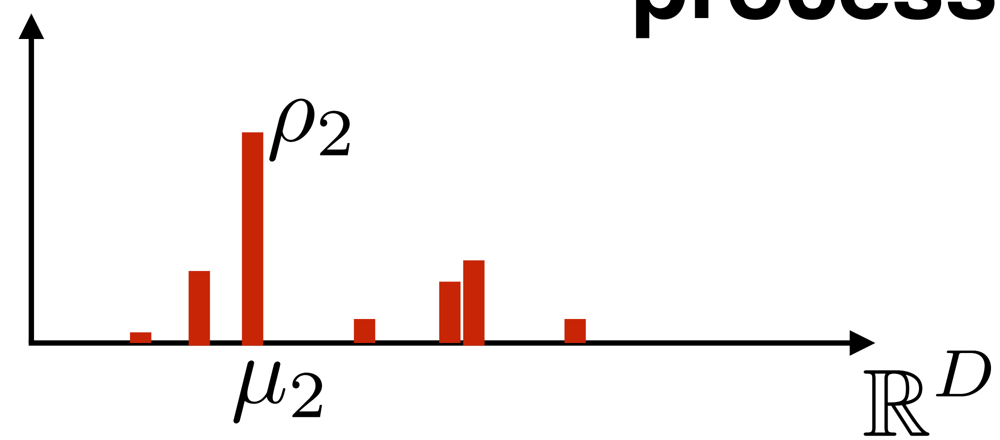
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

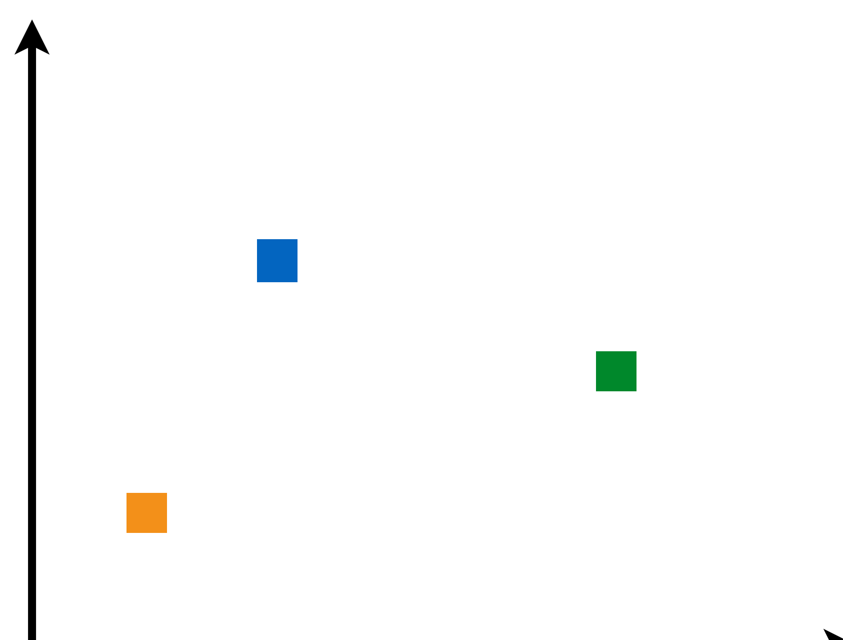
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

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



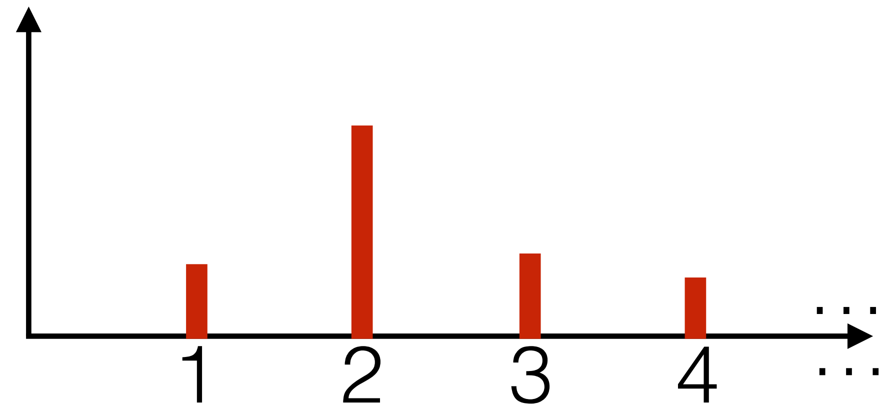
Dirichlet process





Dirichlet process mixture model

- Gaussian mixture model

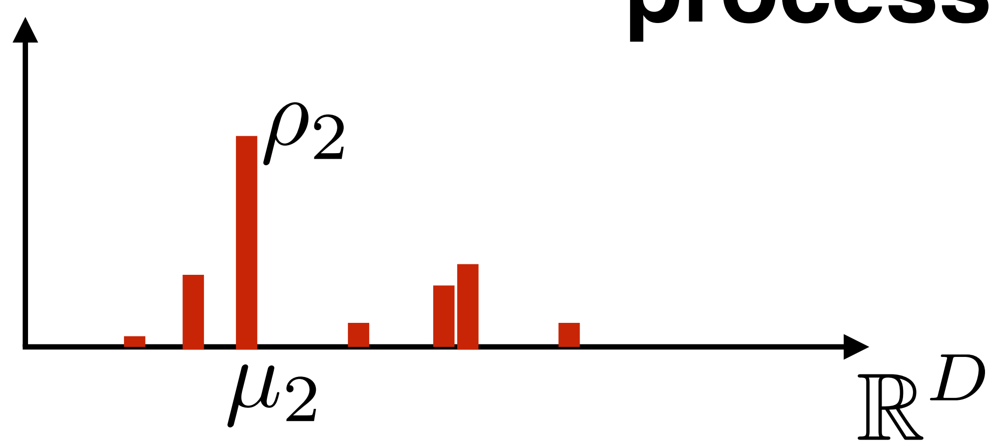


$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

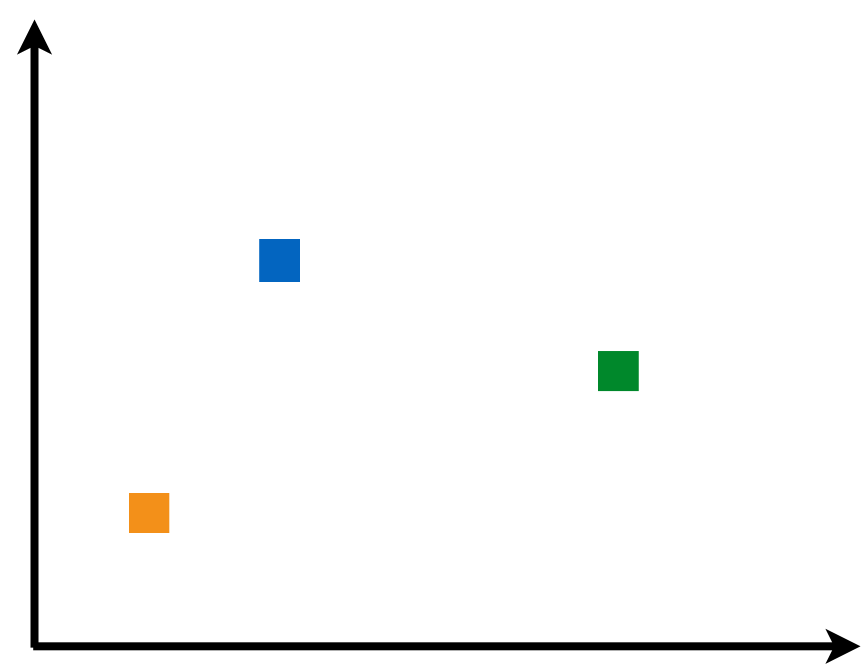
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0), k = 1, 2, \ldots$$

- 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



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



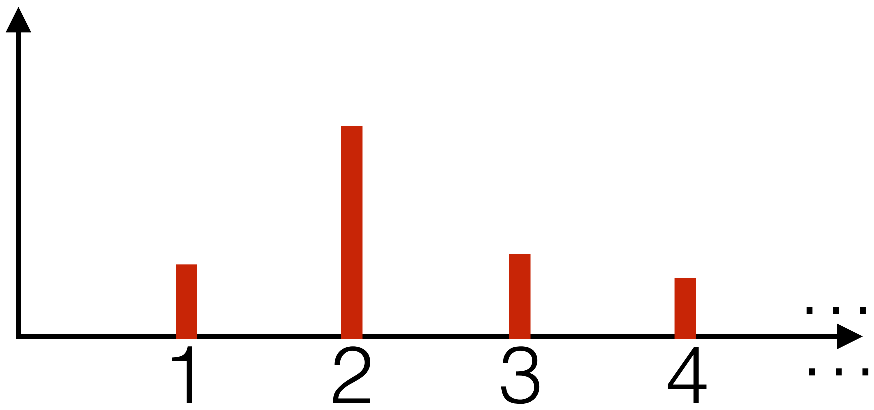
Dirichlet process mixture model

- Gaussian mixture model

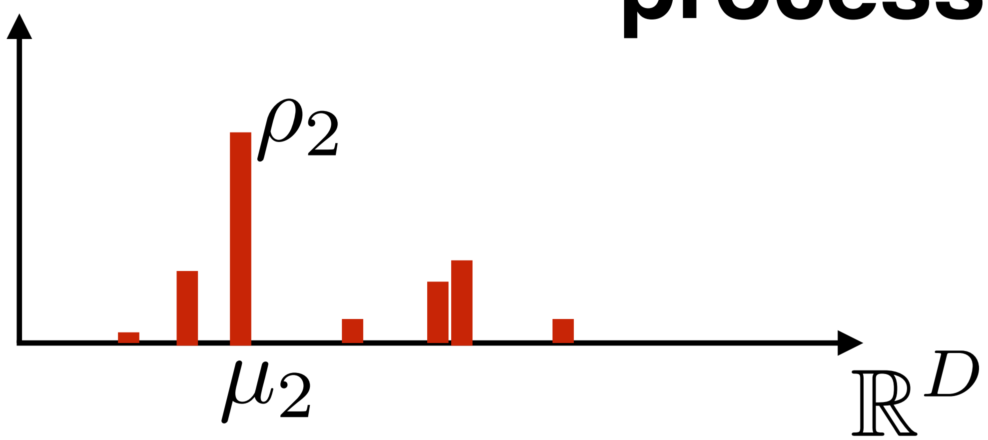
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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

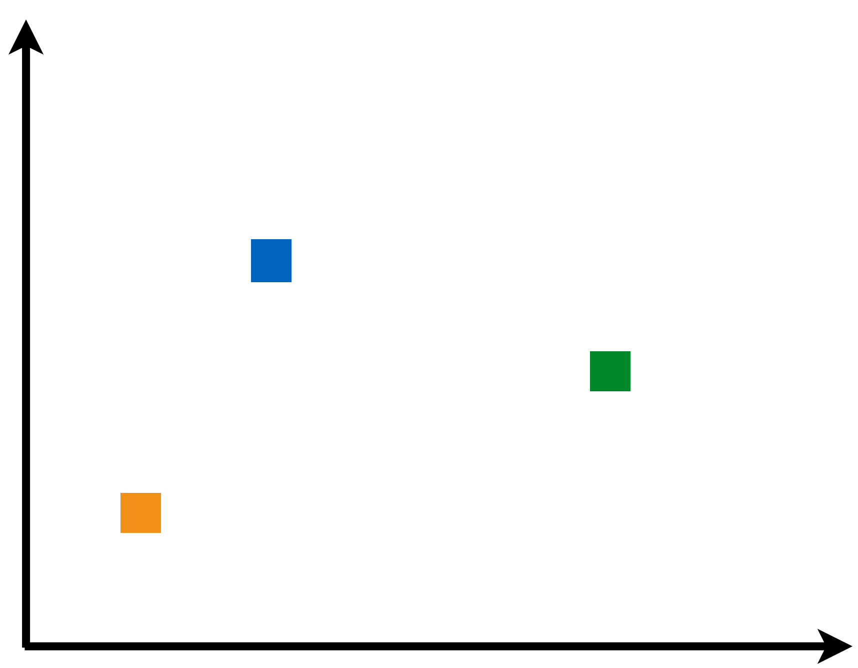


Dirichlet process



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

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



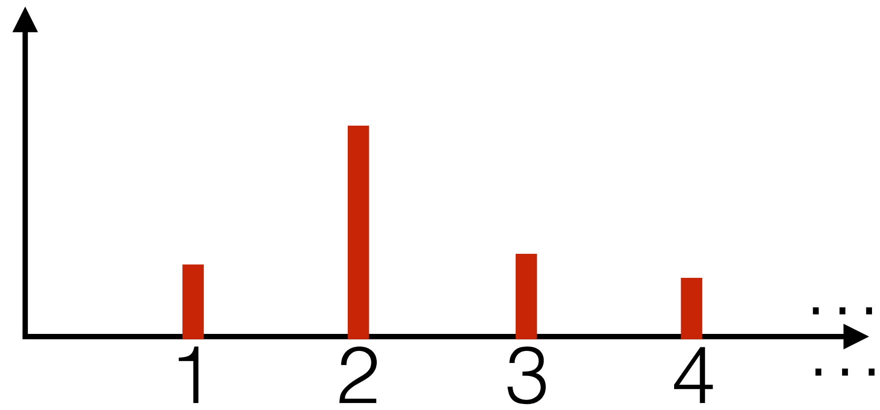
Dirichlet process mixture model

- Gaussian mixture model

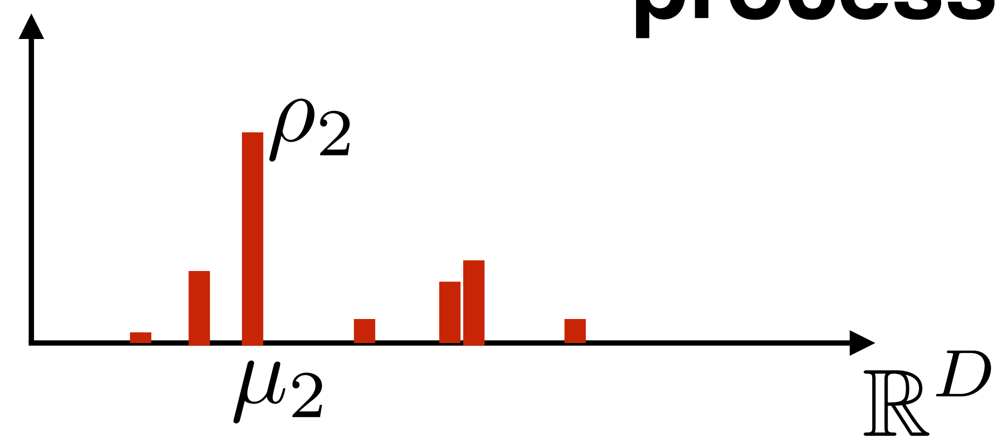
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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



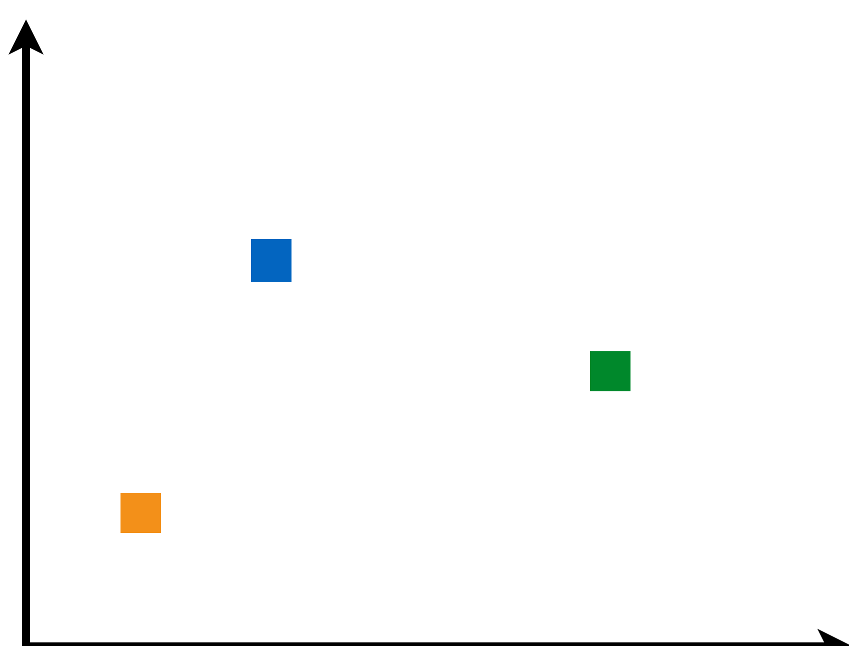
Dirichlet process



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

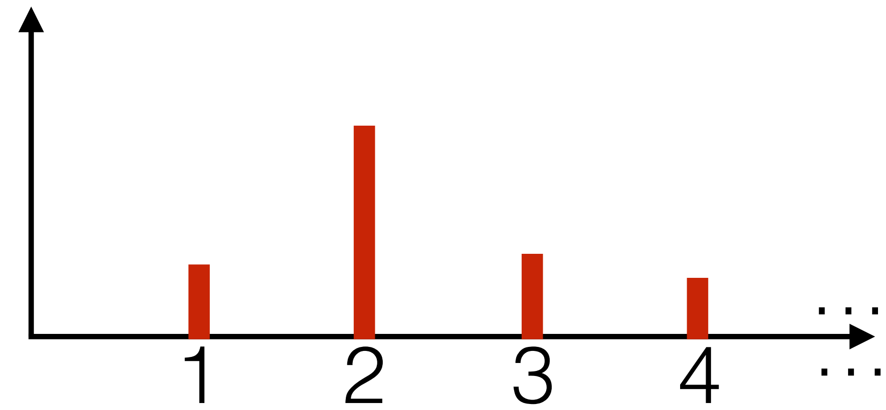
$$\mu_n^* = \mu_{z_n}$$

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



Dirichlet process mixture model

- Gaussian mixture model

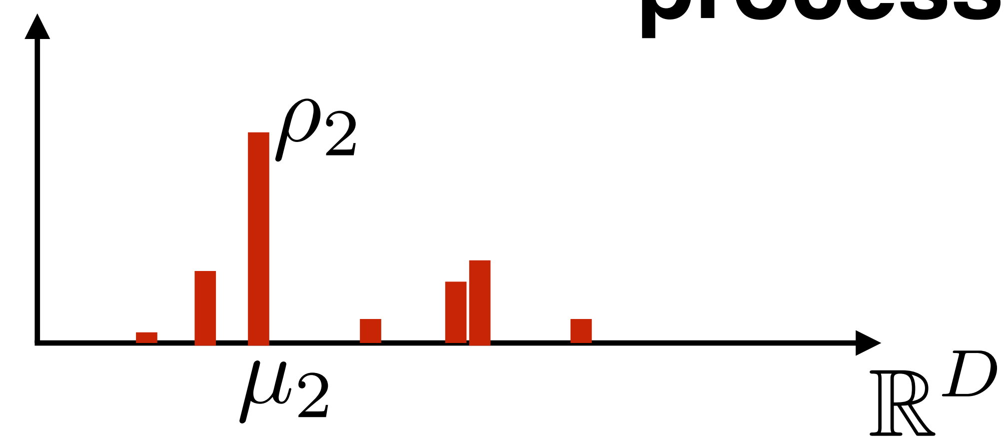


$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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

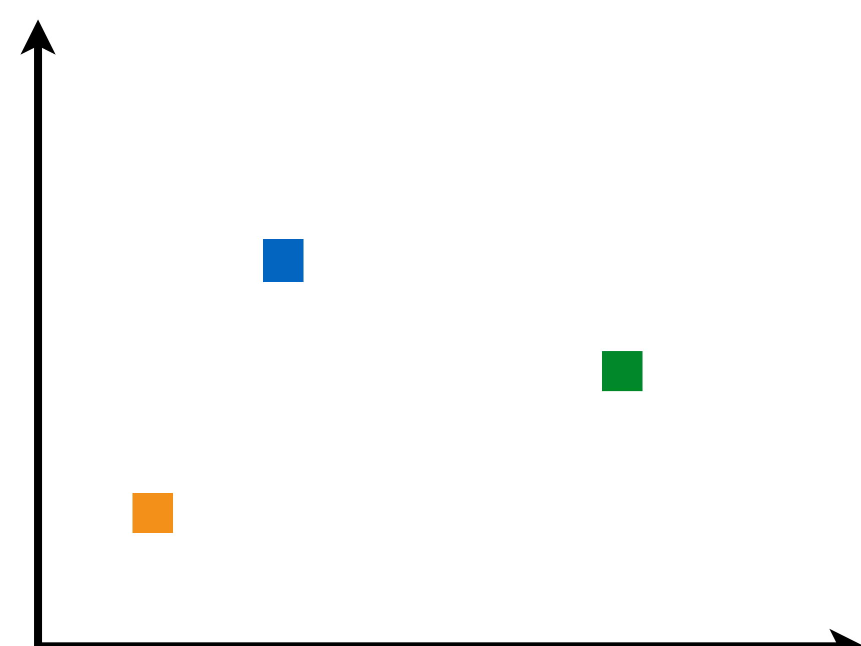
Dirichlet process



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

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

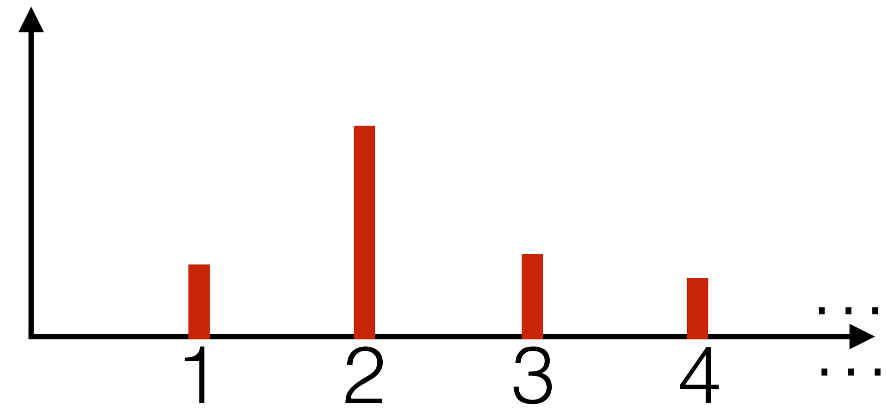
- i.e. $\mu_n^* \overset{iid}{\sim} G$



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

Dirichlet process mixture model

- Gaussian mixture model

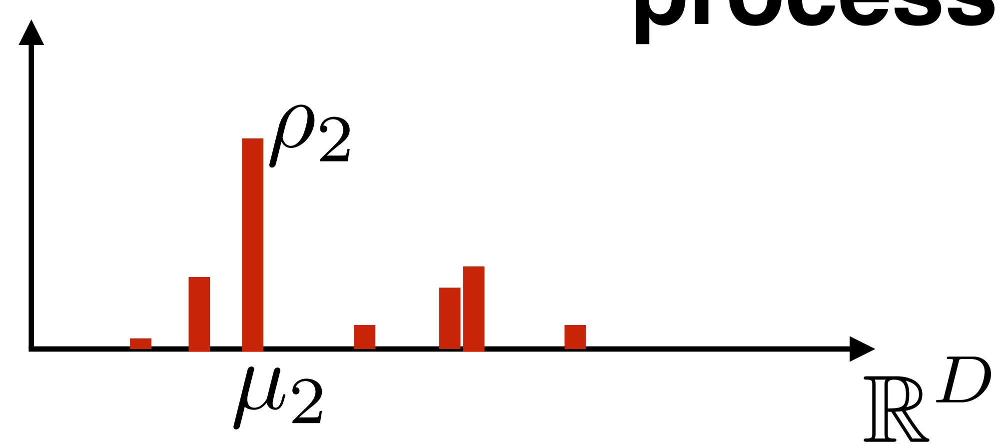


$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$

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

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

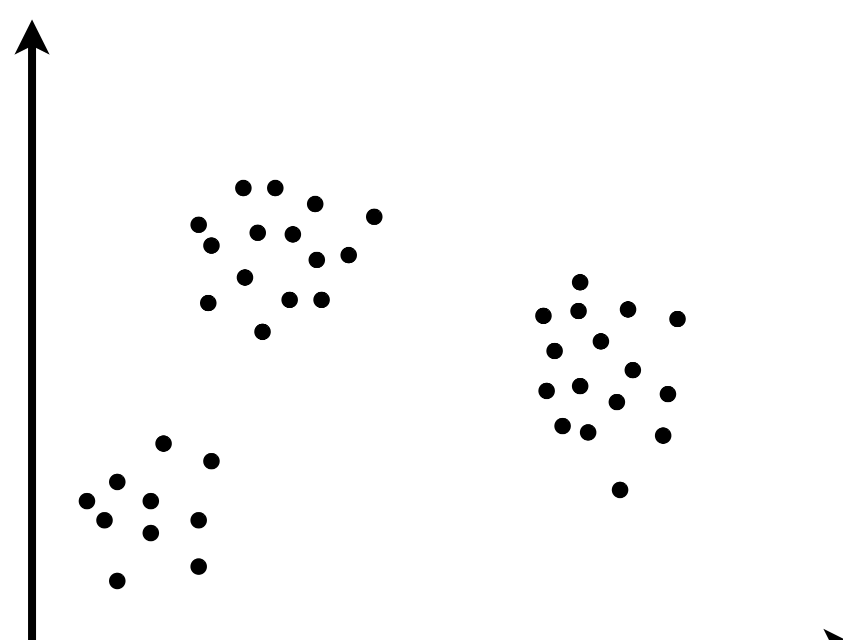
Dirichlet process



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

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

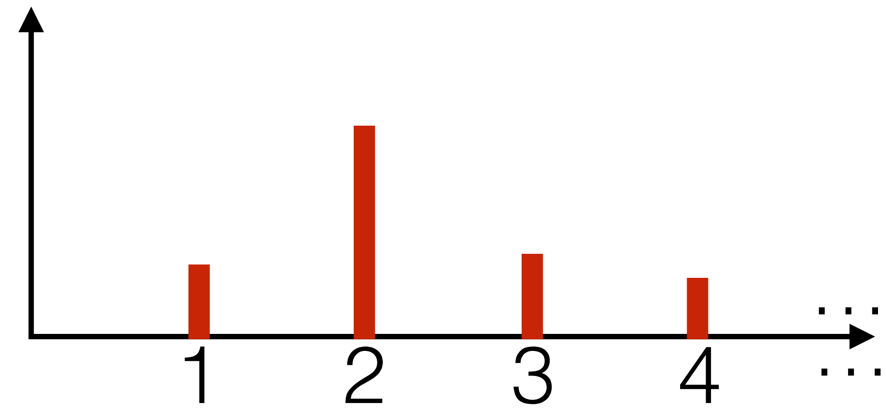
- i.e. $\mu_n^* \overset{iid}{\sim} G$



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

Dirichlet process mixture model

- Gaussian mixture model

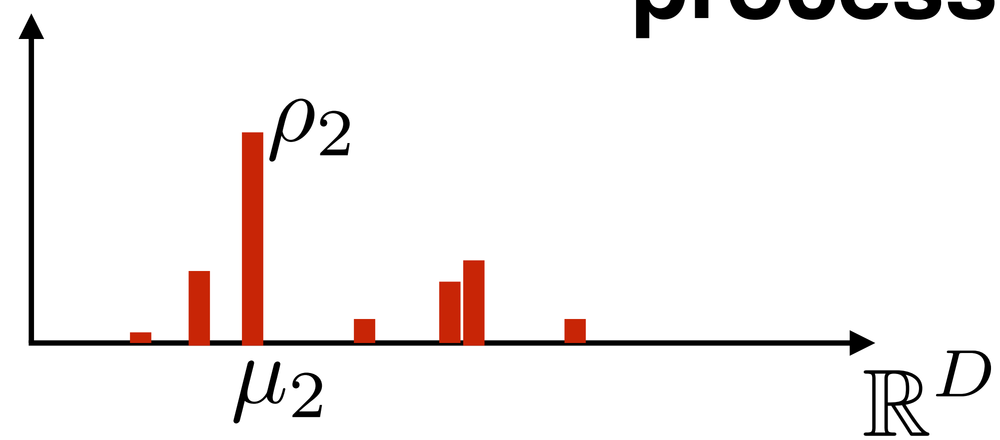


$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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

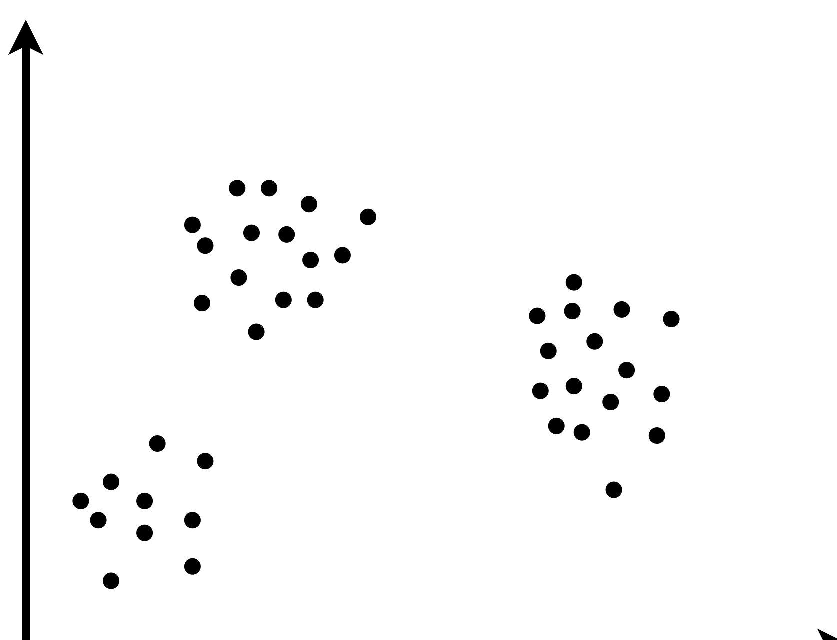
Dirichlet process



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

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

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



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

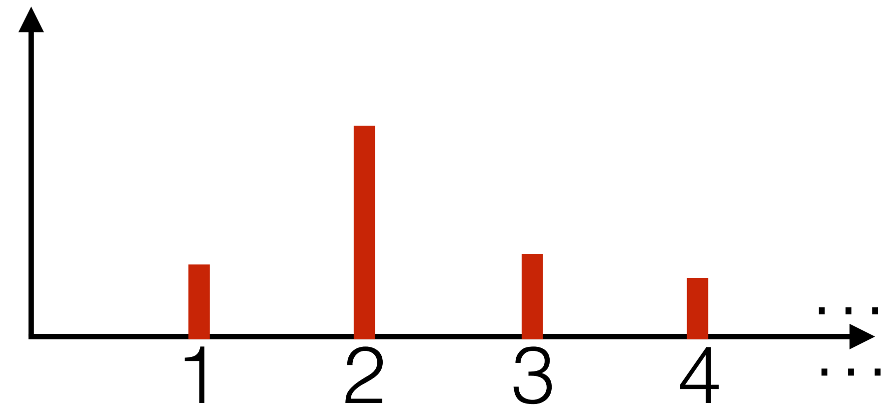
[demo]

Dirichlet process mixture model

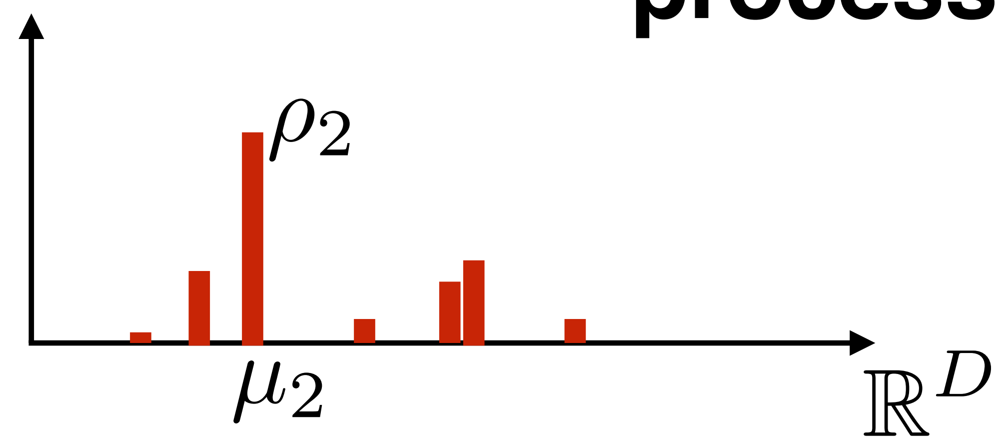
- More generally

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

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



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

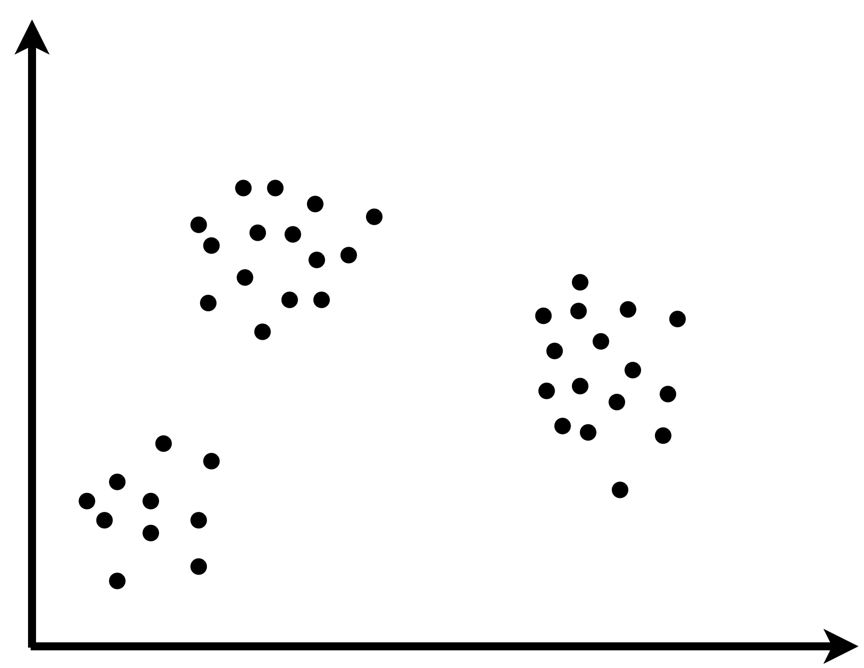


$$z_n \stackrel{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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



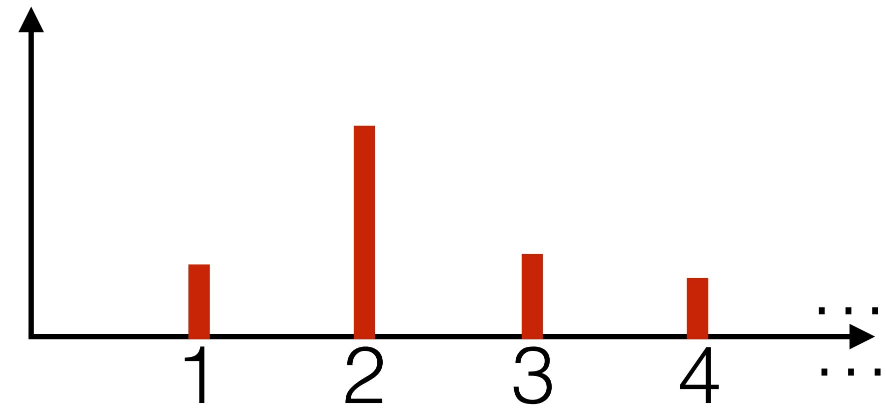
Dirichlet process mixture model

- More generally

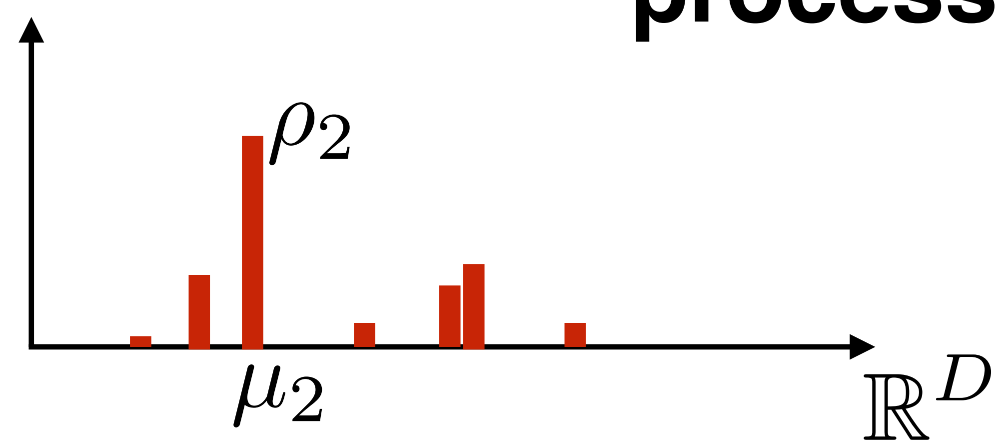
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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



Dirichlet process

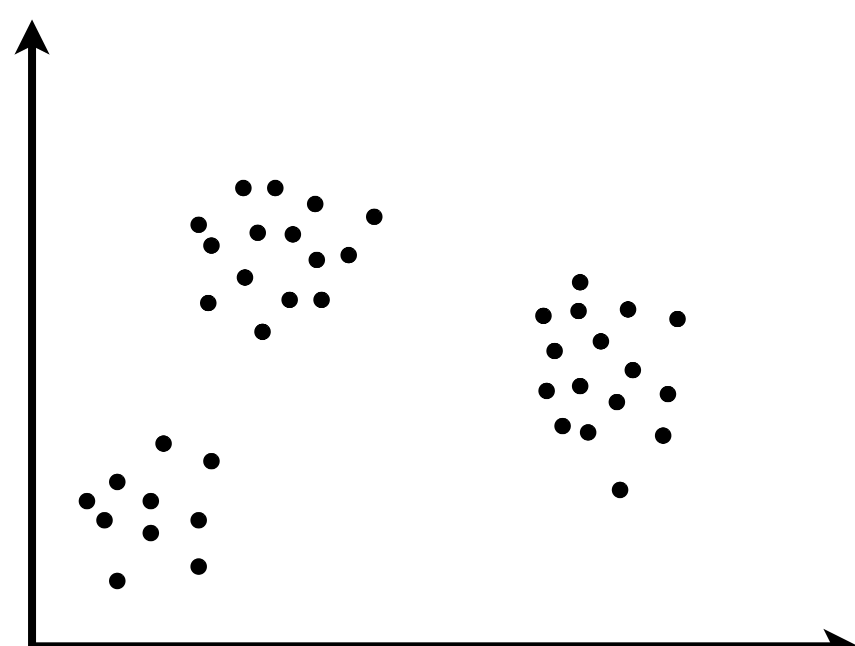


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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

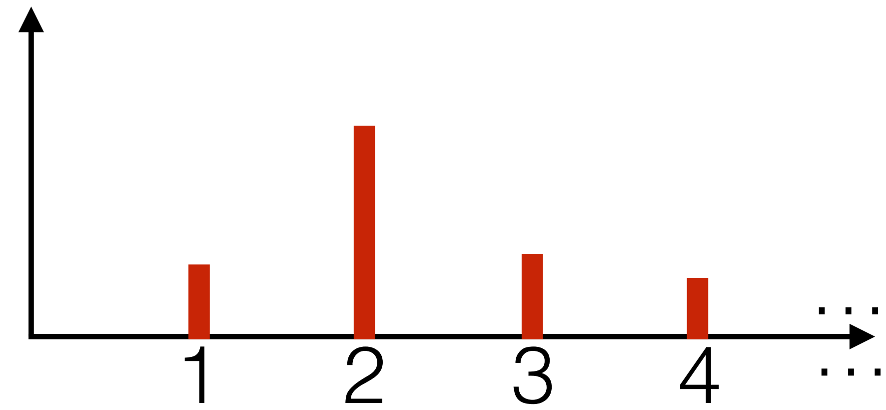


Dirichlet process mixture model

- More generally

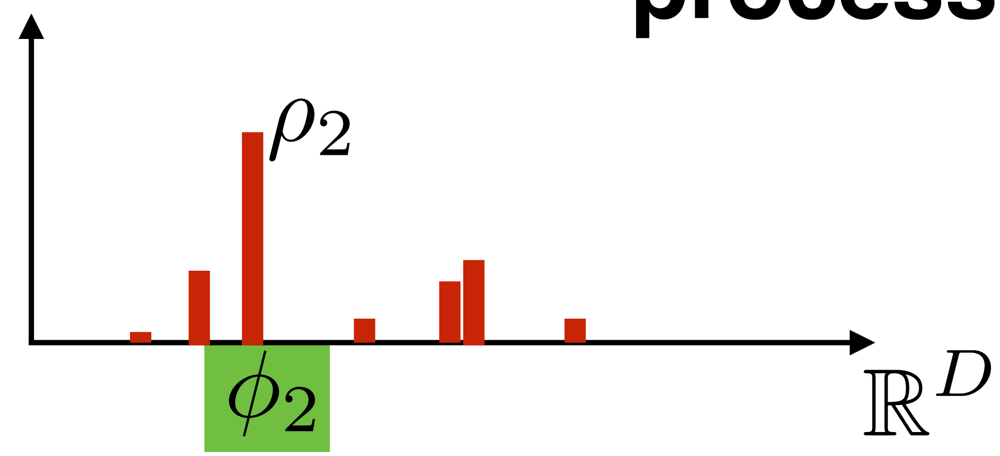
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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



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

Dirichlet process

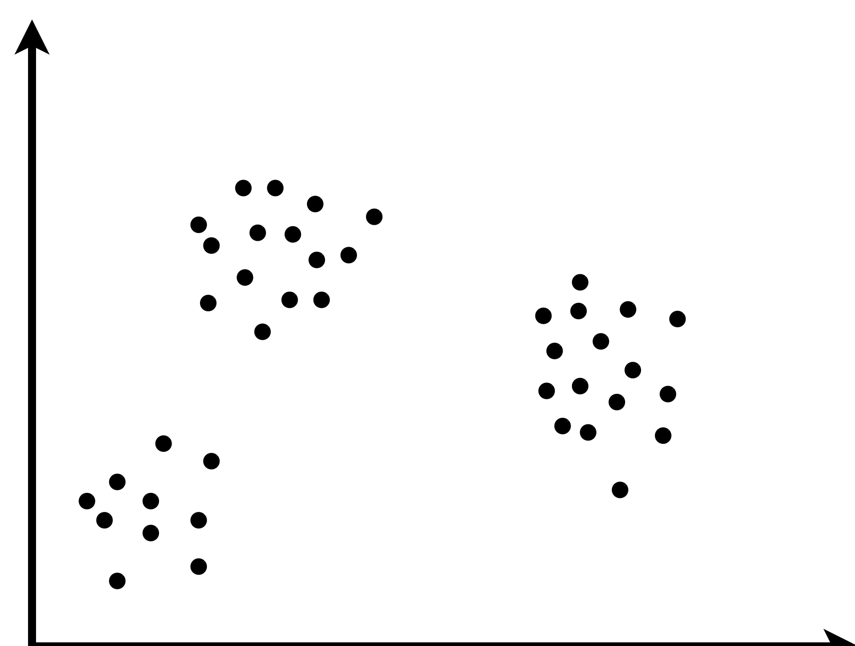


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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

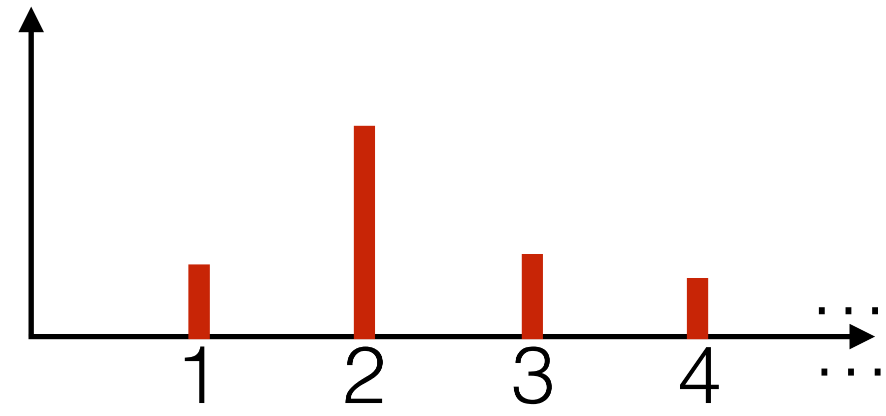


Dirichlet process mixture model

- More generally

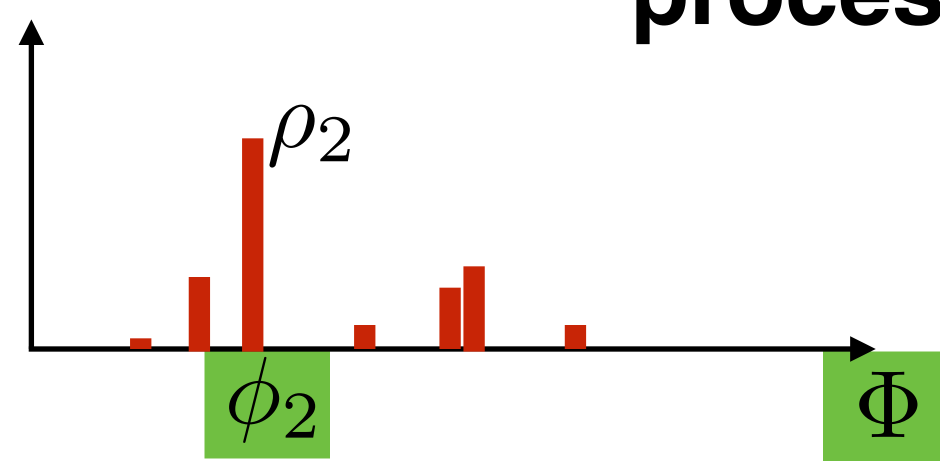
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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



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

Dirichlet process

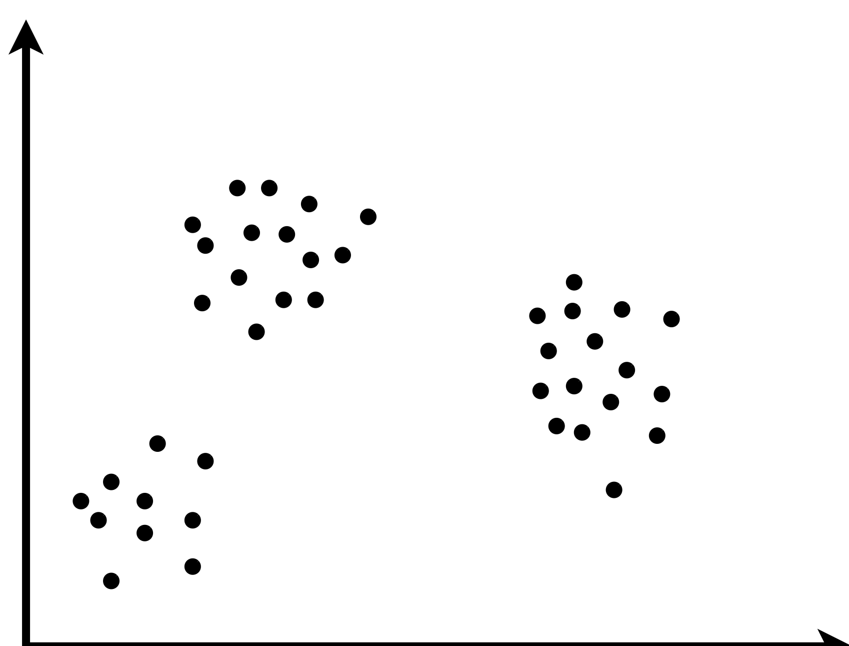


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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

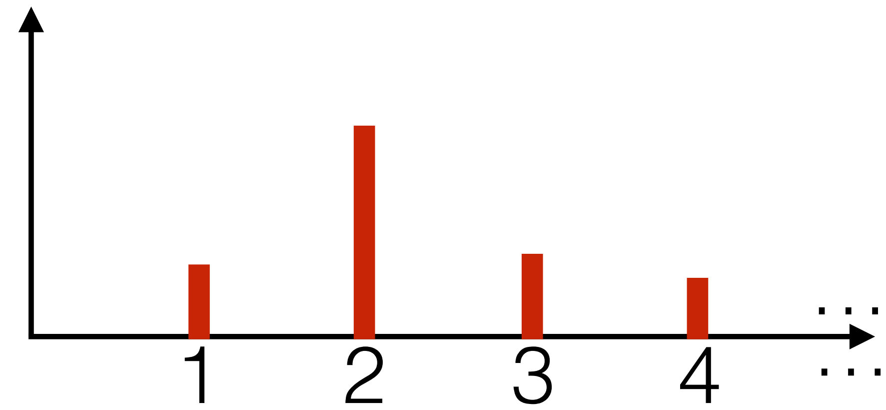


Dirichlet process mixture model

- More generally

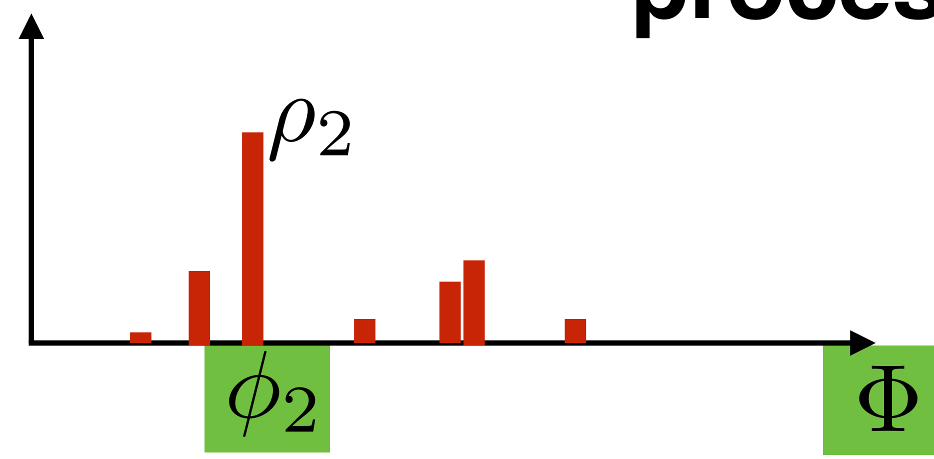
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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



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

Dirichlet process

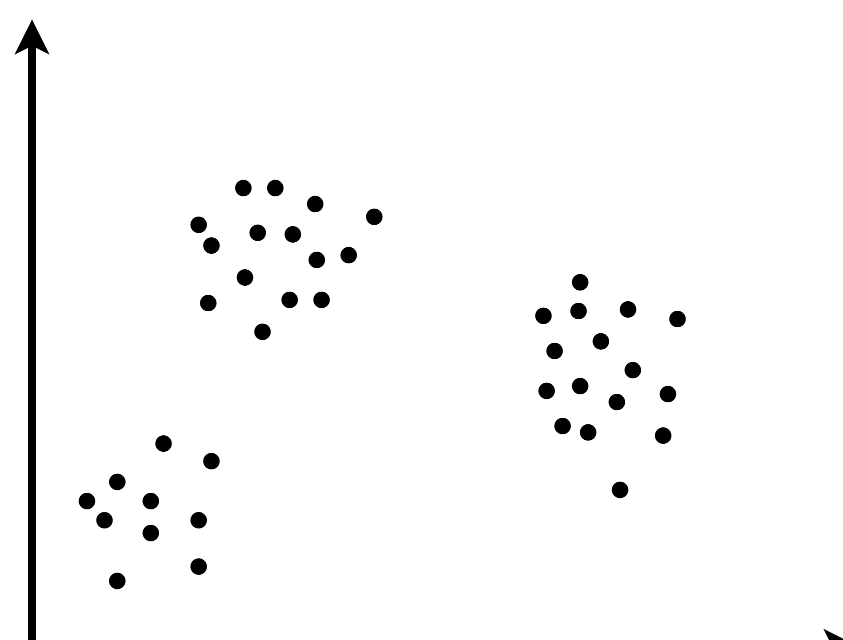


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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

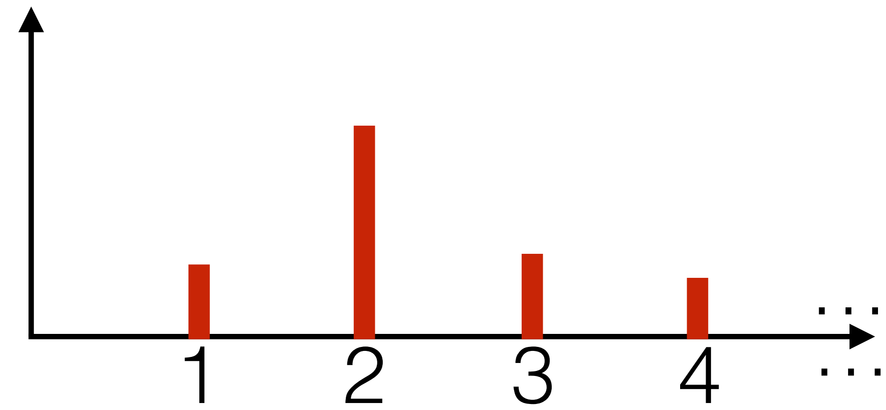


Dirichlet process mixture model

- More generally

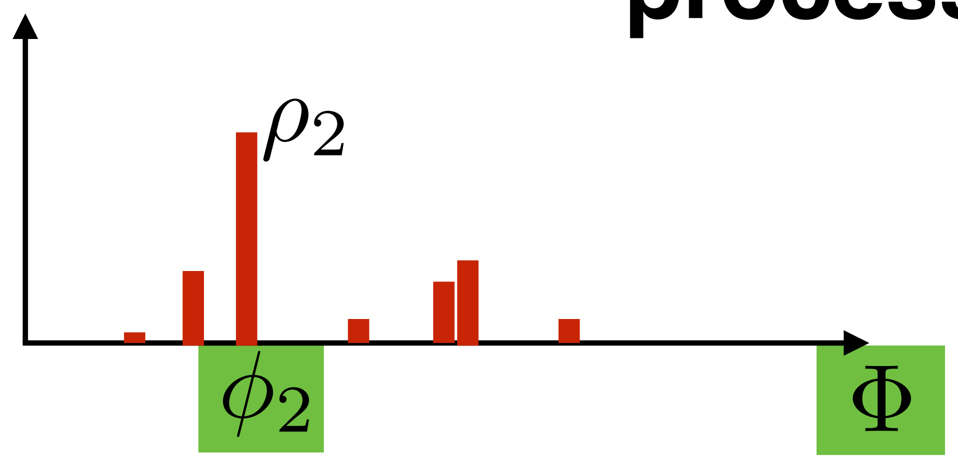
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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



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

Dirichlet process

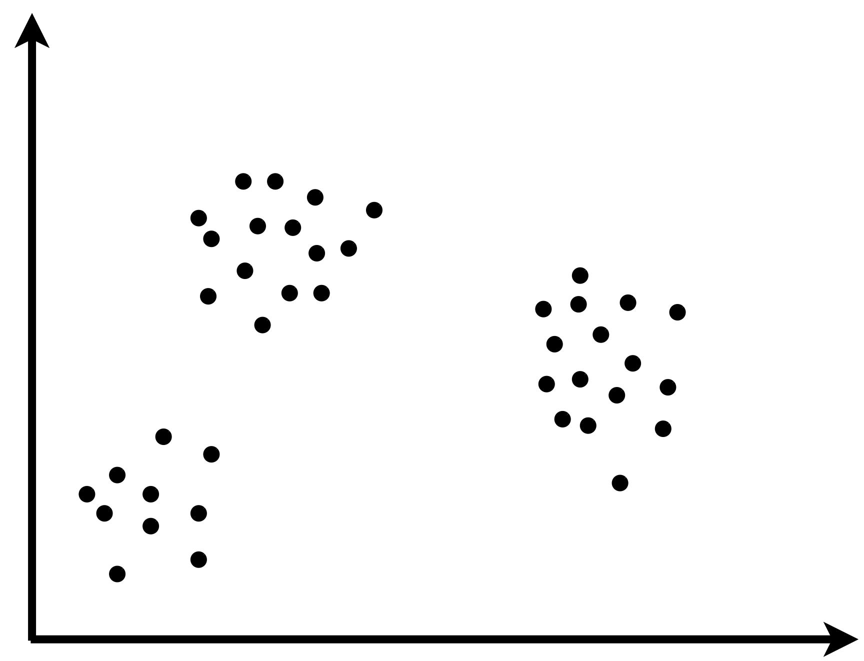


$$z_n \stackrel{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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



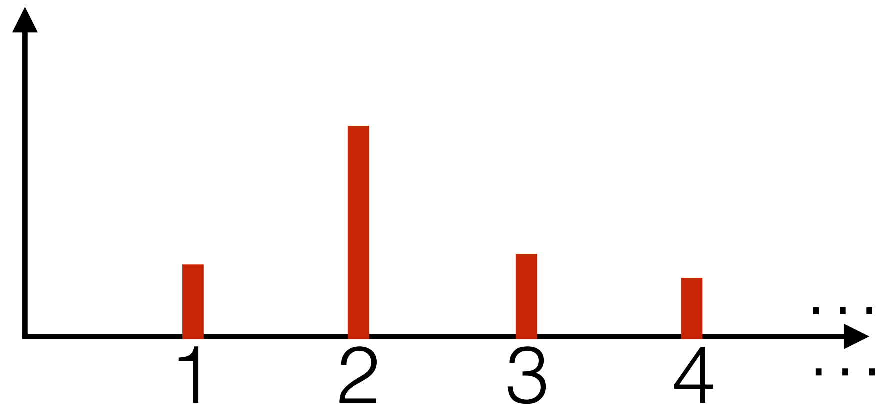
Dirichlet process mixture model

- More generally

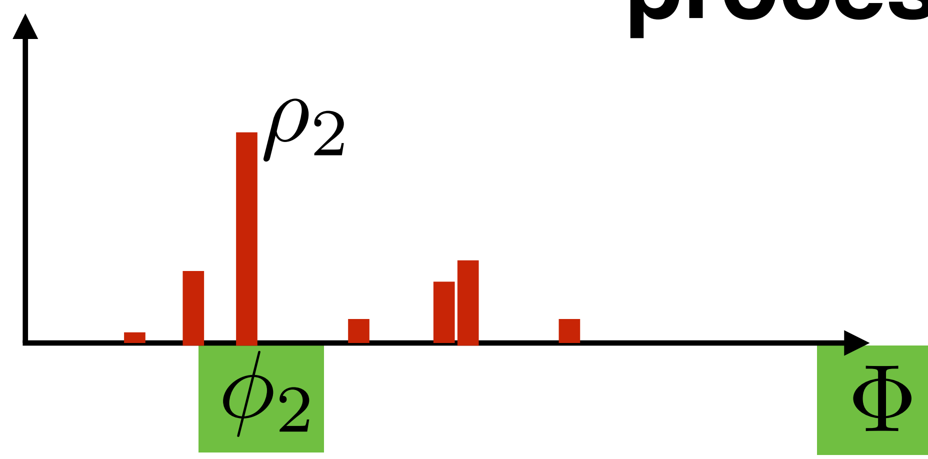
$$\rho = (\rho_1, \rho_2, \ldots) \sim \text{GEM}(\alpha)$$

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

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



Dirichlet process

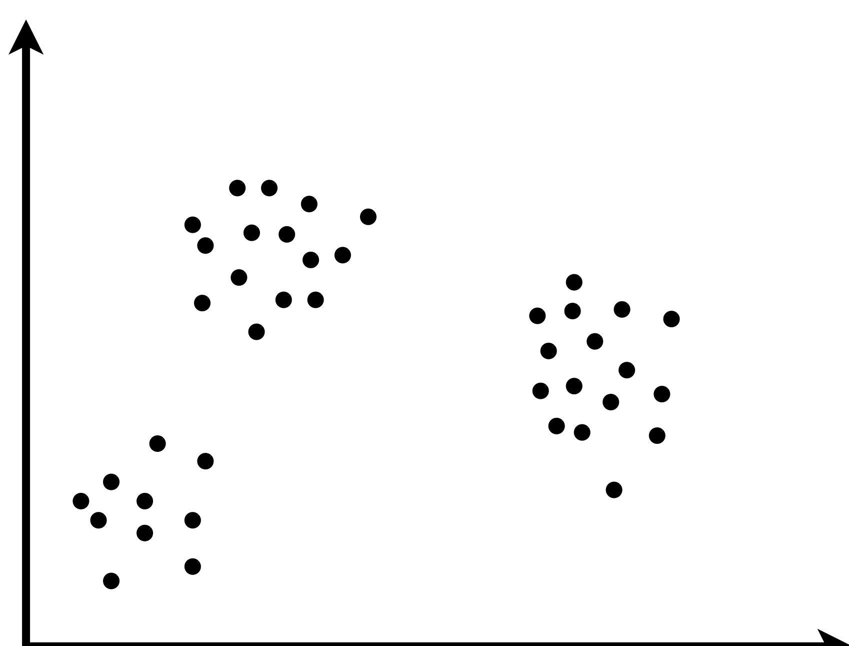


$$z_n \overset{iid}{\sim} \text{Categorical}(\rho)$$

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

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

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



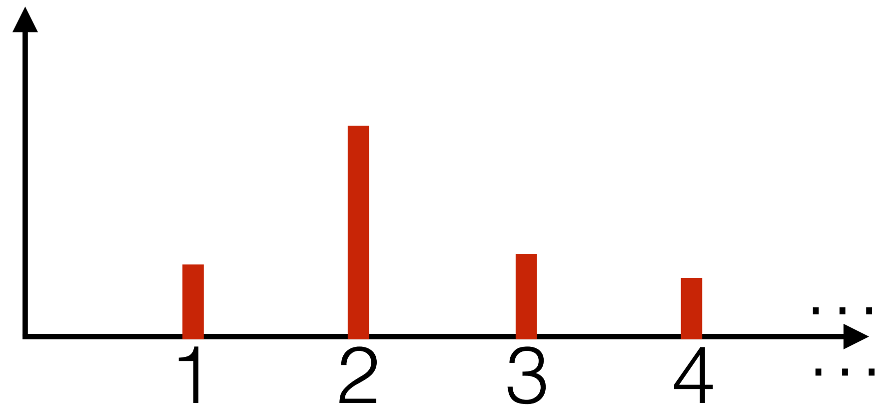
Dirichlet process mixture model

- More generally

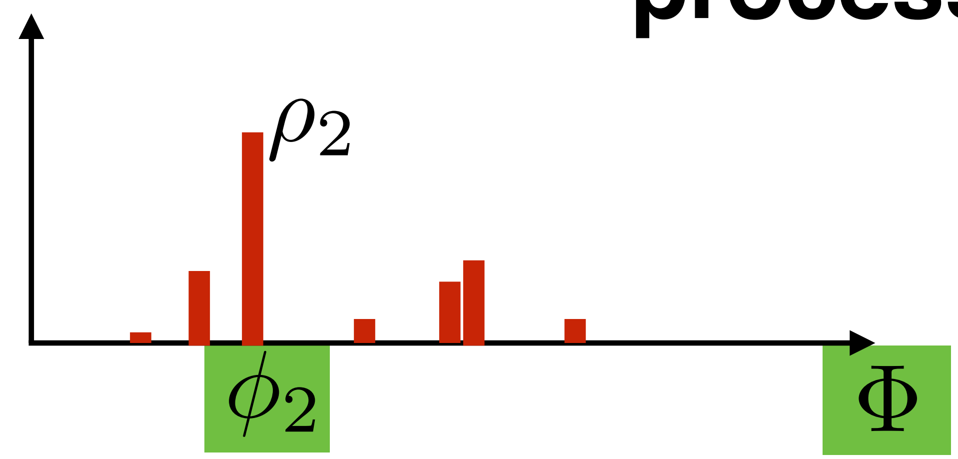
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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



Dirichlet process

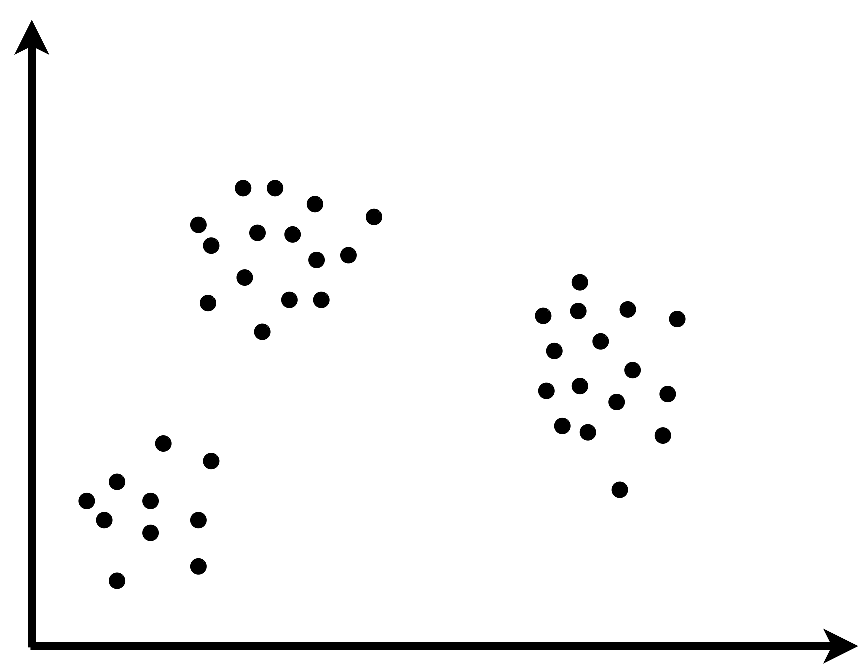


$$z_n \stackrel{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

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



Dirichlet process mixture model

- More generally

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

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

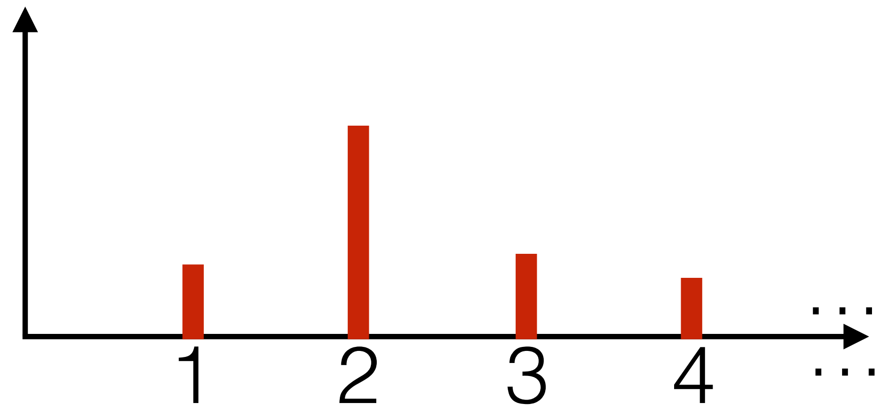
- i.e. $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k} \overset{d}{=} \mathrm{DP}(\alpha, G_0)$

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

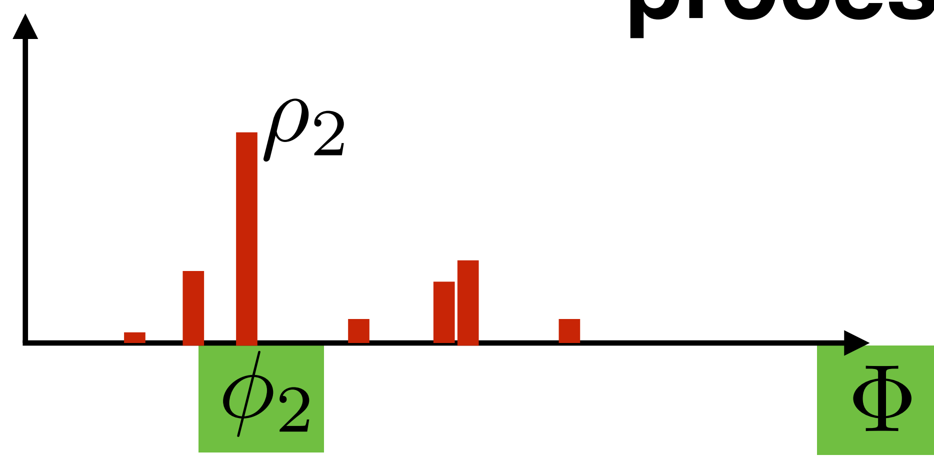
$$\theta_n = \phi_{z_n}$$

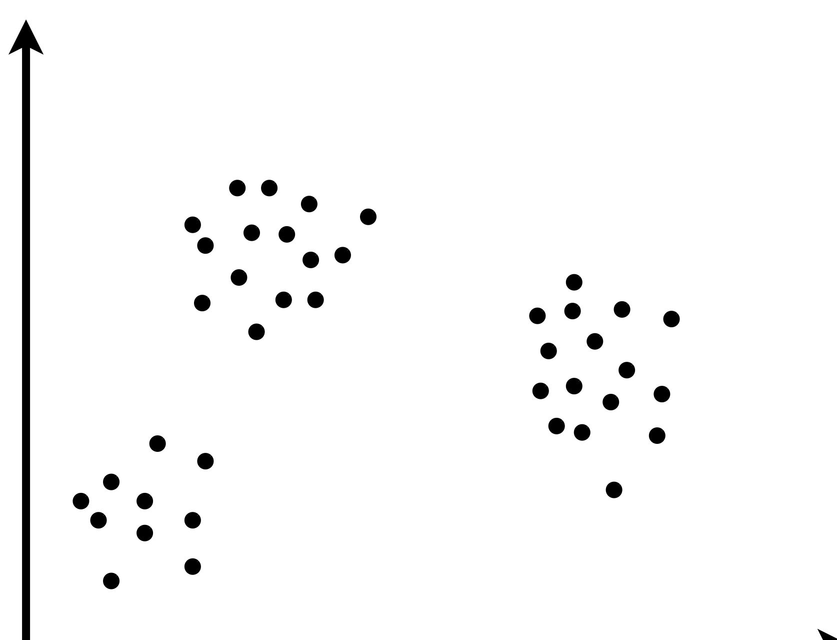
- i.e. $\theta_n \overset{iid}{\sim} G$

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



Dirichlet process





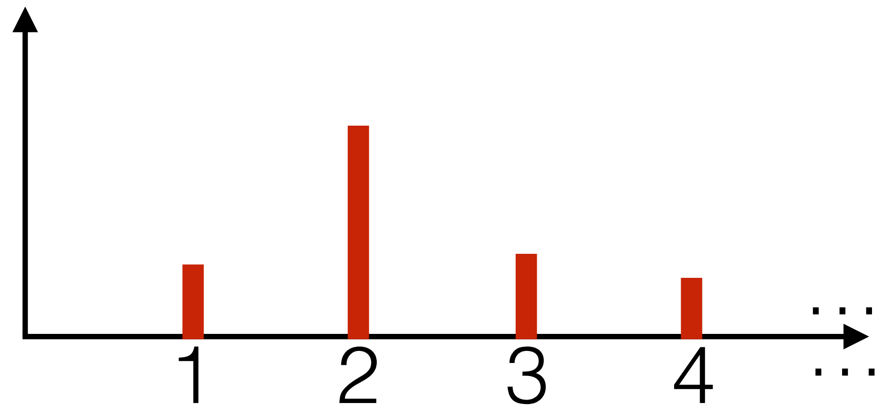
Dirichlet process mixture model

- More generally

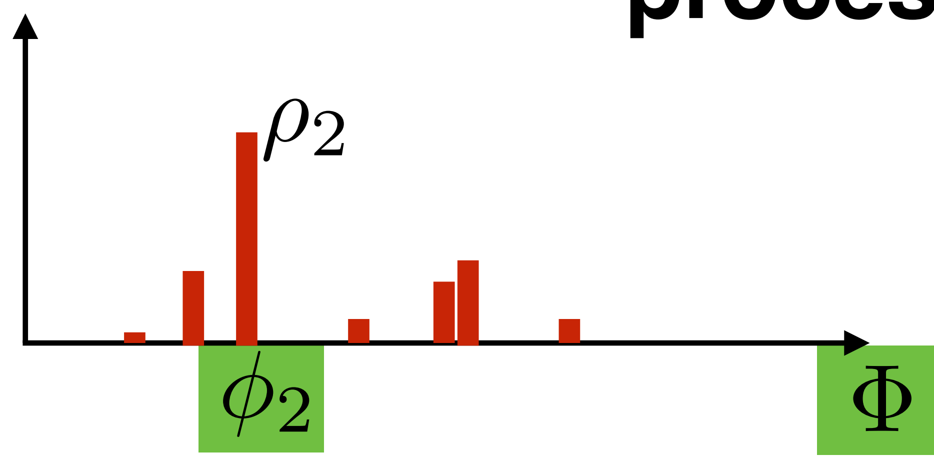
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$

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

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



Dirichlet process

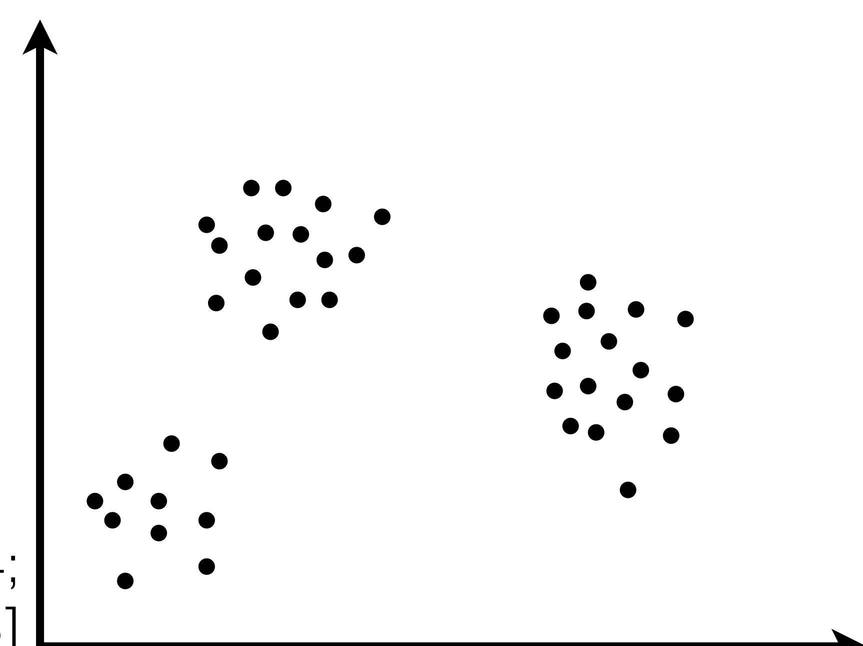


$$z_n \overset{iid}{\sim} \mathrm{Categorical}(\rho)$$

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

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

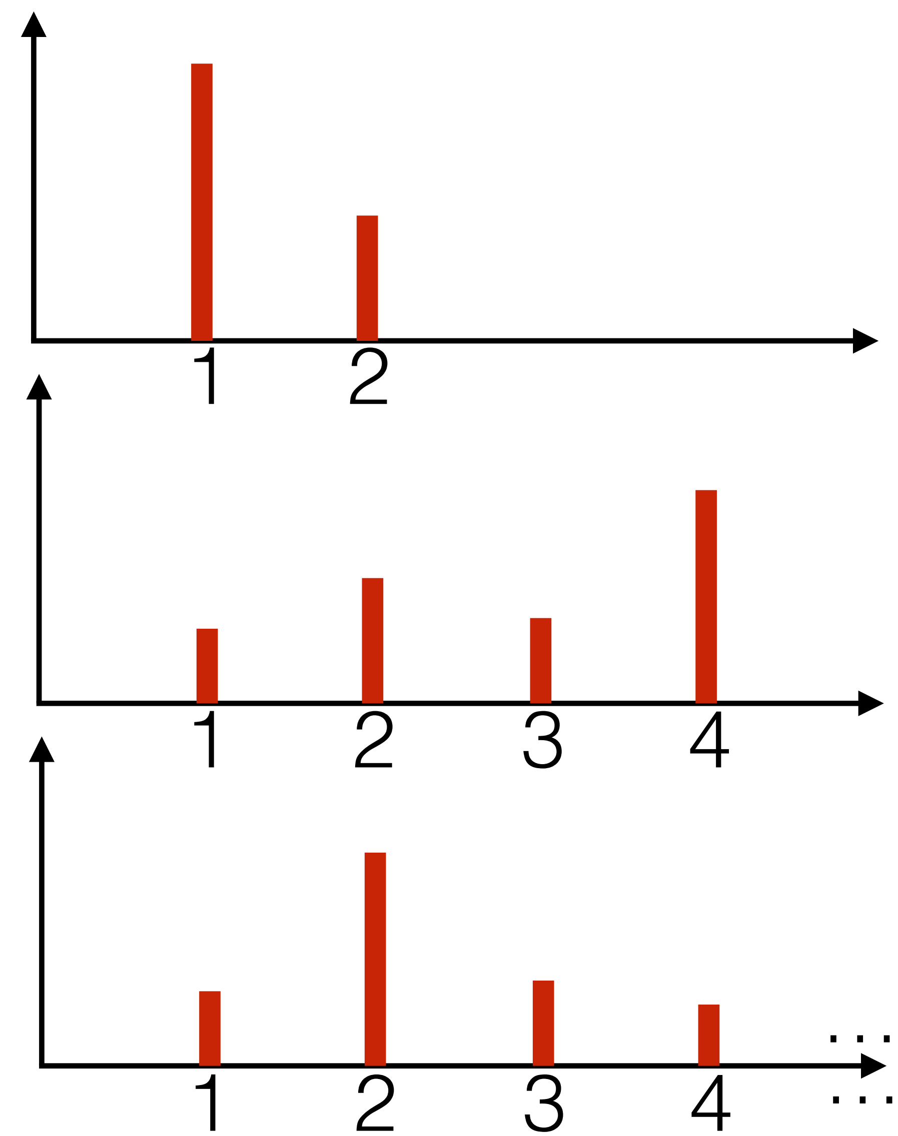
$$x_n \overset{indep}{\sim} F(\theta_n)$$



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

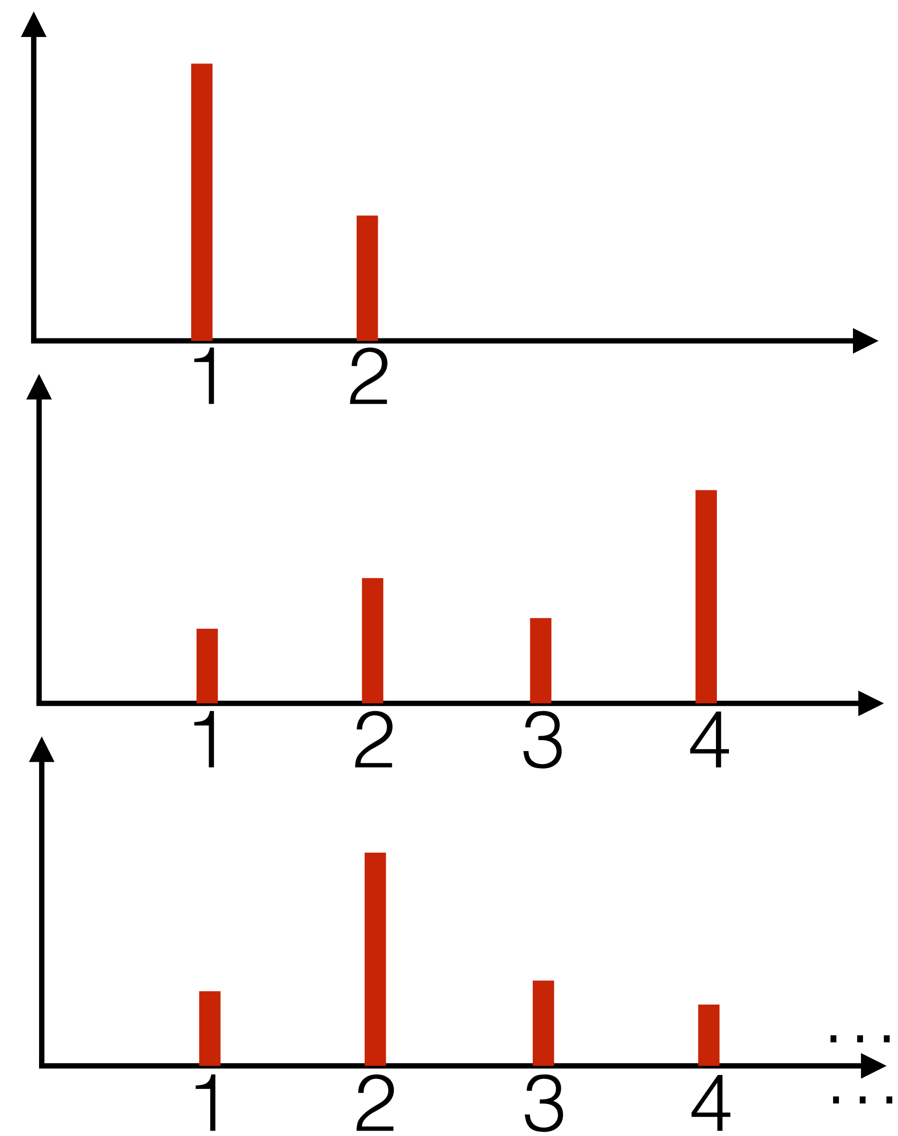
Distributions

- 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

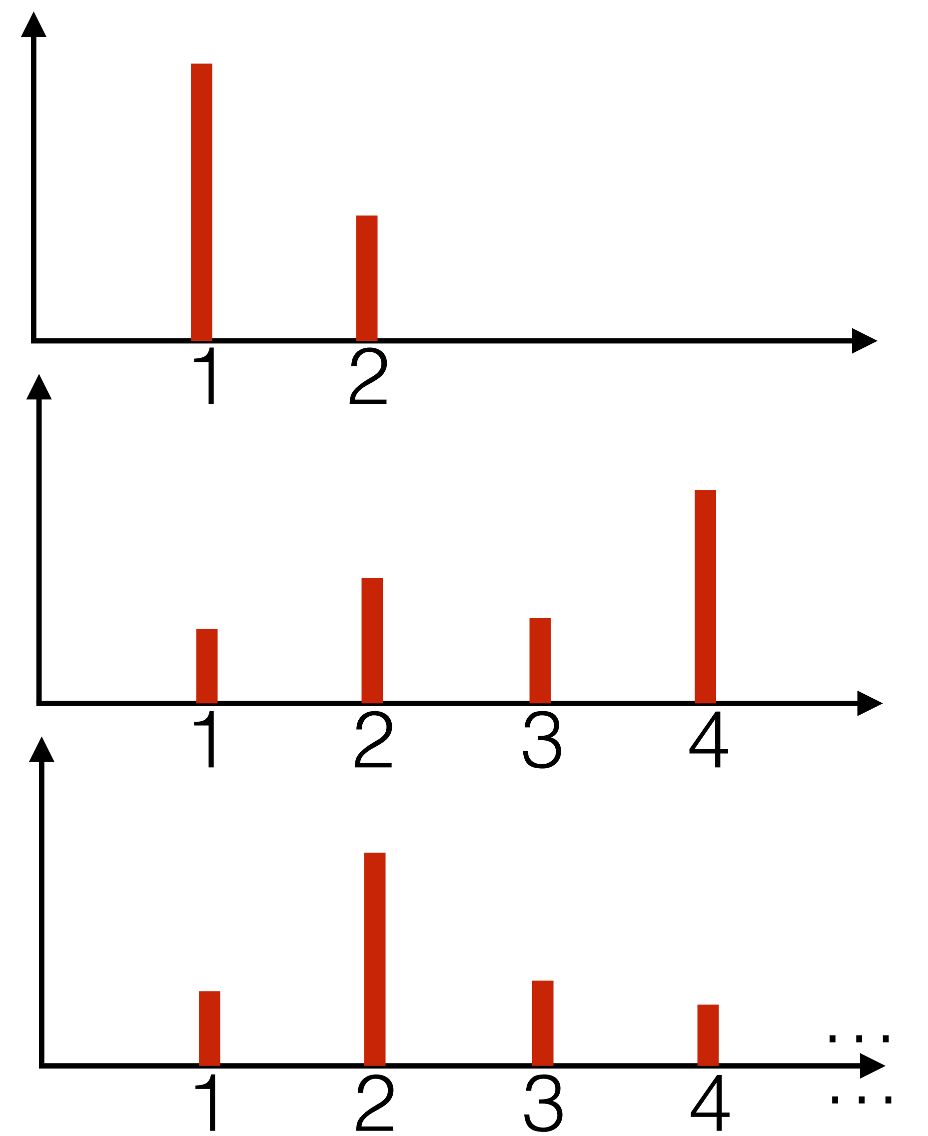
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



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

Distributions

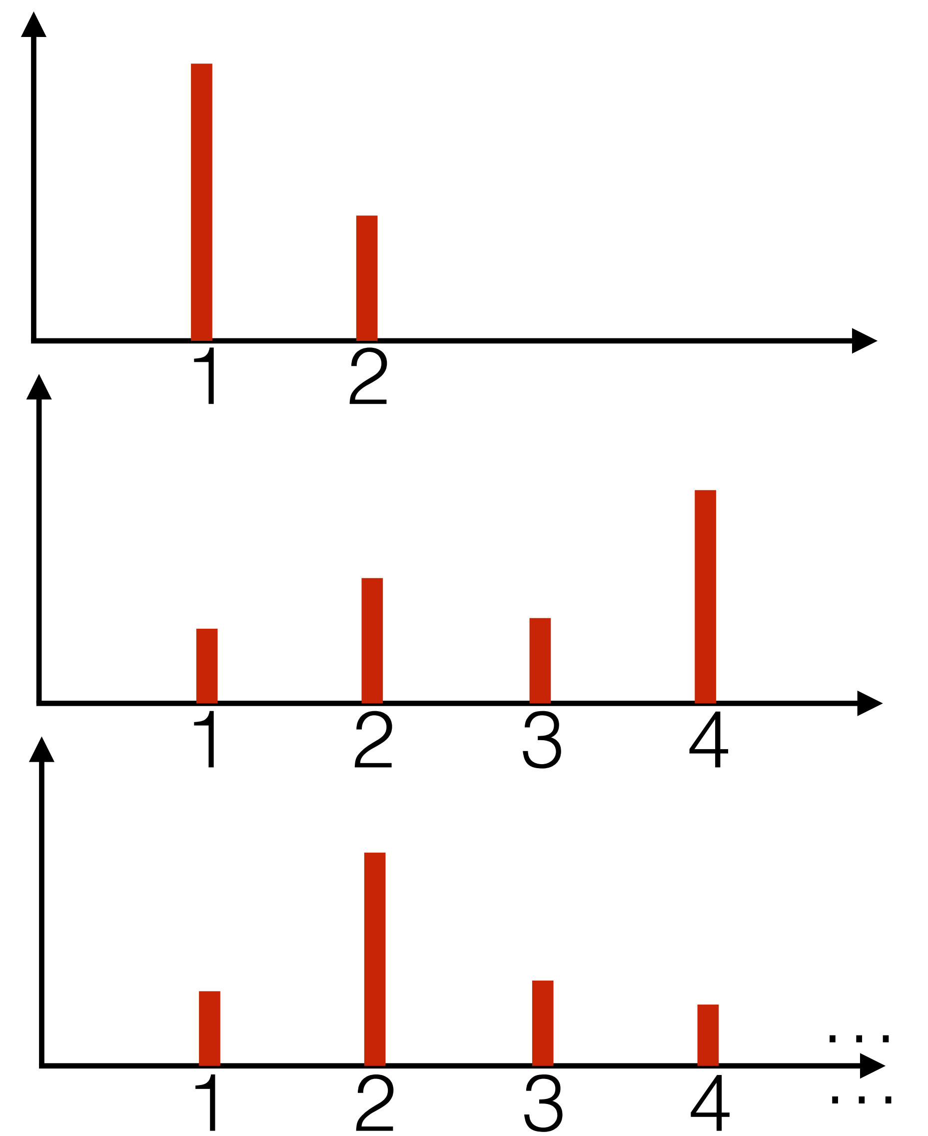
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



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

Distributions

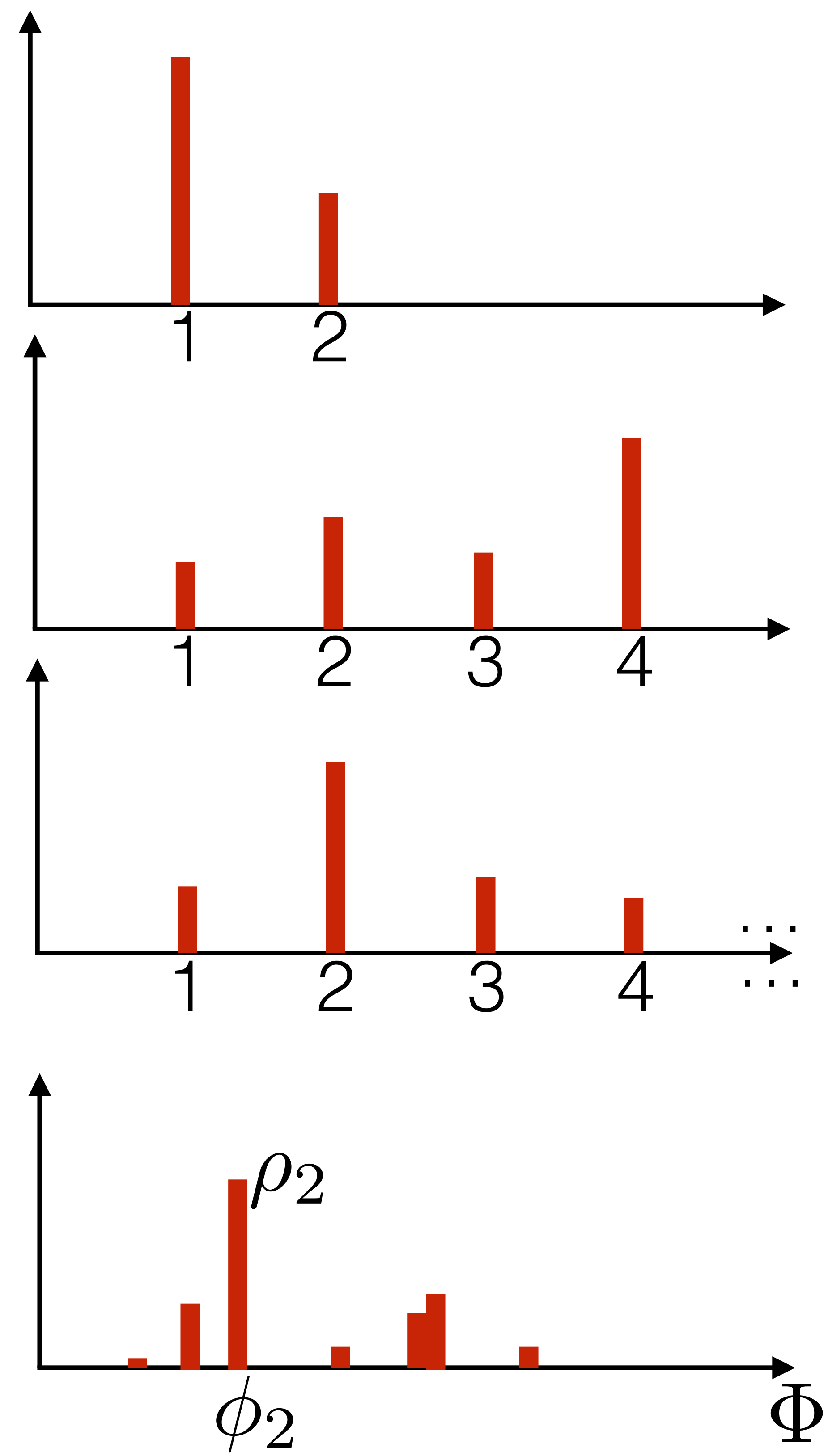
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



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

Distributions

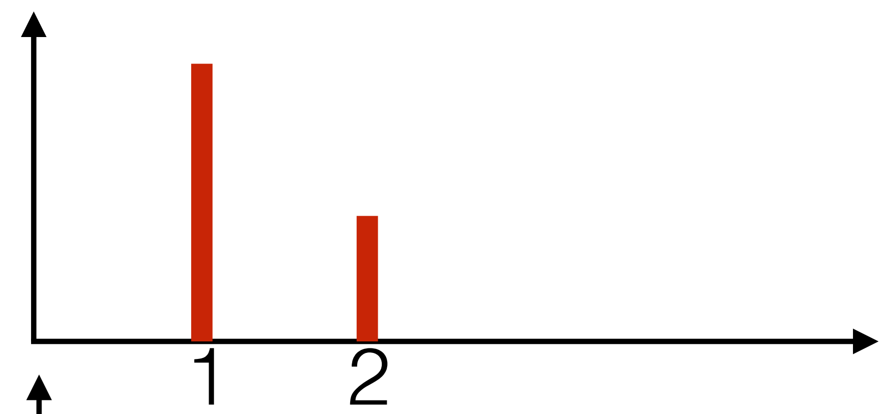
- Beta → random distribution over $1, 2$
- Dirichlet → random distribution over $1, 2, \ldots, K$
- GEM / Dirichlet process stick-breaking → random distribution over $1, 2, \ldots$



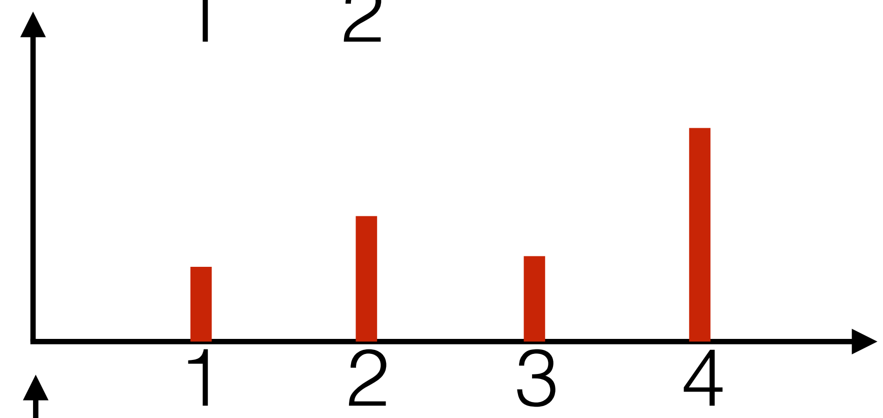
$$\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$$
$$\phi_k \overset{iid}{\sim} G_0$$
$$G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}$$

Distributions

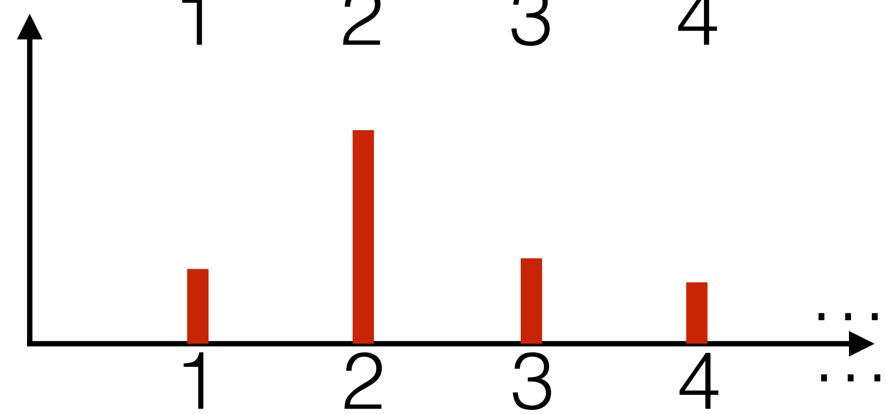
- Beta → random distribution over $1, 2$



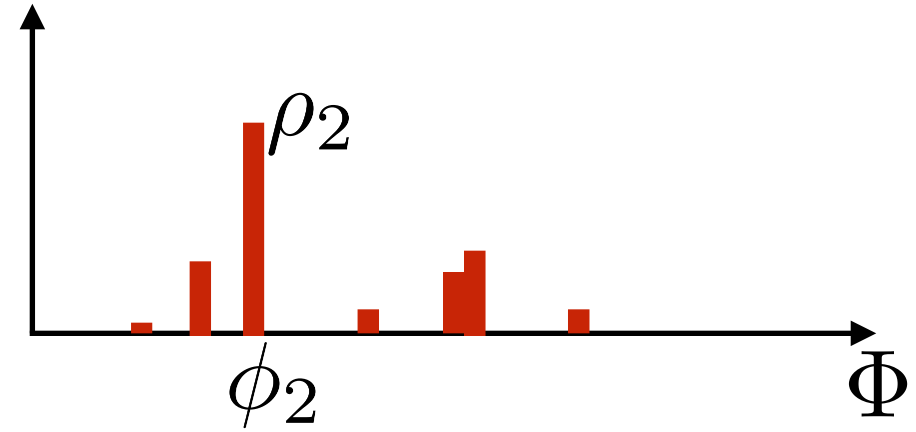
- Dirichlet → random distribution over $1, 2, \ldots, K$



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

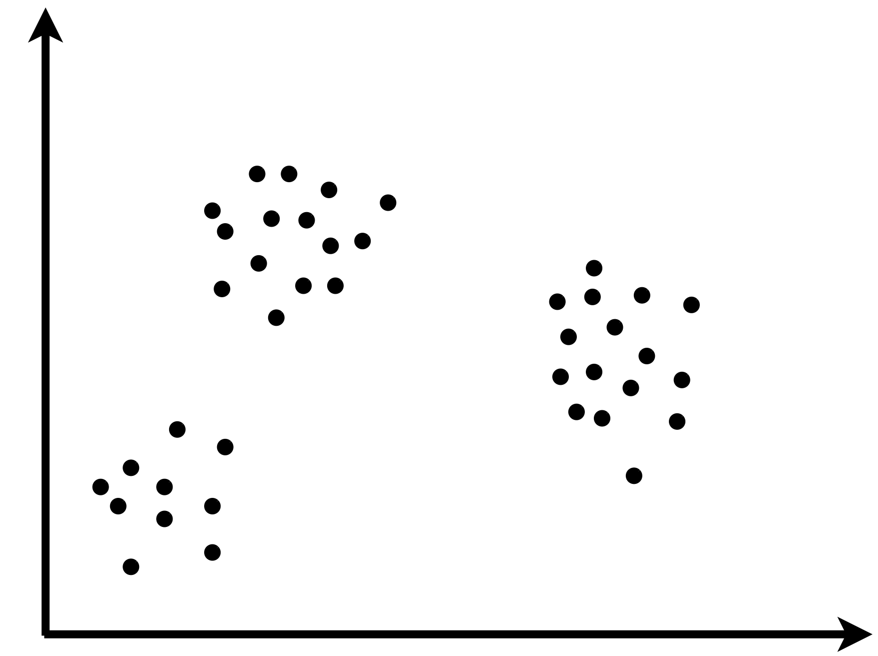


- **Dirichlet process →** random distribution over Φ:
  $\rho = (\rho_1, \rho_2, \ldots) \sim \mathrm{GEM}(\alpha)$
  $\phi_k \stackrel{iid}{\sim} G_0$
  $G = \sum_{k=1}^{\infty} \rho_k \delta_{\phi_k}$



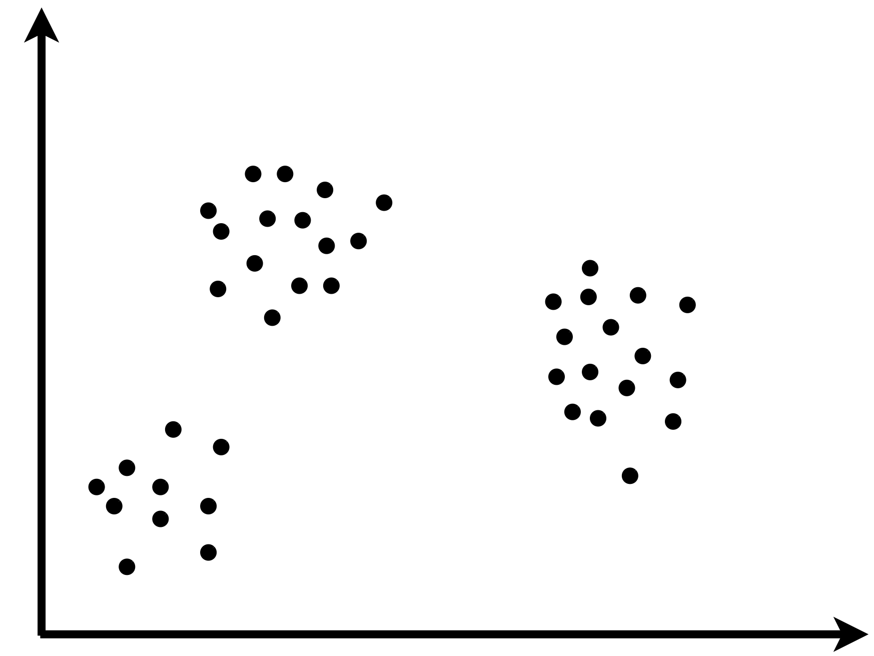
[Ferguson 1973]

DP or not DP, that is the question



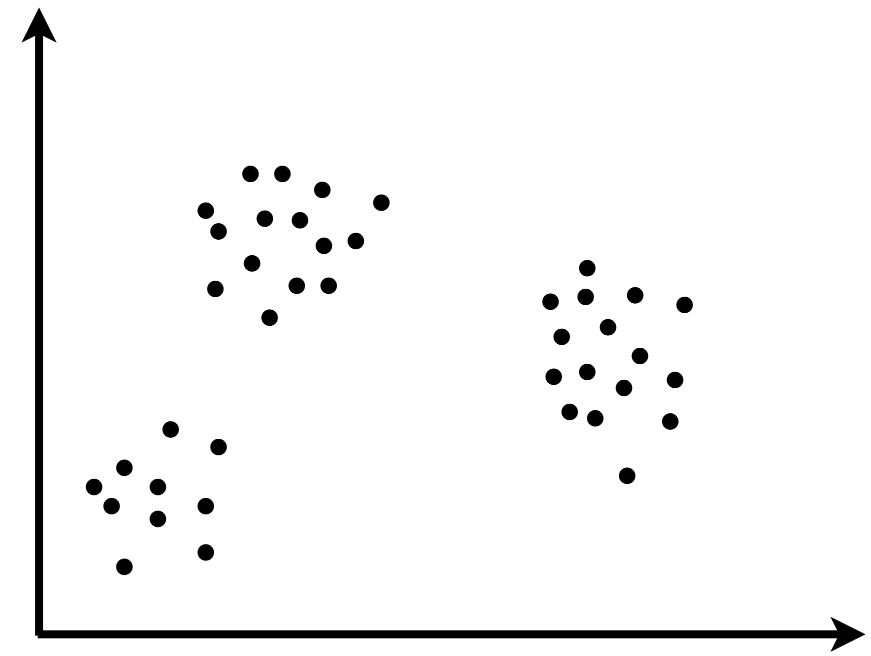
DP or not DP, that is the question

- GEM: 



DP or not DP, that is the question

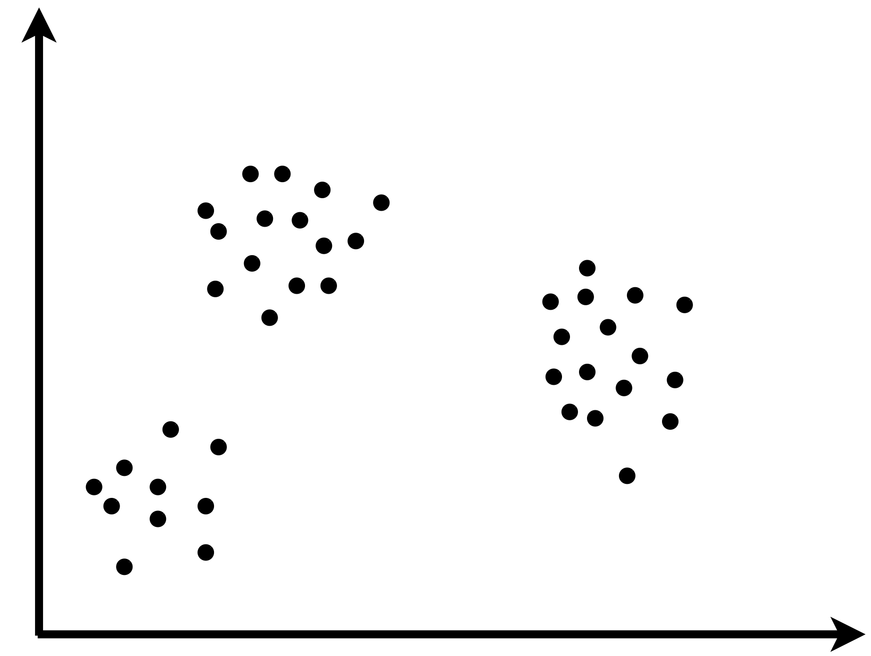
- GEM: 
- Compare to:



DP or not DP, that is the question

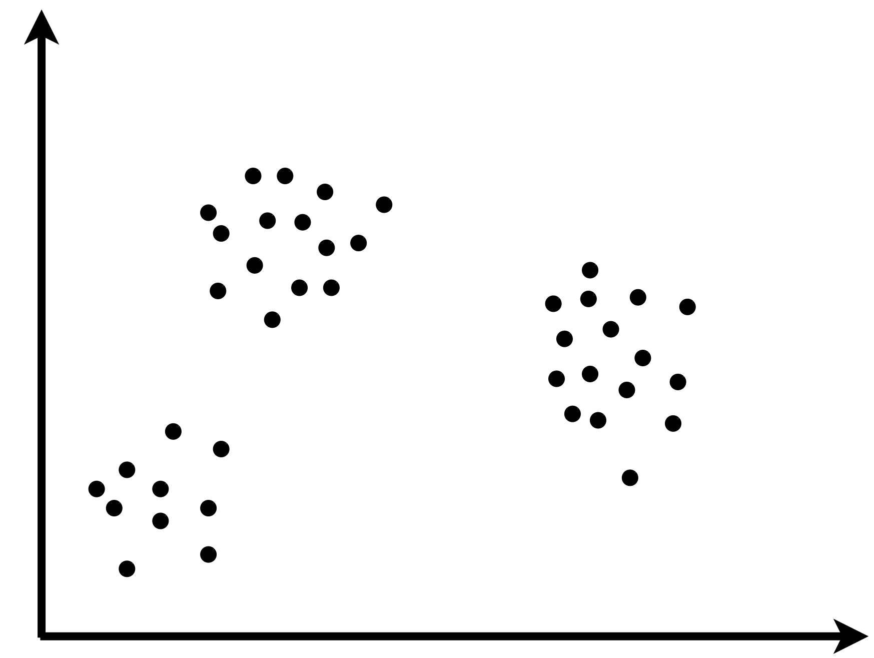
- GEM: 

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



DP or not DP, that is the question

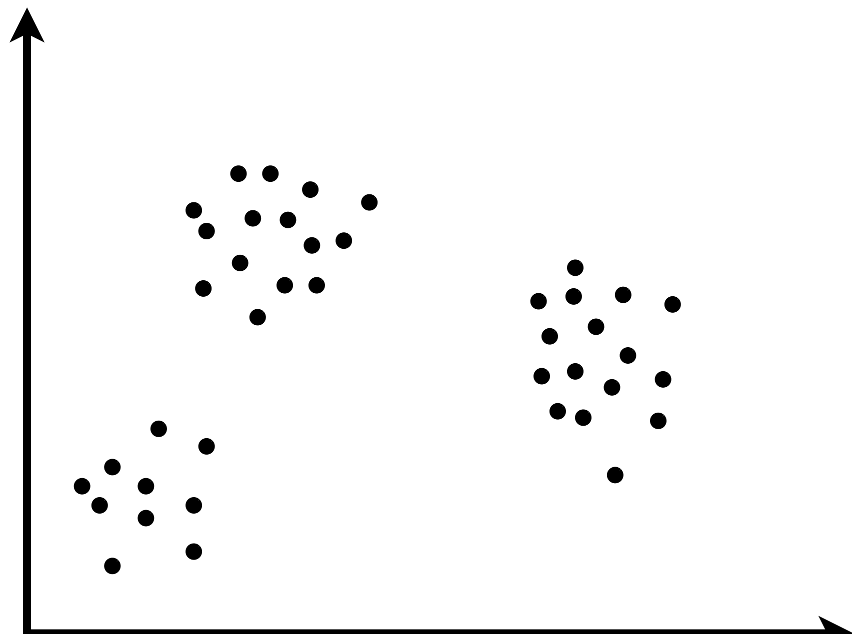
- GEM: 
- Compare to:
  - Finite (small *K*) mixture model



  - Finite (large *K*) mixture model

DP or not DP, that is the question

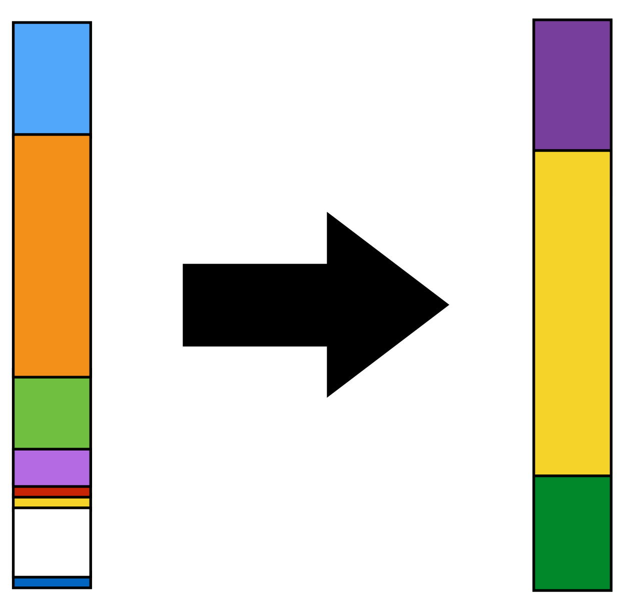
- GEM: 



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

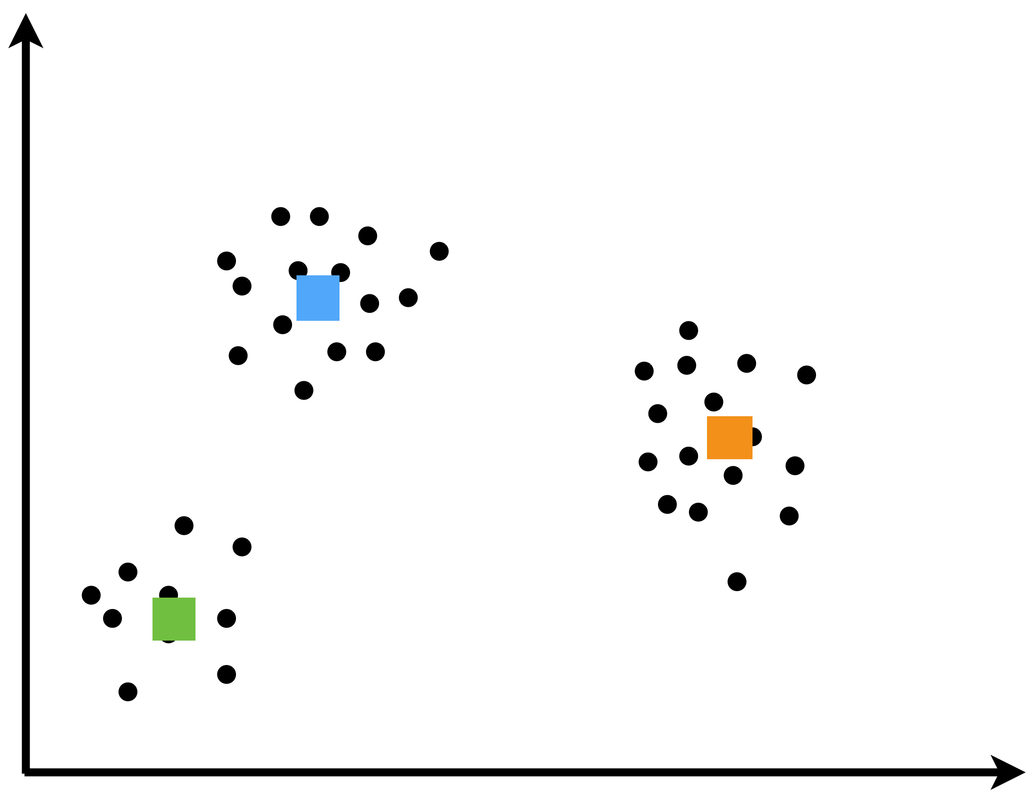
  - Finite (large *K*) mixture model

  - Time series



Calculating the posterior

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



ρ_1 ρ_2 ρ_3

- Finite Gaussian mixture model (*K* clusters)

$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_{1:K})$$

$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$