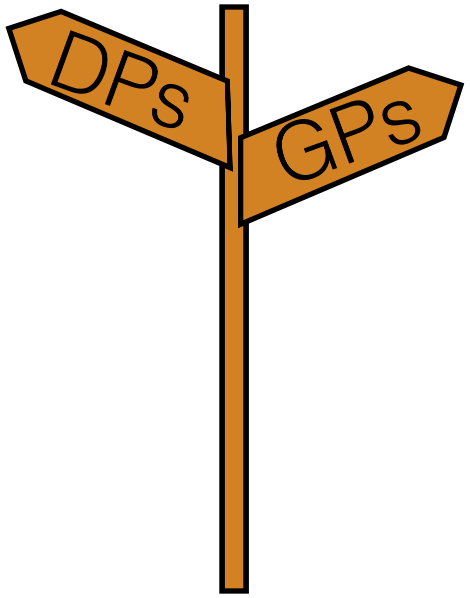
Nonparametric Bayesian Methods: Models, Algorithms, and Applications (Day 2)

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

www.tamarabroderick.com/tutorials.html

Roadmap

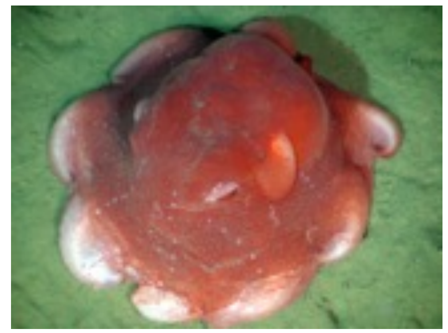


- Bayes Foundations
- Unsupervised Learning
  - Example problem: clustering
  - Example BNP model: Dirichlet process (DP)
  - Chinese restaurant process
- Supervised Learning
  - Example problem: regression
  - Example BNP model: Gaussian process (GP)
- Venture further into the wild world of Nonparametric Bayes

- Big questions
  - Why BNP?
  - What does an infinite/growing number of parameters really mean (in BNP)?
  - Why is BNP challenging but practical?

Applications

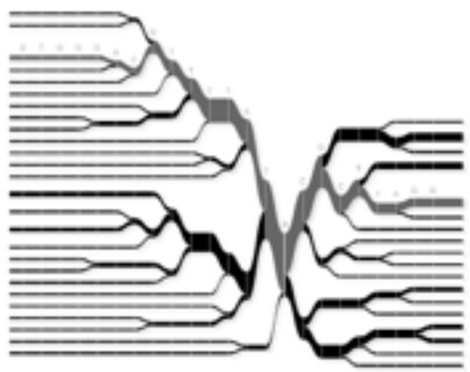
[wikipedia.org]



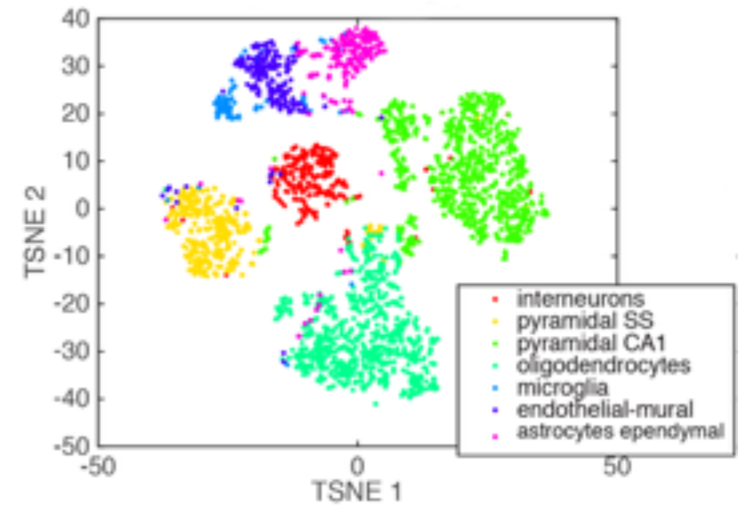
[Ed Bowlby, NOAA]

[Fox et al 2014]

[Lloyd et al 2012; Miller et al 2010]



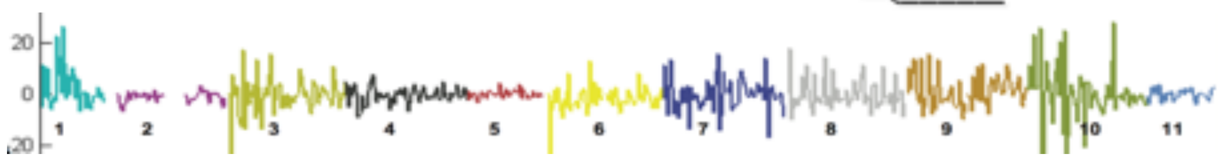
[Ewens 1972; Hartl, Clark 2003]



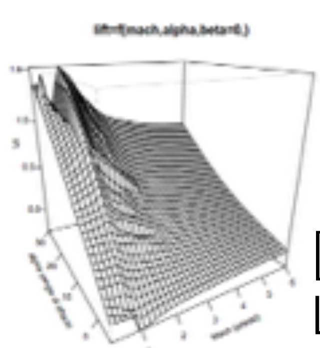
[Prabhakaran, Azizi, Carr, Pe'er 2016]

[Sudderth, Jordan 2009]

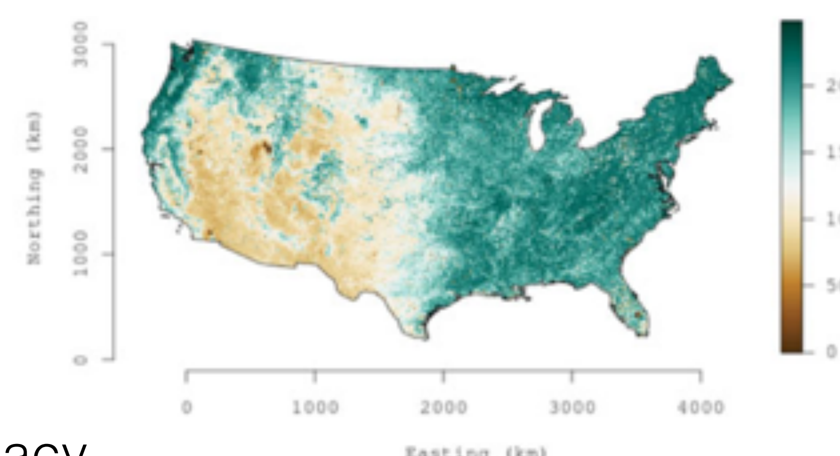
[Saria et al 2010]



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

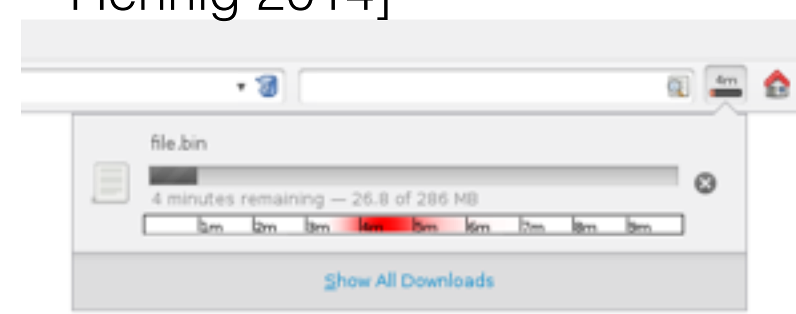


[Gramacy, Lee 2009]



[Datta, Banerjee, Finley, Gelfand 2016]

[Kiefel, Schuler, Hennig 2014]



[Chati, Balakrishnan 2017]

[Deisenroth, Fox, Rasmussen 2015]

E.g., Information retrieval

Snail races usually take place on a circular track with the snails starting in the middle and racing to the perimeter. The track usually takes the form of a damp cloth atop a table. The radius is traditionally set at 13 or
are
l stickers
o

Major constituents of the arts include literature – including poetry, novels and short stories, and epics; performing arts – among them music, dance, and theatre; culinary arts such as baking, chocolatiering, and winemaking; media arts like photography and cinematography, and visual arts – including drawing, painting, ceramics, and s
Some art forms combine
element with performanc
and the written word (e.g

The term economics comes from the Ancient Greek οἰκονομία from οἶκος (oikos, "house") and νόμος (nomos, "custom" or "law"), hence "rules of the house (hold for good management)". 'Political economy' was the earlier name for the subject, but economists in the late 19th century suggested "economics" as a shorter term for "economic science" to establish itself as a separate discipline outside of political science and other social sciences.

ort is generally recognised as
ivities which are based in physical
leticism or physical dexterity, with
largest major competitions such
the Olympic Games admitting only
orts meeting this definition, and
er organisations such as the
uncil of Europe using definitions precluding activities without a physical element from classification as sports.

Architecture has to do with planning, designing and constructing form, space and ambience to reflect functional, technical, social, environmental and aesthetic considerations. It requires the creative manipulation and coordination of materials and technology, and of light and shadow. Often, conflicting requirements must be resolved.

The central premise of Moneyball is that the collected wisdom of baseball insiders (including players, managers, coaches, scouts, and the front office) over the past century is subjective and often flawed. […] The
argues that the Oakland A's'
office took advantage of more
ical gauges of player
mance to field a team that
better compete against richer
etitors in Major League
ball (MLB).

The Tour de France, the Giro d'Italia and Vuelta a España make up cycling's prestigious, three-week-long Grand Tours; the Tour is the oldest and generally considered the most prestigious of the three. Traditionally, the race is held primarily in the month of July. While the route changes each year, the format of the race stays the same with the appearance of time trials, the passage through the mountain chains of the Pyrenees and the Alps, and the finish on the Champs-Élysées in Paris.

The increasing tendency to privilege painting, and to a lesser degree sculpture, above other arts has been a feature of Western art as well as East Asian art. In both regions painting has been seen as relying to the highest degree on the imagination of the artist, and the furthest removed from manual labour - in Chinese painting the most highly valued styles were those of "scholar-painting", at least in theory practiced by gentleman amateurs. The Western hierarchy of genres reflected similar attitudes.

Agricultural economics today includes a variety of applied areas, having considerable overlap with conventional economics. Agricultural economists have made substantial contributions to research in economics, econometrics, development economics, and environmental economics. Agricultural economics influences food policy, agricultural policy, and environmental policy.

The basic tool for econometrics is the linear regression model. In modern econometrics, other statistical tools are frequently used, but linear regression is still the most frequently used starting point for an analysis. Estimating a linear regression on two variables can be visualized as fitting a line through data points representing paired values of the independent and dependent variables.

[wikipedia.org]

E.g., Information retrieval

The term economics comes from the
Ancient Greek οἰκονομία from οἶκος
(oikos, "house") and νόμος (nomos,
"custom" or "law"), hence "rules of the
house (hold
'Political ec
name for th
in the late 1
"economics
"economic
as a separa
political sci
sciences.

The central premise of Moneyball is
that the collected wisdom of baseball
insiders (including players,
managers, coaches, scouts, and the
front office) over the past century is
subjective
book argue
front office
analytical g
performanc
could bette
competitor
Baseball (M

Agricultural economics today
includes a variety of applied areas,
having considerable overlap with
conventional economics. Agricultural
economi
contribut
economi
developm
environm
Agricultu
food poli
environm

The basic tool for econometrics is the
linear regression model. In modern
econometrics, other statistical tools
are frequently used, but linear
regression is still the most frequently
used starting point for an analysis.
Estimating a linear regression on two
variables can be visualized as fitting
a line through data points
representing paired values of the
independent and dependent
variables.

Major constituents of the arts include
literature – including poetry, novels
and short stories, and epics;
performing arts – among them music,
dance, and theatre; culinary arts
such as baking
winemaking;
photography
and visual arts
painting, cera
Some art form
element with p
and the writter

Architecture has to do with planning,
designing and constructing form,
space and ambience to reflect
functional, te
environment
consideratio
creative man
coordination
technology,
Often, confli
be resolved.

The increasing tendency to privilege
painting, and to a lesser degree
sculpture, above other arts has been
a feature of Western art as well as
East Asian art. In both regions
painting has been seen as relying to
the highest degree on the
imagination of the artist, and the
furthest removed from manual labour
- in Chinese painting the most highly
valued styles were those of "scholar-
painting", at least in theory practiced
by gentleman amateurs. The Western
hierarchy of genres reflected similar
attitudes.

Sport is generally recognised as
activities which are based in physical
athleticism or physical dexterity, with
the largest
as the Olym
sports mee
other orga
Council of
precluding
physical el
as sports.

The Tour de France, the Giro d'Italia
and Vuelta a España make up
cycling's prestigious, three-week-
long Grand Tours; the Tour is the
oldest and generally considered the
most prestigious of the three.
Traditionally,
in the month
changes eac
race stays th
appearance
passage thro
chains of the
and the finish
Élysées in Pa

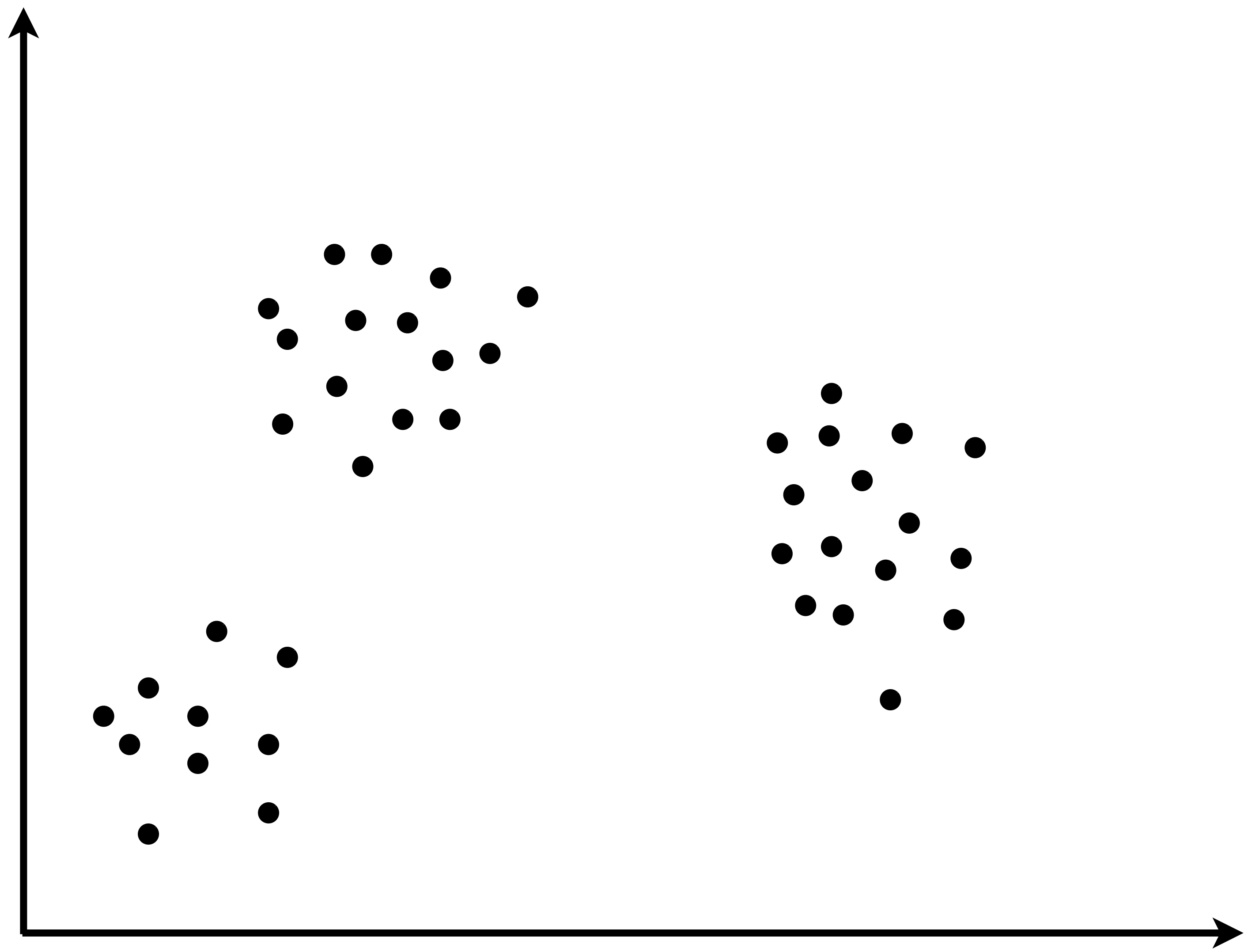
Snail races usually take place on a
circular track with the snails starting
in the middle and racing to the
perimeter. The track usually takes the
form of a damp cloth atop a table.
The radius is traditionally set at 13 or
14 inches. Racing numbers are
painted on the shells or small stickers
or tags are placed on them to
distinguish each competitor.

[wikipedia.org]

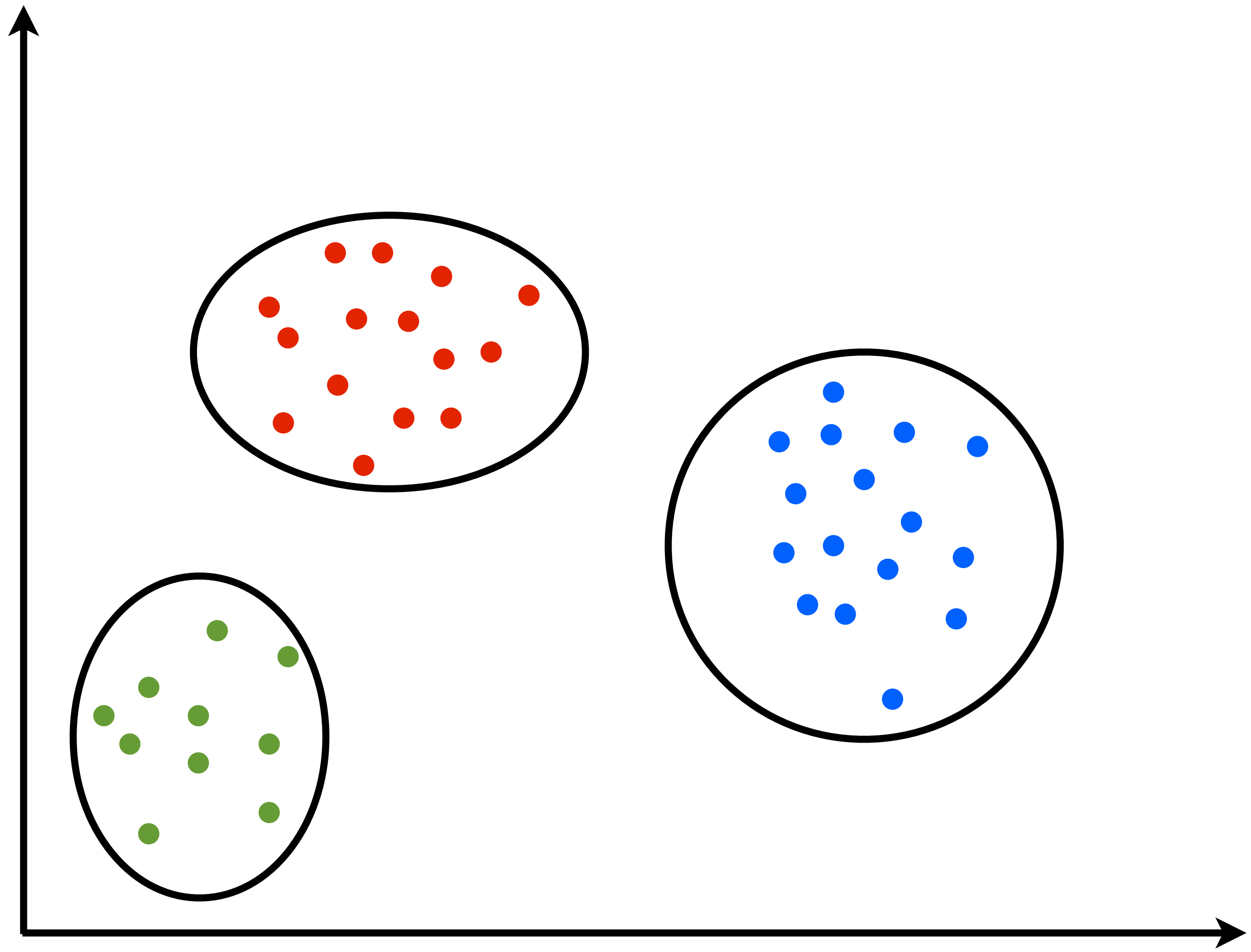
E.g., Information retrieval

[wikipedia.org]

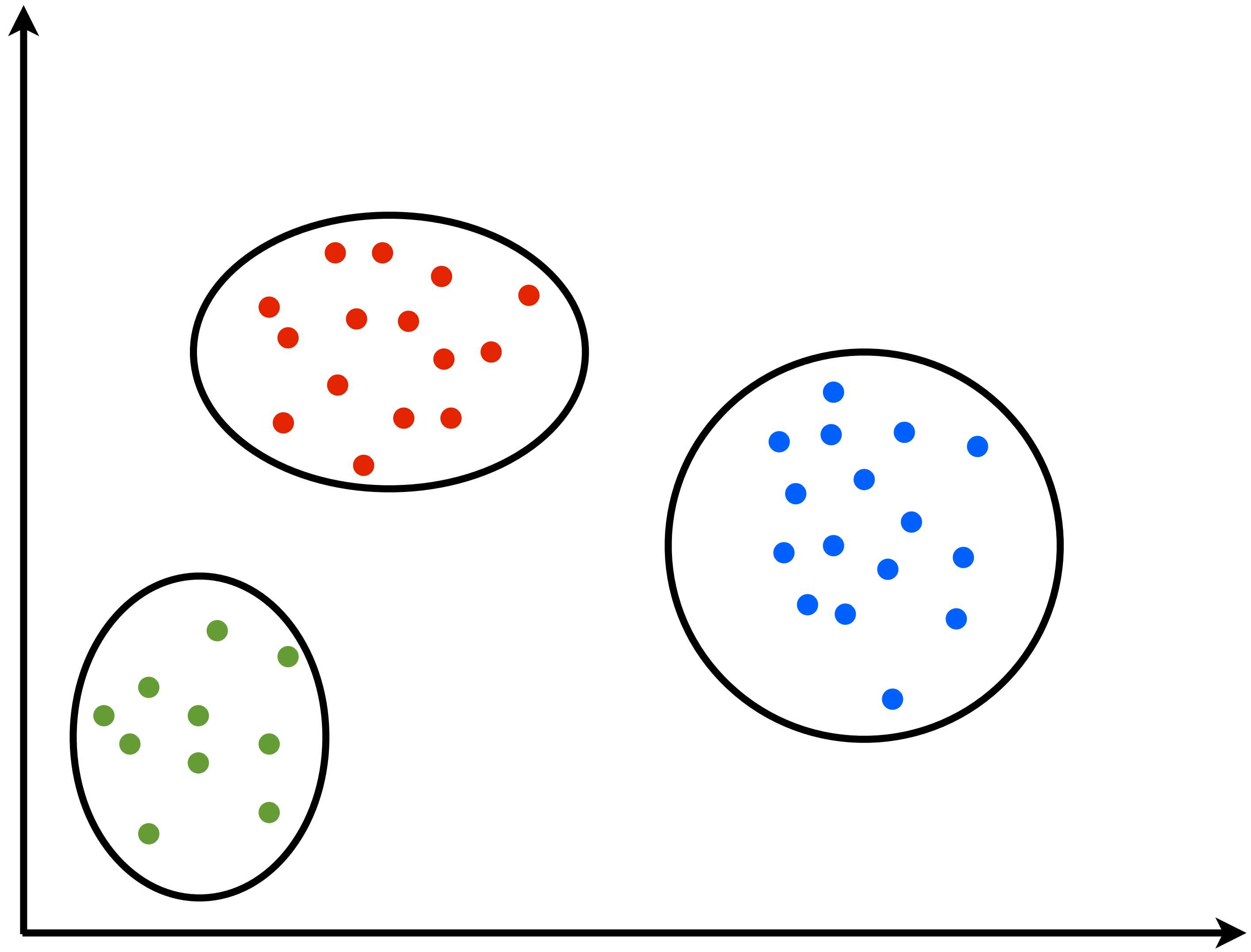
Clustering



Clustering

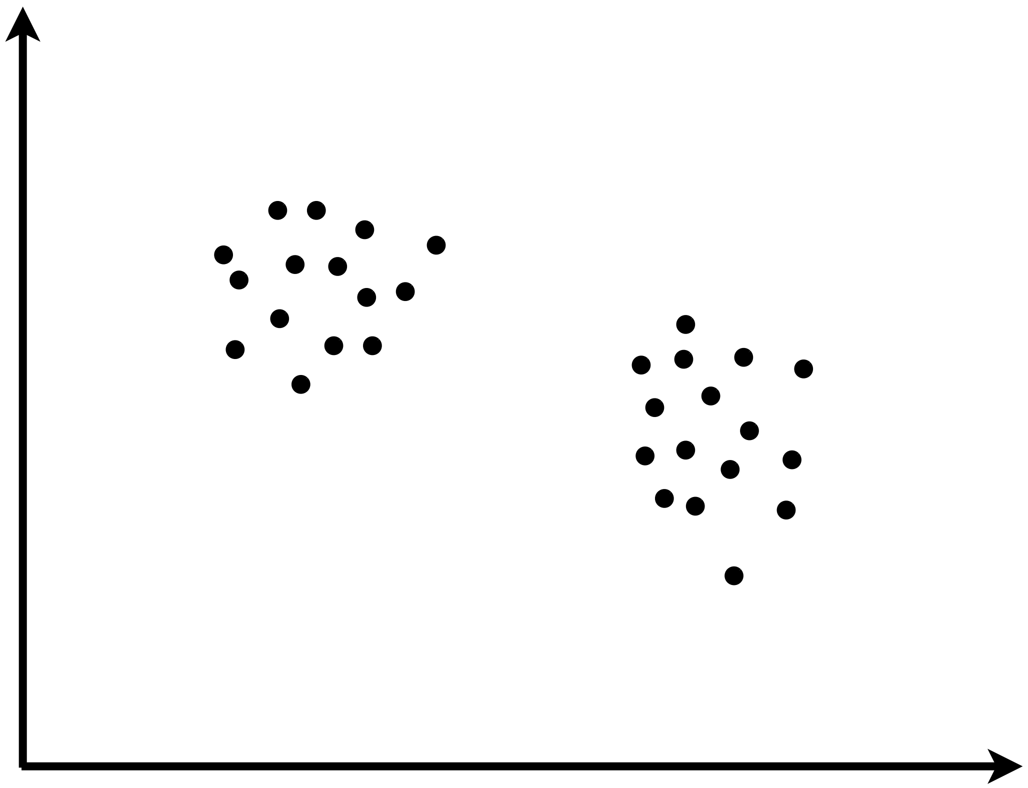


Clustering



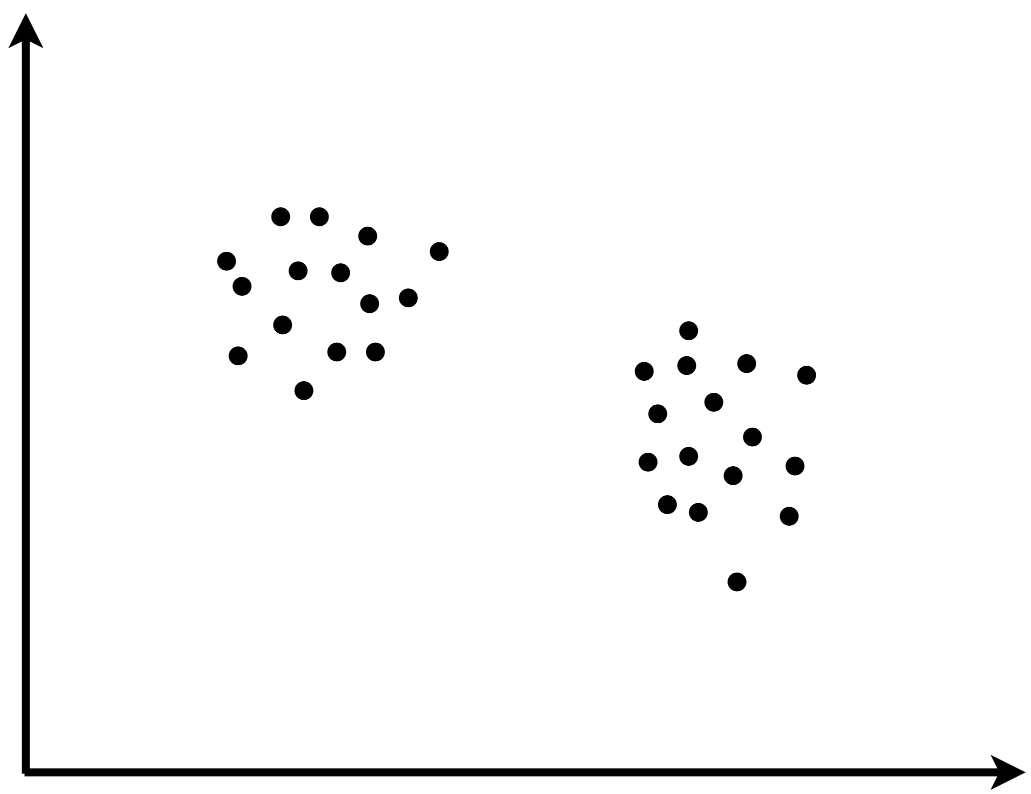
“Clusters”

Generative model



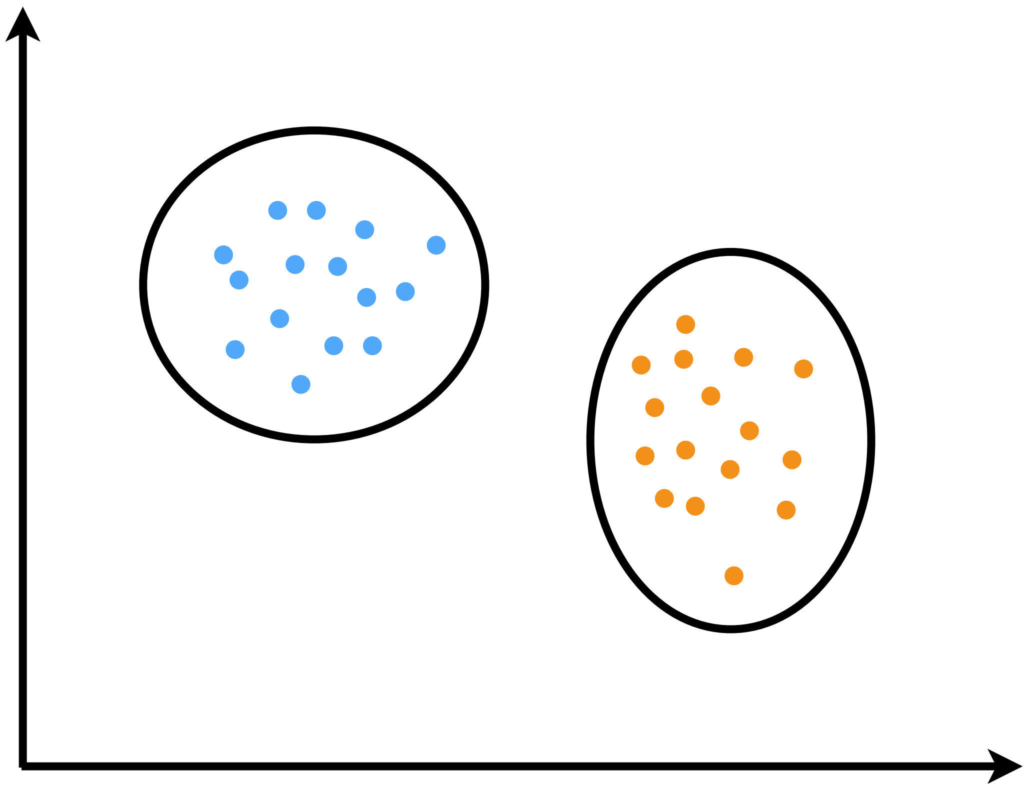
Generative model

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



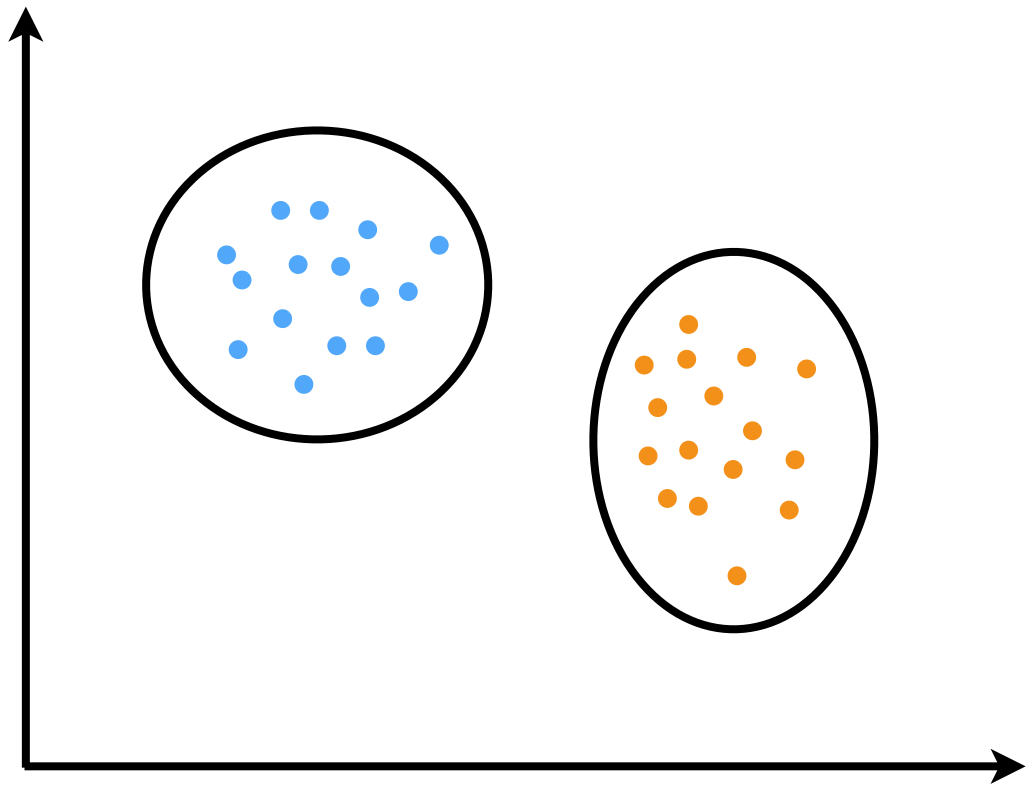
Generative model

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



Generative model

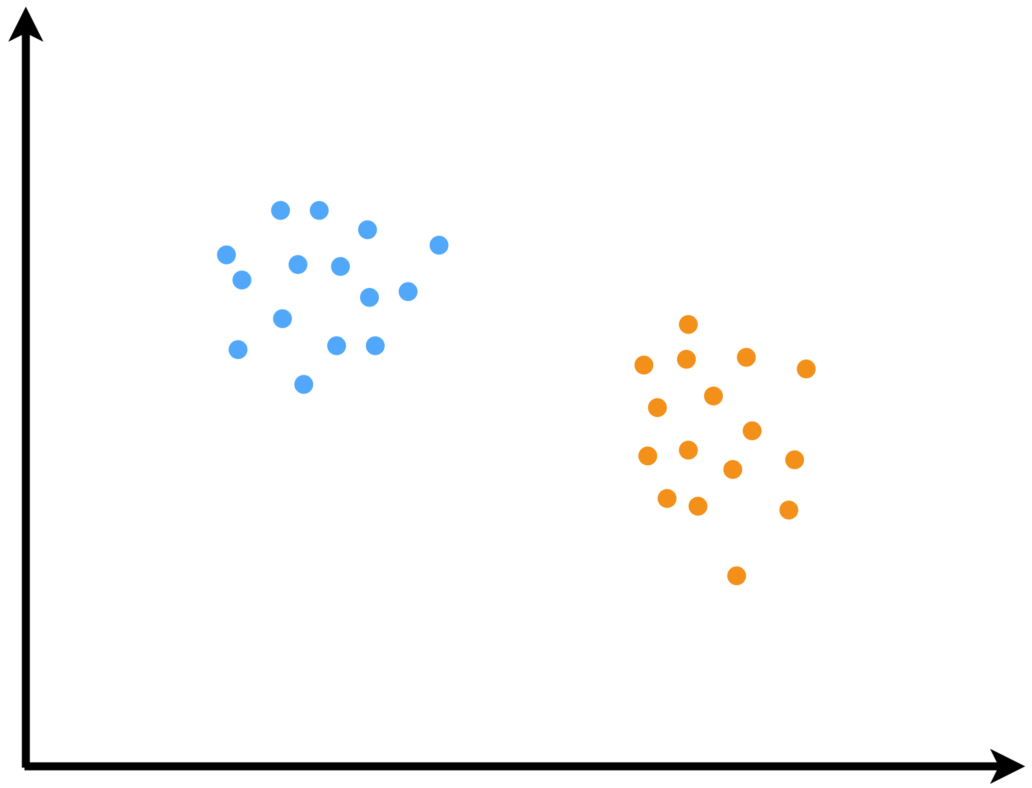
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$



- Inference goal: assignments of data points to clusters

Generative model

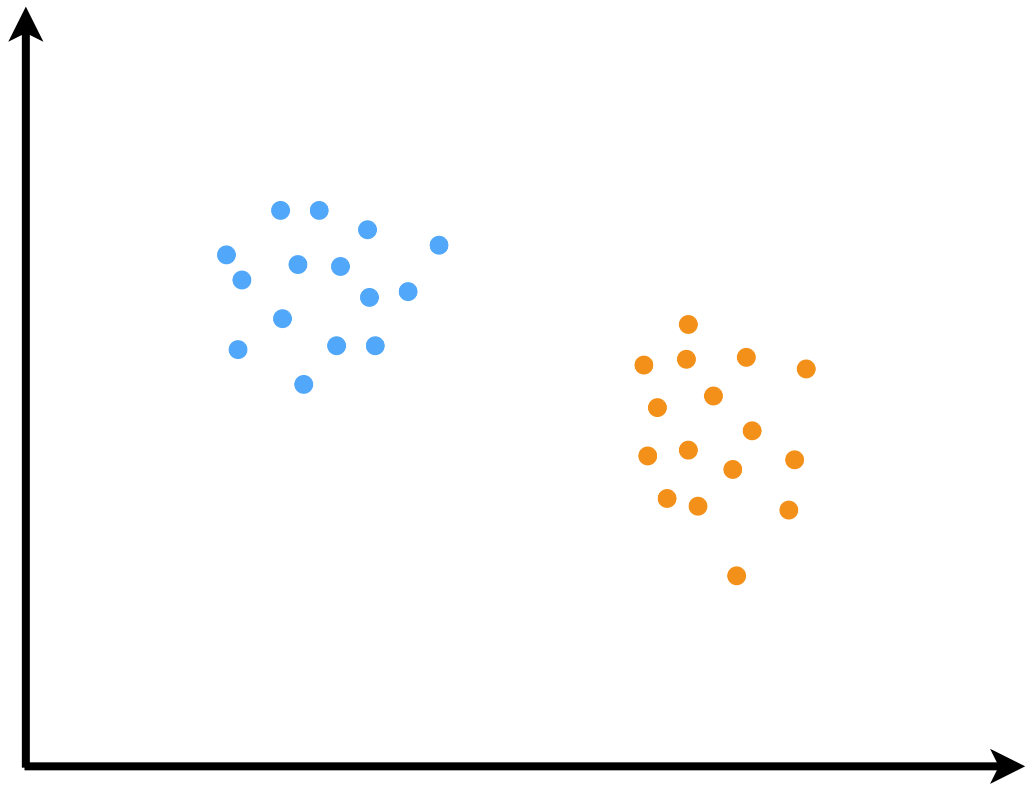
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$



- Inference goal: assignments of data points to clusters

Generative model

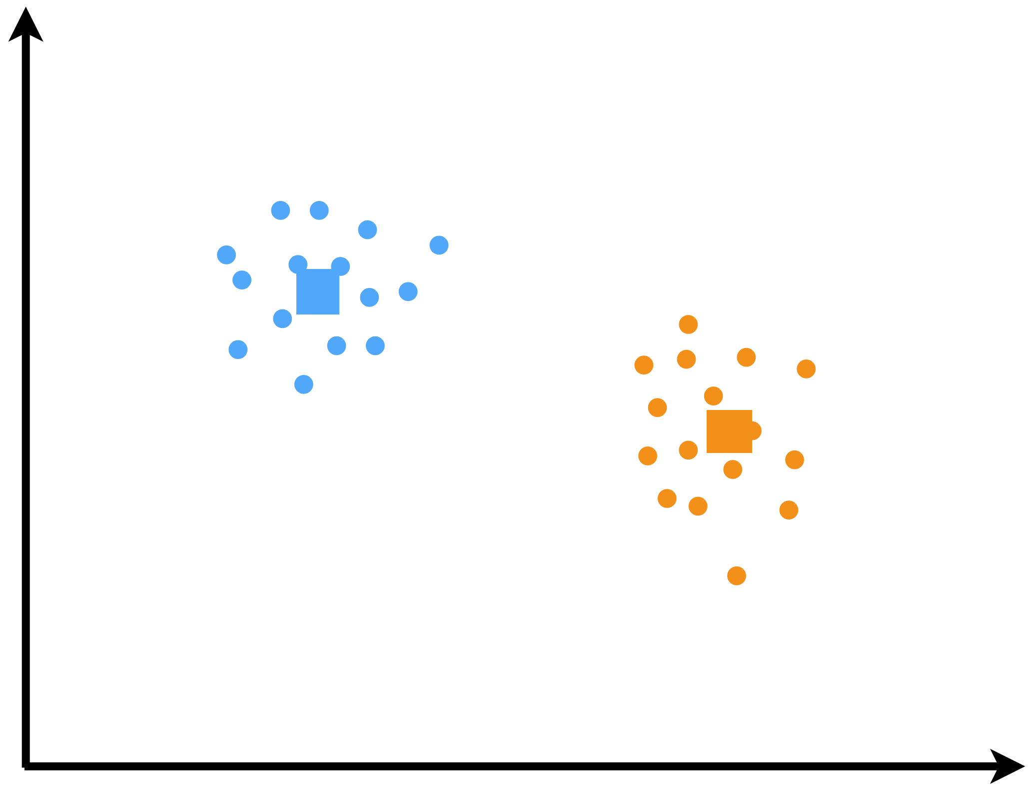
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$



- Inference goal: assignments of data points to clusters, cluster parameters

Generative model

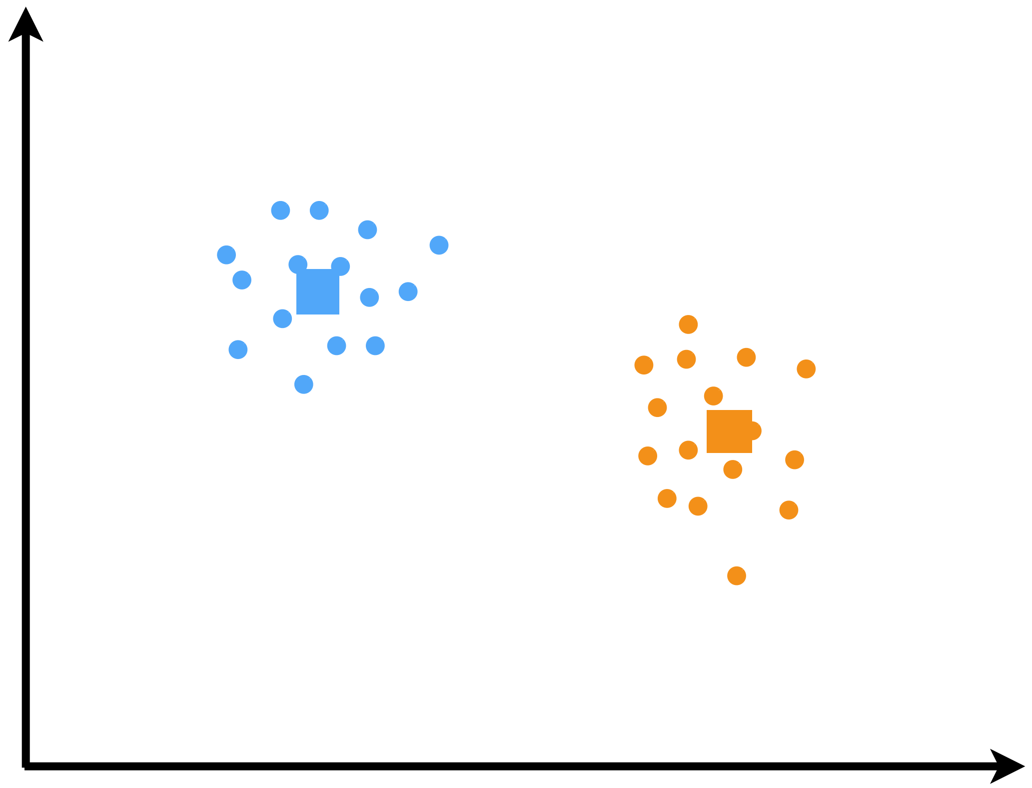
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$



- Inference goal: assignments of data points to clusters, cluster parameters

Generative model

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

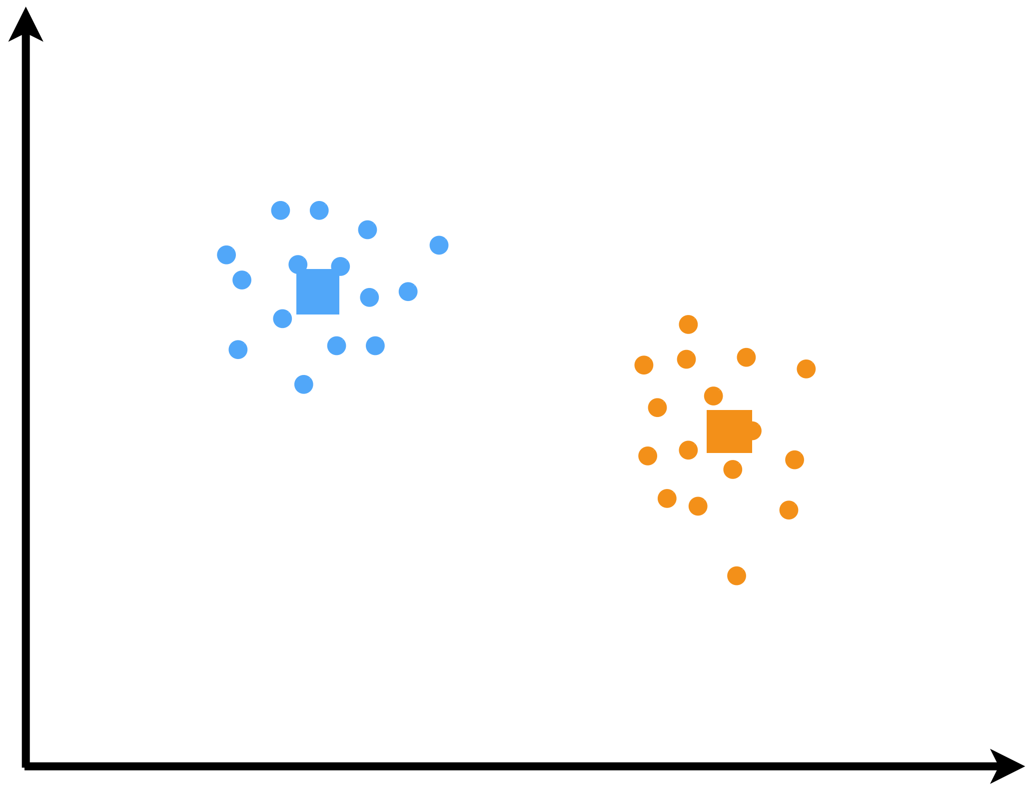


- Inference goal: assignments of data points to clusters, cluster parameters

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

Generative model

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



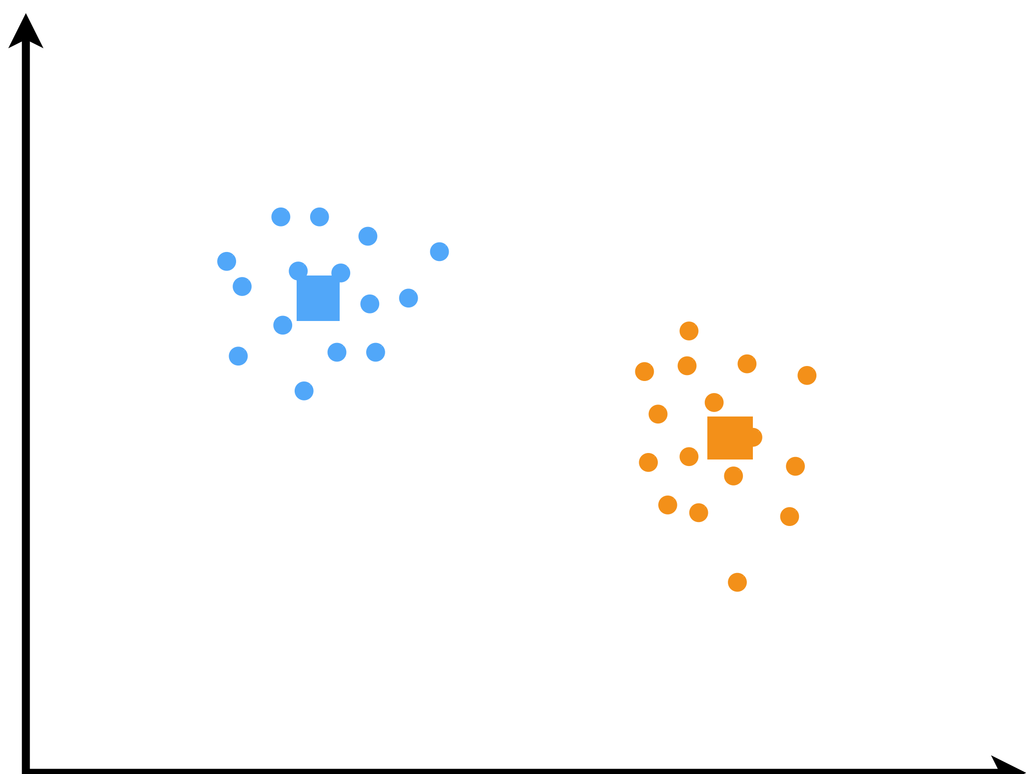
- Inference goal: assignments of data points to clusters, cluster parameters

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

- Don't know $z_{1:N}, \mu_1, \mu_2$

Generative model

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



- Inference goal: assignments of data points to clusters, cluster parameters

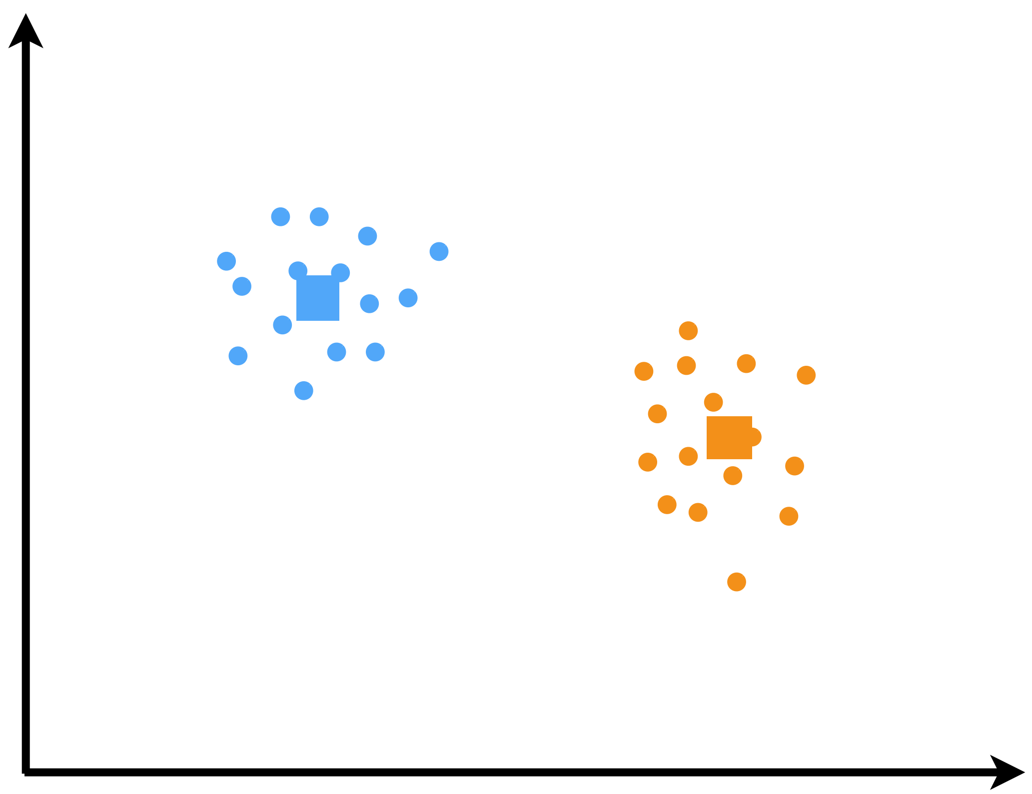
$$x_n \overset{indep}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$

- Don't know $z_{1:N}, \mu_1, \mu_2$

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

Generative model

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

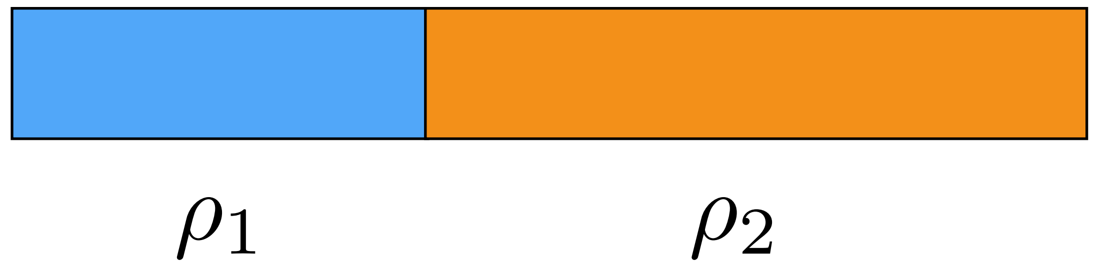


- Inference goal: assignments of data points to clusters, cluster parameters

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

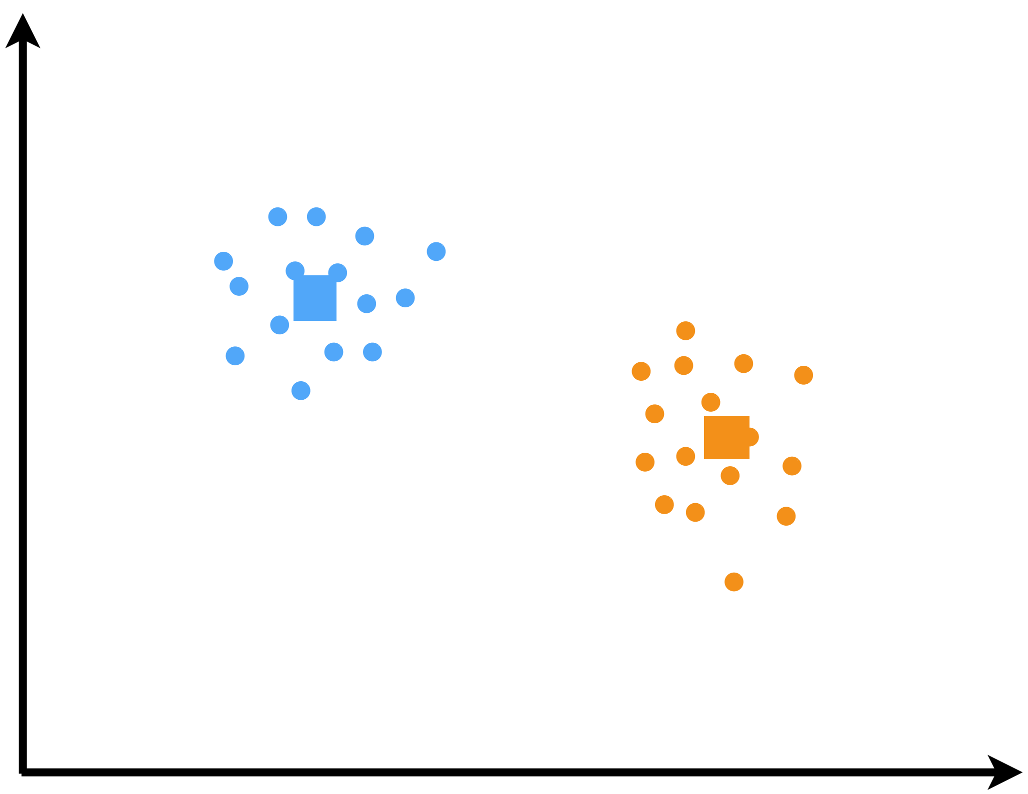
- Don't know $z_{1:N}, \mu_1, \mu_2$

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



Generative model

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

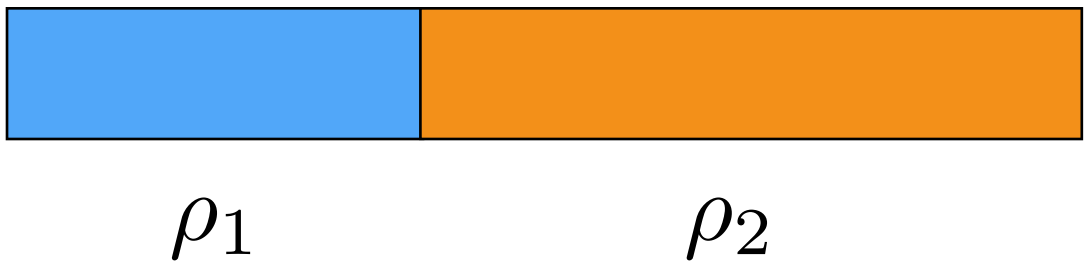


- Inference goal: assignments of data points to clusters, cluster parameters

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

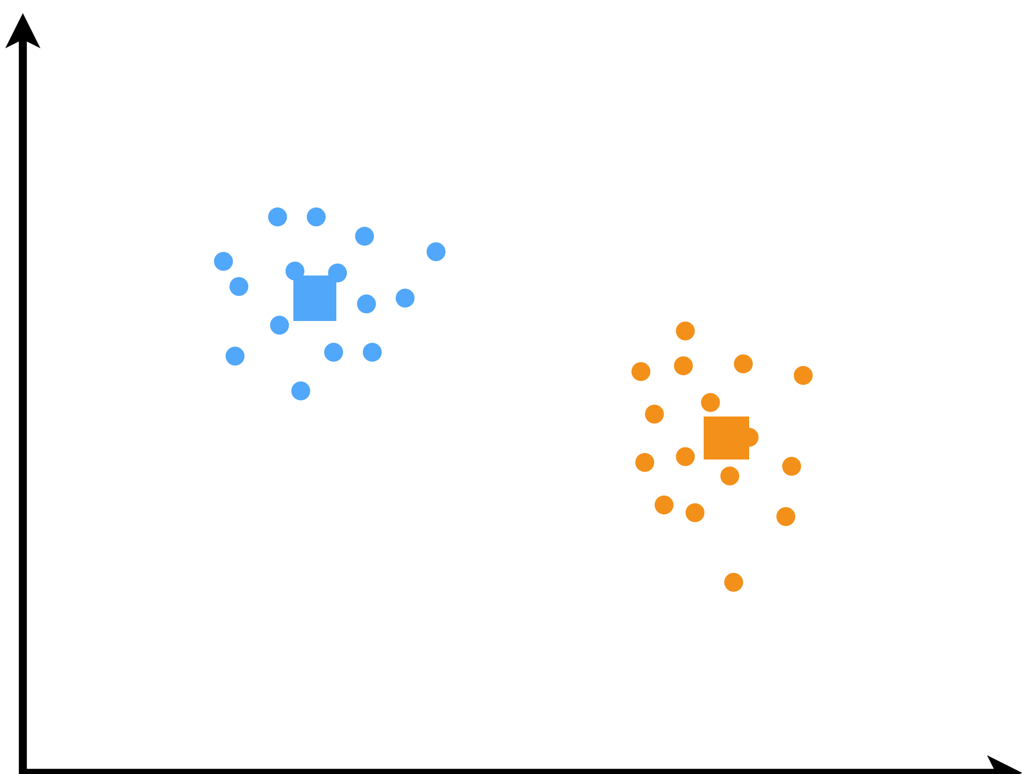
- Don't know $z_{1:N}, \mu_1, \mu_2$

$$z_n \stackrel{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$



Generative model

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



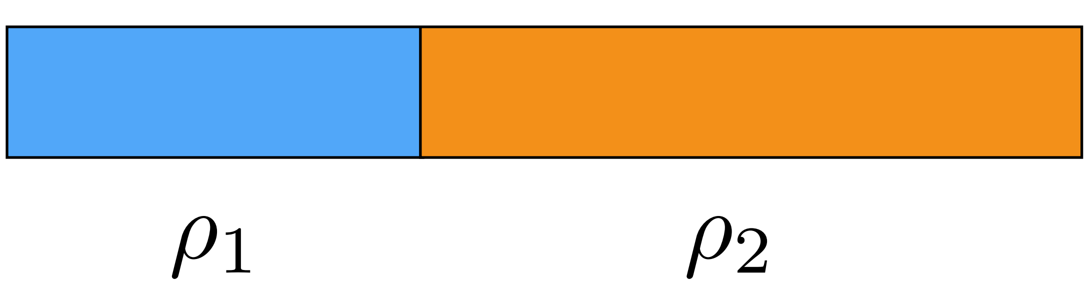
- Inference goal: assignments of data points to clusters, cluster parameters

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

- Don't know $z_{1:N}, \mu_1, \mu_2$

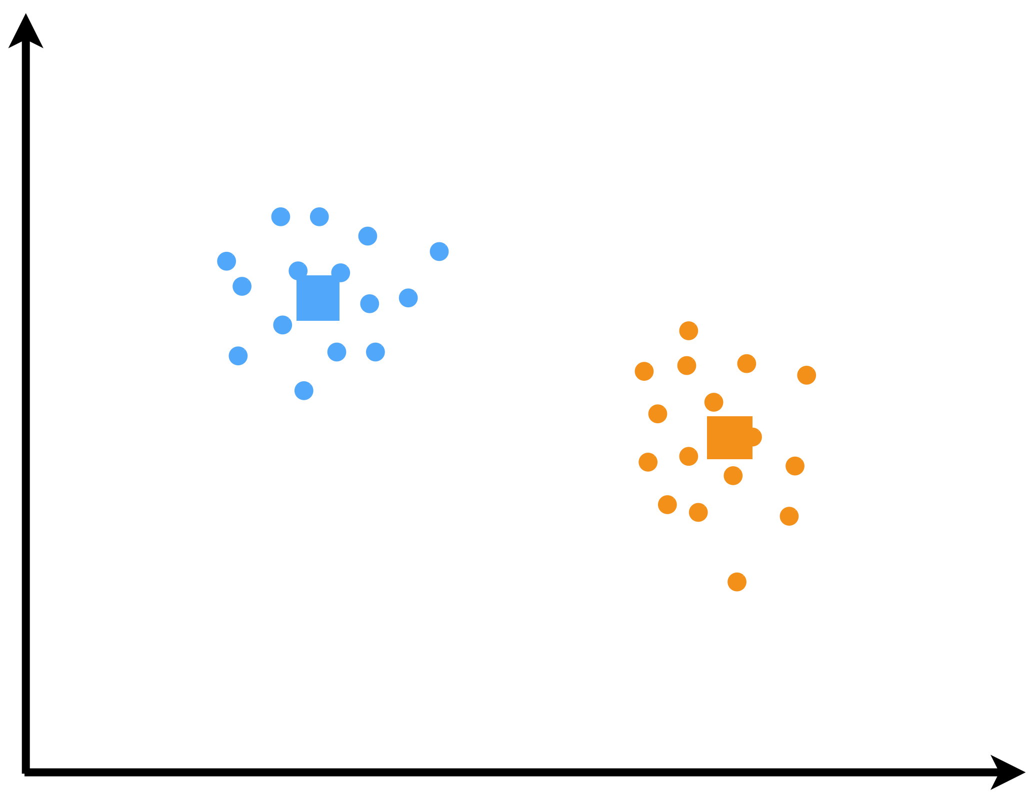
$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

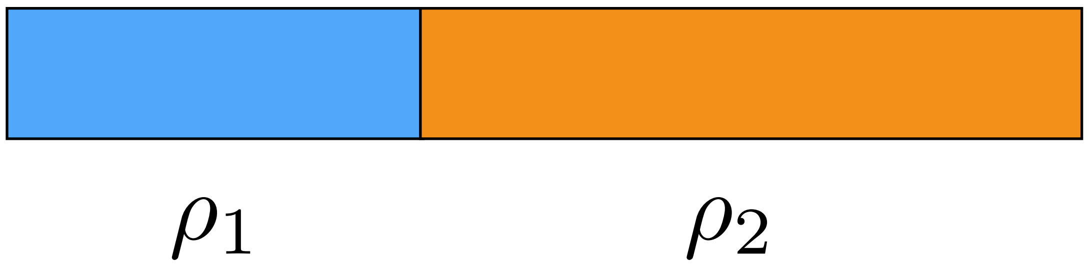
- Don't know ρ_1, ρ_2



Generative model

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





- Inference goal: assignments of data points to clusters, cluster parameters

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

- Don't know $z_{1:N}, \mu_1, \mu_2$

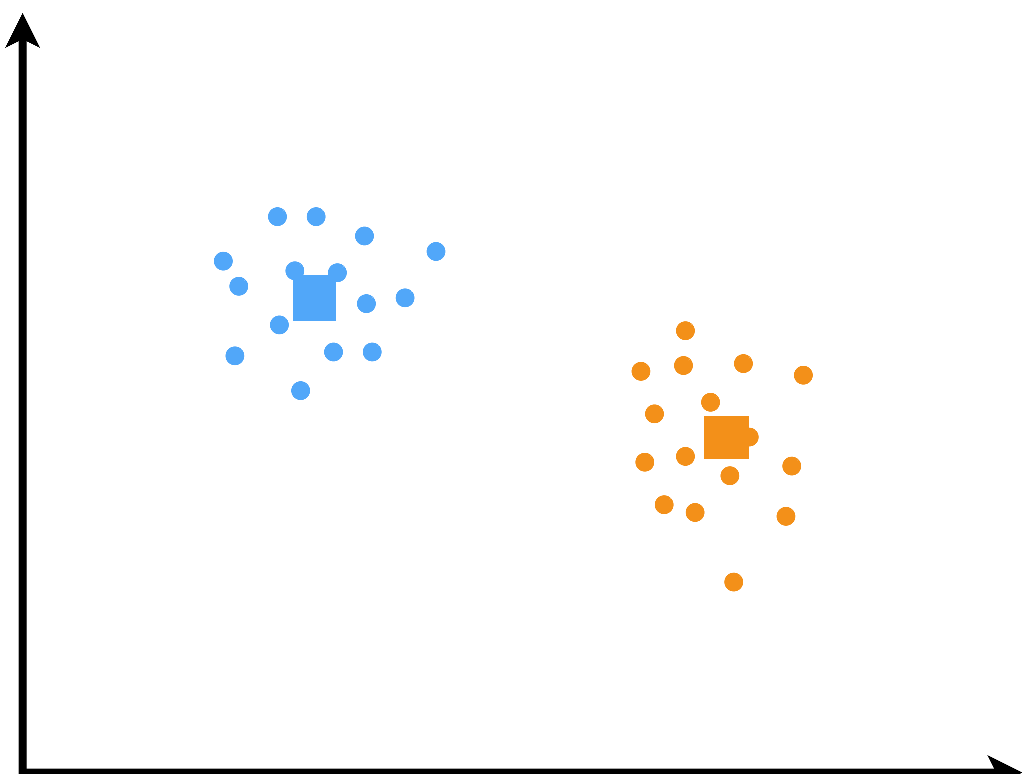
$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

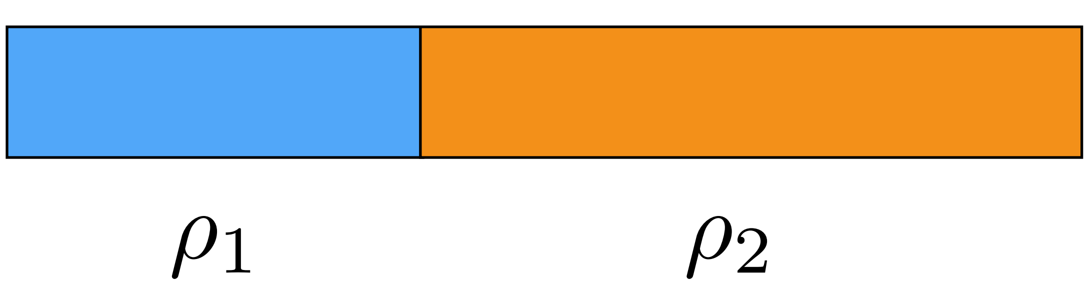
- Don't know ρ_1, ρ_2

$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$

Generative model

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





- Inference goal: assignments of data points to clusters, cluster parameters
- Finite Gaussian mixture model (K=2 clusters)

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

- Don't know $z_{1:N}, \mu_1, \mu_2$

$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

- Don't know ρ_1, ρ_2

$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$

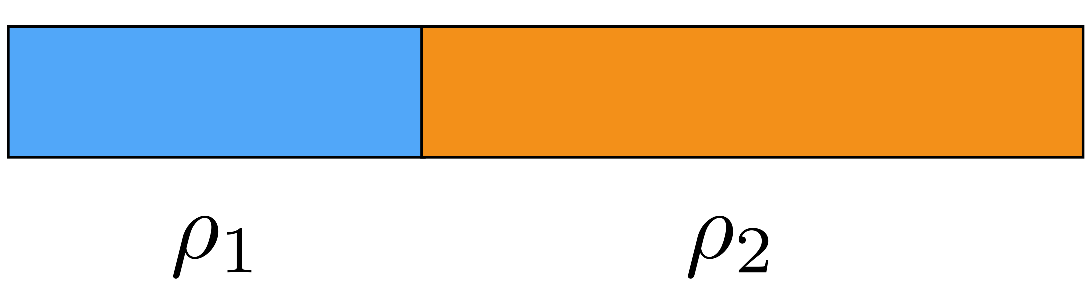
Generative model

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

- Inference goal: assignments of data points to clusters, cluster parameters
- Finite Gaussian mixture model (*K*=2 clusters)
  $$x_n \overset{indep}{\sim} \mathcal{N}(\mu_{z_n}, \Sigma)$$
- Don't know $z_{1:N}, \mu_1, \mu_2$
  $$z_n \overset{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
  $$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$
- Don't know ρ_1, ρ_2
  $$\rho_1 \sim \text{Beta}(a_1, a_2)$$
  $$\rho_2 = 1 - \rho_1$$

Generative model

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



- Inference goal: assignments of data points to clusters, cluster parameters
- Finite Gaussian mixture model (*K*=2 clusters)

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

- Don't know $z_{1:N}, \mu_1, \mu_2$

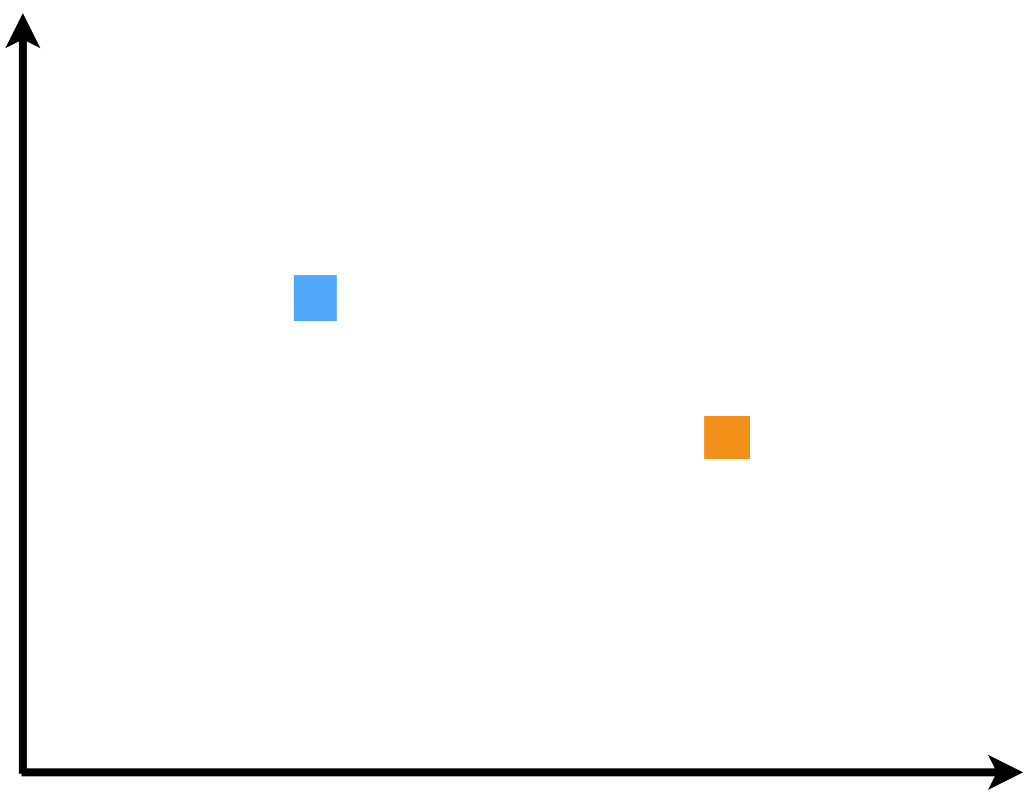
$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

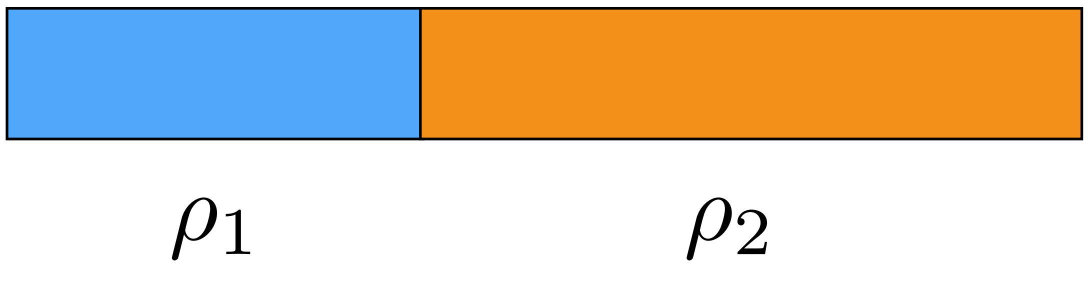
- Don't know ρ_1, ρ_2

$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$

Generative model

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





- Inference goal: assignments of data points to clusters, cluster parameters
- Finite Gaussian mixture model (*K*=2 clusters)

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

- Don't know $z_{1:N}, \mu_1, \mu_2$

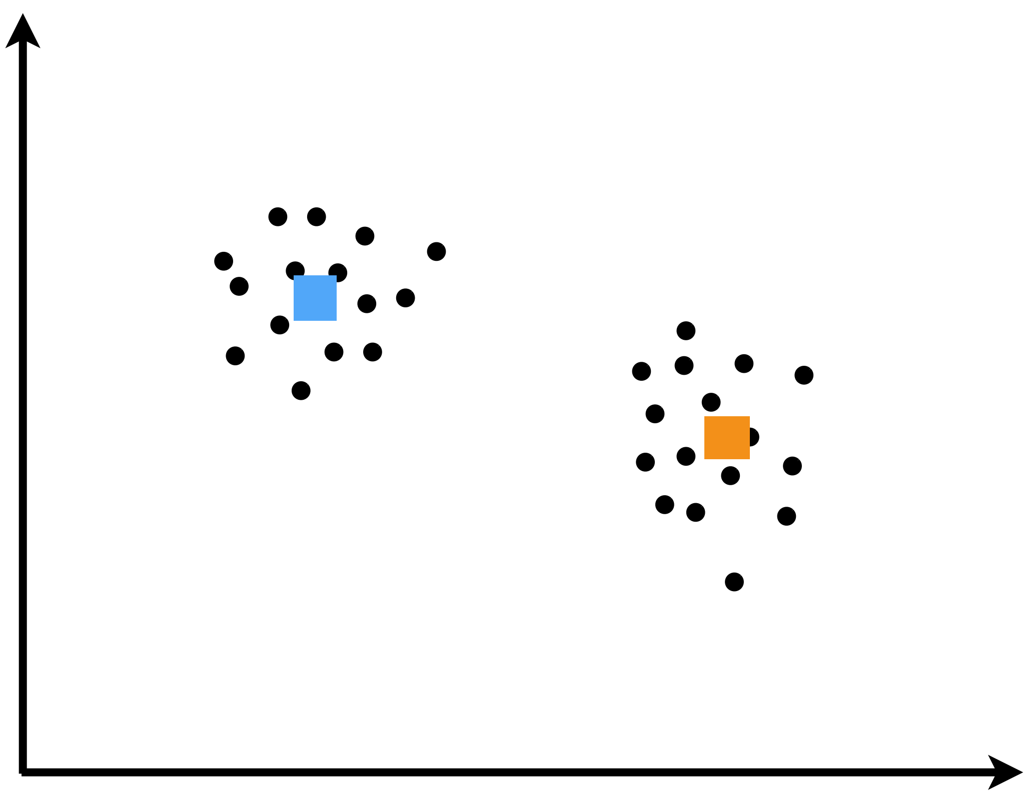
$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

- Don't know ρ_1, ρ_2

$$\rho_1 \sim \text{Beta}(a_1, a_2)$$
$$\rho_2 = 1 - \rho_1$$

Generative model

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



- Inference goal: assignments of data points to clusters, cluster parameters
- Finite Gaussian mixture model (K=2 clusters)

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

- Don't know $z_{1:N}, \mu_1, \mu_2$

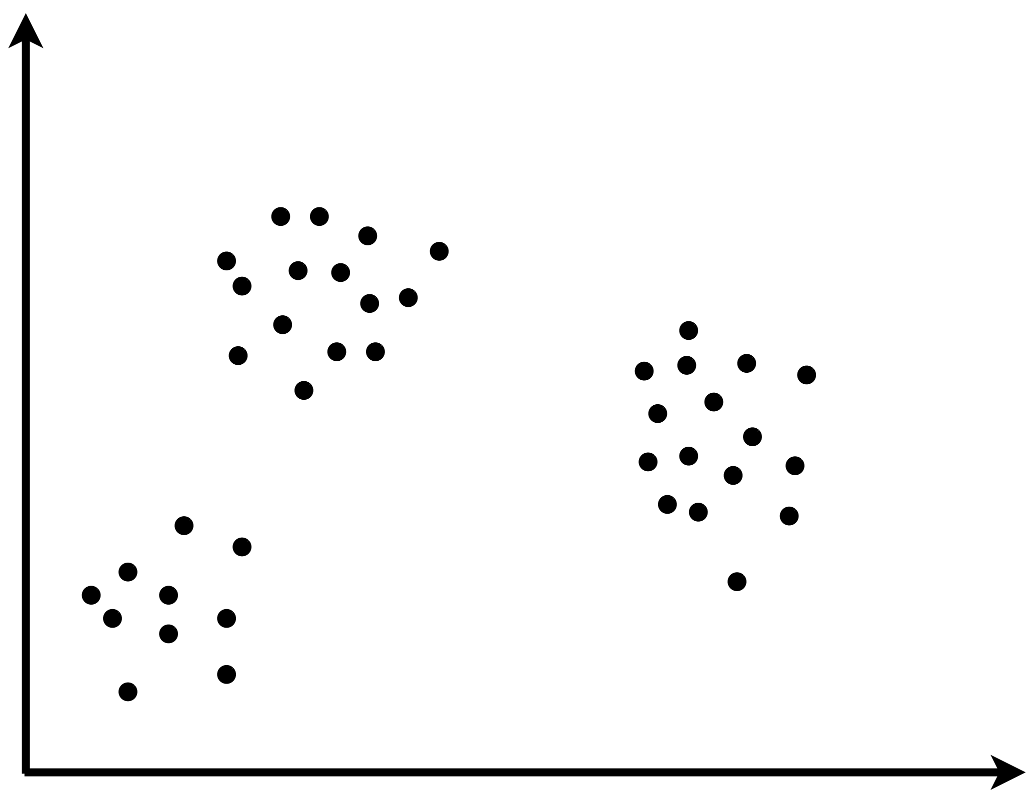
$$z_n \overset{iid}{\sim} \text{Categorical}(\rho_1, \rho_2)$$
$$\mu_k \overset{iid}{\sim} \mathcal{N}(\mu_0, \Sigma_0)$$

- 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 clusters)

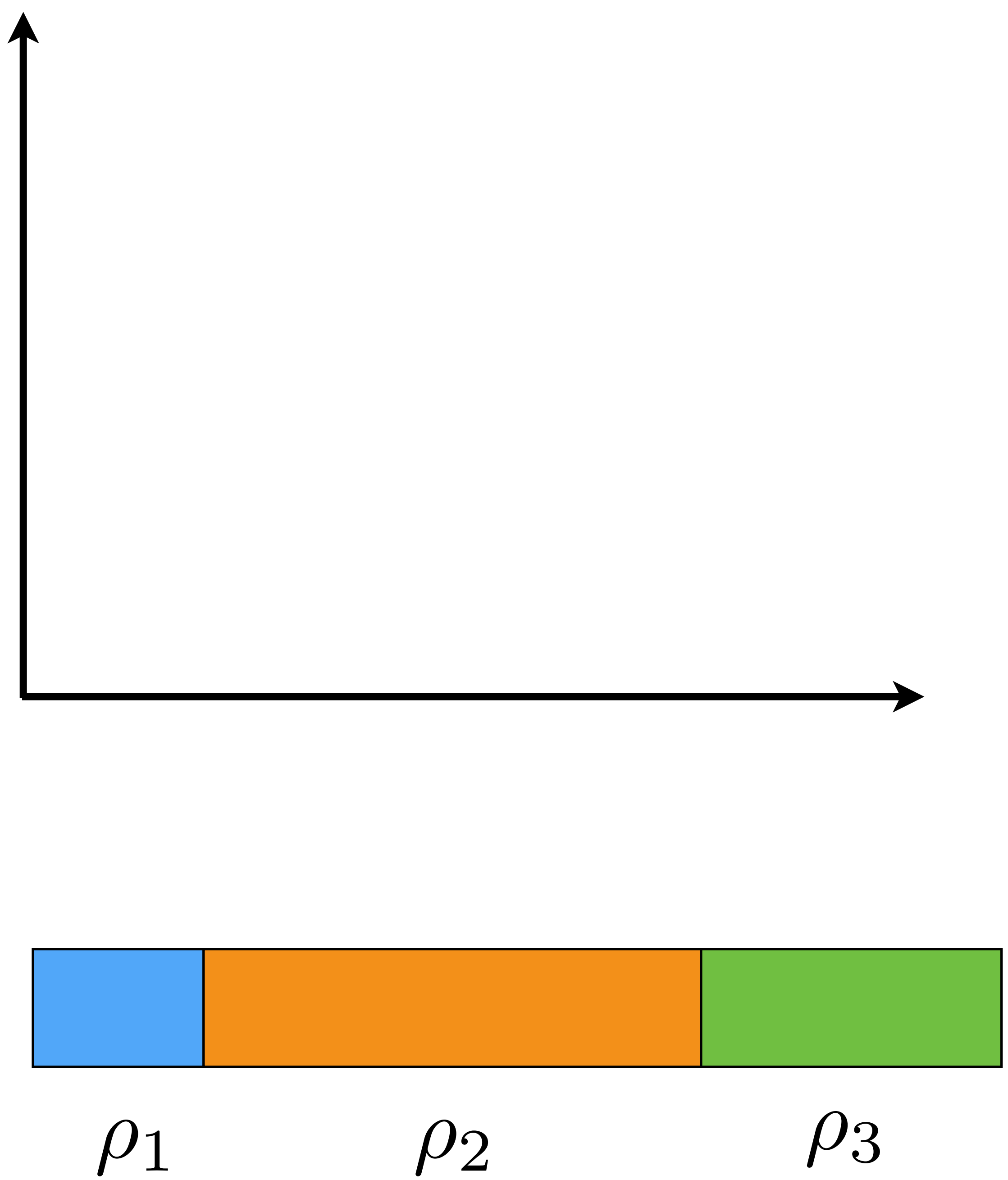
Generative model

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

- Finite Gaussian mixture model (K clusters)

Generative model

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



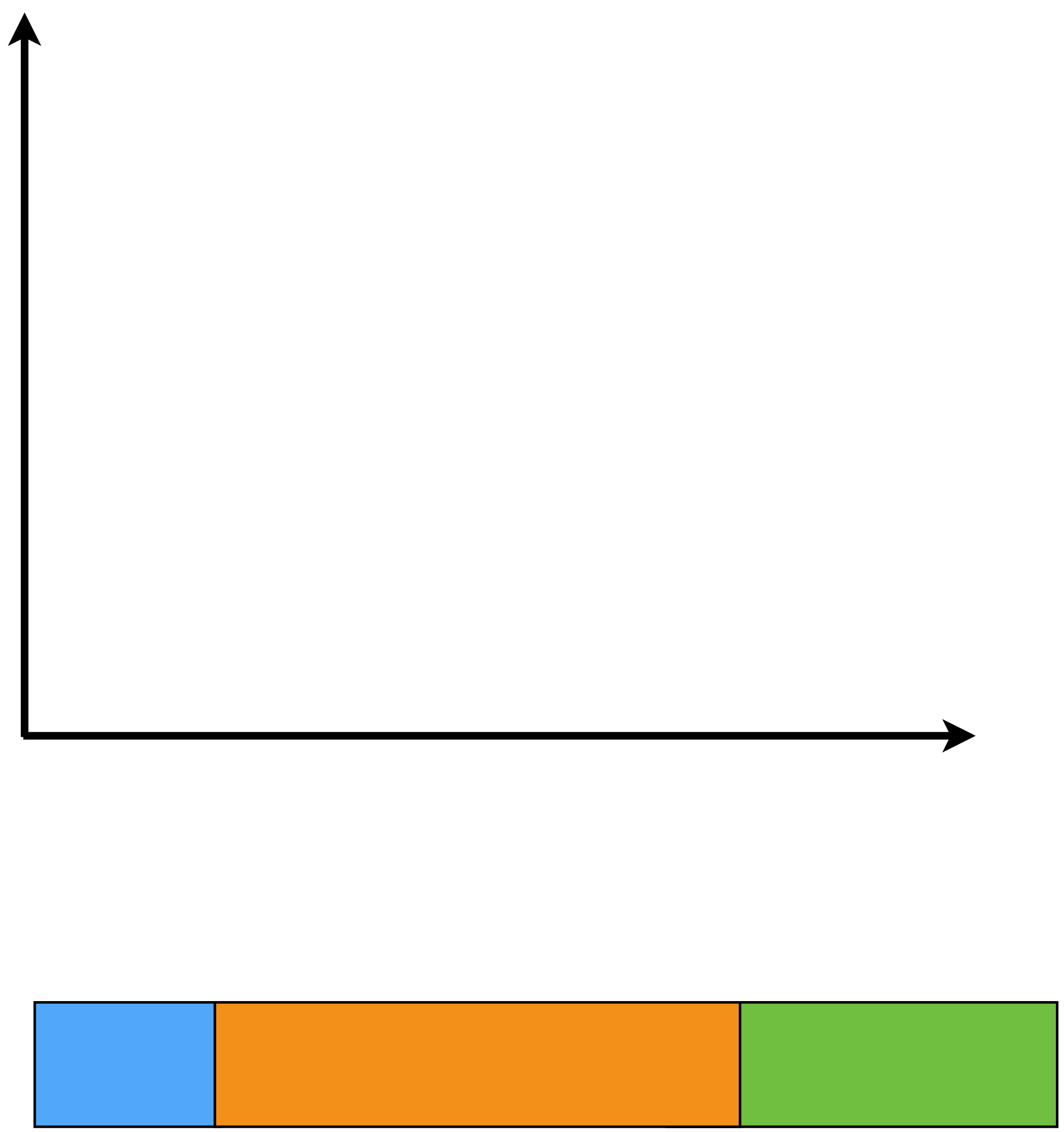
- Finite Gaussian mixture model (K clusters)

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)

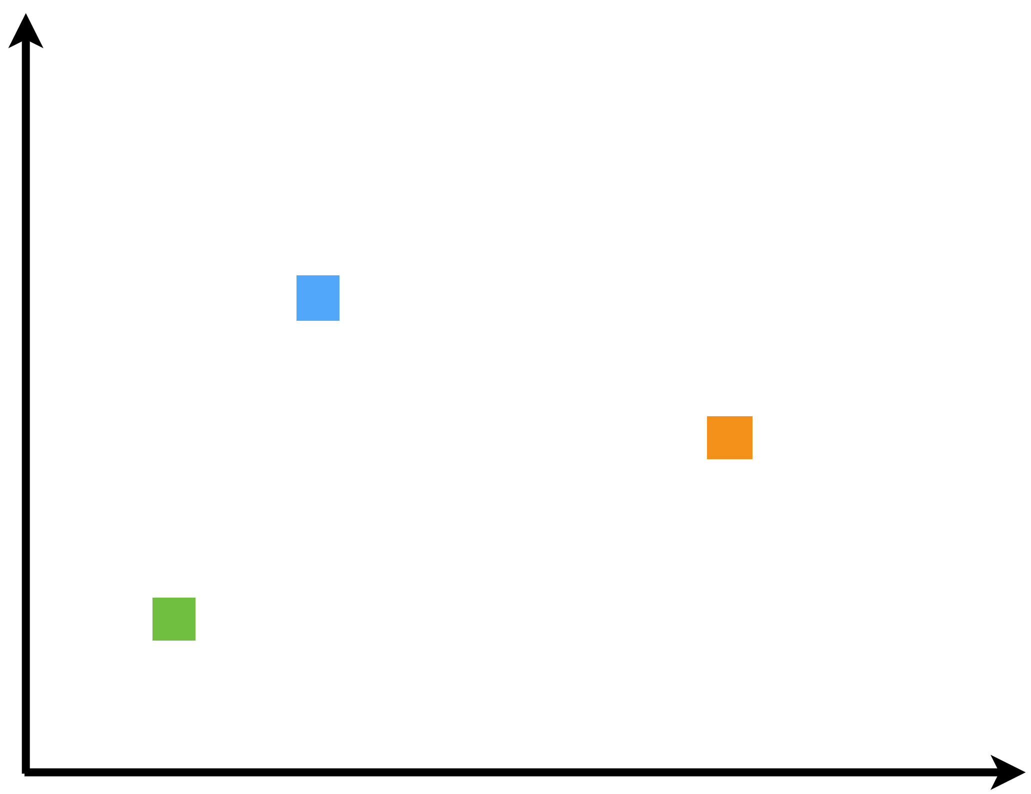
$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$$



ρ_1 ρ_2 ρ_3

Generative model

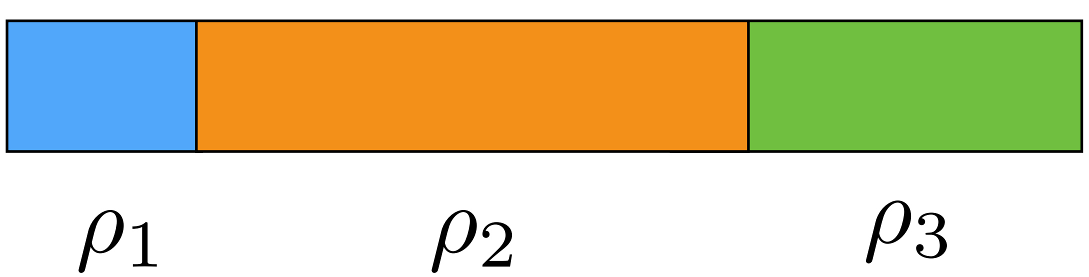
$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$



- Finite Gaussian mixture model (*K* clusters)

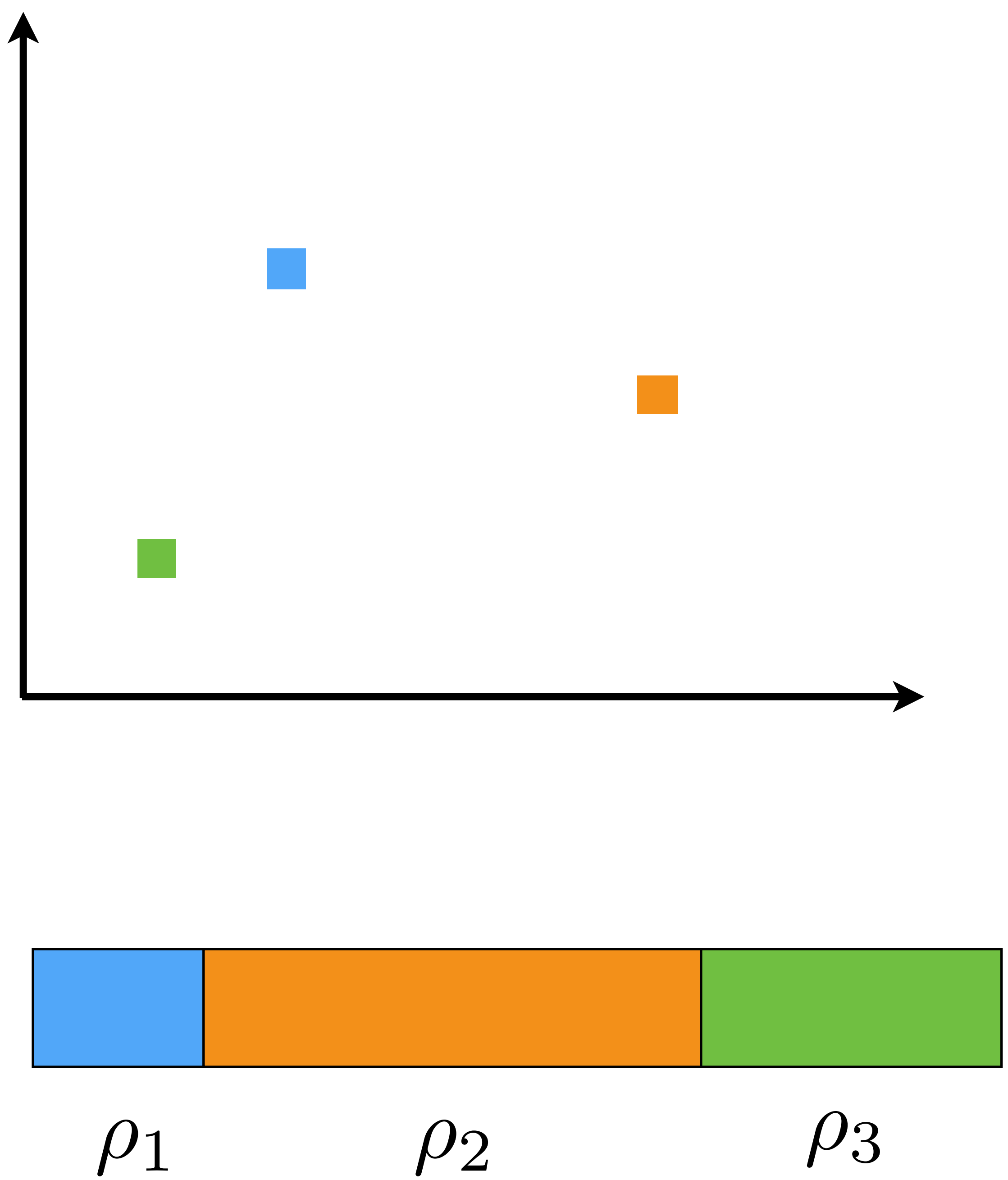
$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$$

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



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)

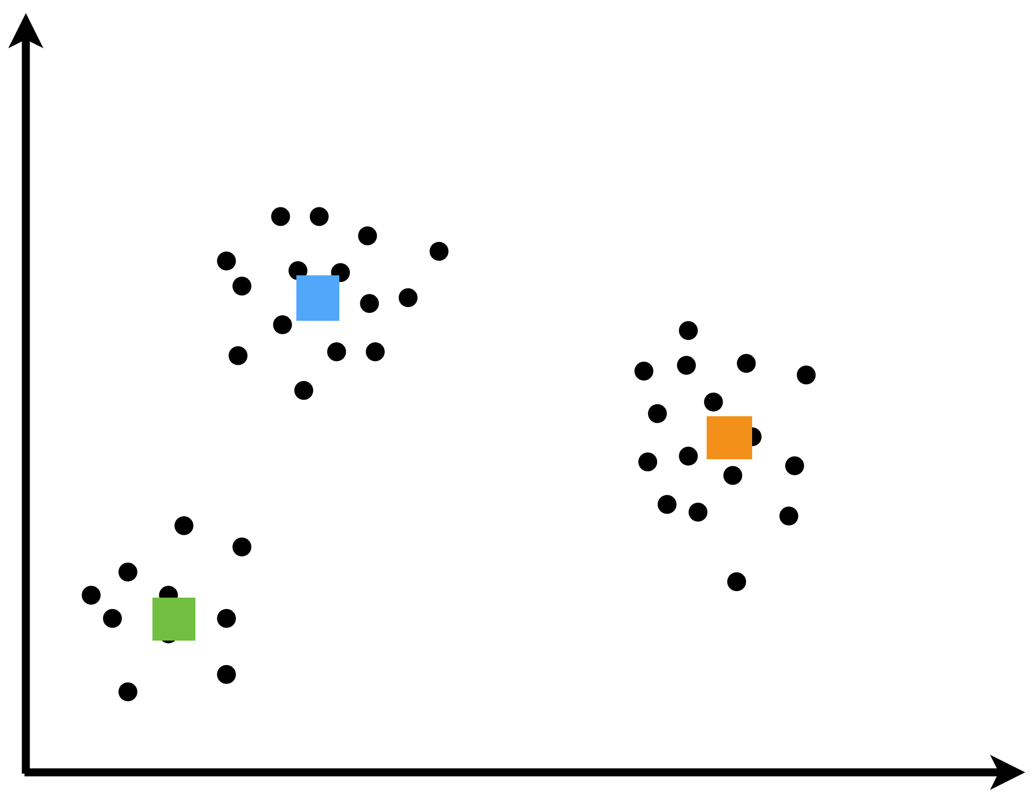
$$\rho_{1:K} \sim \text{Dirichlet}(a_{1:K})$$

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

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

Generative model

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



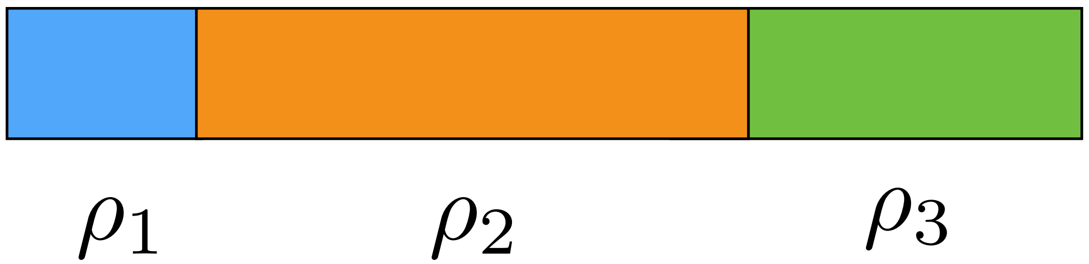
- Finite Gaussian mixture model (*K* clusters)

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

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

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

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



Dirichlet distribution review

$$\mathrm{Dirichlet}(\rho_{1:K}|a_{1:K}) = \frac{\Gamma(\sum_{k=1}^{K} a_k)}{\prod_{k=1}^{K}\Gamma(a_k)} \prod_{k=1}^{K} \rho_k^{a_k - 1} \qquad a_k > 0$$

Dirichlet distribution review

$$a_k > 0$$

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

$$\sum_k \rho_k = 1$$

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

Dirichlet distribution review

$$a_k > 0$$

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

$$\sum_k \rho_k = 1$$

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

- What happens?

Dirichlet distribution review

$$a_k > 0$$

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

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

$$\sum_k \rho_k = 1$$

a = (0.5,0.5,0.5)

- What happens?

Dirichlet distribution review

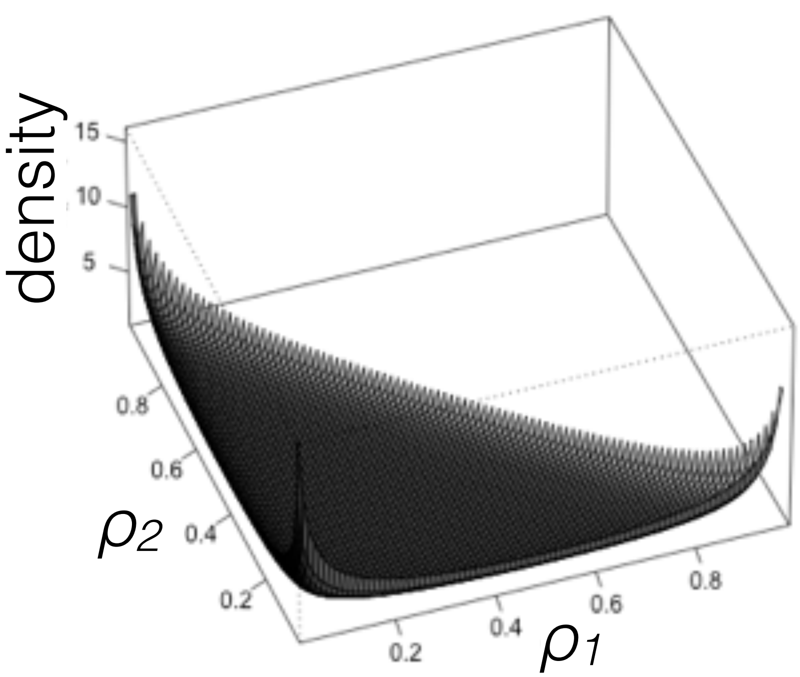
$$a_k > 0$$

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

$$\sum_k \rho_k = 1$$

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

a = (0.5,0.5,0.5)



- What happens?

Dirichlet distribution review

$$a_k > 0$$

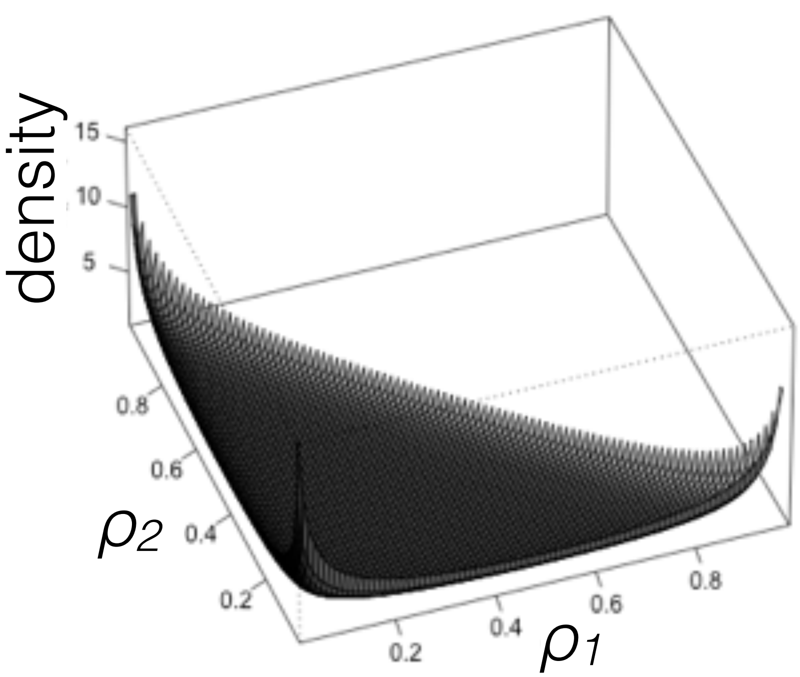
$$\rho_k \in (0,1)$$

$$\sum_k \rho_k = 1$$

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

a = (0.5,0.5,0.5)

a = (5,5,5)



- What happens?

Dirichlet distribution review

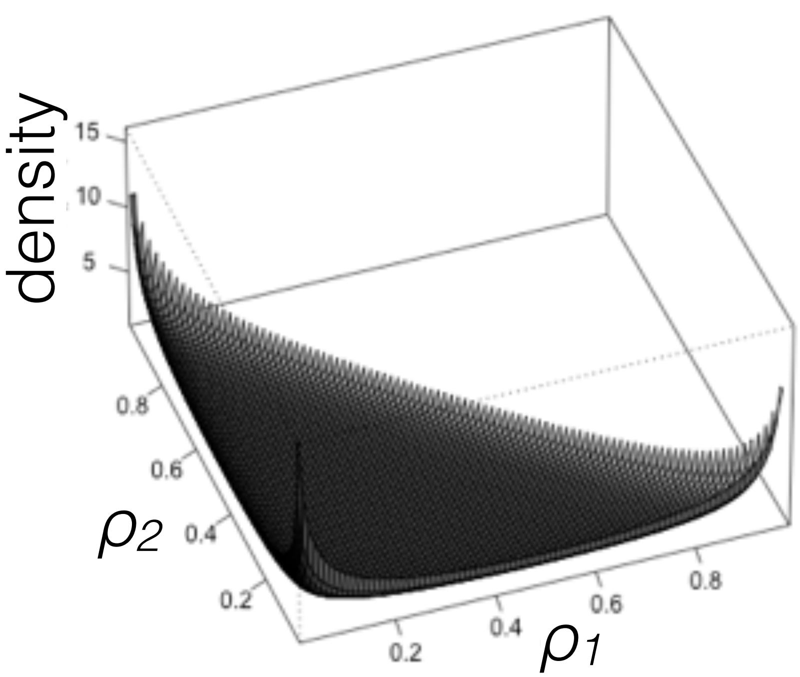
$$a_k > 0$$

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

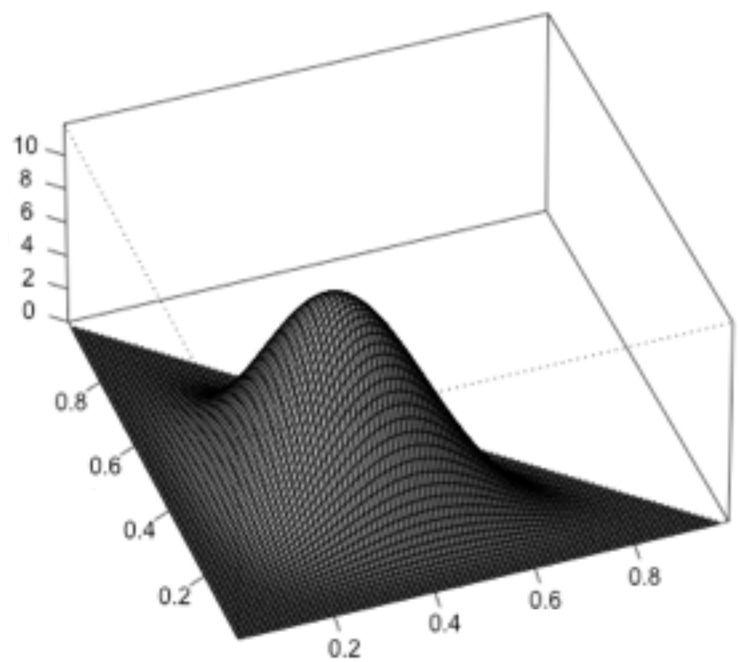
$$\sum_k \rho_k = 1$$

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

a = (0.5,0.5,0.5)



a = (5,5,5)



- What happens?

Dirichlet distribution review

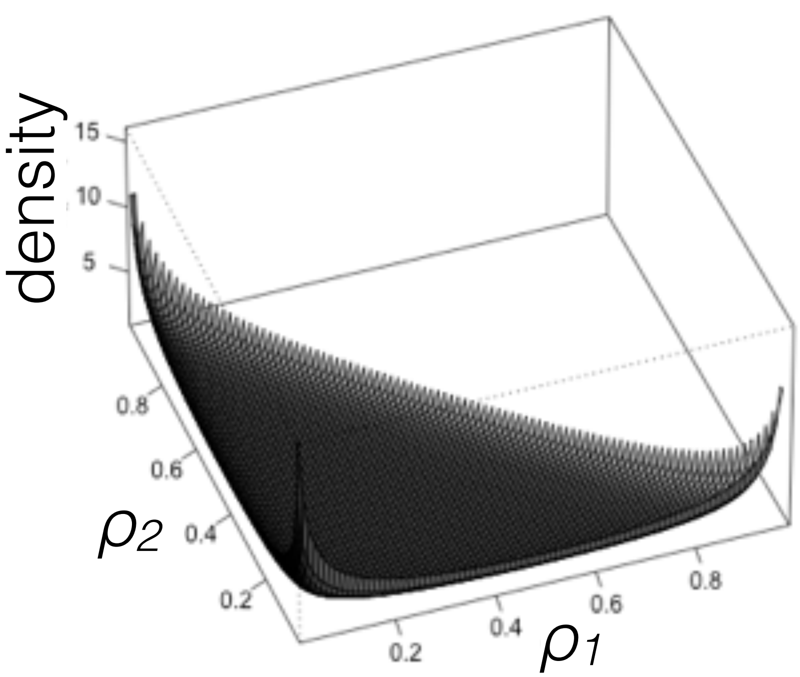
$$a_k > 0$$

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

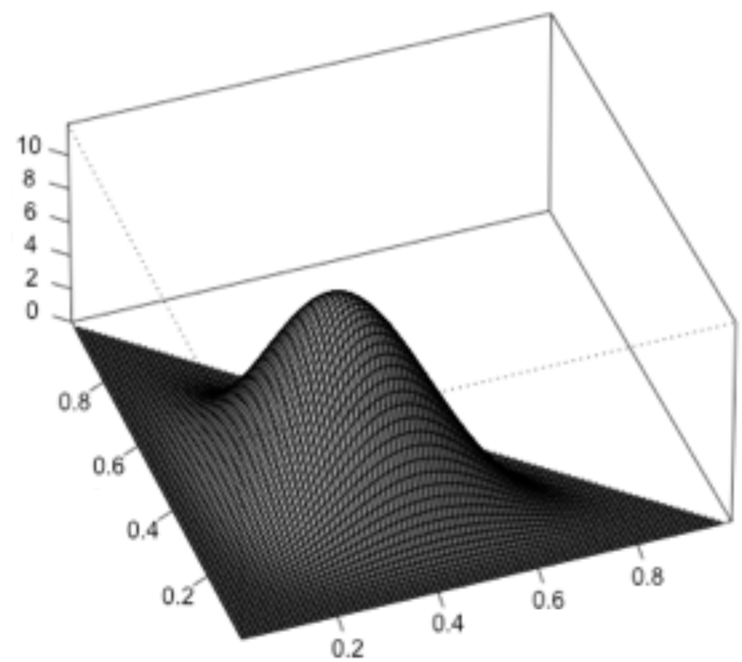
$$\sum_k \rho_k = 1$$

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

a = (0.5,0.5,0.5)



a = (5,5,5)



a = (40,10,10)

- What happens?

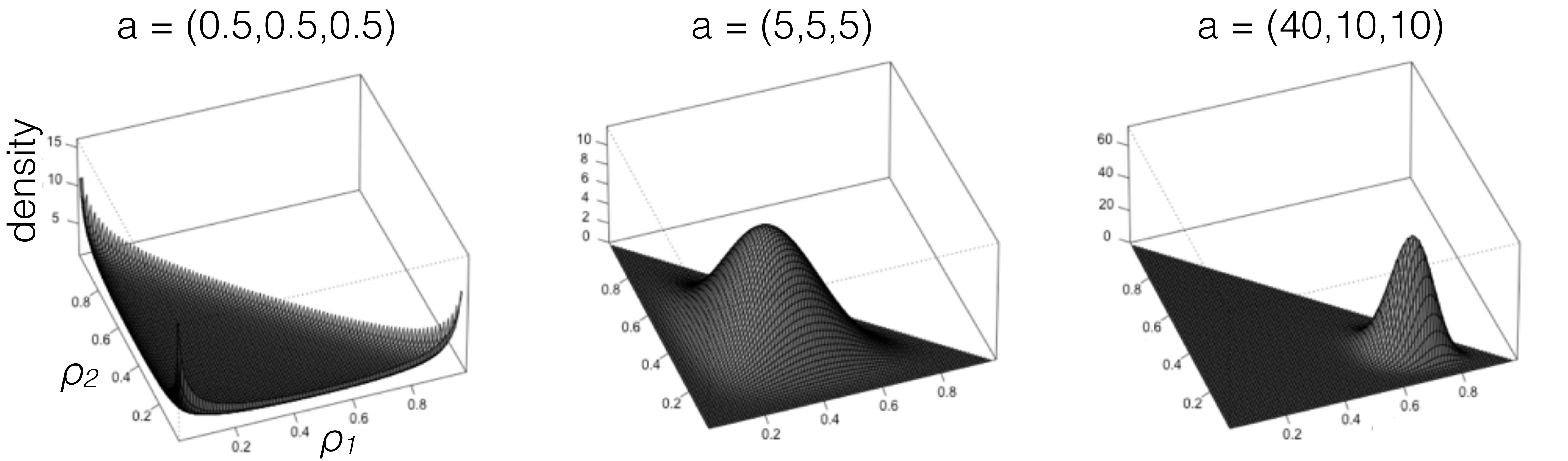
Dirichlet distribution review

$$a_k > 0$$

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

$$\sum_k \rho_k = 1$$

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



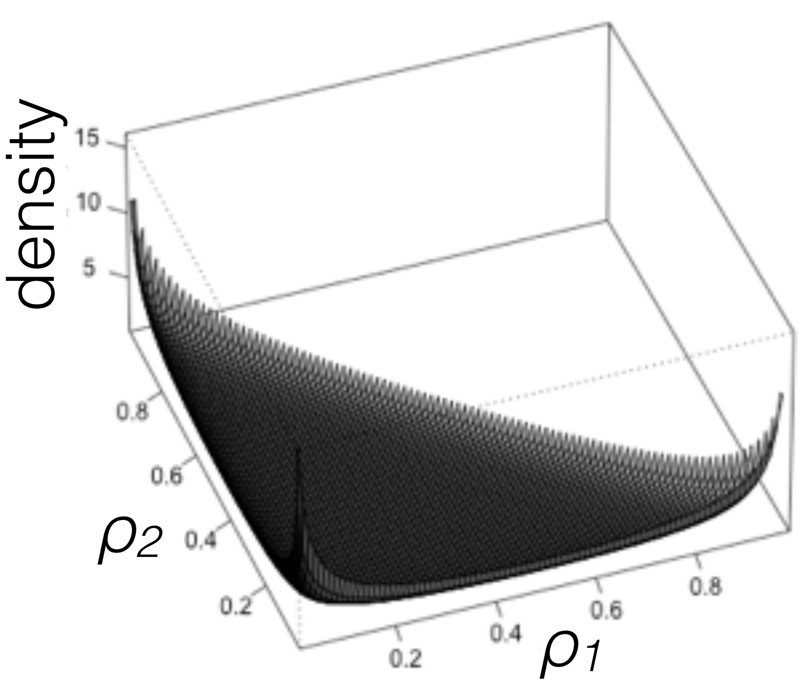
- What happens?

Dirichlet distribution review

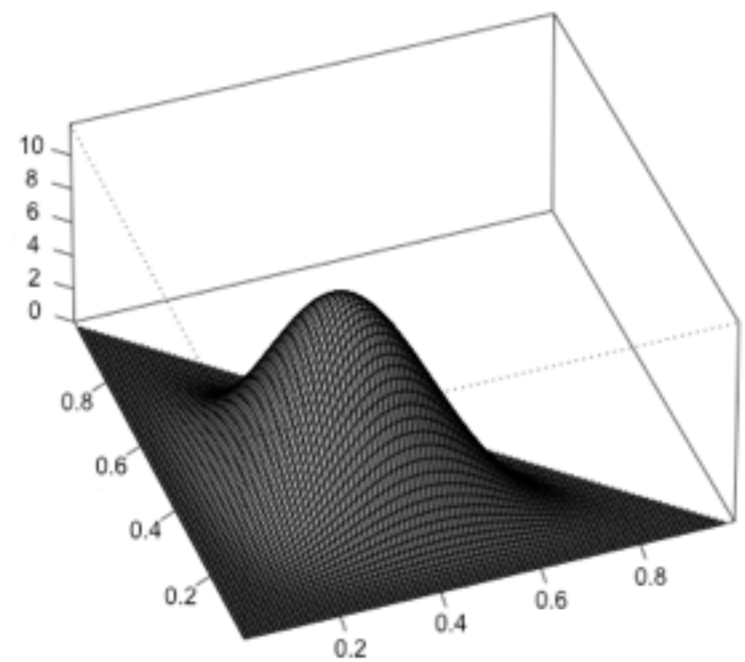
$$\mathrm{Dirichlet}(\rho_{1:K}|a_{1:K}) = \frac{\Gamma(\sum_{k=1}^{K} a_k)}{\prod_{k=1}^{K}\Gamma(a_k)} \prod_{k=1}^{K} \rho_k^{a_k - 1}$$

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

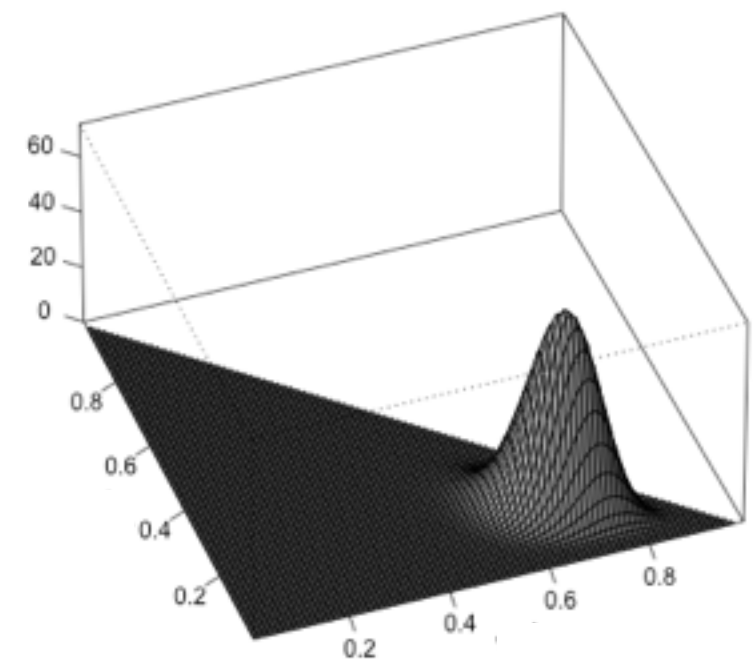
a = (0.5,0.5,0.5)



a = (5,5,5)



a = (40,10,10)



- What happens? $a = a_k = 1$

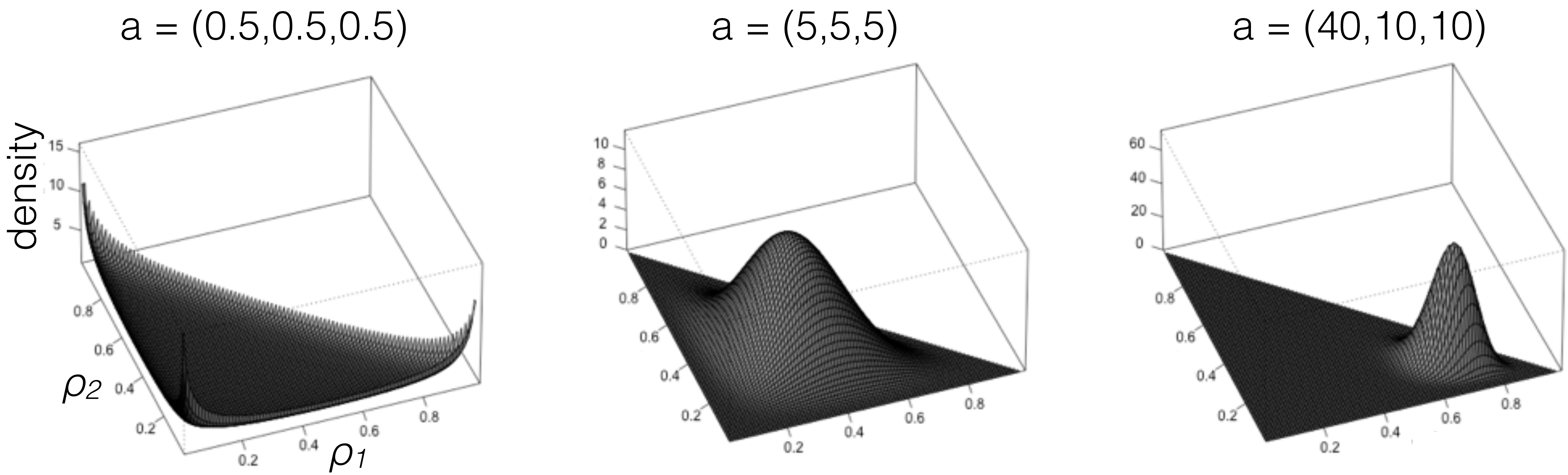
Dirichlet distribution review

$$a_k > 0$$

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

$$\sum_k \rho_k = 1$$

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



- What happens? $a = a_k = 1$

[demo]

Dirichlet distribution review

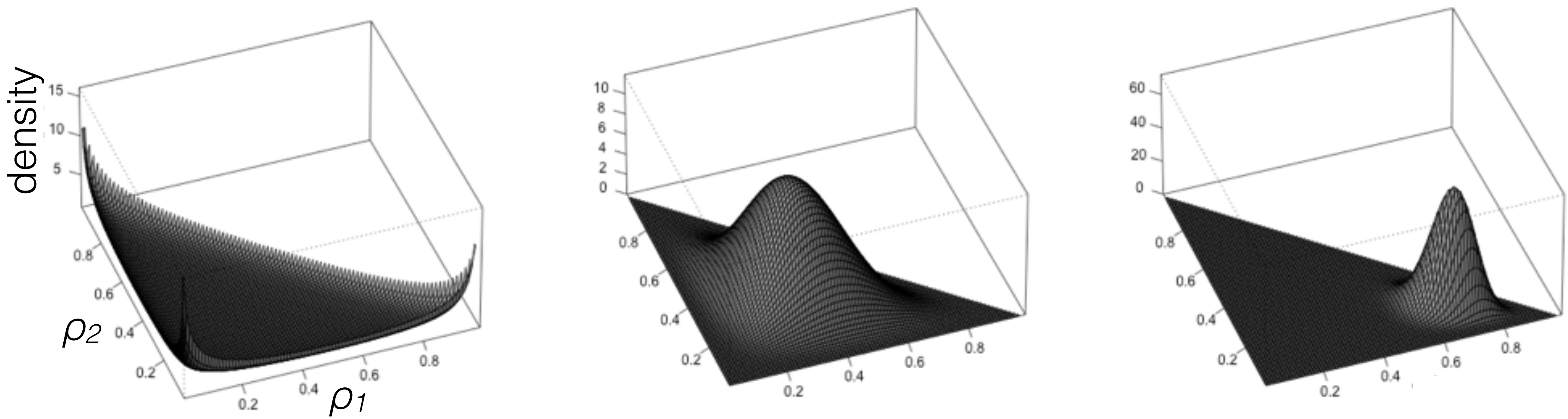
$$a_k > 0$$

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

$$\sum_k \rho_k = 1$$

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

a = (0.5,0.5,0.5) a = (5,5,5) a = (40,10,10)



- What happens? $a = a_k = 1$ $a = a_k \to 0$

[demo]

Dirichlet distribution review

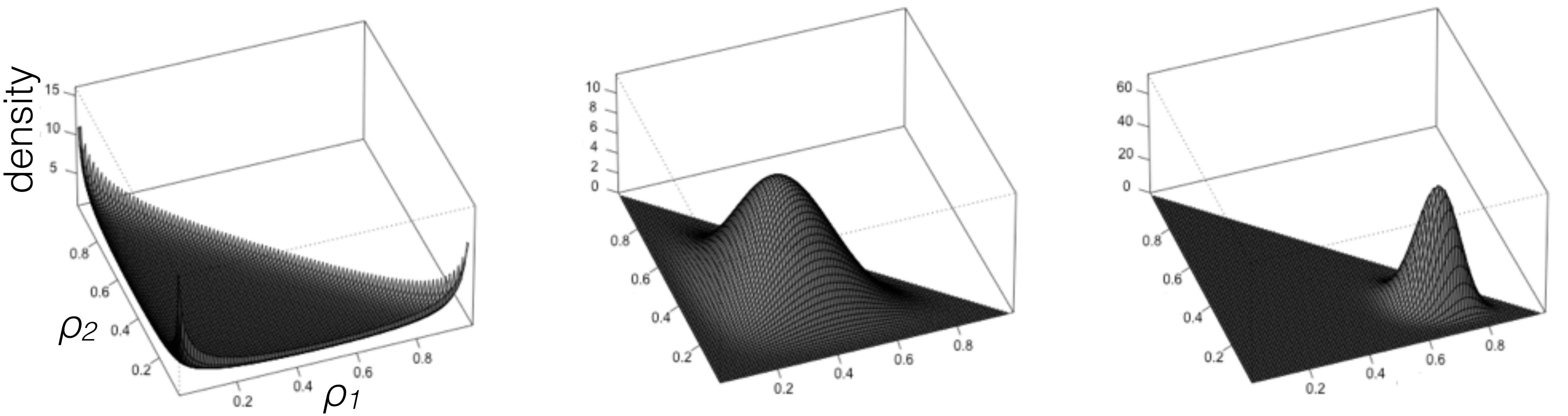
$$a_k > 0$$

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

$$\sum_k \rho_k = 1$$

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

a = (0.5,0.5,0.5) a = (5,5,5) a = (40,10,10)



- What happens? $a = a_k = 1$ $a = a_k \to 0$ $a = a_k \to \infty$

[demo]

Dirichlet distribution review

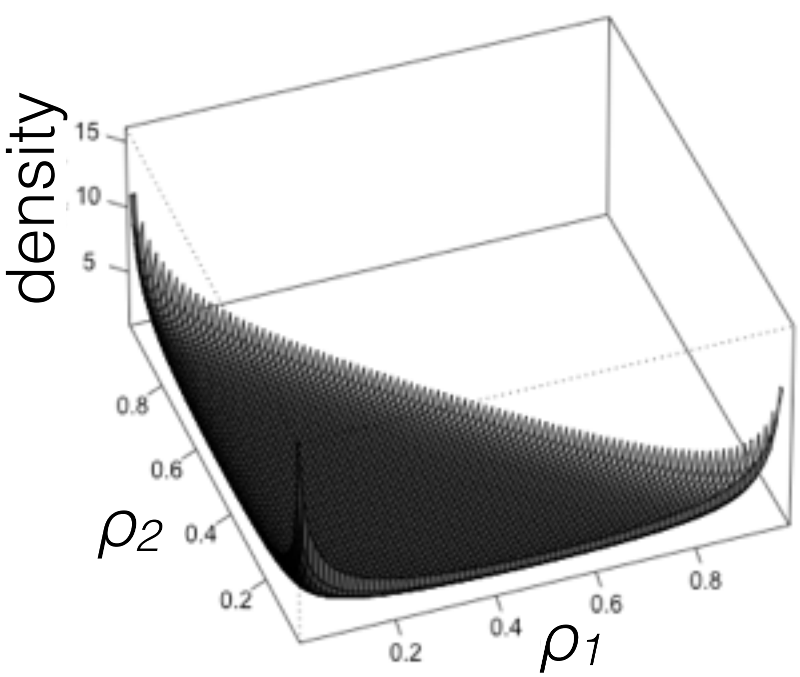
$$a_k > 0$$

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

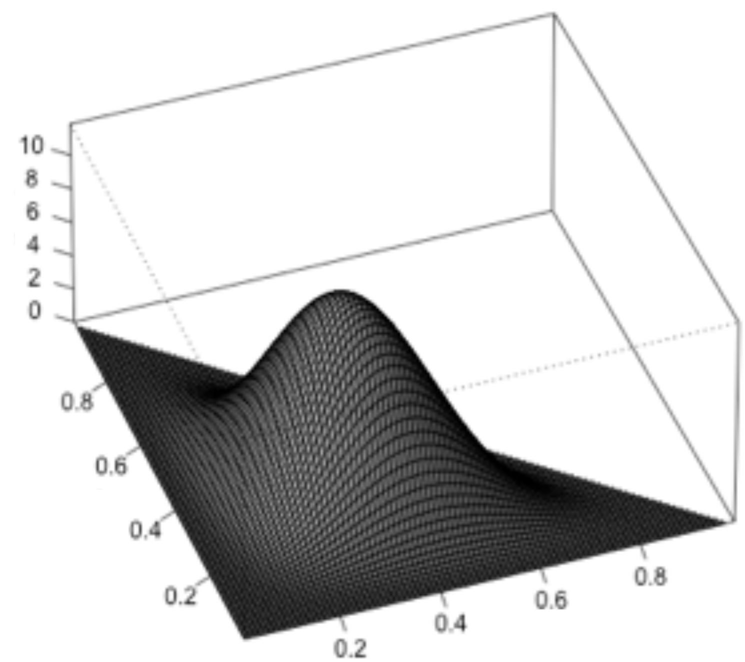
$$\mathrm{Dirichlet}(\rho_{1:K} | a_{1:K}) = \frac{\Gamma(\sum_{k=1}^{K} a_k)}{\prod_{k=1}^{K} \Gamma(a_k)} \prod_{k=1}^{K} \rho_k^{a_k - 1}$$

$$\sum_k \rho_k = 1$$

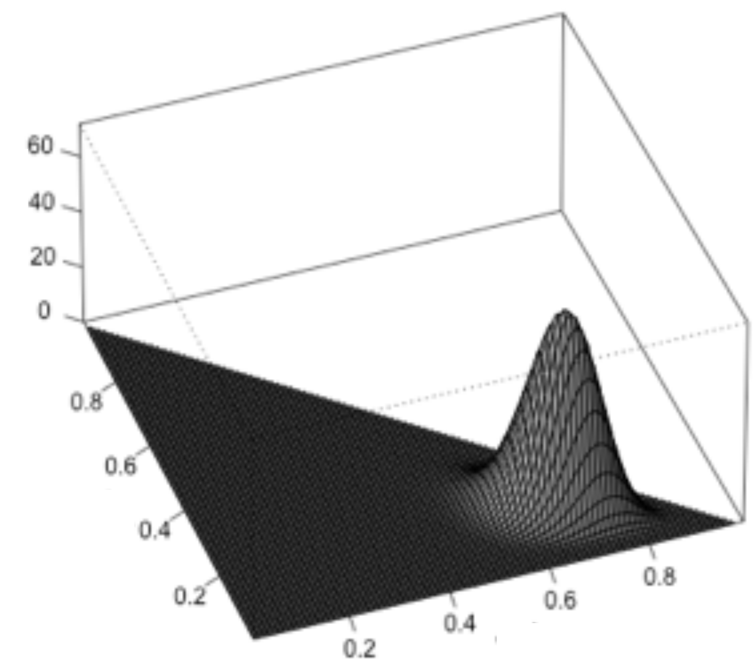
a = (0.5,0.5,0.5)



a = (5,5,5)



a = (40,10,10)

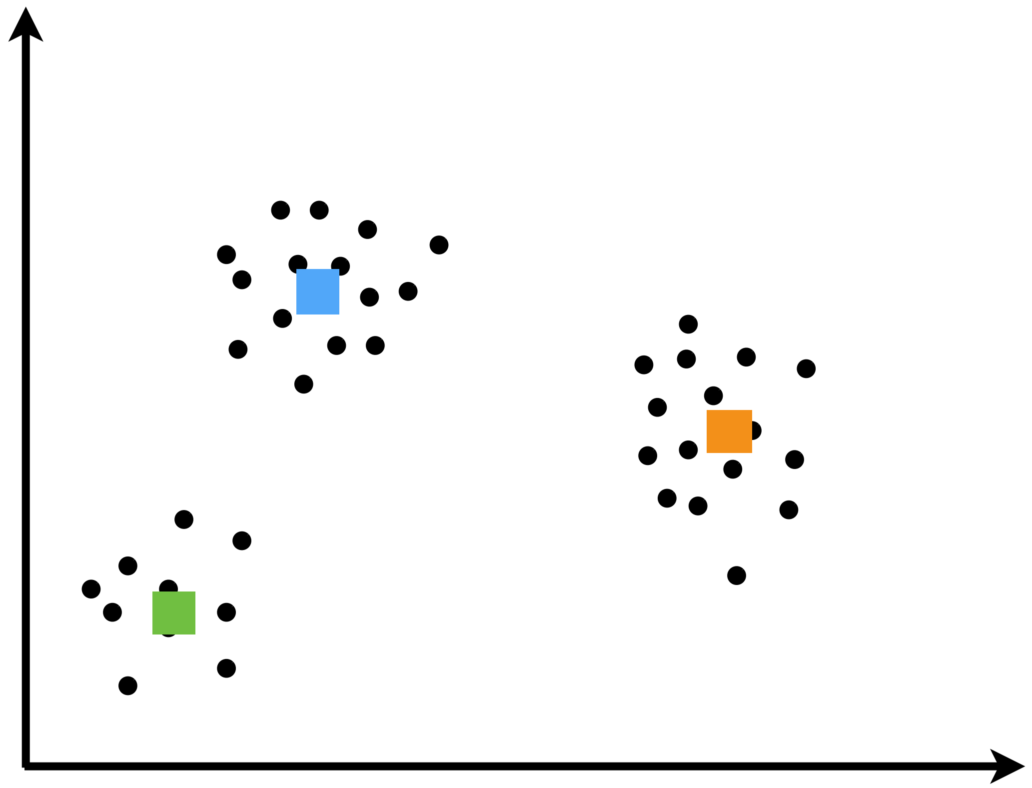


- What happens? $a = a_k = 1$ $\quad a = a_k \to 0$ $\quad a = a_k \to \infty$
- Dirichlet is conjugate to Categorical

[demo]

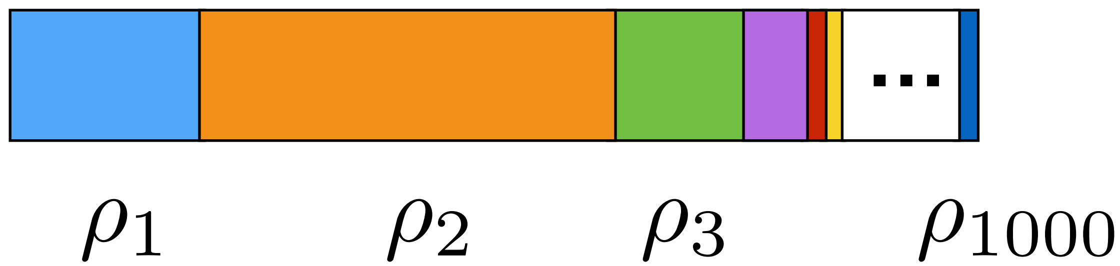
What if $K > N$?

What if $K > N$?



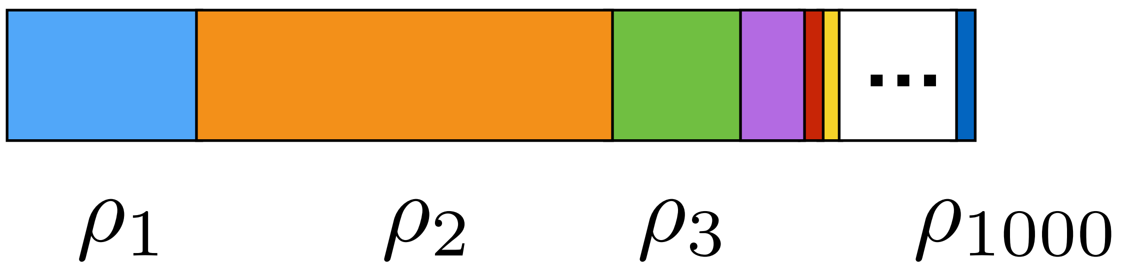
What if $K > N$?

What if $K > N$?



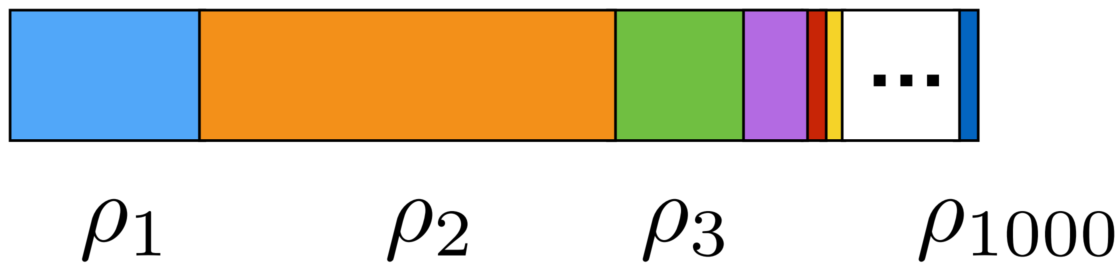
What if $K > N$?

- 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: number of latent groups

What if $K > N$?

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

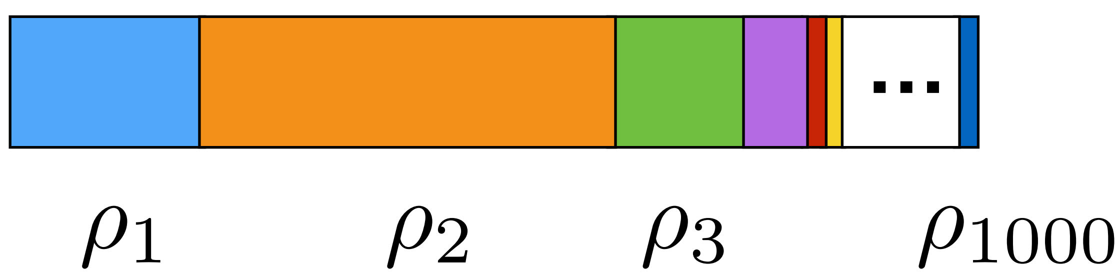
ρ_1 ρ_2 ρ_3 ρ_{1000}

- Components: number of latent groups

- Clusters: number of components represented in the data

What if $K > N$?

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



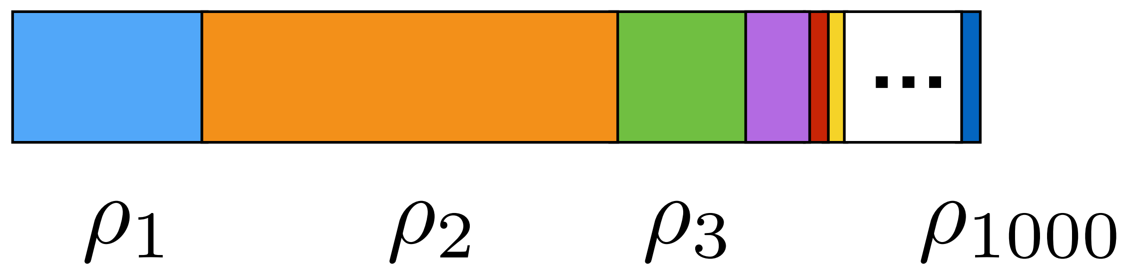
- Components: number of latent groups

- Clusters: number of components represented in the data

- [demo 1, demo 2]

What if $K > N$?

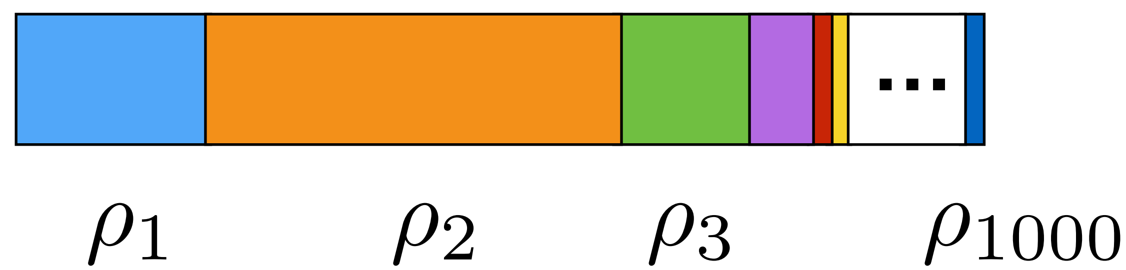
- e.g. species sampling, topic modeling, groups on a social network, etc.



- 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

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