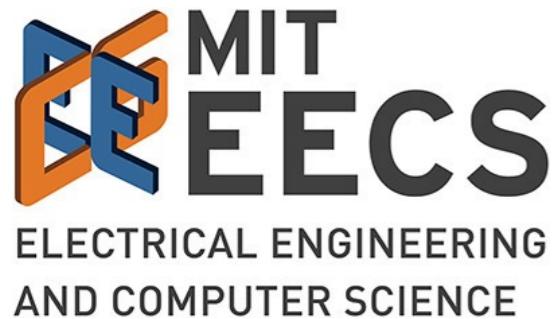


Nonparametric Bayes and Exchangeability

Tamara Broderick

Associate Professor
Electrical Engineering & Computer Science
MIT



Nonparametric Bayes and Exchangeability

Tamara Broderick

Associate Professor
Electrical Engineering & Computer Science
MIT

“goal of presenting the basic techniques, definitions and goals in
several of the communities”

Nonparametric Bayes

Nonparametric Bayes

- Bayesian

Nonparametric Bayes

- Bayesian

$\mathbb{P}(\text{parameters})$

Nonparametric Bayes

- Bayesian

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

Nonparametric Bayes

- Bayesian

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

Nonparametric Bayes

- Bayesian

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

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

Nonparametric Bayes

- Bayesian

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

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

WIKIPEDIA

The screenshot shows the main page of Wikipedia with language statistics and a search bar.

Language	Native Name	Description	Number of Articles
English	The Free Encyclopedia	4 853 000+ articles	
Deutsch	Die freie Enzyklopädie	1 806 000+ Artikel	
日本語	フリー百科事典	962 000+ 記事	
中文	自由的百科全書	814 000+ 條目	
Polski	Wolna encyklopedia	1 106 000+ hasł	
Español	La enciclopedia libre	1 172 000+ artículos	
Rусский	Свободная энциклопедия	1 213 000+ статей	
Français	L'encyclopédie libre	1 614 000+ articles	
Italiano	L'encyclopédia libera	1 193 000+ voci	
Português	A enciclopédia livre	871 000+ artigos	

[wikipedia.org]

Nonparametric Bayes

- Bayesian

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

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

WIKIPEDIA



“Wikipedia phenomenon”

[wikipedia.org]

Nonparametric Bayes

- Bayesian

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

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

WIKIPEDIA

The screenshot shows the main page of Wikipedia with language statistics for various editions:

Language	Native Name	Full Name	Articles
English	English	The Free Encyclopedia	4 853 000+ articles
Deutsch	Deutsch	Die freie Enzyklopädie	1 806 000+ Artikel
日本語	日本語	フリー百科事典	962 000+ 記事
中文	中文	自由的百科全書	814 000+ 條目
Polski	Polski	Wolna encyklopedia	1 106 000+ hasł
Español	Español	La enciclopedia libre	1 172 000+ artículos
Rусский	Русский	Свободная энциклопедия	1 213 000+ статей
Français	Français	L'encyclopédie libre	1 614 000+ articles
Italiano	Italiano	L'encyclopédia libera	1 193 000+ voci
Português	Português	A encyclopédia livre	871 000+ artigos

At the bottom, there is a search bar with the text "wikipedia.org".

[wikipedia.org]

Nonparametric Bayes

- Bayesian

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

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

WIKIPEDIA

English <i>The Free Encyclopedia</i> 4 853 000+ articles	Español <i>La enciclopedia libre</i> 1 172 000+ artículos	Rусский <i>Свободная энциклопедия</i> 1 213 000+ статей
Deutsch <i>Die freie Enzyklopädie</i> 1 806 000+ Artikel	Français <i>L'encyclopédie libre</i> 1 614 000+ articles	Italiano <i>L'encyclopédia libera</i> 1 193 000+ voci
日本語 <i>フリー百科事典</i> 962 000+ 記事	Português <i>A enciclopédia livre</i> 871 000+ artigos	
中文 <i>自由的百科全書</i> 814 000+ 條目		
Polski <i>Wolna encyklopedia</i> 1 106 000+ hasł		



[Ed Bowlby, NOAA]

Q English ▾ →

[wikipedia.org]

Nonparametric Bayes

- Bayesian

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

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

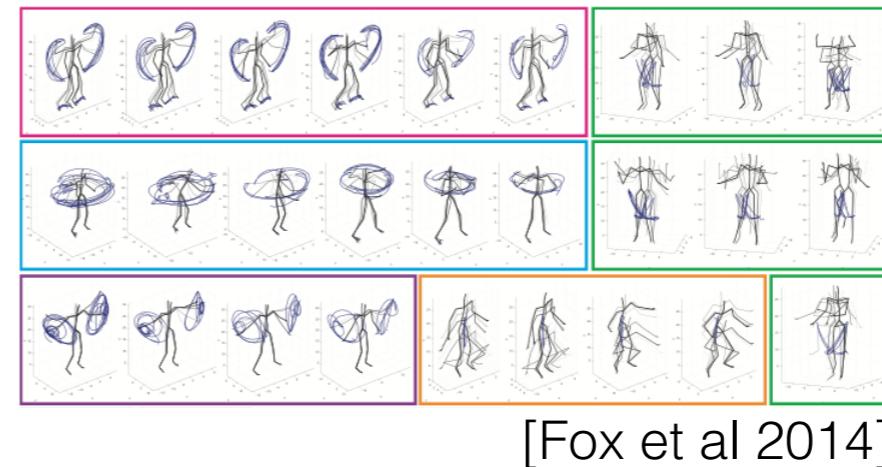
WIKIPEDIA

English The Free Encyclopedia 4 853 000+ articles	Español La enciclopedia libre 1 172 000+ artículos
Deutsch Die freie Enzyklopädie 1 806 000+ Artikel	Русский Свободная энциклопедия 1 213 000+ статей
日本語 フリー百科事典 962 000+ 記事	Français L'encyclopédie libre 1 614 000+ articles
中文 自由的百科全書 814 000+ 條目	Italiano L'encyclopédia libera 1 193 000+ voci
Polski Wolna encyklopedia 1 106 000+ hasł	Português A enciclopédia livre 871 000+ artigos

[wikipedia.org]



[Ed Bowlby, NOAA]

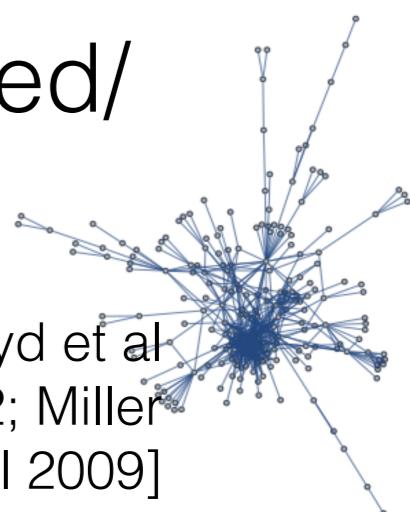


Nonparametric Bayes

- Bayesian

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

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



[Lloyd et al
2012; Miller
et al 2009]

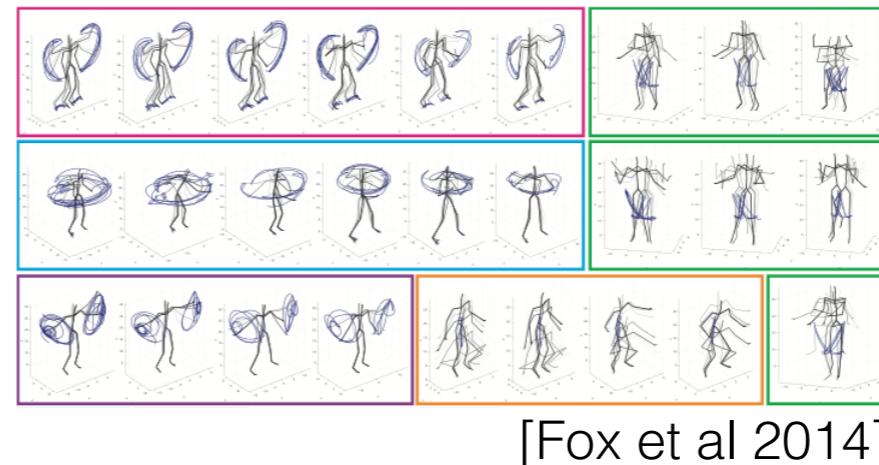
WIKIPEDIA

English <i>The Free Encyclopedia</i> 4 853 000+ articles	Español <i>La enciclopedia libre</i> 1 172 000+ artículos
Deutsch <i>Die freie Enzyklopädie</i> 1 806 000+ Artikel	Русский <i>Свободная энциклопедия</i> 1 213 000+ статей
日本語 <i>フリー百科事典</i> 962 000+記事	Français <i>L'encyclopédie libre</i> 1 614 000+ articles
中文 <i>自由的百科全書</i> 814 000+條目	Italiano <i>L'encyclopédia libera</i> 1 193 000+ voci
Polski <i>Wolna encyklopedia</i> 1 106 000+ haset	Português <i>A enciclopédia livre</i> 871 000+ artigos

English



[Ed Bowlby, NOAA]



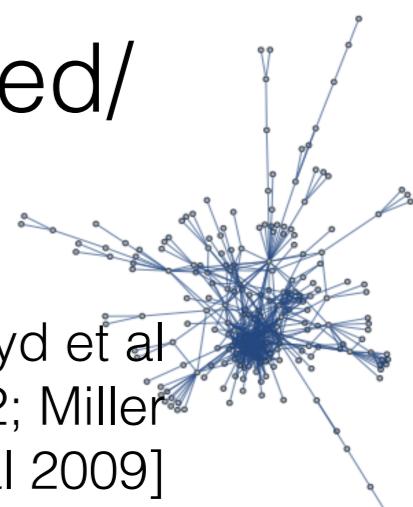
[wikipedia.org]

Nonparametric Bayes

- Bayesian

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

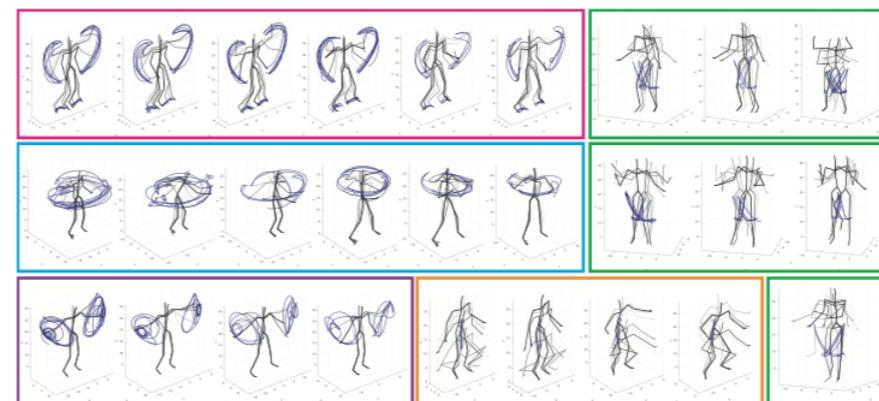
- Not parametric (i.e. not finite parameter, unbounded/growing/infinite number of parameters)



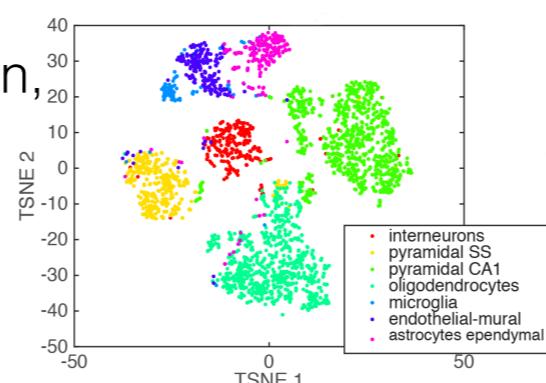
[Lloyd et al
2012; Miller
et al 2009]



[Ed Bowlby, NOAA]
[Prabhakaran,
Azizi, Carr,
Pe'er 2016]



[Fox et al 2014]

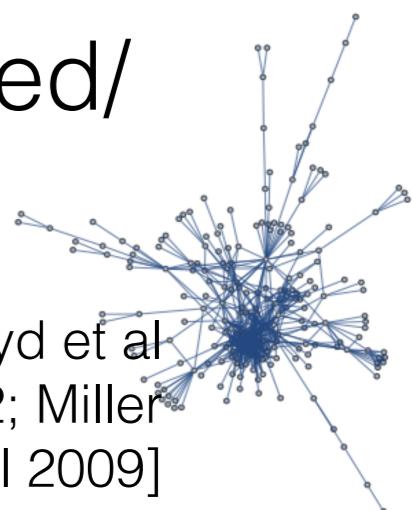


Nonparametric Bayes

- Bayesian

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

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

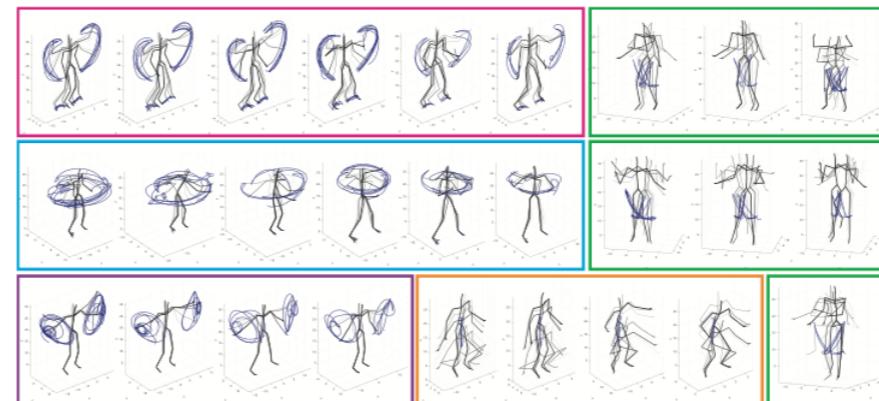


[Lloyd et al
2012; Miller
et al 2009]

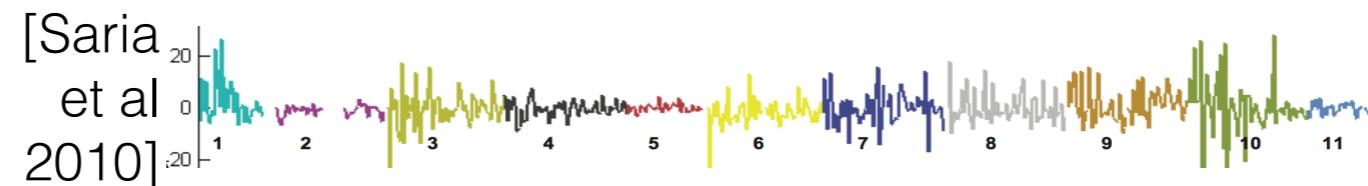
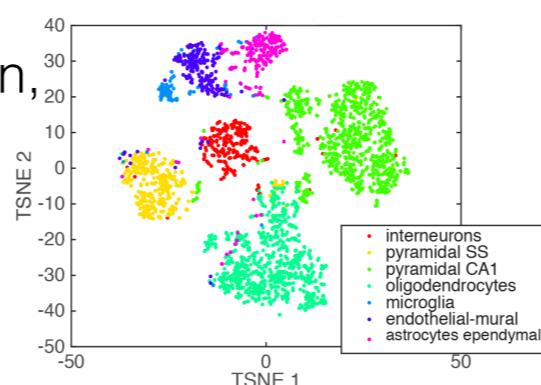


[Ed Bowlby, NOAA]

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



[Fox et al 2014]

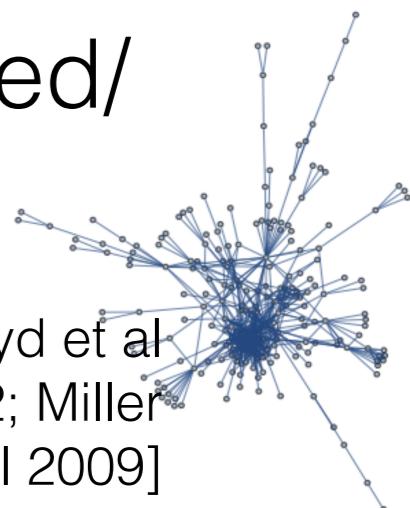


Nonparametric Bayes

- Bayesian

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

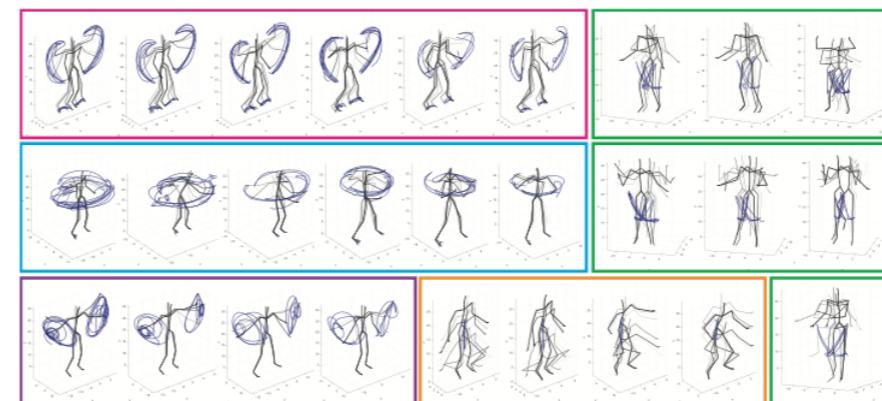
- Not parametric (i.e. not finite parameter, unbounded/growing/infinite number of parameters)



[Lloyd et al
2012; Miller
et al 2009]

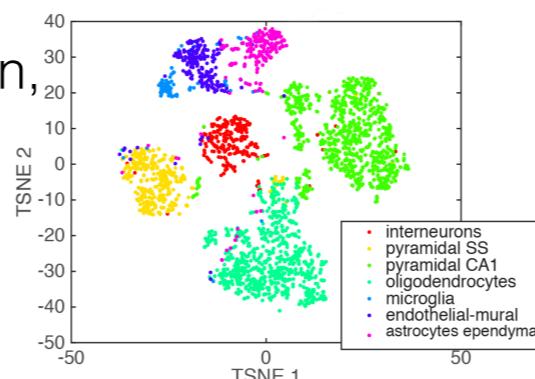


[Ed Bowlby, NOAA]

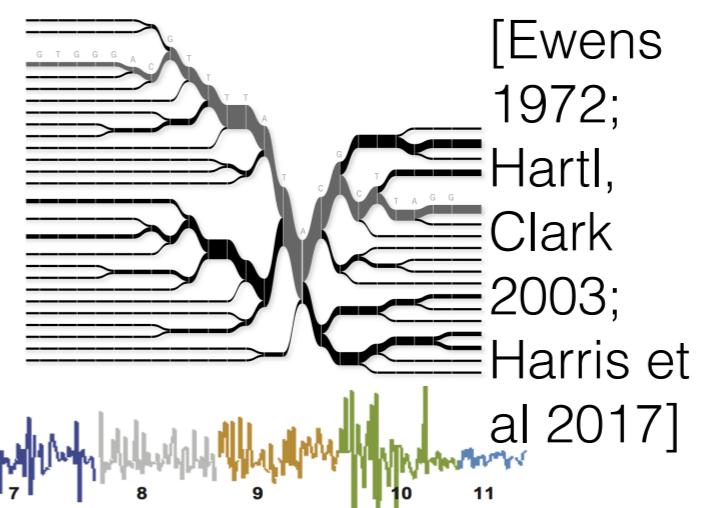


[Fox et al 2014]

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



[Saria
et al
2010]

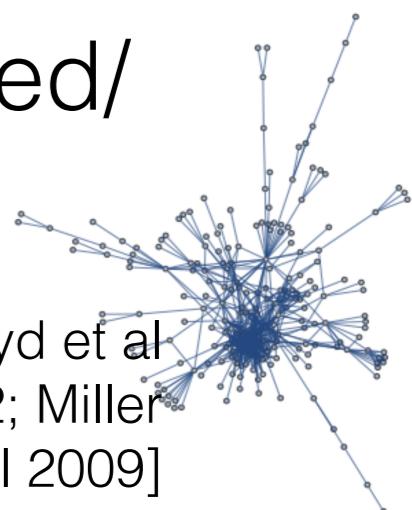


Nonparametric Bayes

- Bayesian

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

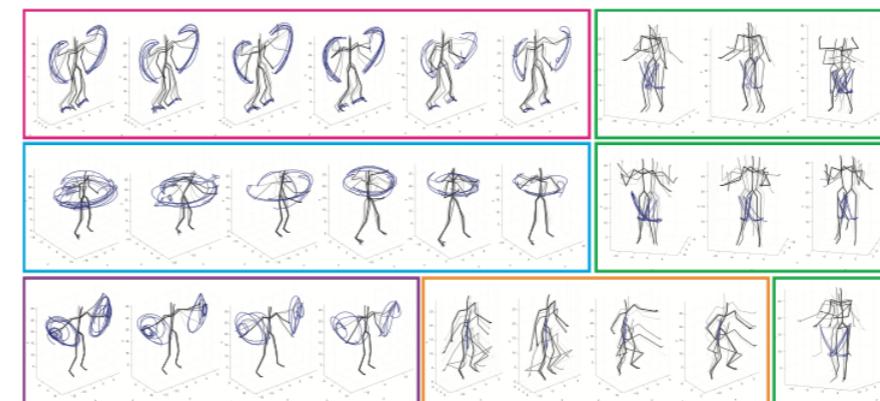
- Not parametric (i.e. not finite parameter, unbounded/growing/infinite number of parameters)



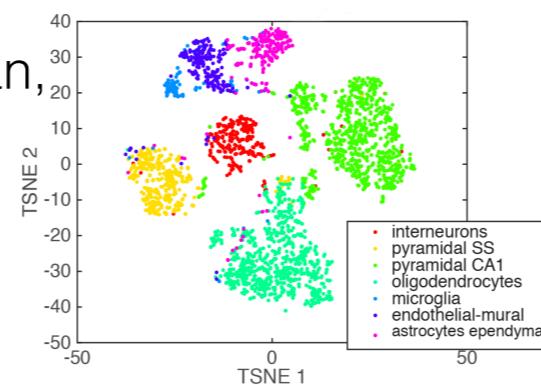
[Lloyd et al
2012; Miller
et al 2009]



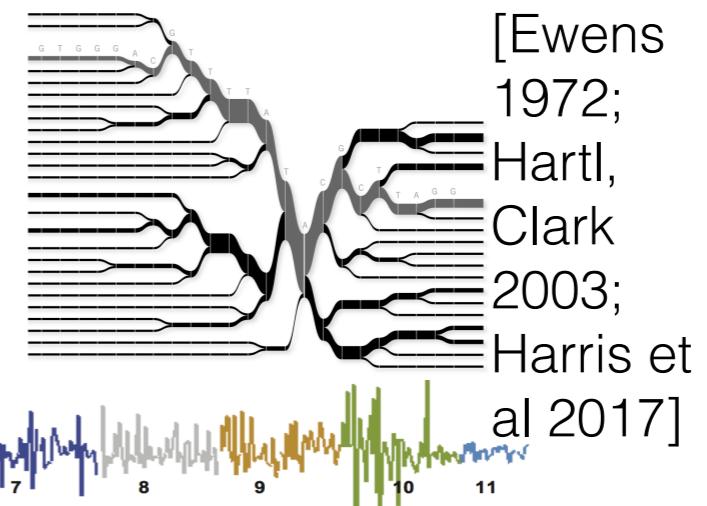
[Ed Bowlby, NOAA]



[Fox et al 2014]



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



[Ewens
1972;
Hartl,
Clark
2003;
Harris et
al 2017]



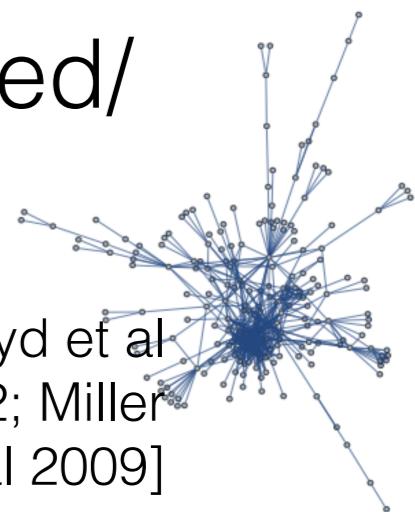
[Saria
et al
2010]

Nonparametric Bayes

- Bayesian

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

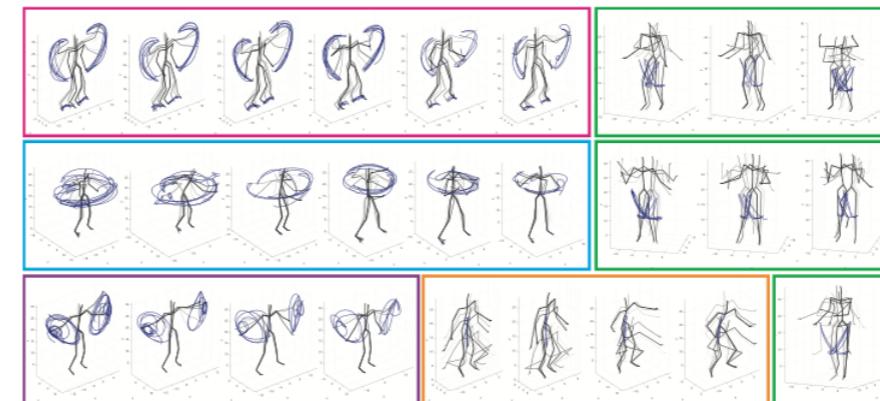
- Not parametric (i.e. not finite parameter, unbounded/growing/infinite number of parameters)



[Lloyd et al
2012; Miller
et al 2009]



[Ed Bowlby, NOAA]

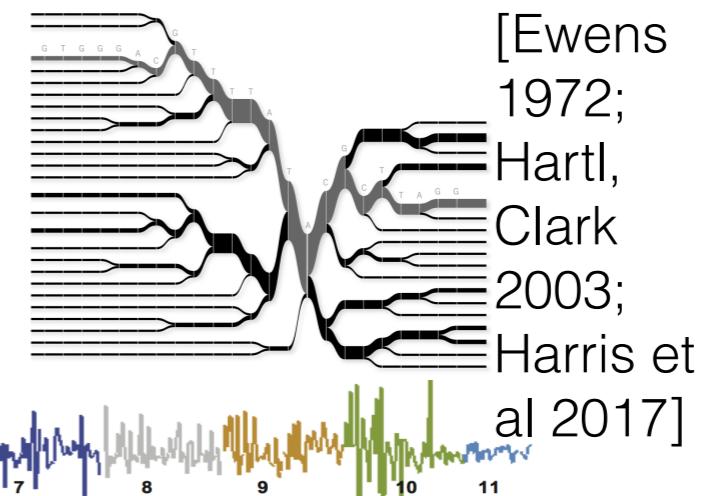
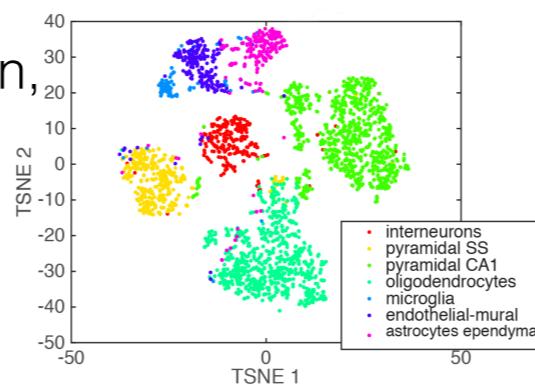


[Fox et al 2014]

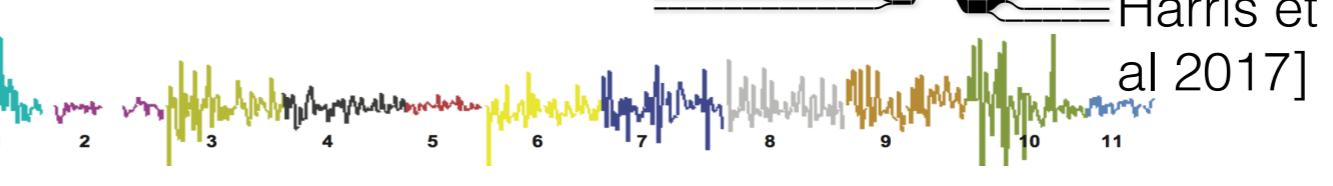


[Lan et al 2015]

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



[Saria
et al
2010]

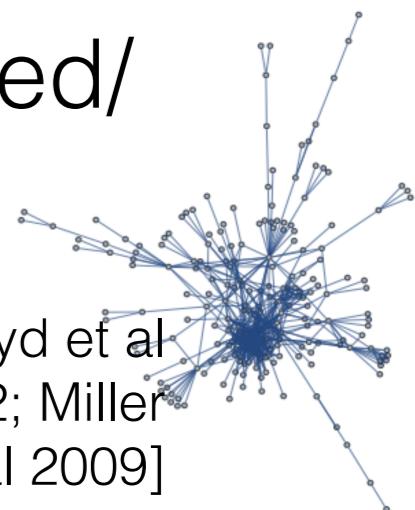


Nonparametric Bayes

- Bayesian

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

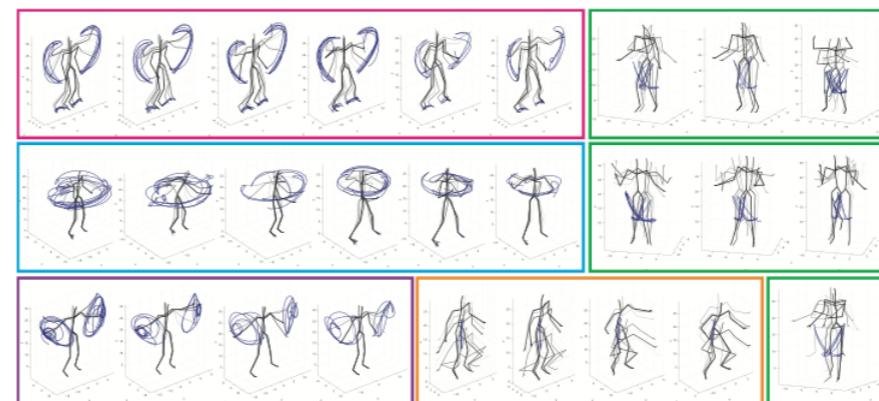
- Not parametric (i.e. not finite parameter, unbounded/growing/infinite number of parameters)



[Lloyd et al
2012; Miller
et al 2009]



[Ed Bowlby, NOAA]

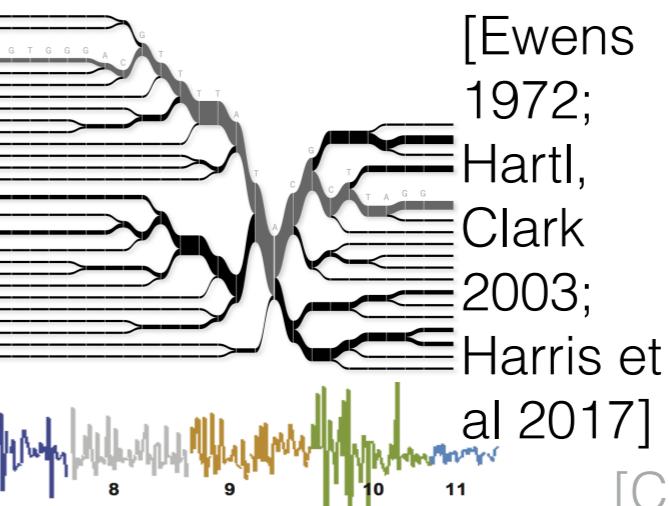
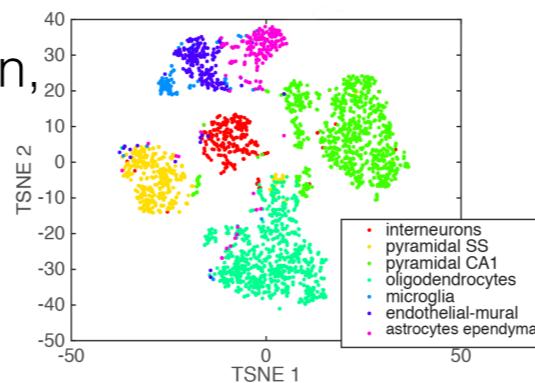


[Fox et al 2014]

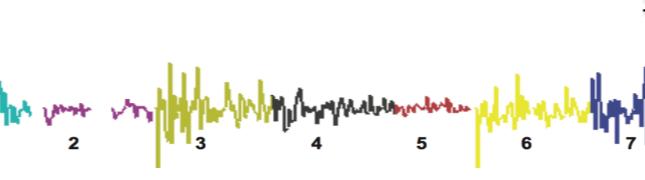


[Lan et al 2015]

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



[Saria
et al
2010]



[Xu et al 2015]
[Cassidy et al 2015]

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem
- A data sequence is *infinitely exchangeable* if the distribution of any N data points doesn't change when permuted: $p(X_1, \dots, X_N) = p(X_{\sigma(1)}, \dots, X_{\sigma(N)})$

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem
- A data sequence is *infinitely exchangeable* if the distribution of any N data points doesn't change when permuted: $p(X_1, \dots, X_N) = p(X_{\sigma(1)}, \dots, X_{\sigma(N)})$
- *De Finetti's Theorem* (roughly): A sequence X_1, X_2, \dots is infinitely exchangeable if and only if, for all N and some distribution P :

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem
- A data sequence is *infinitely exchangeable* if the distribution of any N data points doesn't change when permuted: $p(X_1, \dots, X_N) = p(X_{\sigma(1)}, \dots, X_{\sigma(N)})$
- *De Finetti's Theorem* (roughly): A sequence X_1, X_2, \dots is infinitely exchangeable if and only if, for all N and some distribution P :

$$p(X_1, \dots, X_N) = \int_{\theta} \prod_{n=1}^N p(X_n | \theta) P(d\theta)$$

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem
- A data sequence is *infinitely exchangeable* if the distribution of any N data points doesn't change when permuted: $p(X_1, \dots, X_N) = p(X_{\sigma(1)}, \dots, X_{\sigma(N)})$
- *De Finetti's Theorem* (roughly): A sequence X_1, X_2, \dots is infinitely exchangeable if and only if, for all N and some distribution P :
$$p(X_1, \dots, X_N) = \int_{\theta} \prod_{n=1}^N p(X_n | \theta) P(d\theta)$$
- Motivates:

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem
- A data sequence is *infinitely exchangeable* if the distribution of any N data points doesn't change when permuted: $p(X_1, \dots, X_N) = p(X_{\sigma(1)}, \dots, X_{\sigma(N)})$
- *De Finetti's Theorem* (roughly): A sequence X_1, X_2, \dots is infinitely exchangeable if and only if, for all N and some distribution P :
$$p(X_1, \dots, X_N) = \int_{\theta} \prod_{n=1}^N p(X_n | \theta) P(d\theta)$$
- Motivates:
 - Parameters and likelihoods

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem
- A data sequence is *infinitely exchangeable* if the distribution of any N data points doesn't change when permuted: $p(X_1, \dots, X_N) = p(X_{\sigma(1)}, \dots, X_{\sigma(N)})$
- *De Finetti's Theorem* (roughly): A sequence X_1, X_2, \dots is infinitely exchangeable if and only if, for all N and some distribution P :
$$p(X_1, \dots, X_N) = \int_{\theta} \prod_{n=1}^N p(X_n | \theta) P(d\theta)$$
- Motivates:
 - Parameters and likelihoods
 - Priors

Nonparametric Bayes

- A theoretical motivation: De Finetti's Theorem
- A data sequence is *infinitely exchangeable* if the distribution of any N data points doesn't change when permuted: $p(X_1, \dots, X_N) = p(X_{\sigma(1)}, \dots, X_{\sigma(N)})$
- *De Finetti's Theorem* (roughly): A sequence X_1, X_2, \dots is infinitely exchangeable if and only if, for all N and some distribution P :
$$p(X_1, \dots, X_N) = \int_{\theta} \prod_{n=1}^N p(X_n | \theta) P(d\theta)$$
- Motivates:
 - Parameters and likelihoods
 - Priors
 - “Nonparametric Bayesian” priors

Roadmap

Roadmap

- Example problem: clustering

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- Big questions

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- Big questions
 - Why NPBayes?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- Big questions
 - Why NPBayes?
 - What does an infinite/growing number of parameters really mean (in NPBayes)?

Roadmap

- Example problem: clustering
- Example NPBayes model: Dirichlet process (DP)
- Big questions
 - Why NPBayes?
 - What does an infinite/growing number of parameters really mean (in NPBayes)?
 - Why is NPBayes challenging but practical?

Roadmap

- Example problem: clustering
- Example NPYbayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- Big questions
 - Why NPYbayes?
 - What does an infinite/growing number of parameters really mean (in NPYbayes)?
 - Why is NPYbayes challenging but practical?

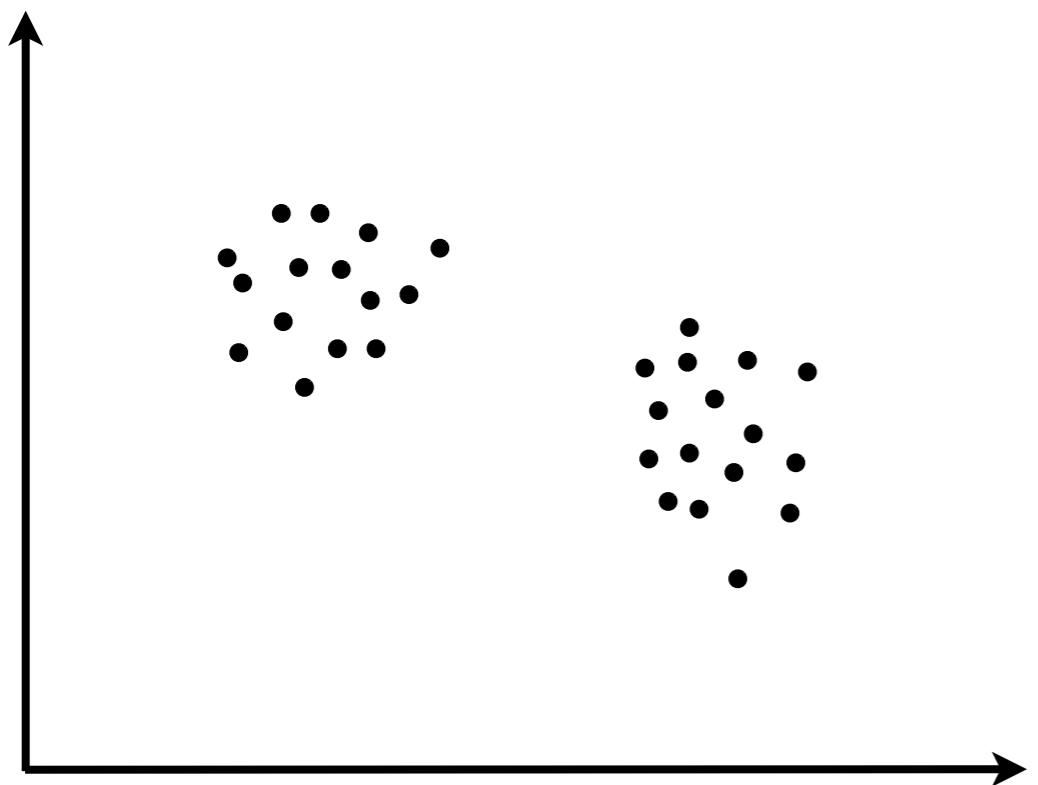
Roadmap

- Example problem: clustering
- Example NPYBayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs
- Big questions
 - Why NPYBayes?
 - What does an infinite/growing number of parameters really mean (in NPYBayes)?
 - Why is NPYBayes challenging but practical?

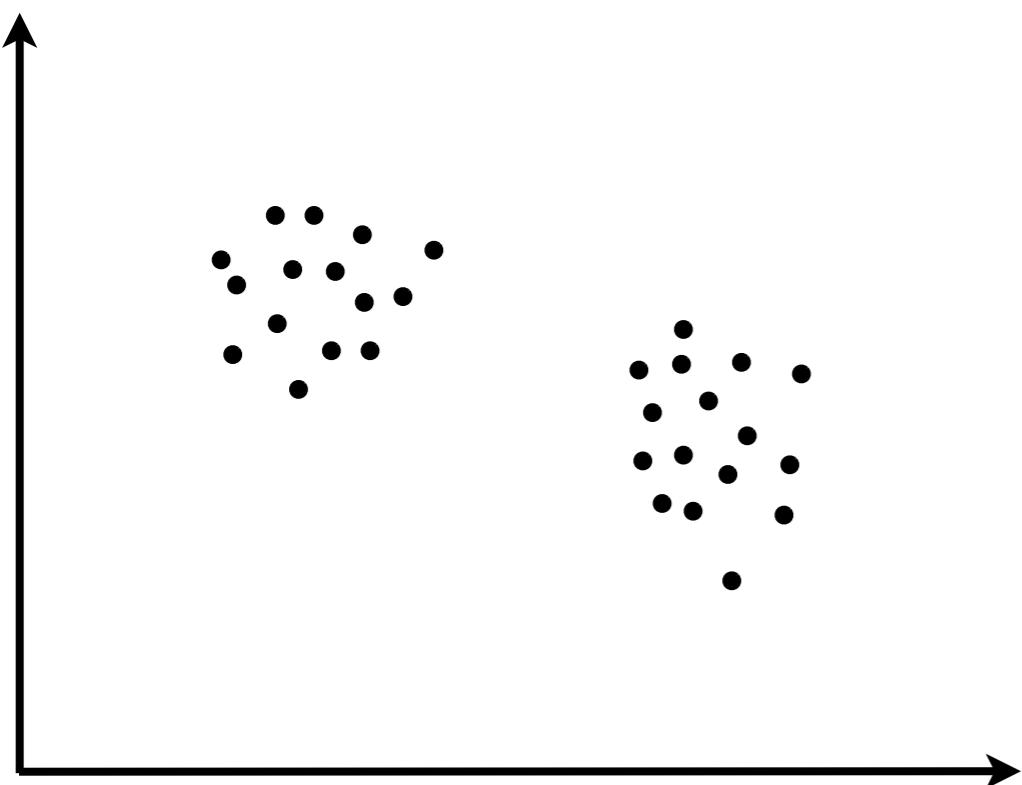
Roadmap

- Example problem: clustering
- Example NPYbayes model: Dirichlet process (DP)
- De Finetti for clustering: Kingman Paintbox
- De Finetti for networks/graphs
- Big questions
 - Why NPYbayes?
 - What does an infinite/growing number of parameters really mean (in NPYbayes)?
 - Why is NPYbayes challenging but practical?
 - How does thinking about exchangeability help us use NPYbayes in practice?

Generative model



Generative model

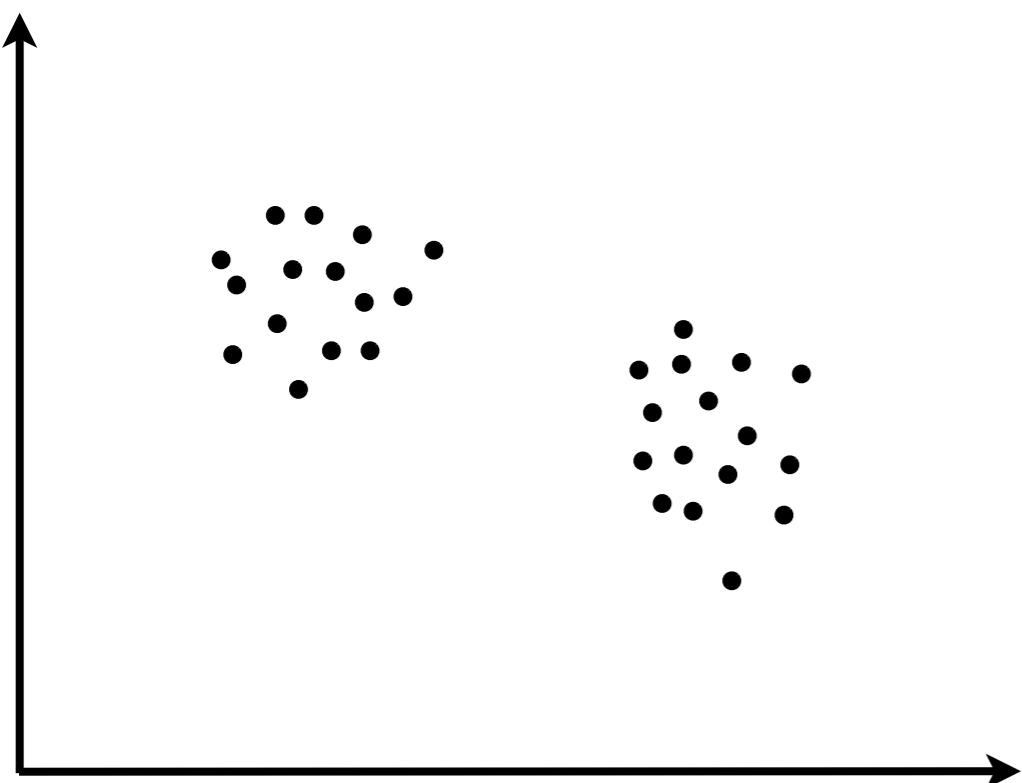


- Finite Gaussian mixture model ($K=2$ clusters)

Generative model

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

- Finite Gaussian mixture model ($K=2$ clusters)

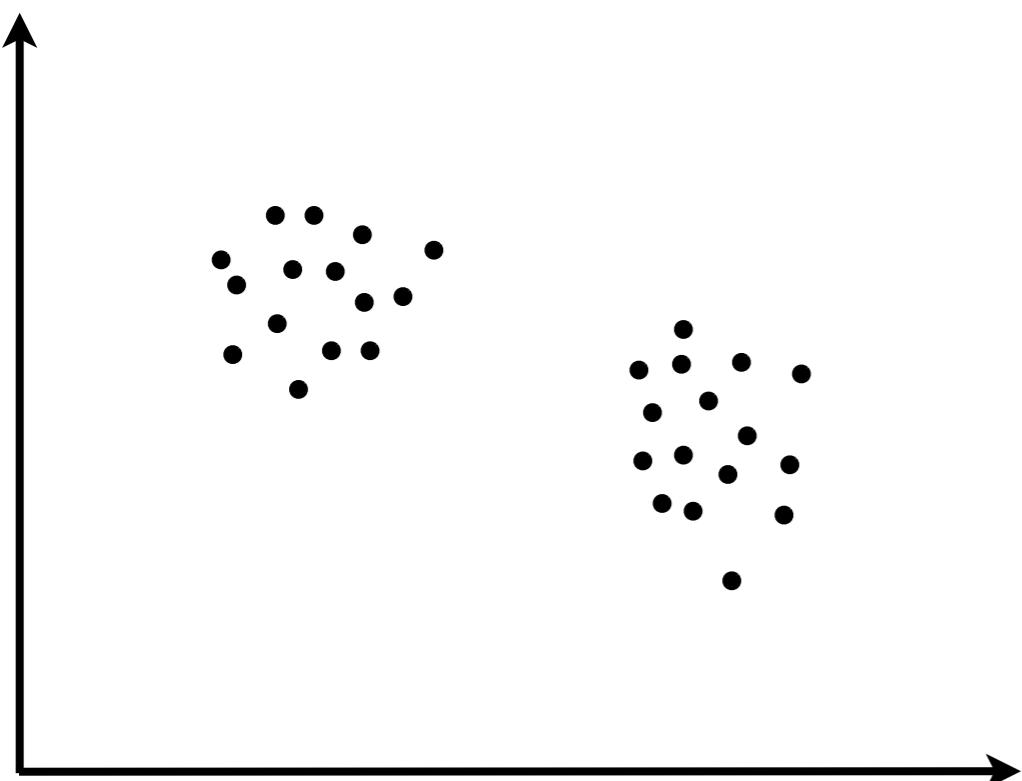


Generative model

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

- Finite Gaussian mixture model ($K=2$ clusters)

$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$

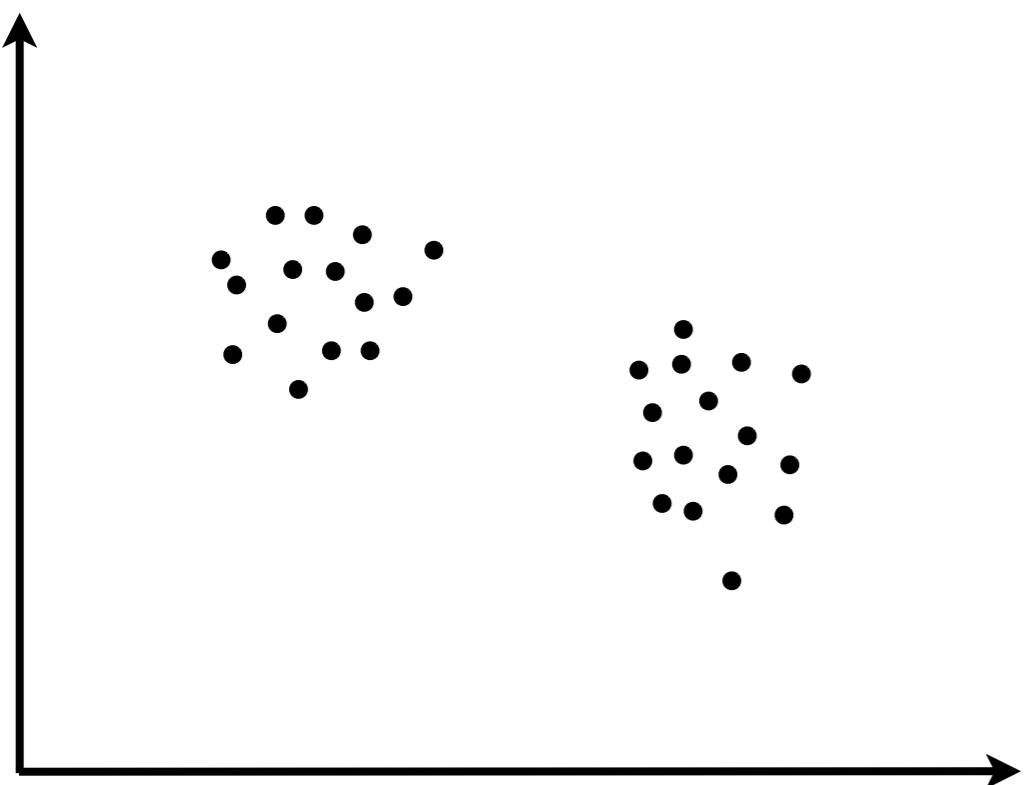


Generative model

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

- Finite Gaussian mixture model ($K=2$ clusters)

$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$



ρ_1

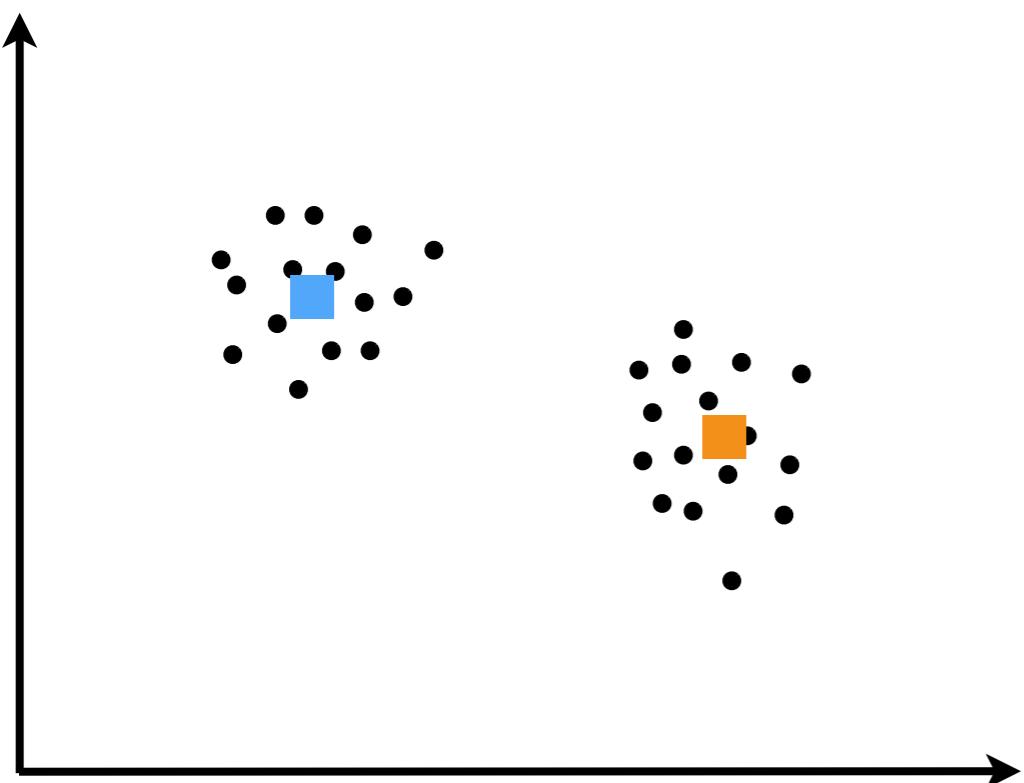
ρ_2

Generative model

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

- Finite Gaussian mixture model ($K=2$ clusters)

$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$



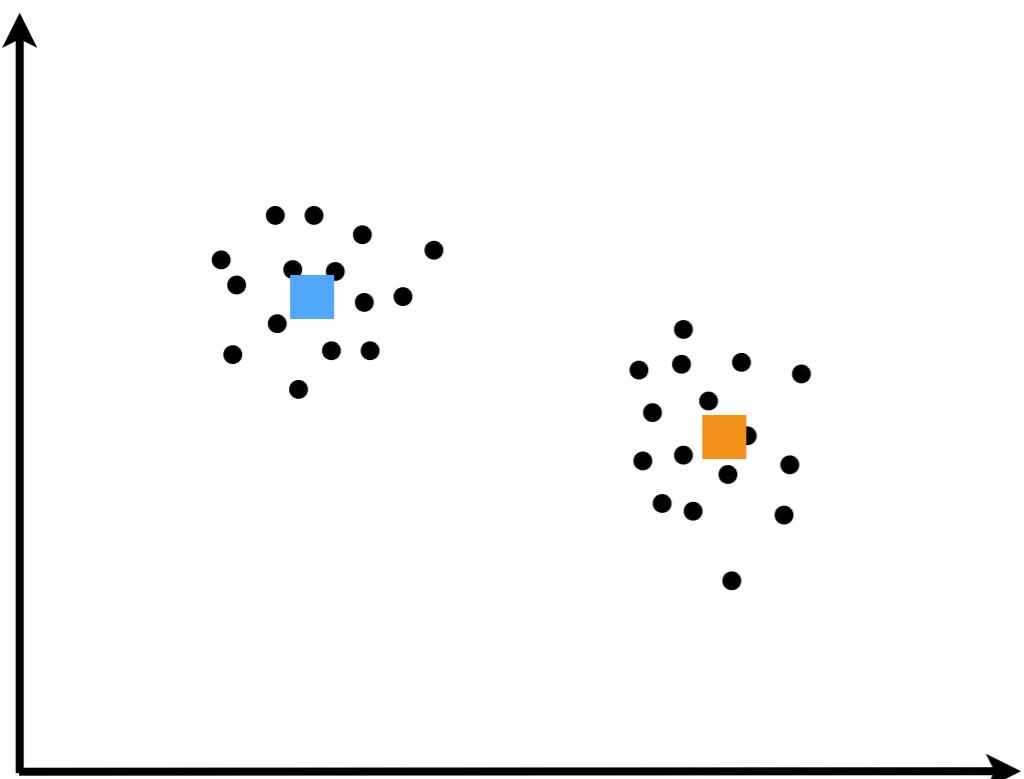
ρ_1

ρ_2

Generative model

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

- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2



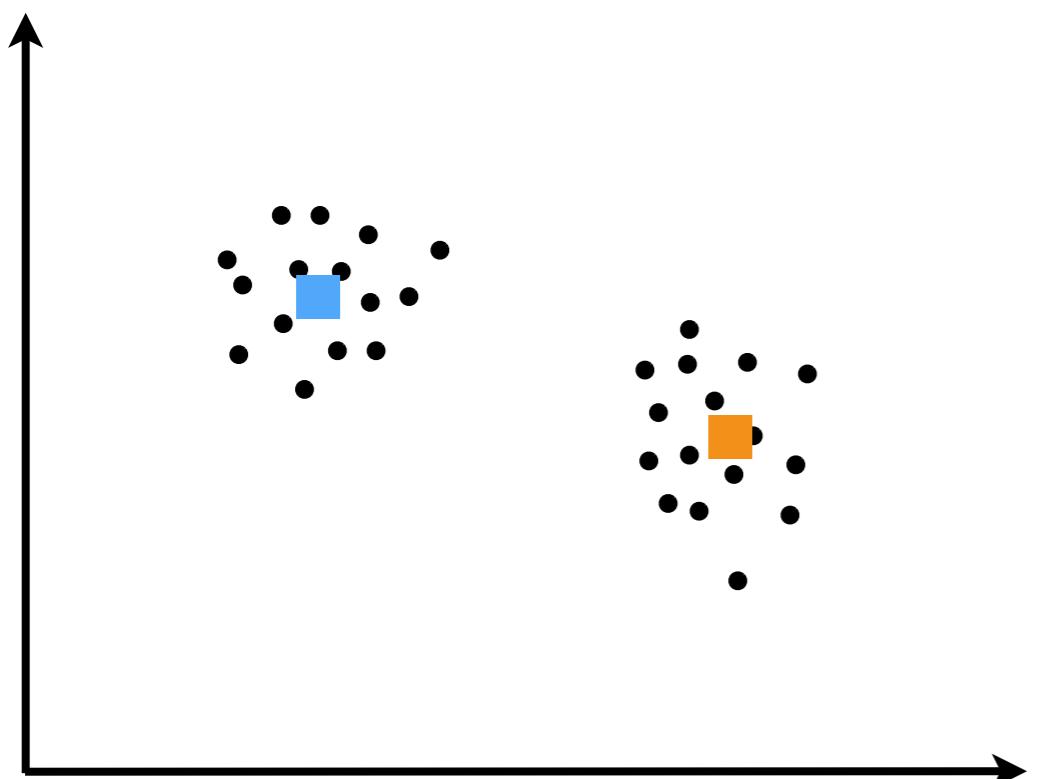
ρ_1

ρ_2

Generative model

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

- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$



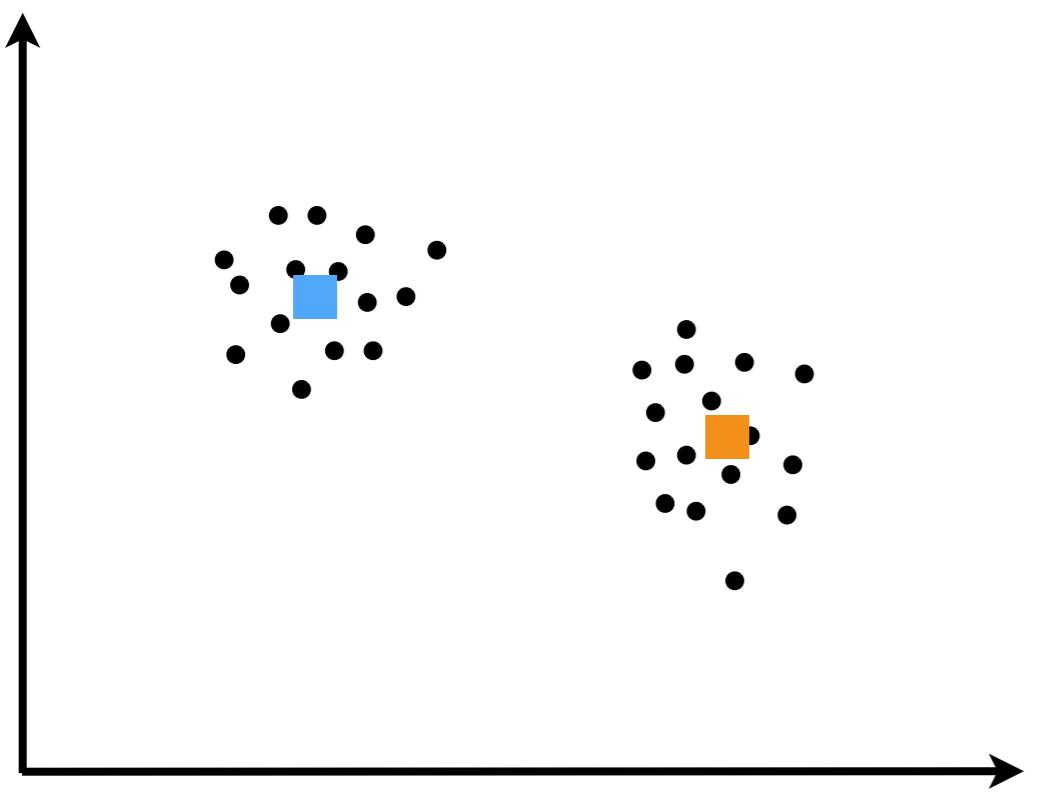
ρ_1

ρ_2

Generative model

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

- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2



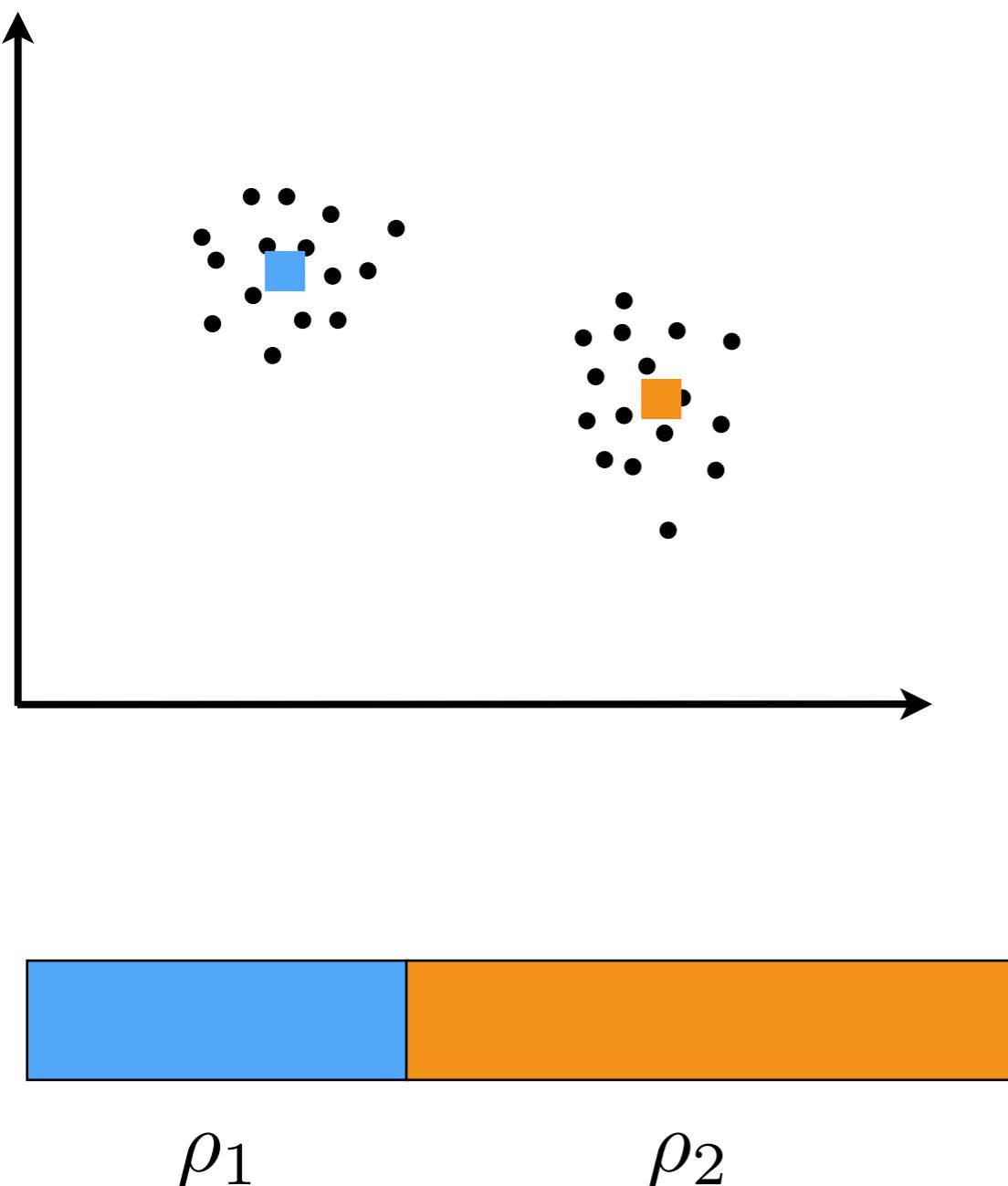
ρ_1

ρ_2

Generative model

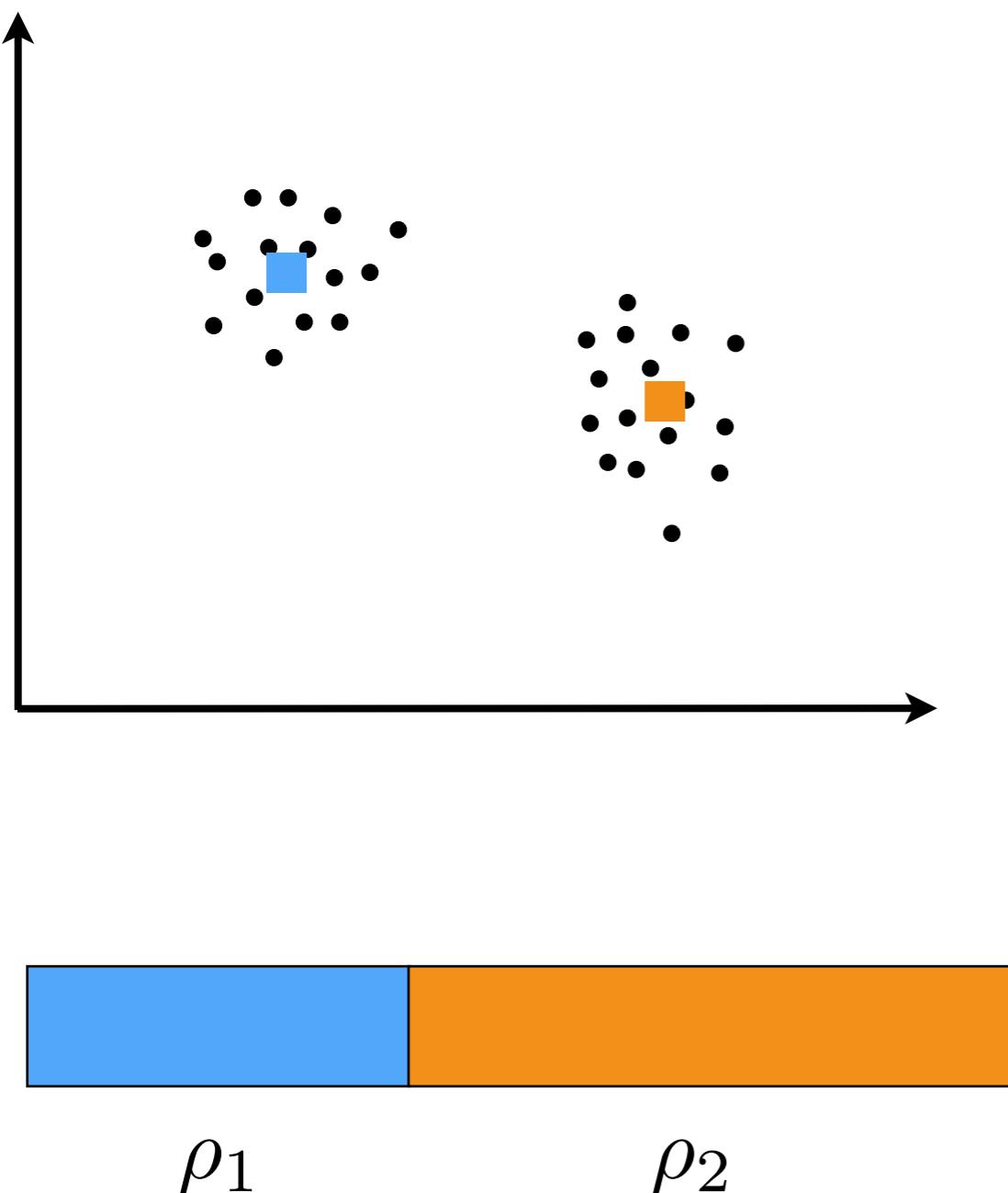
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2
$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$



Generative model

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



- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2
$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$
- Inference goal: assignments of data points to clusters, cluster parameters

Generative model

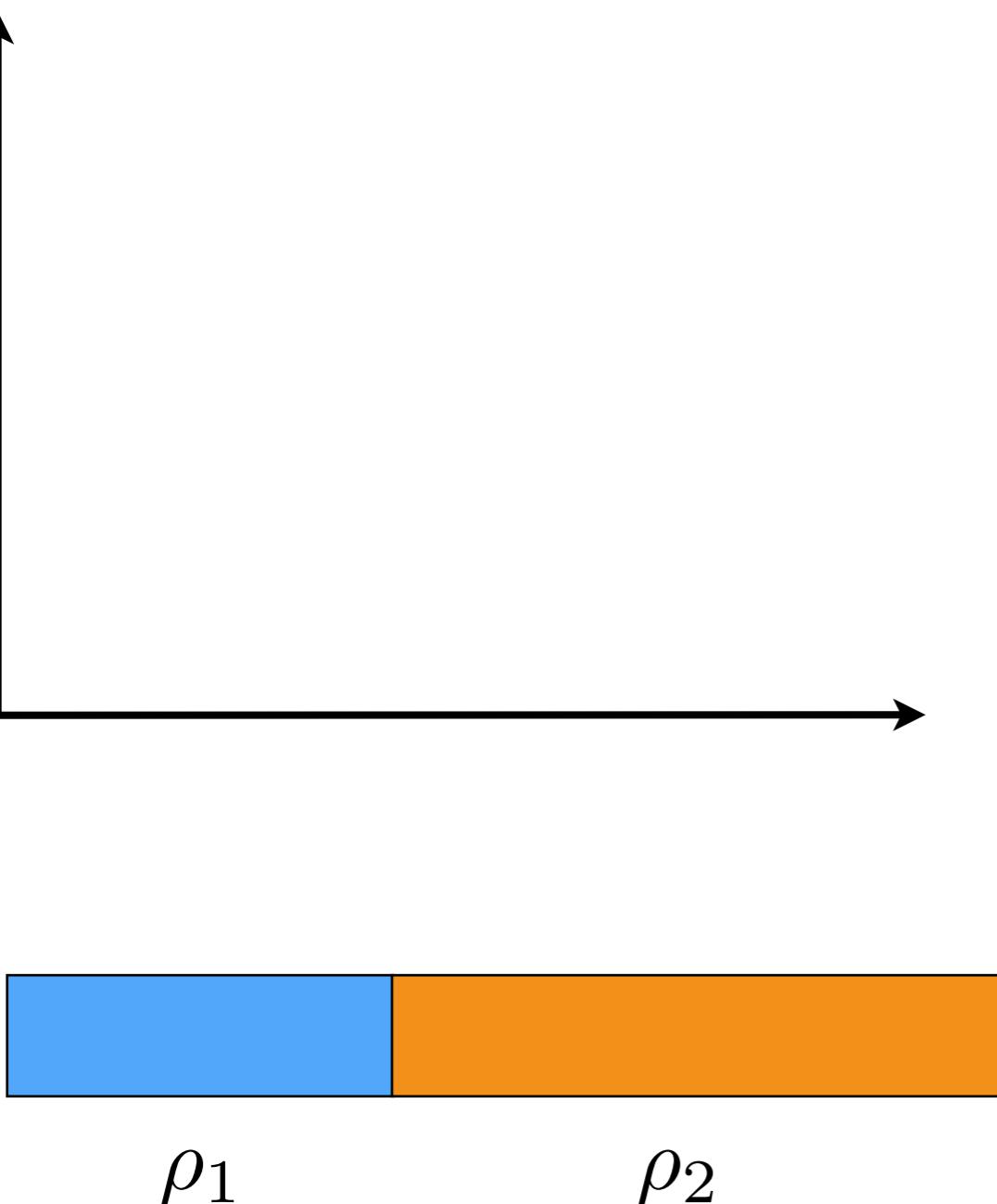
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2
$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$
- Inference goal: assignments of data points to clusters, cluster parameters



Generative model

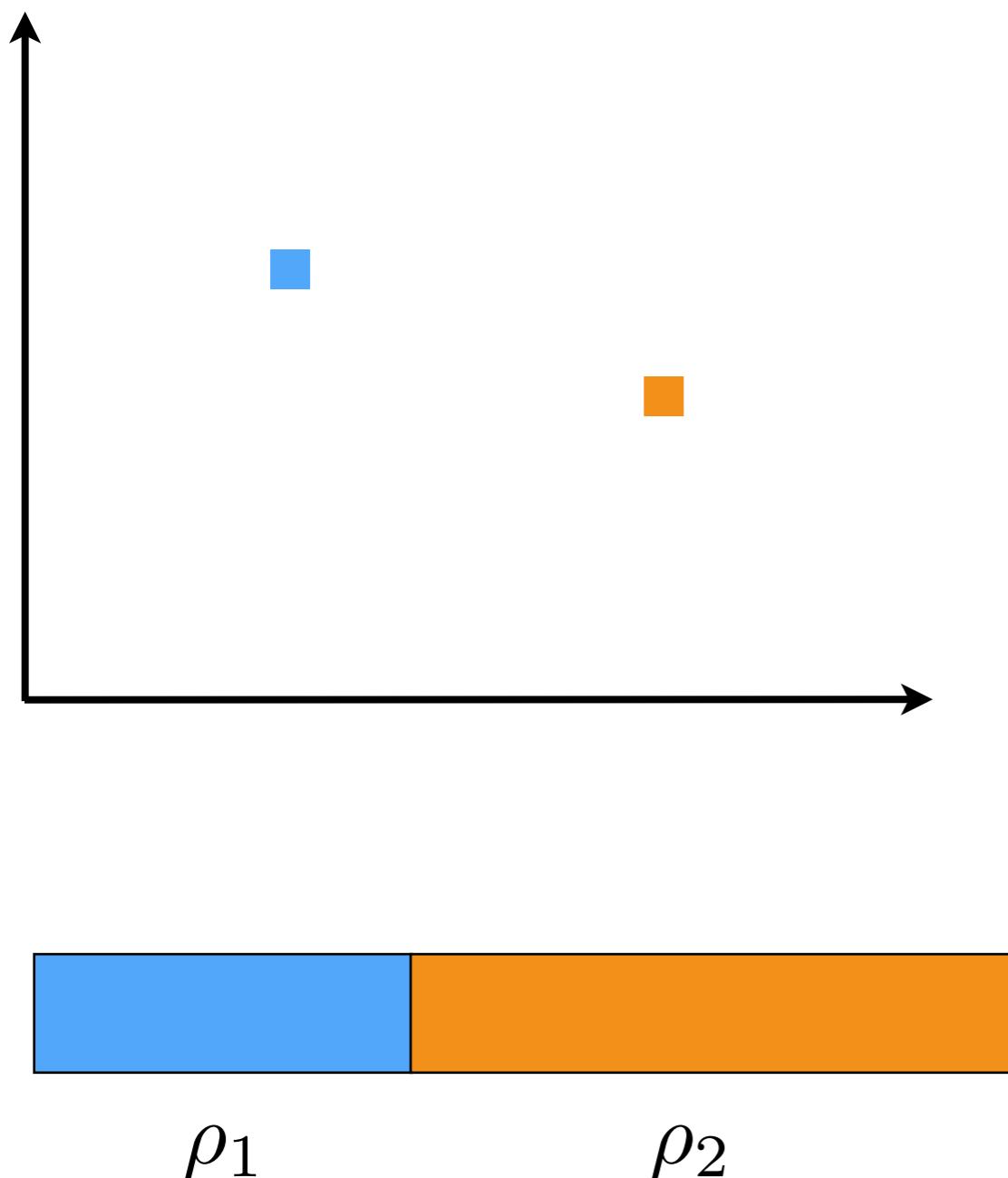
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$



- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2
$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$
- Inference goal: assignments of data points to clusters, cluster parameters

Generative model

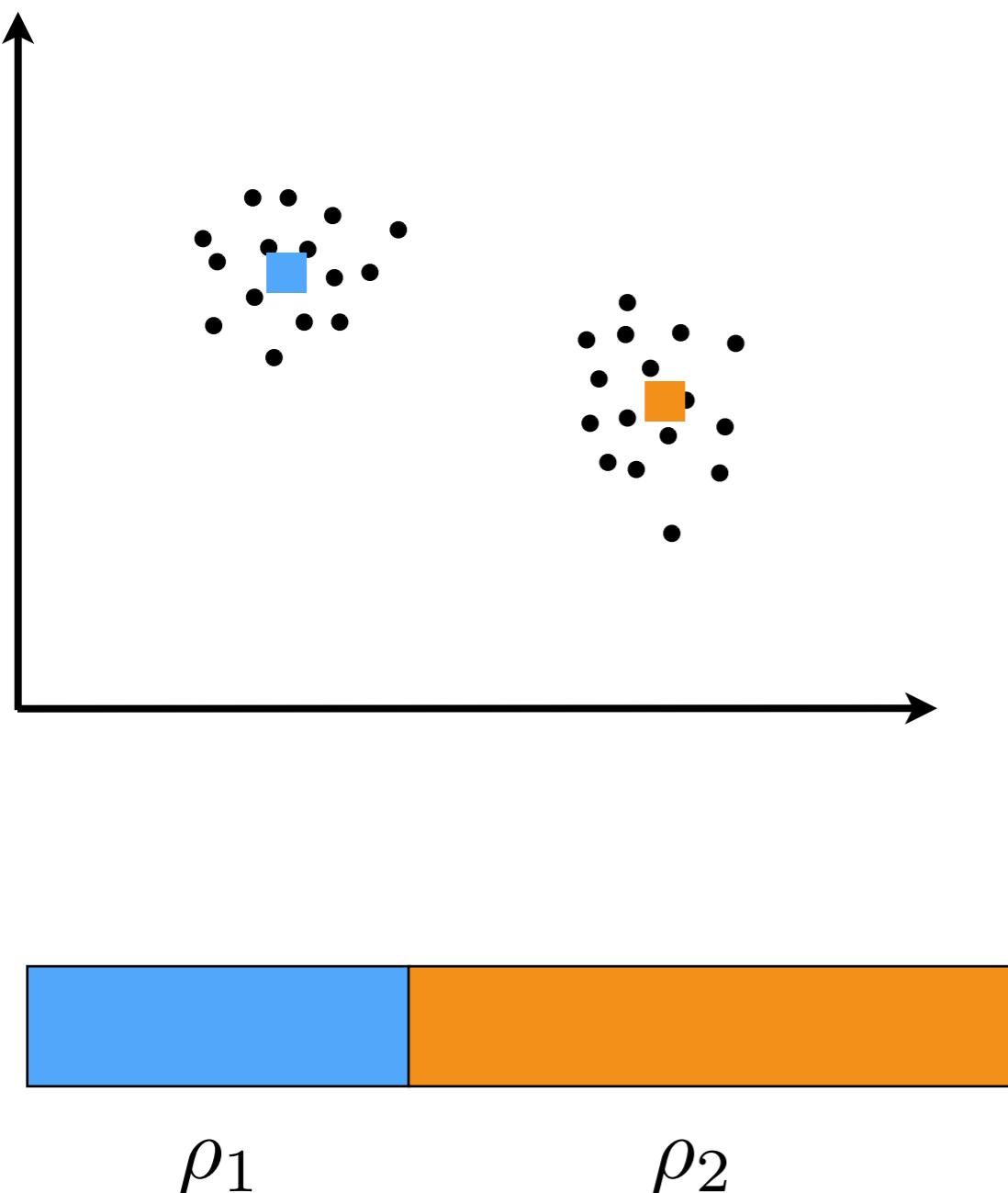
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$



- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2
$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$
- Inference goal: assignments of data points to clusters, cluster parameters

Generative model

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



- Finite Gaussian mixture model ($K=2$ clusters)
$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$x_n \stackrel{\text{indep}}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know μ_1, μ_2
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2
$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$
- Inference goal: assignments of data points to clusters, cluster parameters

Beta distribution review

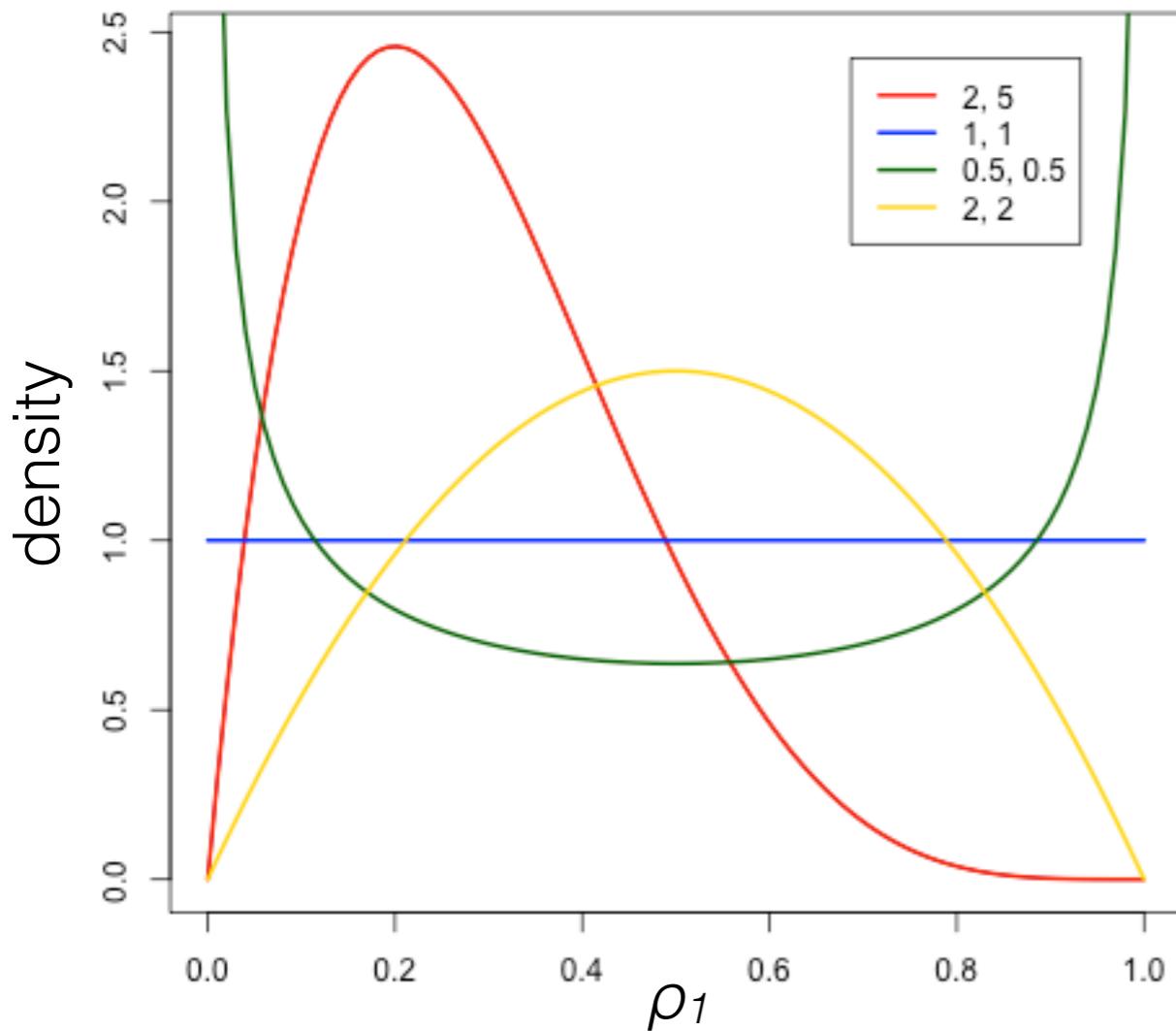
$$\text{Beta}(\rho_1 | a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \rho_1^{a_1 - 1} (1 - \rho_1)^{a_2 - 1}$$

$\rho_1 \in (0, 1)$
 $a_1, a_2 > 0$

Beta distribution review

$$\text{Beta}(\rho_1 | a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \rho_1^{a_1-1} (1 - \rho_1)^{a_2-1}$$

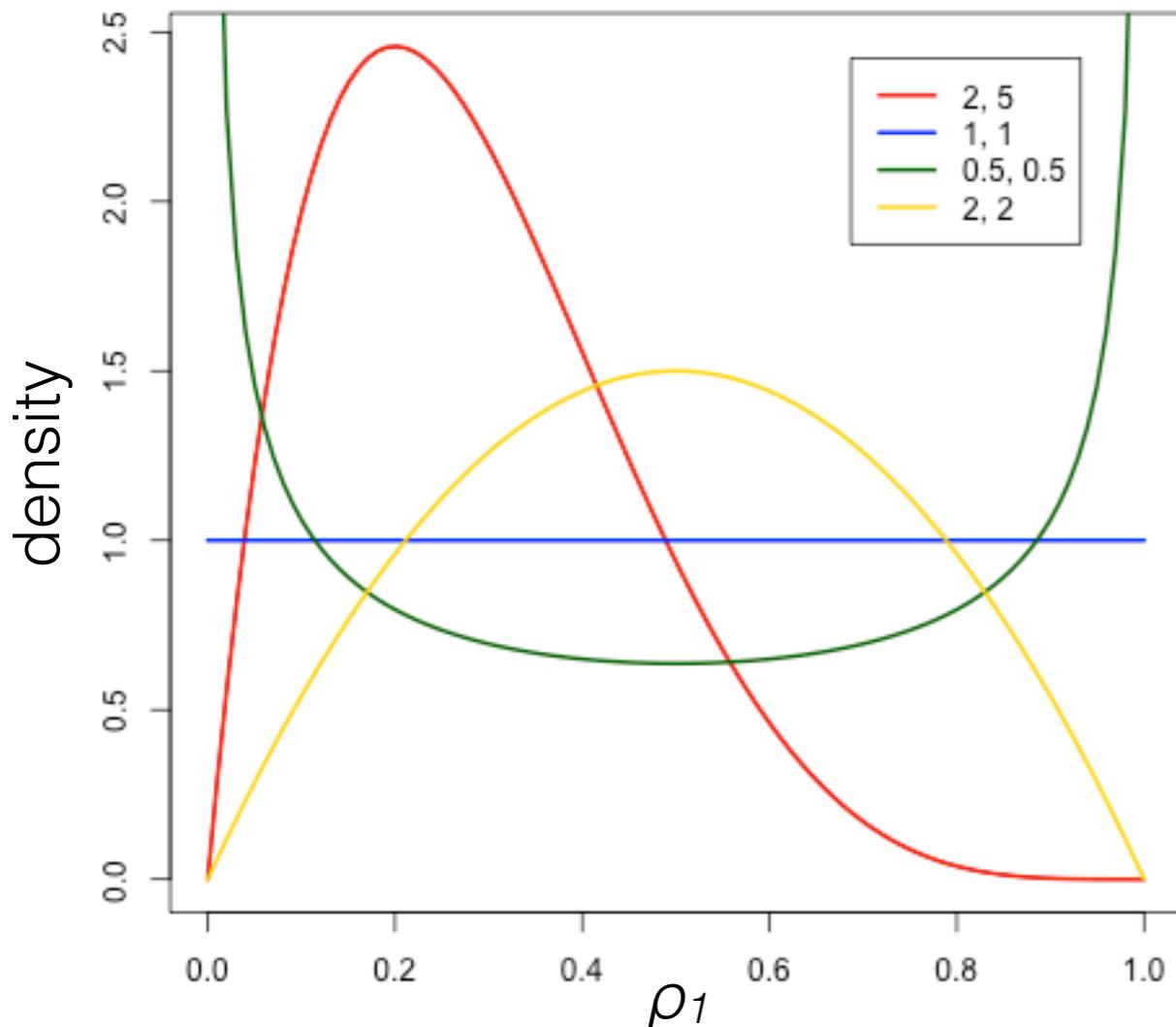
$$\begin{aligned}\rho_1 &\in (0, 1) \\ a_1, a_2 &> 0\end{aligned}$$



Beta distribution review

$$\text{Beta}(\rho_1 | a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \rho_1^{a_1-1} (1 - \rho_1)^{a_2-1}$$

$$\begin{aligned}\rho_1 &\in (0, 1) \\ a_1, a_2 &> 0\end{aligned}$$

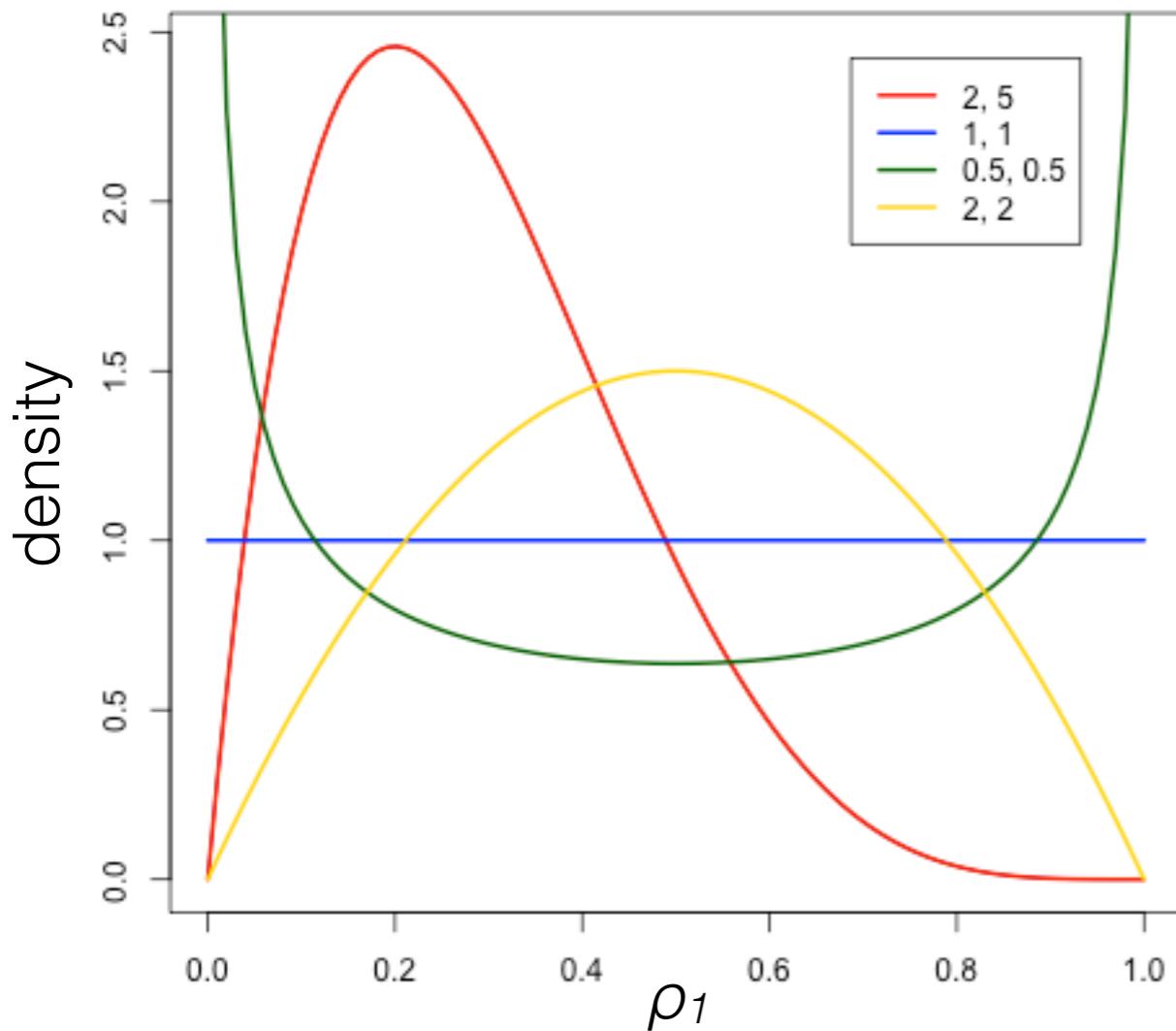


- What happens?

Beta distribution review

$$\text{Beta}(\rho_1 | a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \rho_1^{a_1-1} (1 - \rho_1)^{a_2-1}$$

$$\begin{aligned}\rho_1 &\in (0, 1) \\ a_1, a_2 &> 0\end{aligned}$$

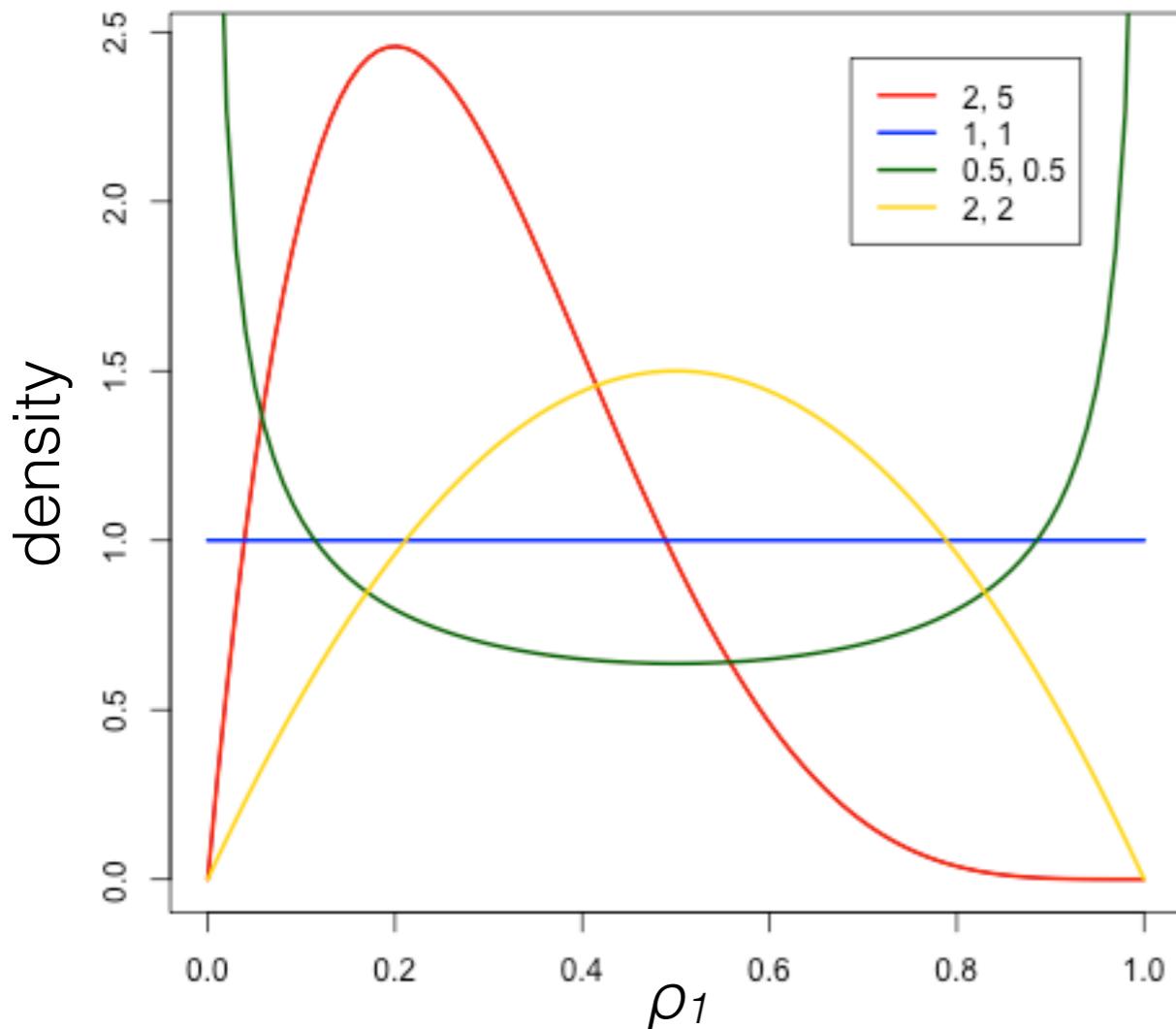


- What happens?
 - $a = a_1 = a_2 \rightarrow 0$
 - $a = a_1 = a_2 \rightarrow \infty$
 - $a_1 > a_2$

Beta distribution review

$$\text{Beta}(\rho_1 | a_1, a_2) = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \rho_1^{a_1-1} (1 - \rho_1)^{a_2-1}$$

$$\begin{aligned}\rho_1 &\in (0, 1) \\ a_1, a_2 &> 0\end{aligned}$$

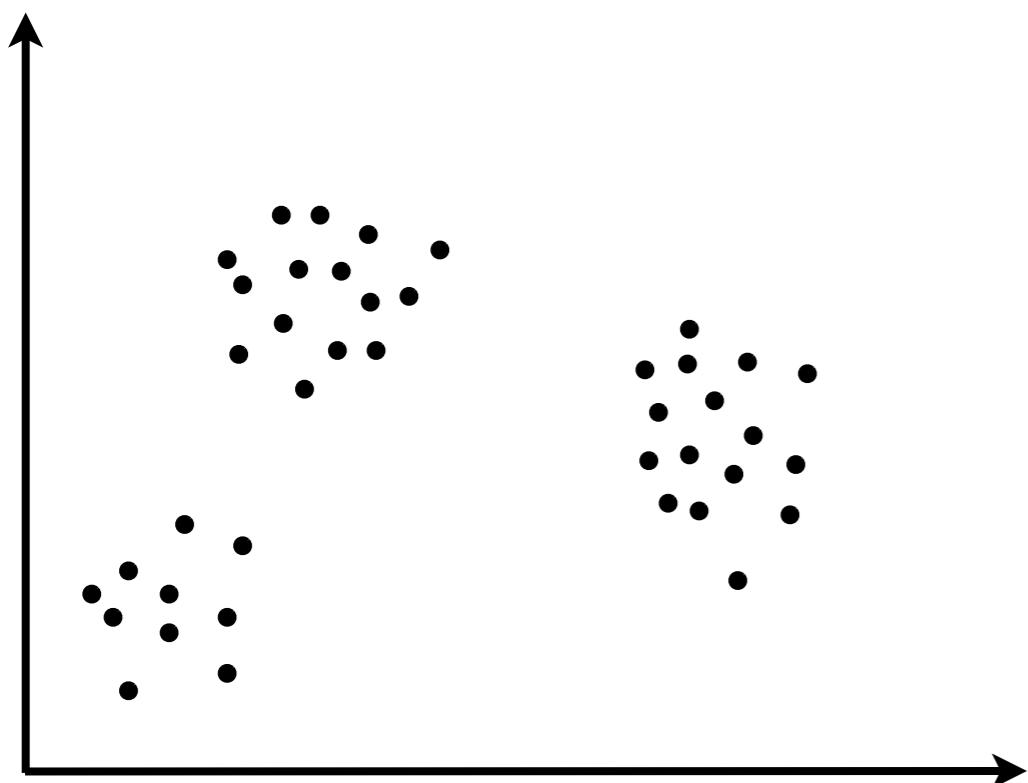


- What happens?
 - $a = a_1 = a_2 \rightarrow 0$
 - $a = a_1 = a_2 \rightarrow \infty$
 - $a_1 > a_2$

[demo]

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)

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)

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)

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



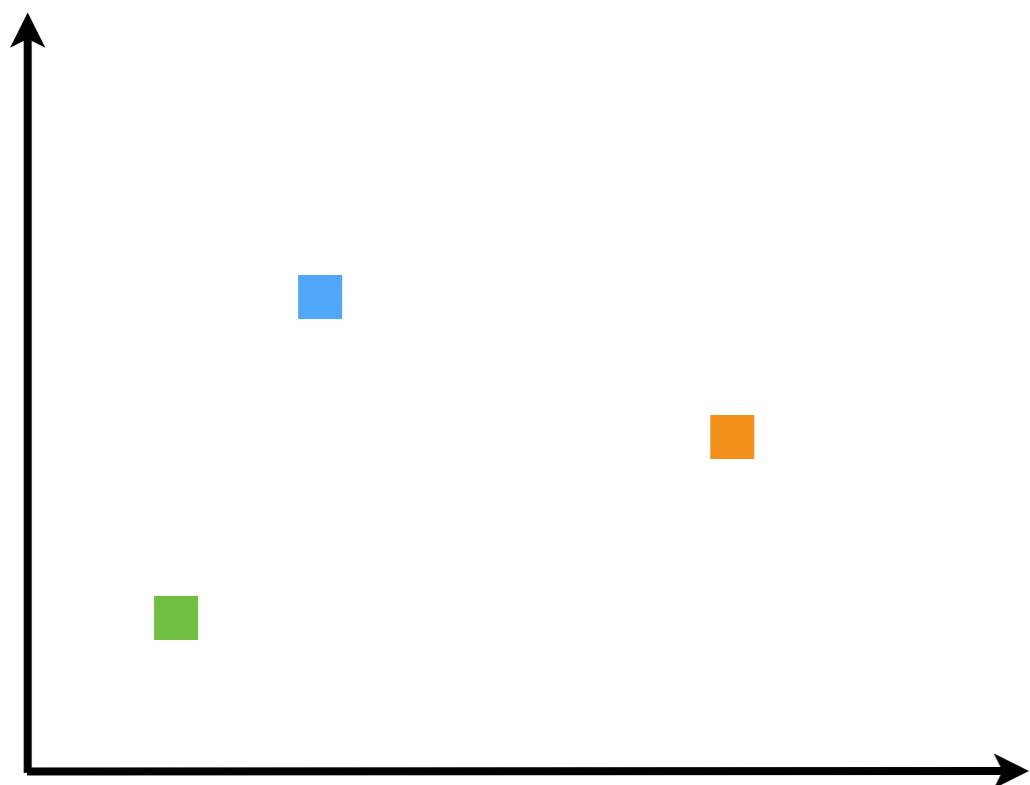
ρ_1

ρ_2

ρ_3

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)

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

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



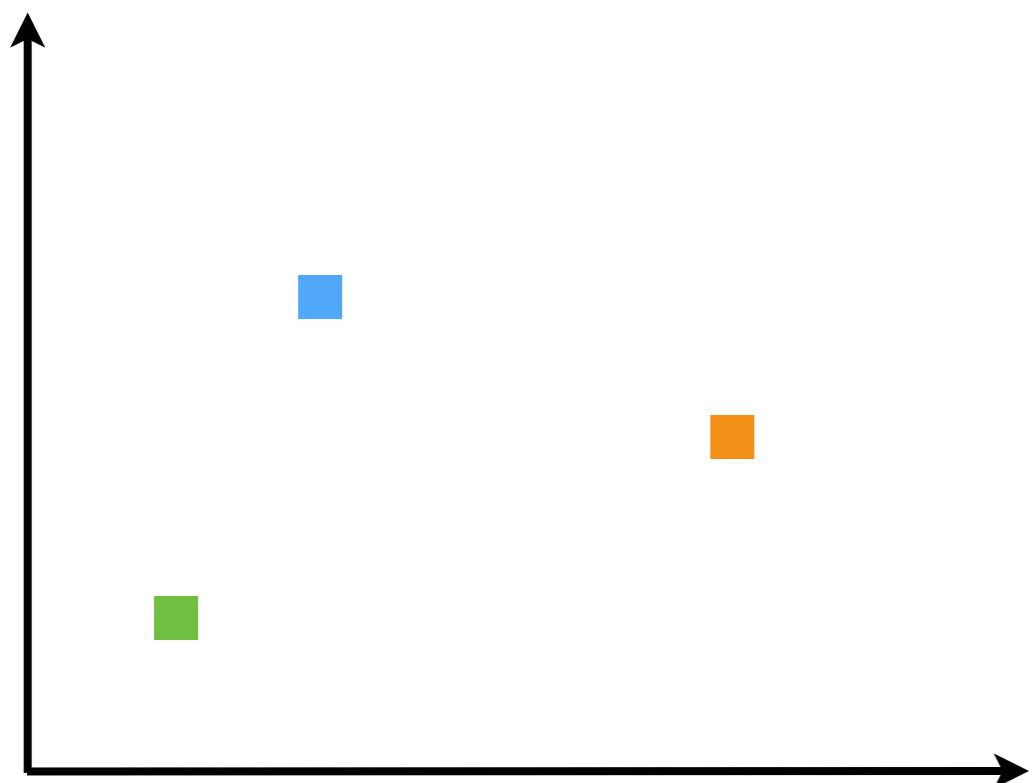
ρ_1

ρ_2

ρ_3

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)

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

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

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



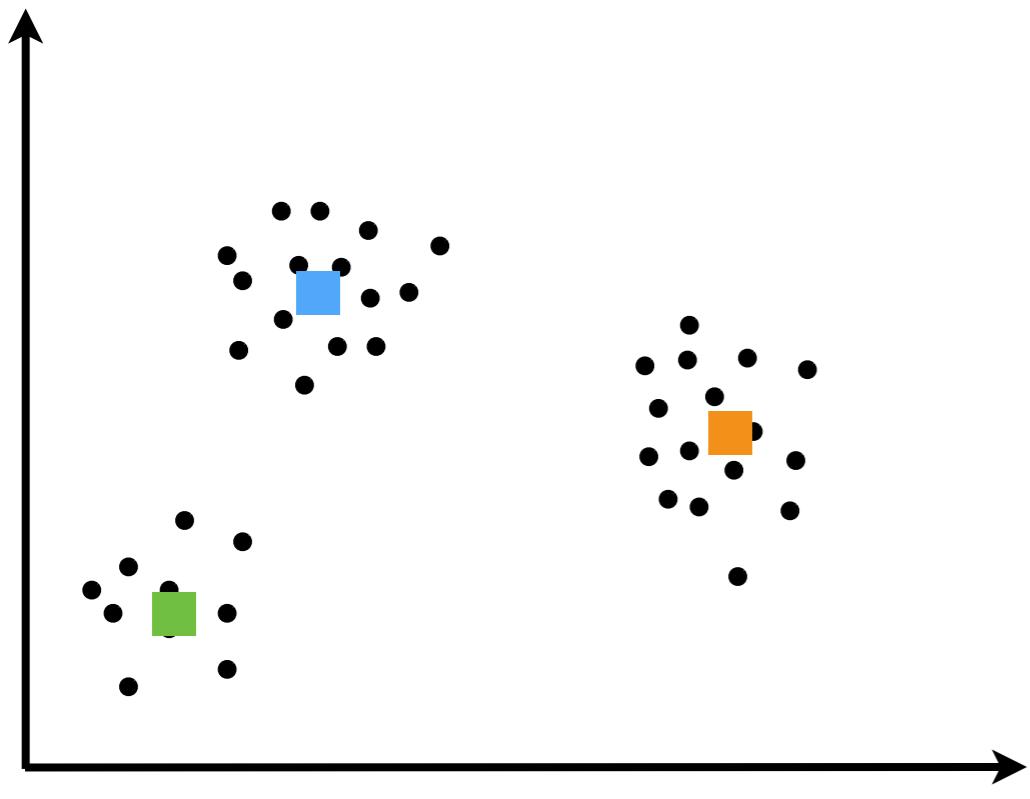
ρ_1

ρ_2

ρ_3

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)

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

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

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

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



ρ_1

ρ_2

ρ_3

Dirichlet distribution review

$$\text{Dirichlet}(\rho_{1:K} | a_{1:K}) = \frac{\Gamma(\sum_{k=1}^K a_k)}{\prod_{k=1}^K \Gamma(a_k)} \prod_{k=1}^K \rho_k^{a_k - 1}$$

Dirichlet distribution review

$$a_k > 0$$

$$\text{Dirichlet}(\rho_{1:K} | a_{1:K}) = \frac{\Gamma(\sum_{k=1}^K a_k)}{\prod_{k=1}^K \Gamma(a_k)} \prod_{k=1}^K \rho_k^{a_k - 1}$$

Dirichlet distribution review

$$\text{Dirichlet}(\rho_{1:K} | a_{1:K}) = \frac{\Gamma(\sum_{k=1}^K a_k)}{\prod_{k=1}^K \Gamma(a_k)} \prod_{k=1}^K \rho_k^{a_k - 1}$$

$$a_k > 0$$

$$\rho_k \in (0, 1)$$

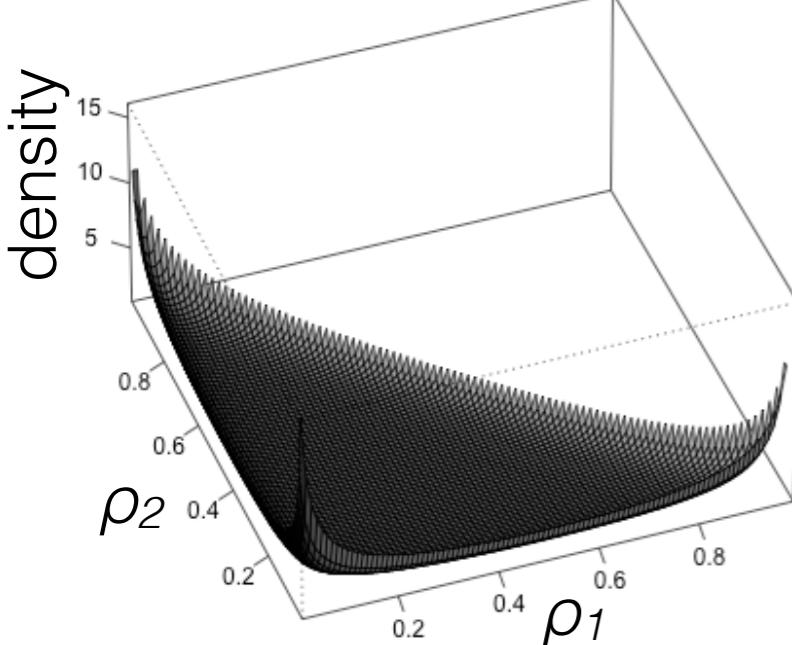
$$\sum_k \rho_k = 1$$

Dirichlet distribution review

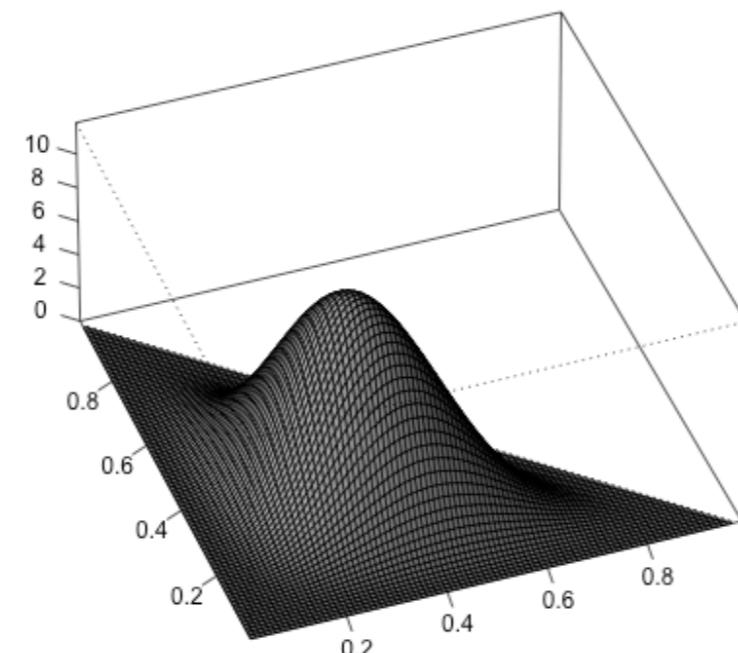
$$\text{Dirichlet}(\rho_{1:K} | a_{1:K}) = \frac{\Gamma(\sum_{k=1}^K a_k)}{\prod_{k=1}^K \Gamma(a_k)} \prod_{k=1}^K \rho_k^{a_k - 1}$$

$a_k > 0$
 $\rho_k \in (0, 1)$
 $\sum_k \rho_k = 1$

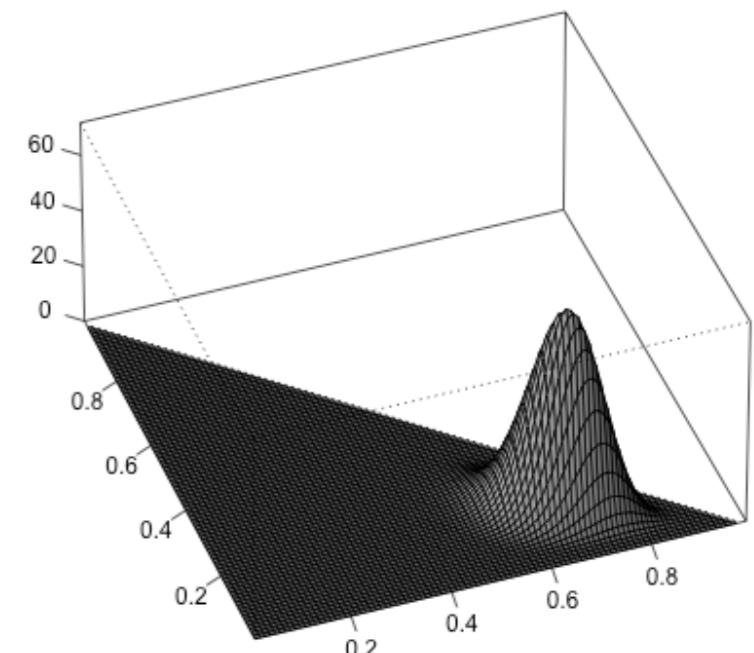
$a = (0.5, 0.5, 0.5)$



$a = (5, 5, 5)$



$a = (40, 10, 10)$

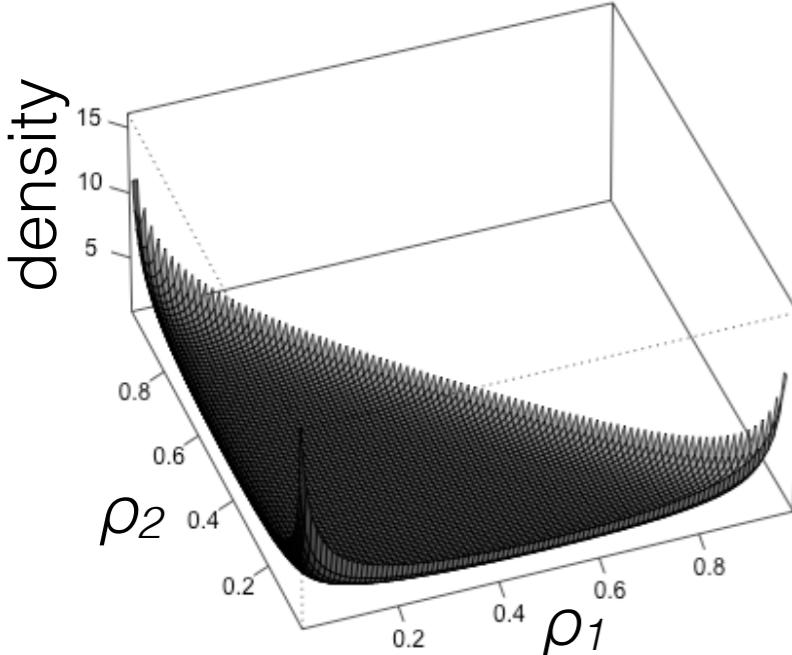


Dirichlet distribution review

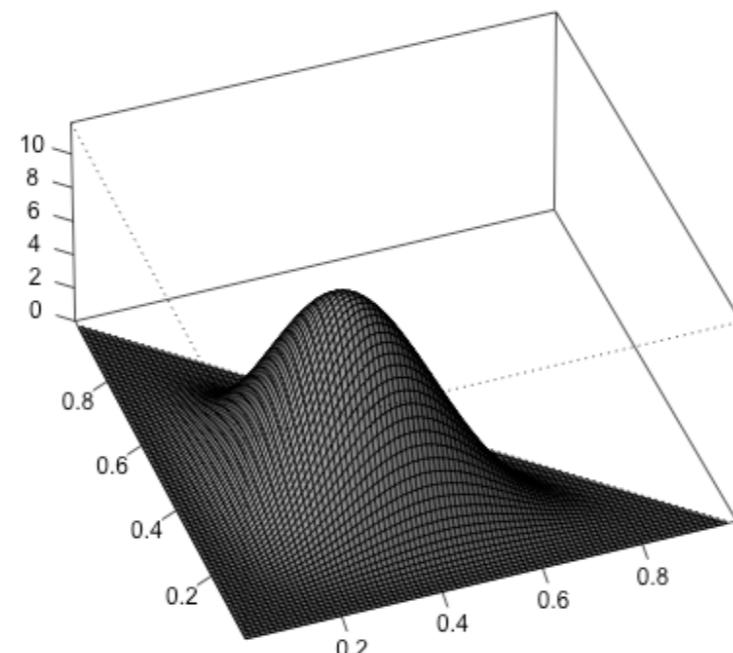
$$\text{Dirichlet}(\rho_{1:K} | a_{1:K}) = \frac{\Gamma(\sum_{k=1}^K a_k)}{\prod_{k=1}^K \Gamma(a_k)} \prod_{k=1}^K \rho_k^{a_k - 1}$$

$$a_k > 0$$
$$\rho_k \in (0, 1)$$
$$\sum_k \rho_k = 1$$

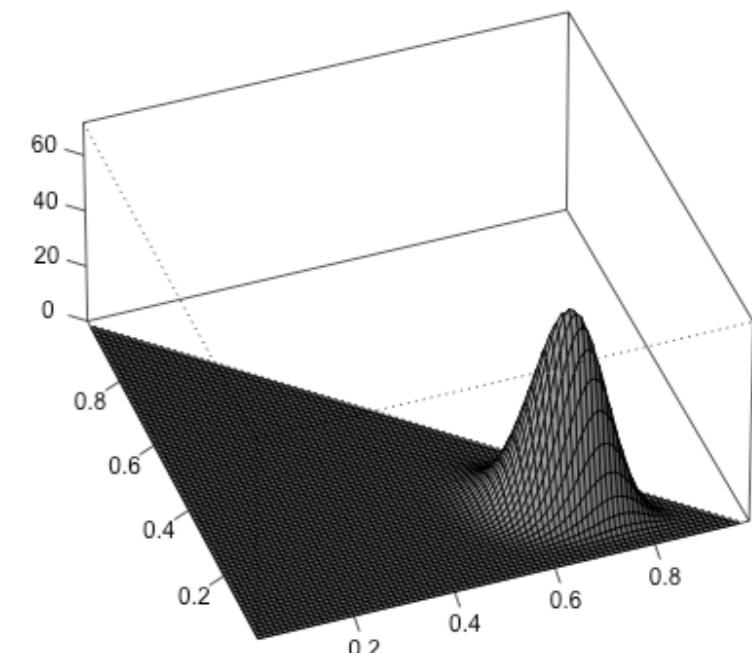
$a = (0.5, 0.5, 0.5)$



$a = (5, 5, 5)$



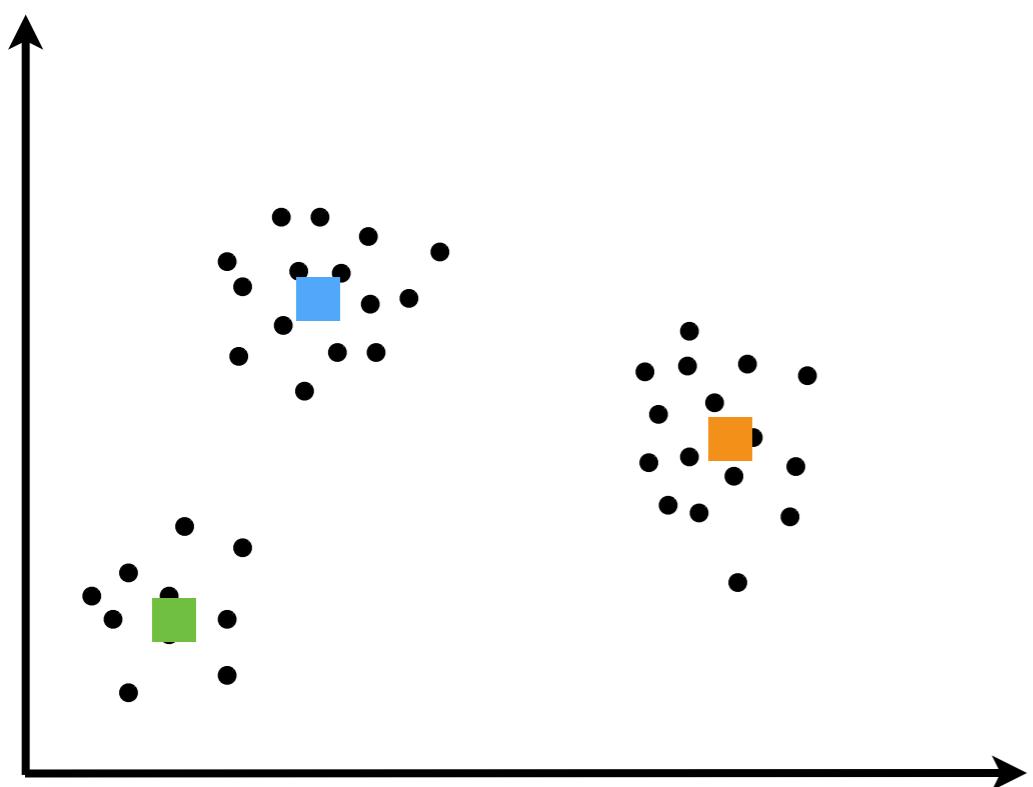
$a = (40, 10, 10)$



- What happens? $a = a_k = 1$ $a = a_k \rightarrow 0$ $a = a_k \rightarrow \infty$ [demo]

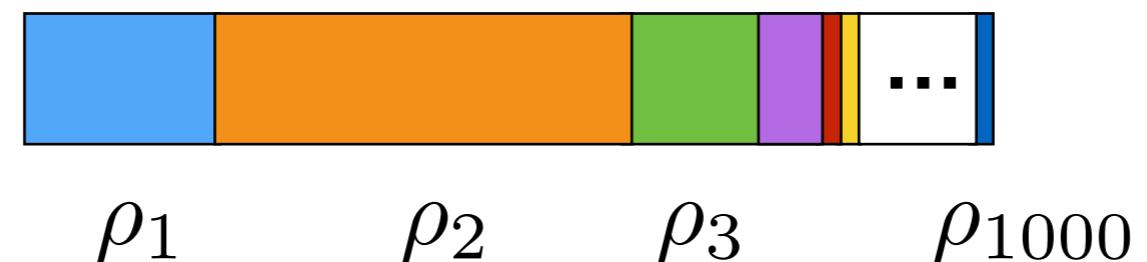
What if $K > N$?

What if $K > N$?



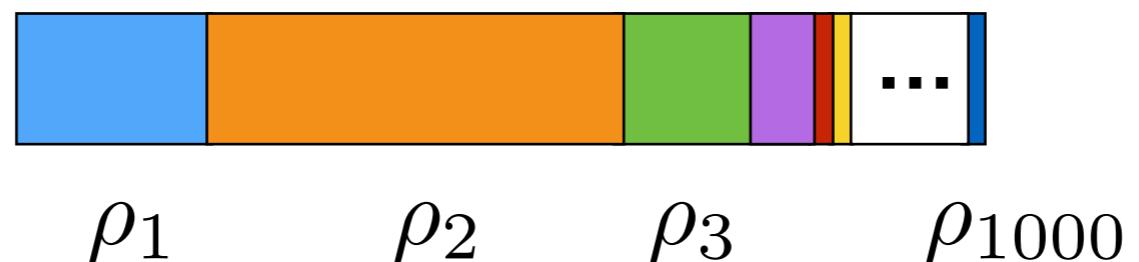
What if $K > N$?

What if $K > N$?



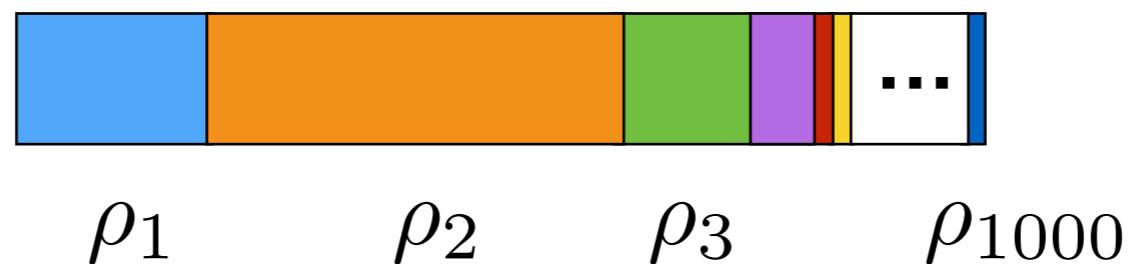
What if $K > N$?

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



What if $K > N$?

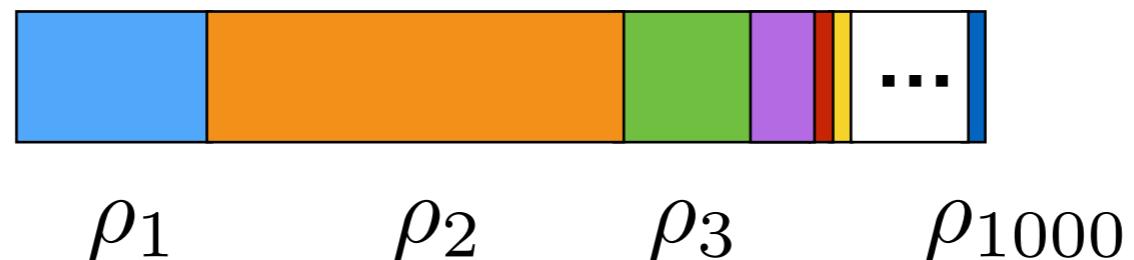
- e.g. species sampling, topic modeling, groups on a social network, etc.



- Components: latent groups

What if $K > N$?

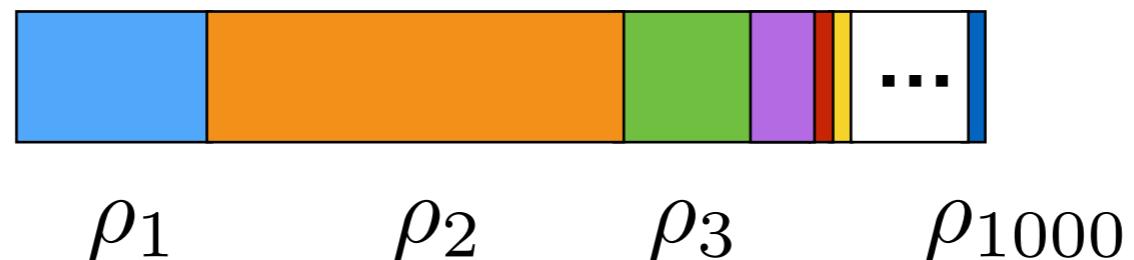
- e.g. species sampling, topic modeling, groups on a social network, etc.



- Components: latent groups
- Clusters: components represented in the data

What if $K > N$?

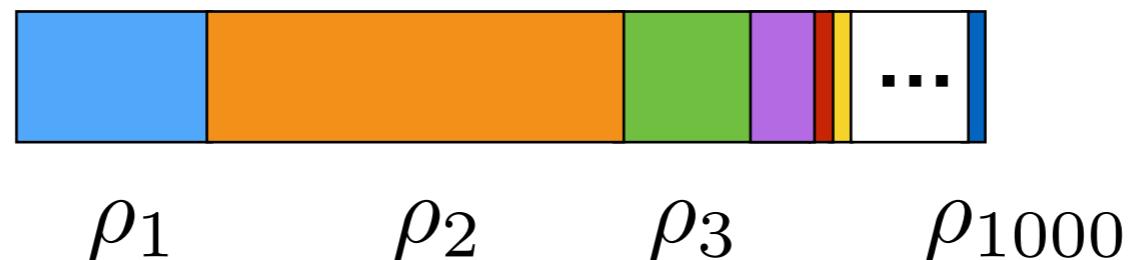
- e.g. species sampling, topic modeling, groups on a social network, etc.



- Components: latent groups
- Clusters: components represented in the data
- [demo 1, demo 2]

What if $K > N$?

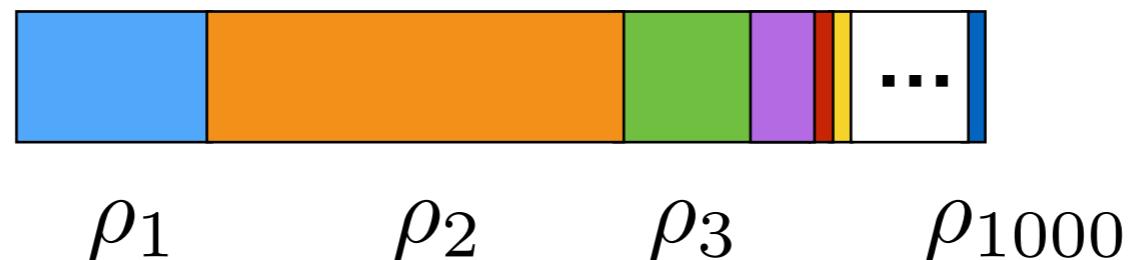
- e.g. species sampling, topic modeling, groups on a social network, etc.



- Components: latent groups
- Clusters: components represented in the data
- [demo 1, demo 2]
- Number of clusters for N data points is $< K$ and random

What if $K > N$?

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



- Components: latent groups
- Clusters: components represented in the data
- [demo 1, demo 2]
- Number of clusters for N data points is $< K$ and 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: $\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$

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



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

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

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

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



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

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



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

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

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

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

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

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

- “Stick breaking”

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

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

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

- “Stick breaking”

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

$$\rho_1 = V_1$$

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

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

- “Stick breaking”

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

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

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

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

- “Stick breaking”

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

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

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

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

- “Stick breaking”


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

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

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

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

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



- “Stick breaking”



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

$$\rho_1 = V_1$$



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

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



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

$$\rho_3 = (1 - V_1)(1 - V_2)V_3$$

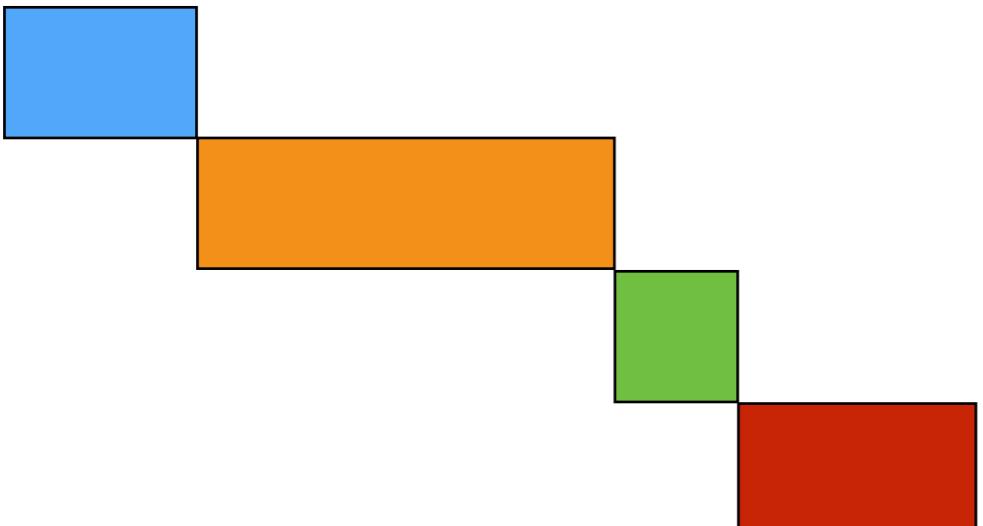
Choosing $K = \infty$

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

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



- “Stick breaking”



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

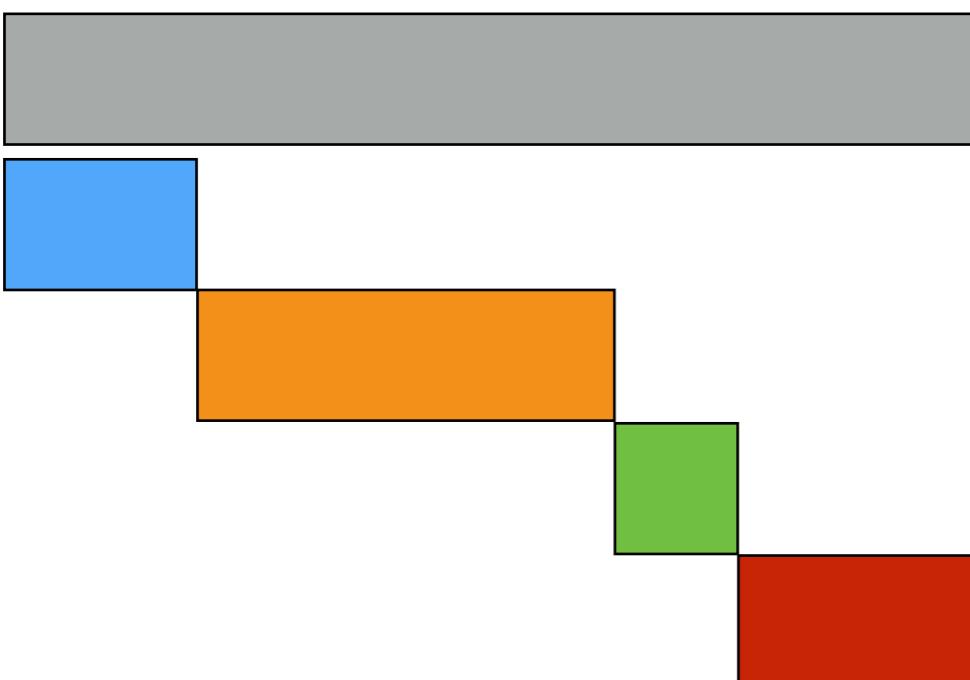
$$V_2 \sim \text{Beta}(a_2, a_3 + a_4) \quad \rho_2 = (1 - V_1)V_2$$

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

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

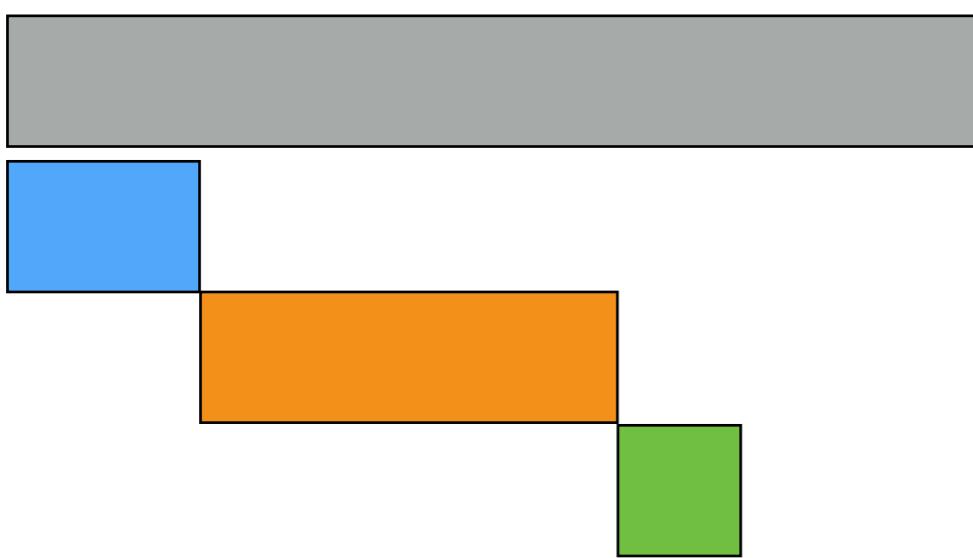
Choosing $K = \infty$

- 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?



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

Choosing $K = \infty$

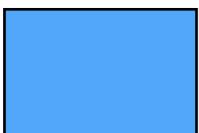
- Here, difficult to choose finite K in advance (contrast with small K): don't know K , difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



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

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?



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

$$V_2 \sim \text{Beta}(a_2, b_2)$$

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?



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

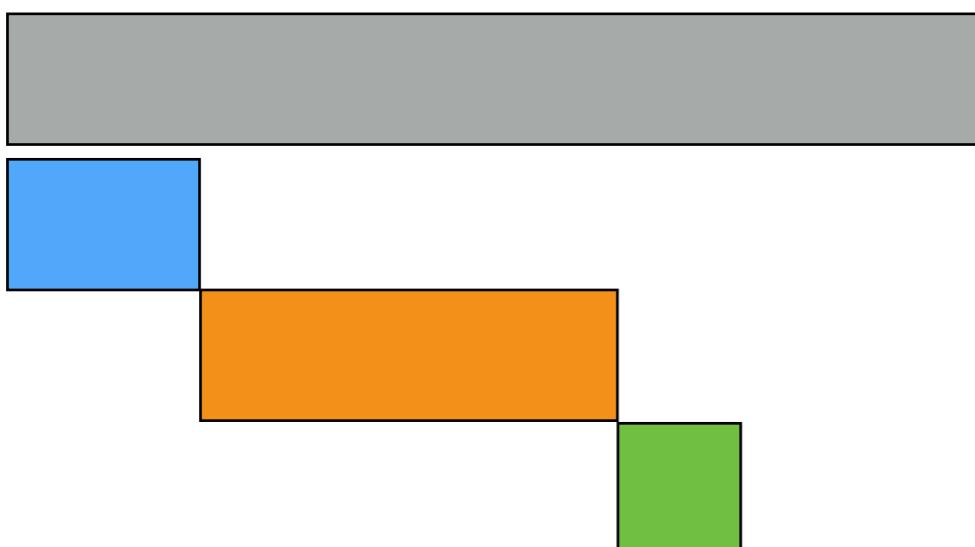
$$\rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, b_2)$$

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

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?



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

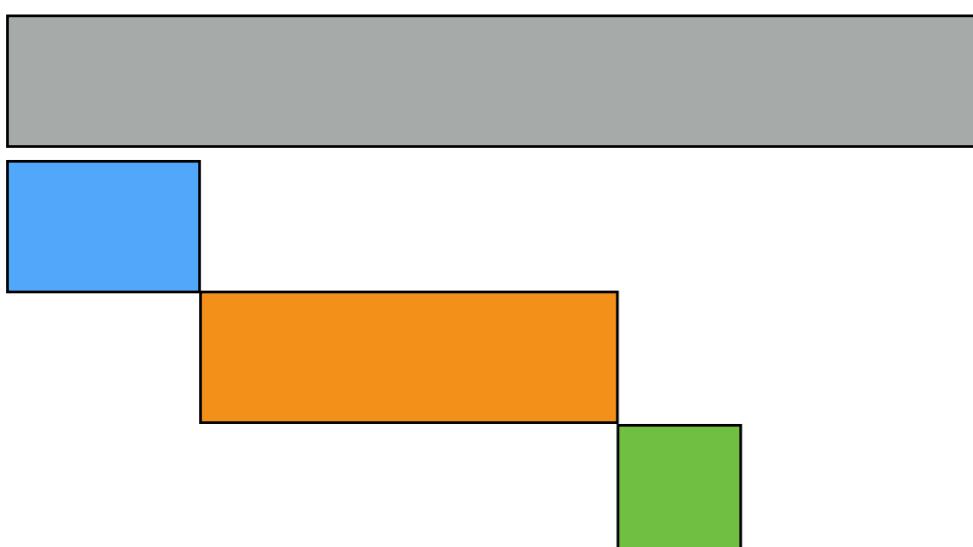
$$V_2 \sim \text{Beta}(a_2, b_2)$$

$$\rho_1 = V_1$$

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

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?



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

$$V_2 \sim \text{Beta}(a_2, b_2)$$

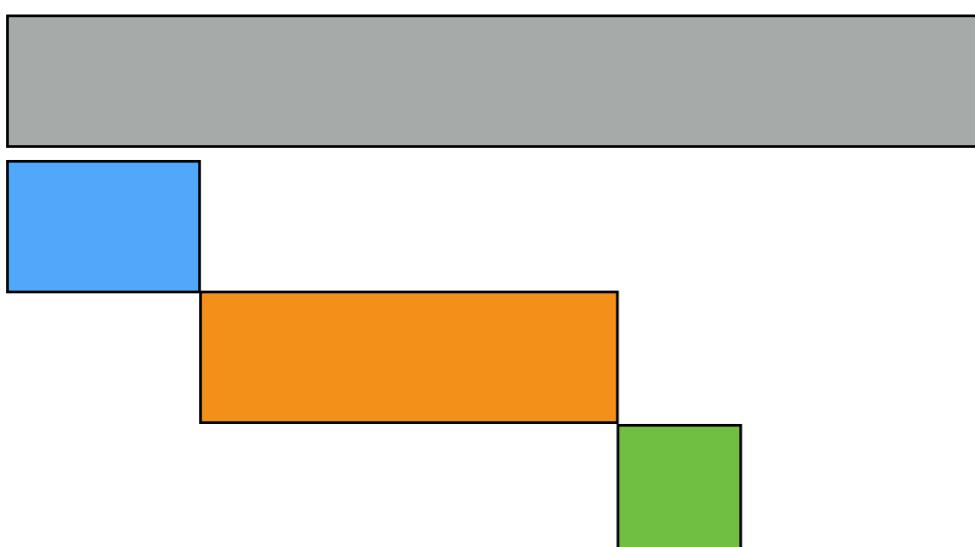
$$\rho_1 = V_1$$

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

...

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?



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

$$\rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, b_2)$$

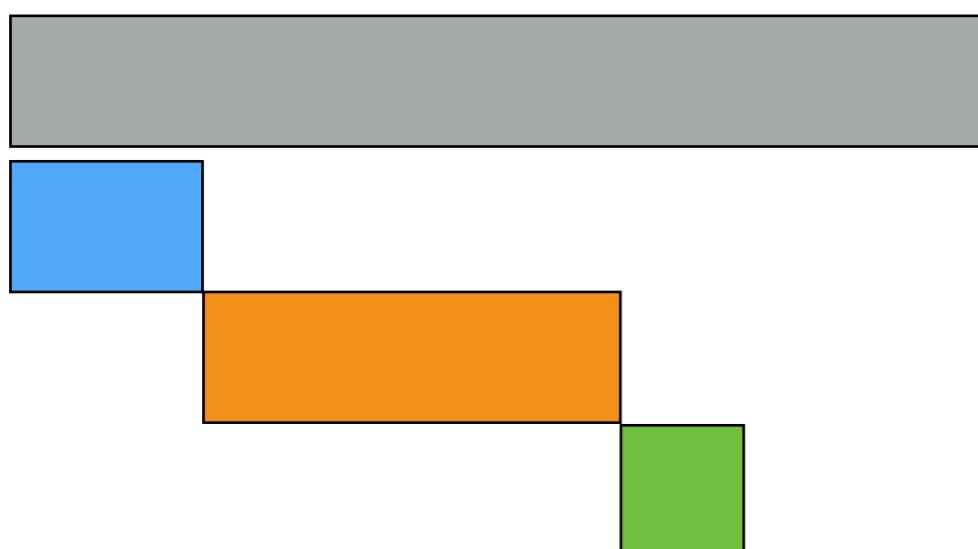
$$\rho_2 = (1 - V_1)V_2$$

...

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

Choosing $K = \infty$

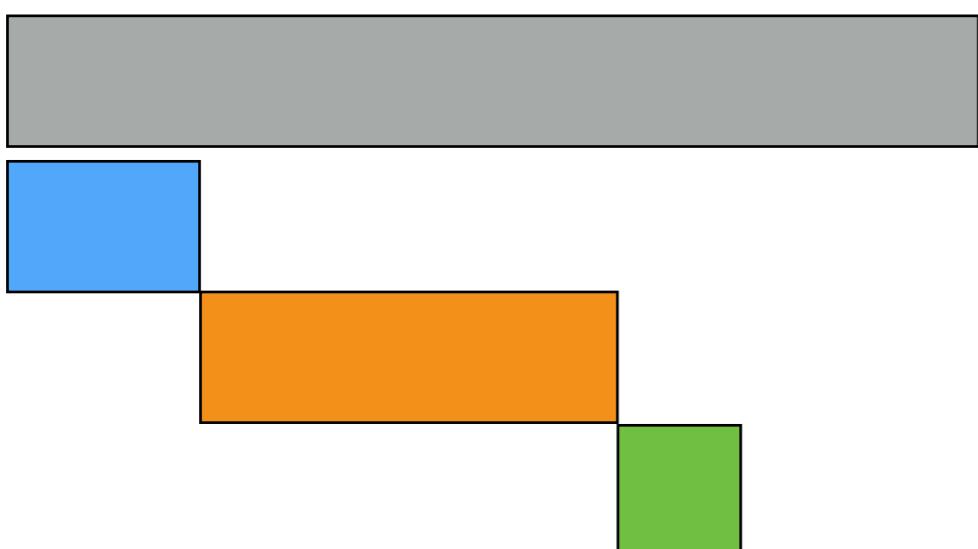
- Here, difficult to choose finite K in advance (contrast with small K): don't know K , difficult to infer, streaming data
- How to generate $K = \infty$ strictly positive frequencies that sum to one?



$$\begin{array}{lll} V_1 \sim \text{Beta}(a_1, b_1) & & \rho_1 = V_1 \\ V_2 \sim \text{Beta}(a_2, b_2) & & \rho_2 = (1 - V_1)V_2 \\ \cdots & & \\ V_k \sim \text{Beta}(a_k, b_k) & & \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j) \right] V_k \end{array}$$

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?



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

$$V_2 \sim \text{Beta}(a_2, b_2)$$

⋮

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

$$\rho_1 = V_1$$

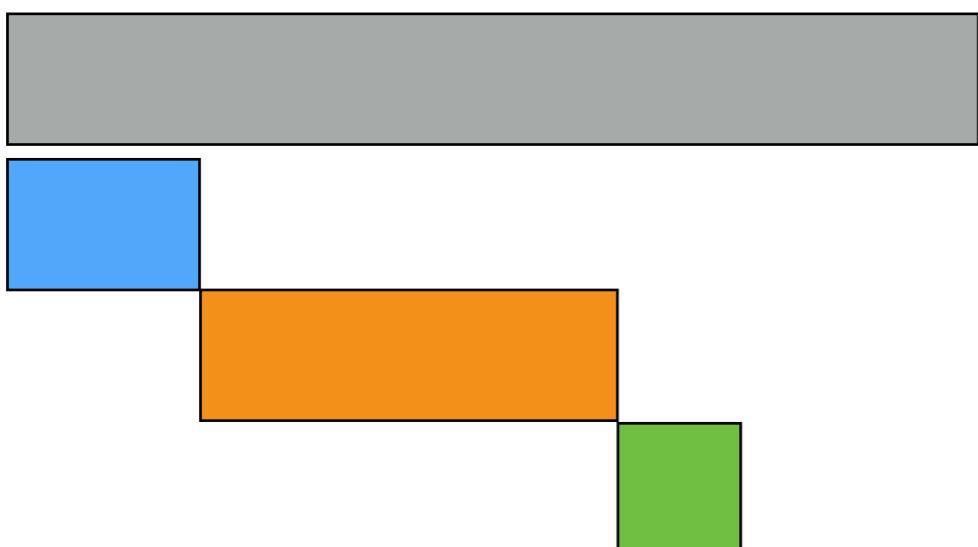
$$\rho_2 = (1 - V_1)V_2$$

$$\rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j) \right] V_k$$

[van der Vaart, Ghosal 2017]

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:** $a_k = 1, b_k = \alpha > 0$



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

$$\rho_1 = V_1$$

$$V_2 \sim \text{Beta}(a_2, b_2)$$

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

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

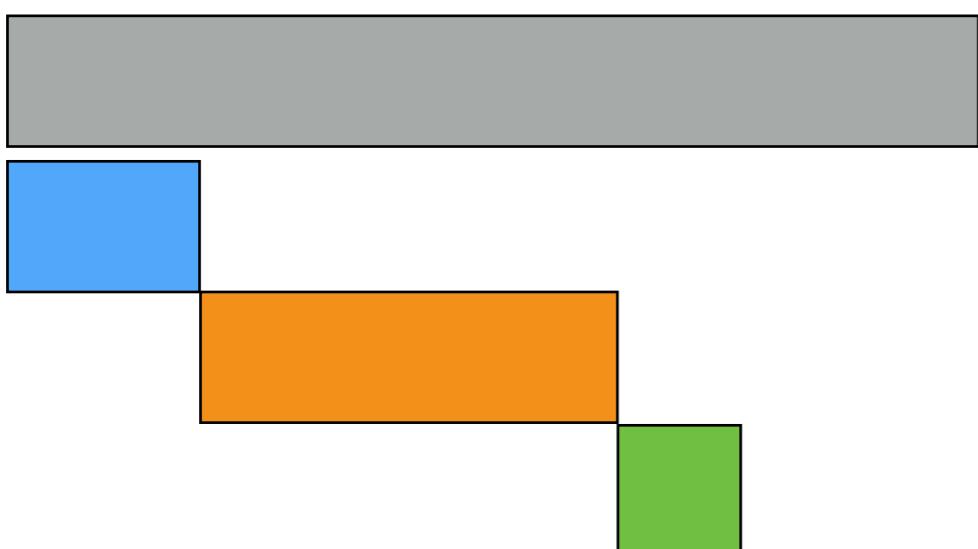
$$\rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j) \right] V_k$$

[van der Vaart, Ghosal 2017]

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:

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

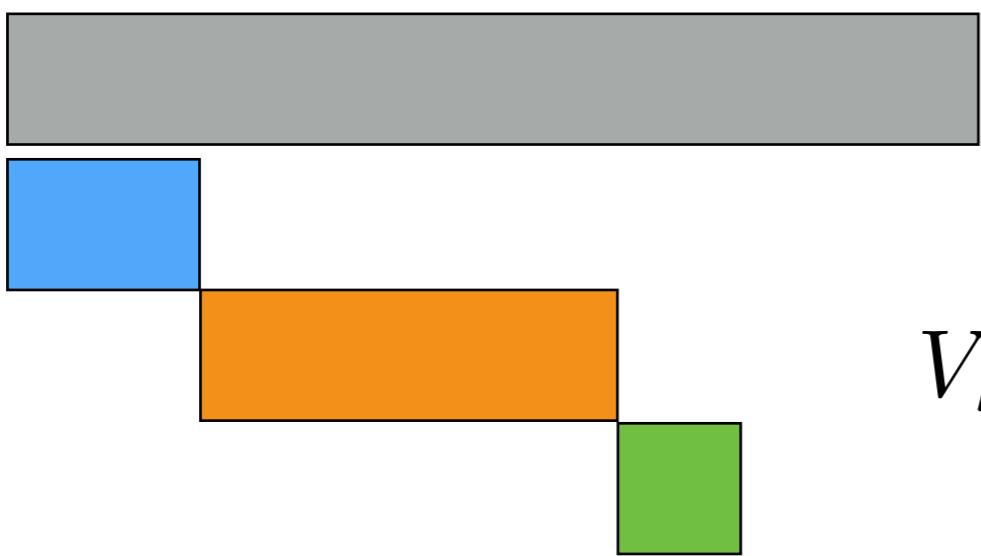


$$\begin{array}{lll} V_1 \sim \text{Beta}(a_1, b_1) & & \rho_1 = V_1 \\ V_2 \sim \text{Beta}(a_2, b_2) & & \rho_2 = (1 - V_1)V_2 \\ \cdots & & \\ V_k \sim \text{Beta}(a_k, b_k) & & \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j) \right] V_k \end{array}$$

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:

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



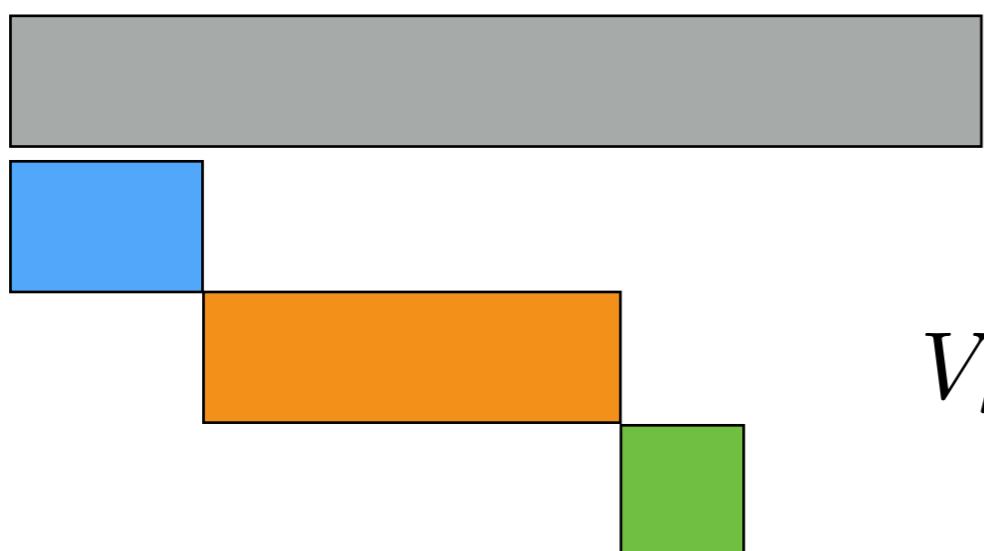
$$V_k \stackrel{iid}{\sim} \text{Beta}(1, \alpha) \quad \rho_k = \left[\prod_{j=1}^{k-1} (1 - V_j) \right] V_k$$

...

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?
 - **Dirichlet process stick-breaking:** $a_k = 1, b_k = \alpha > 0$
 - Griffiths-Engen-McCloskey (**GEM**) distribution:

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



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

...

[demo]

References

- See Part II for full information