MIT EECS
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

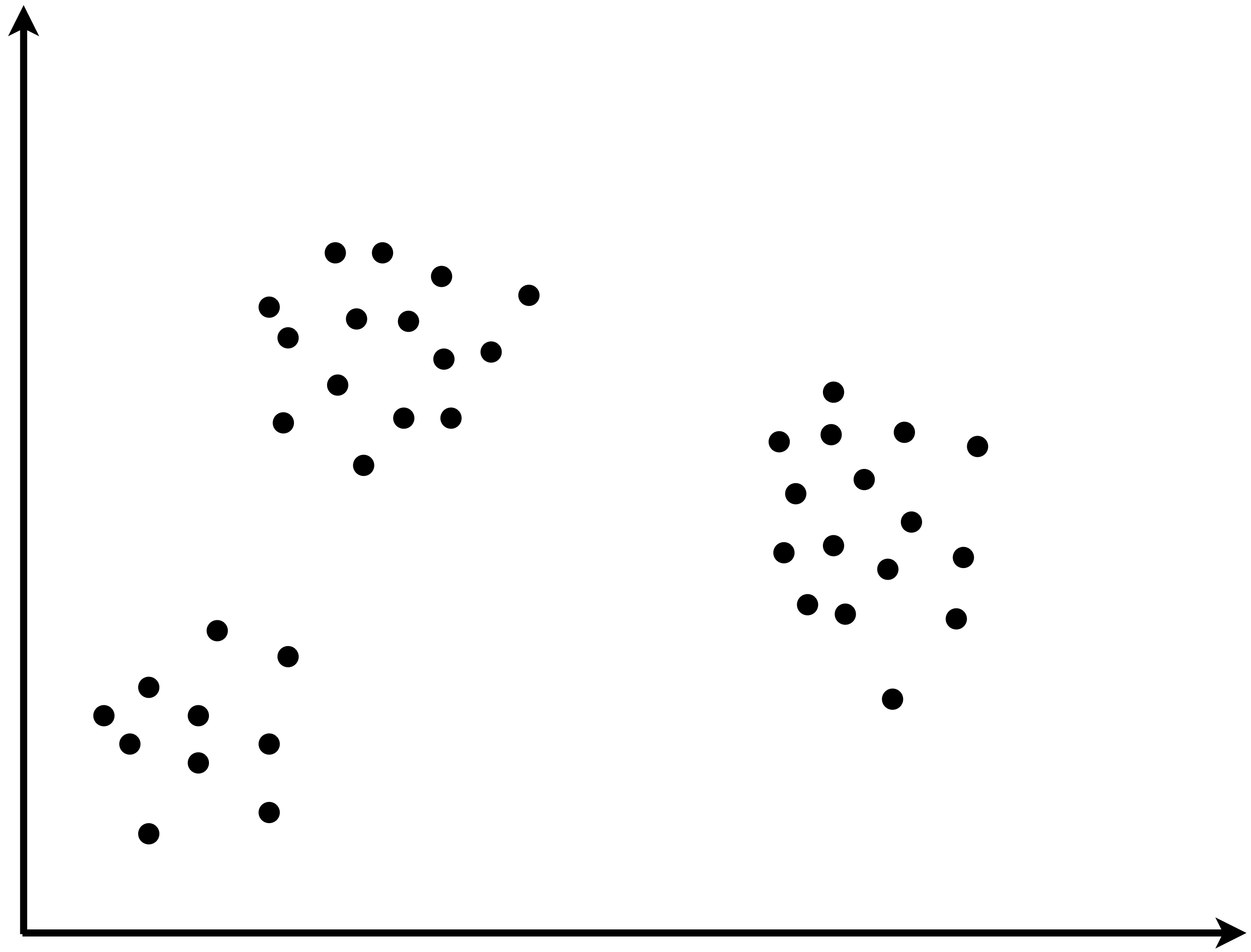
CSAIL

Feature allocations, probability functions, and paintboxes

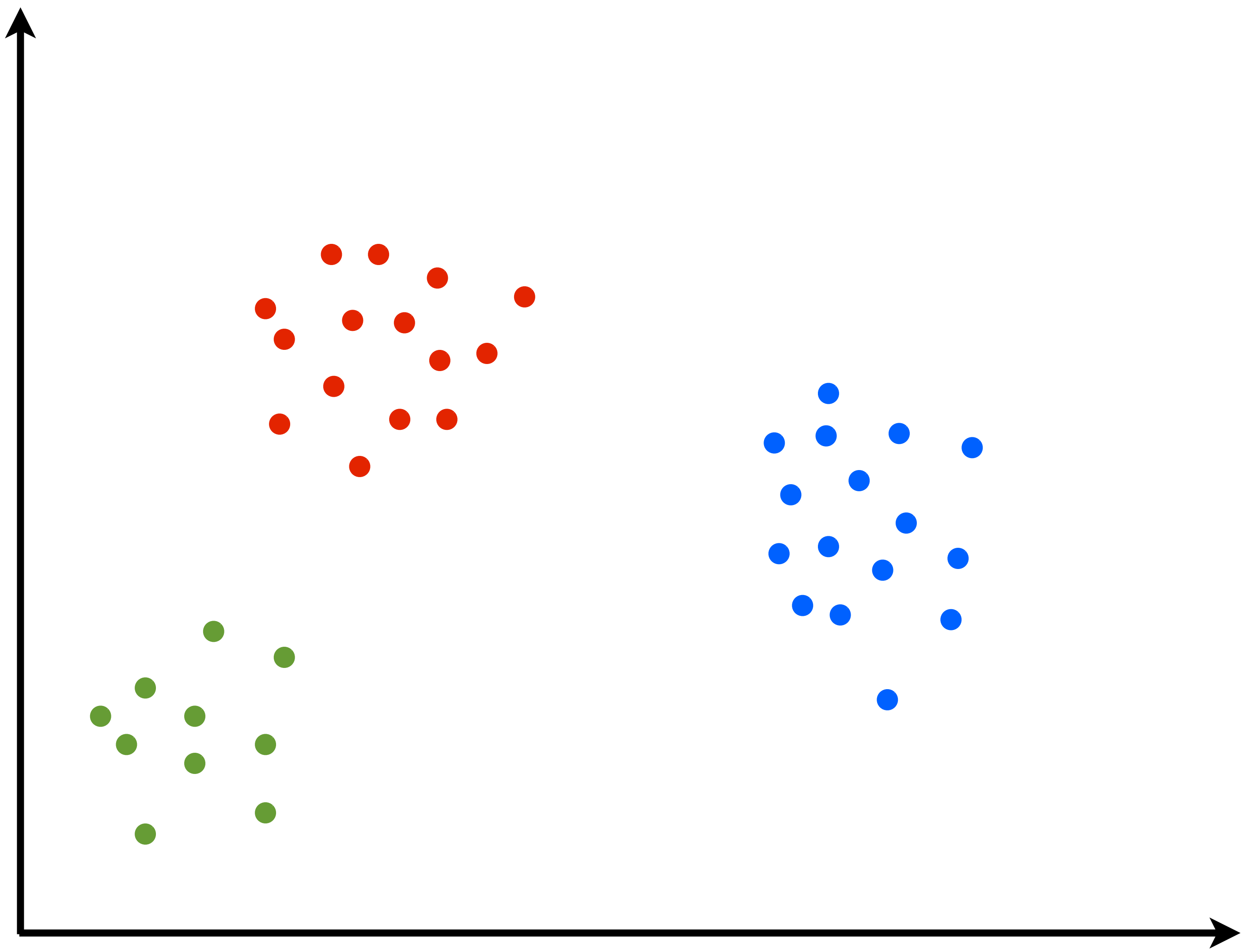
Tamara Broderick
ITT Career Development
Assistant Professor,
MIT

IDSS
MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Clustering/Partition

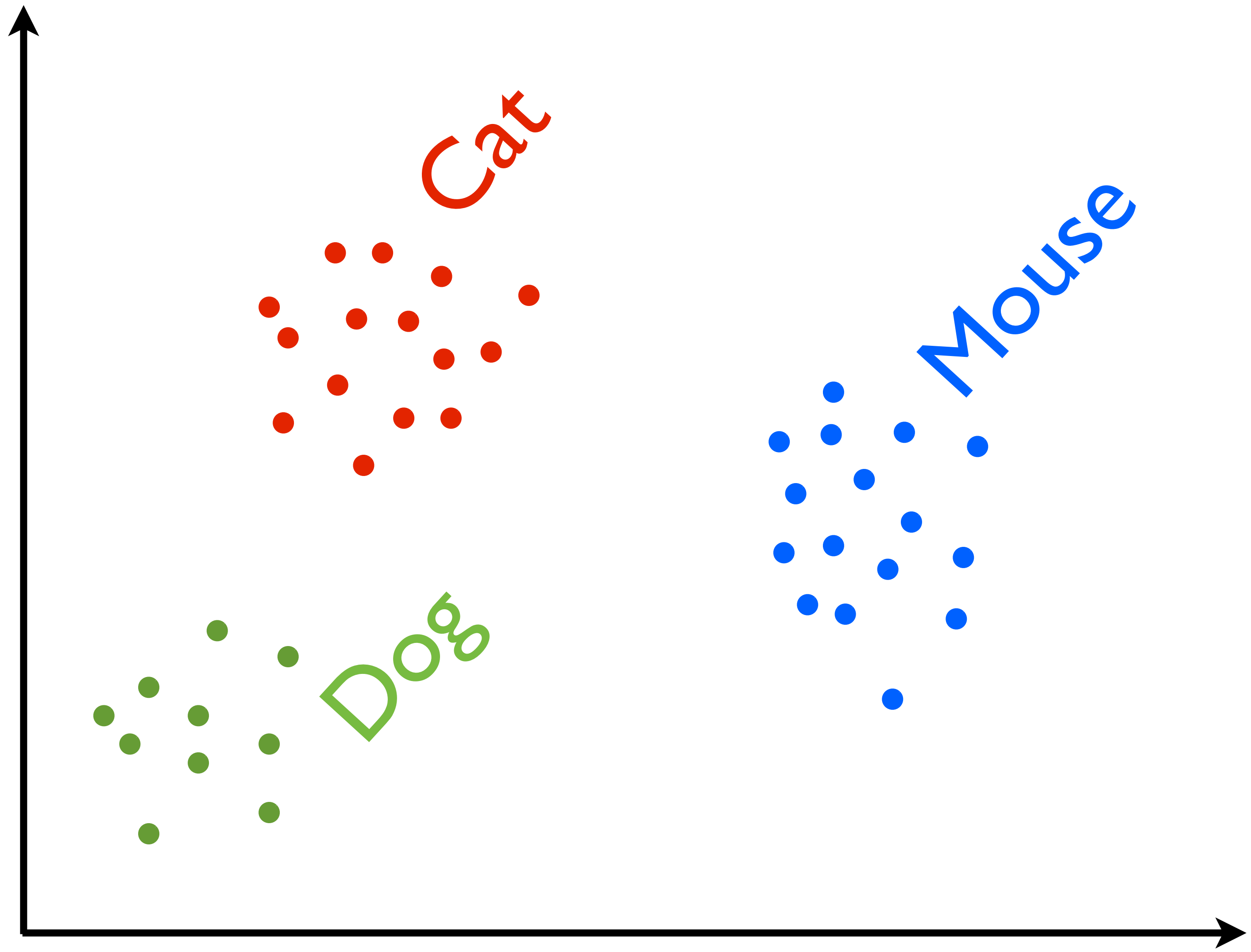


Clustering/Partition



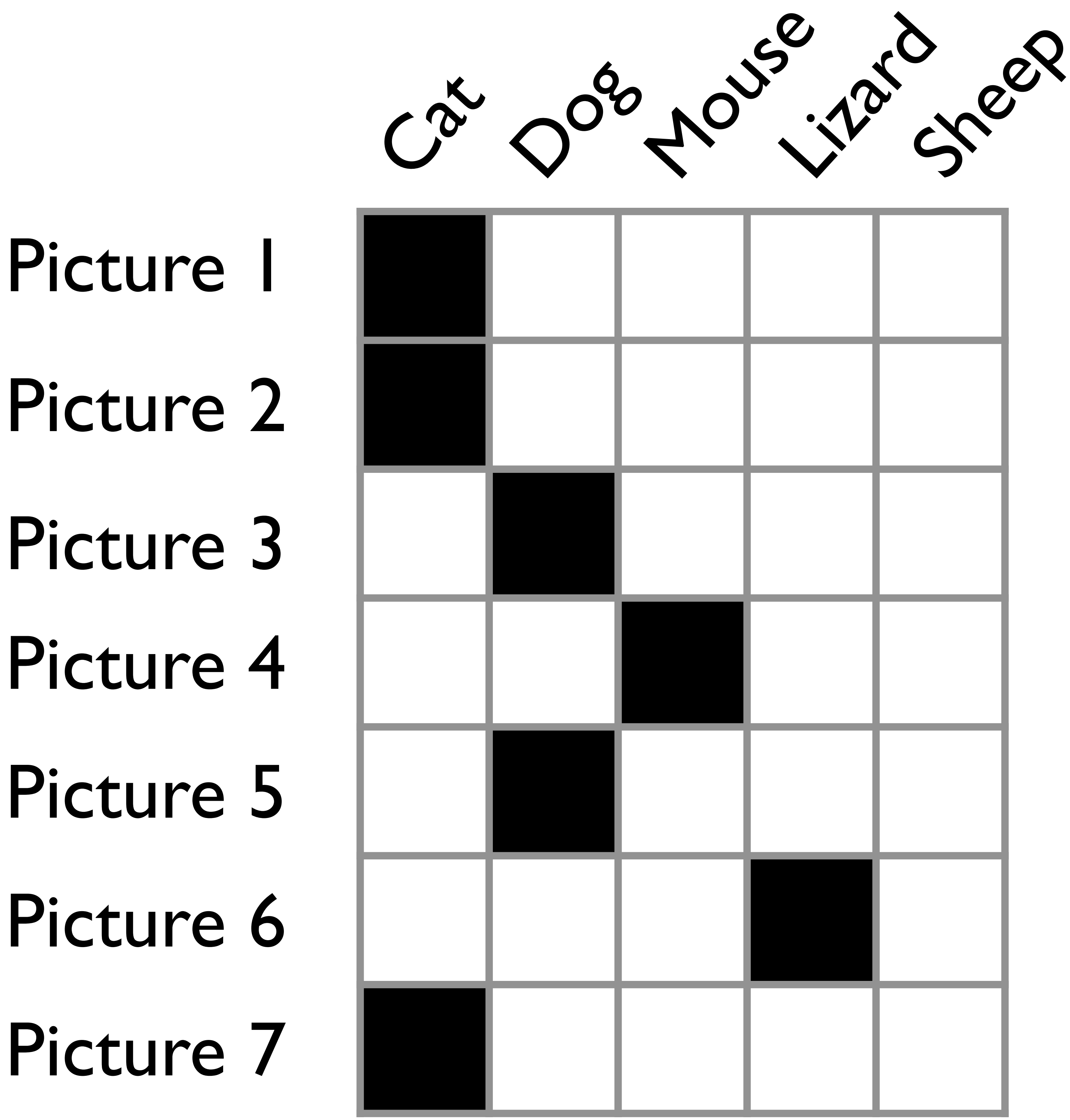
"clusters",
"classes",
"blocks (of a partition)"

Clustering/Partition

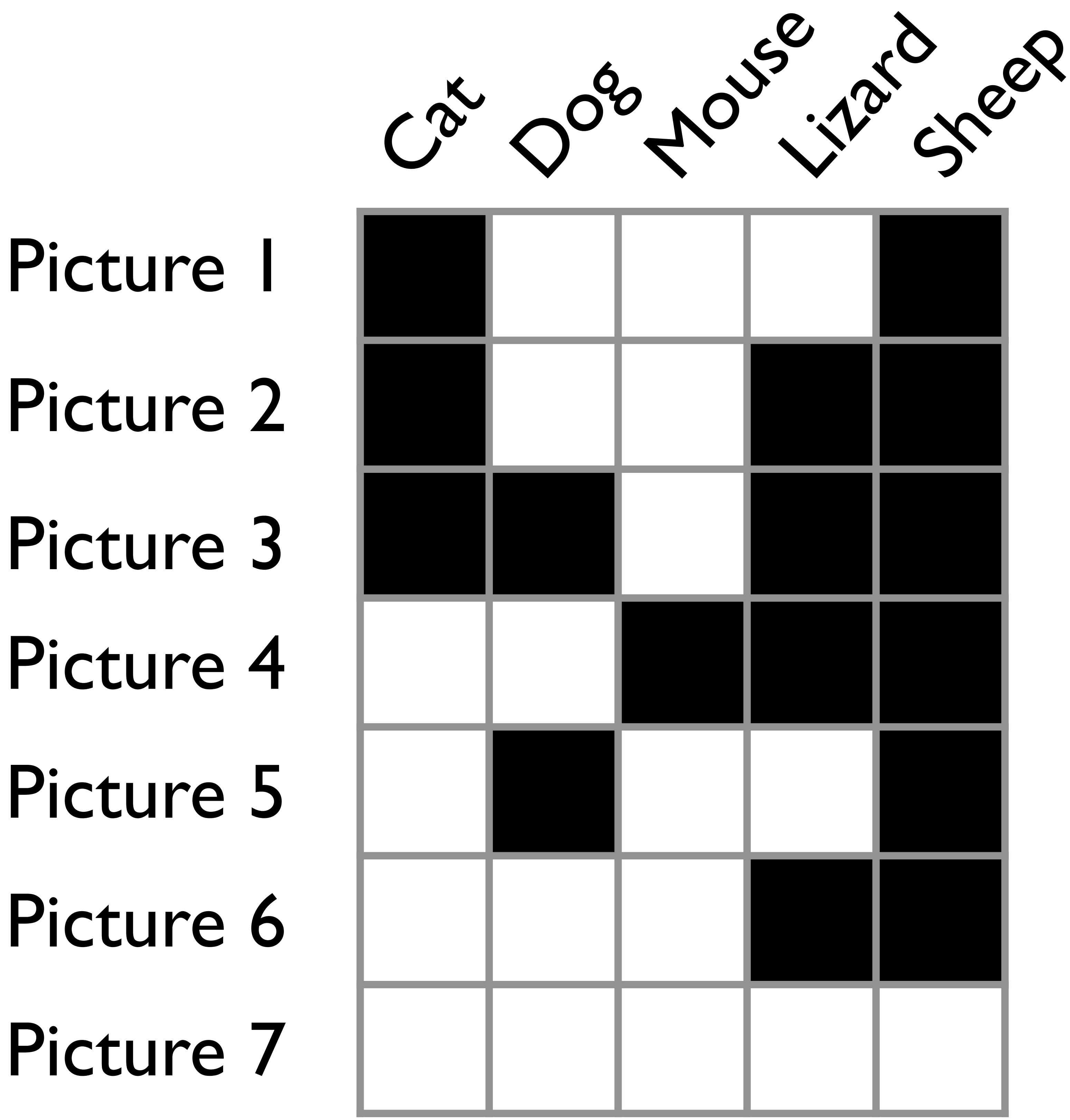


“clusters”,
“classes”,
“blocks (of a partition)”

Clustering/Partition

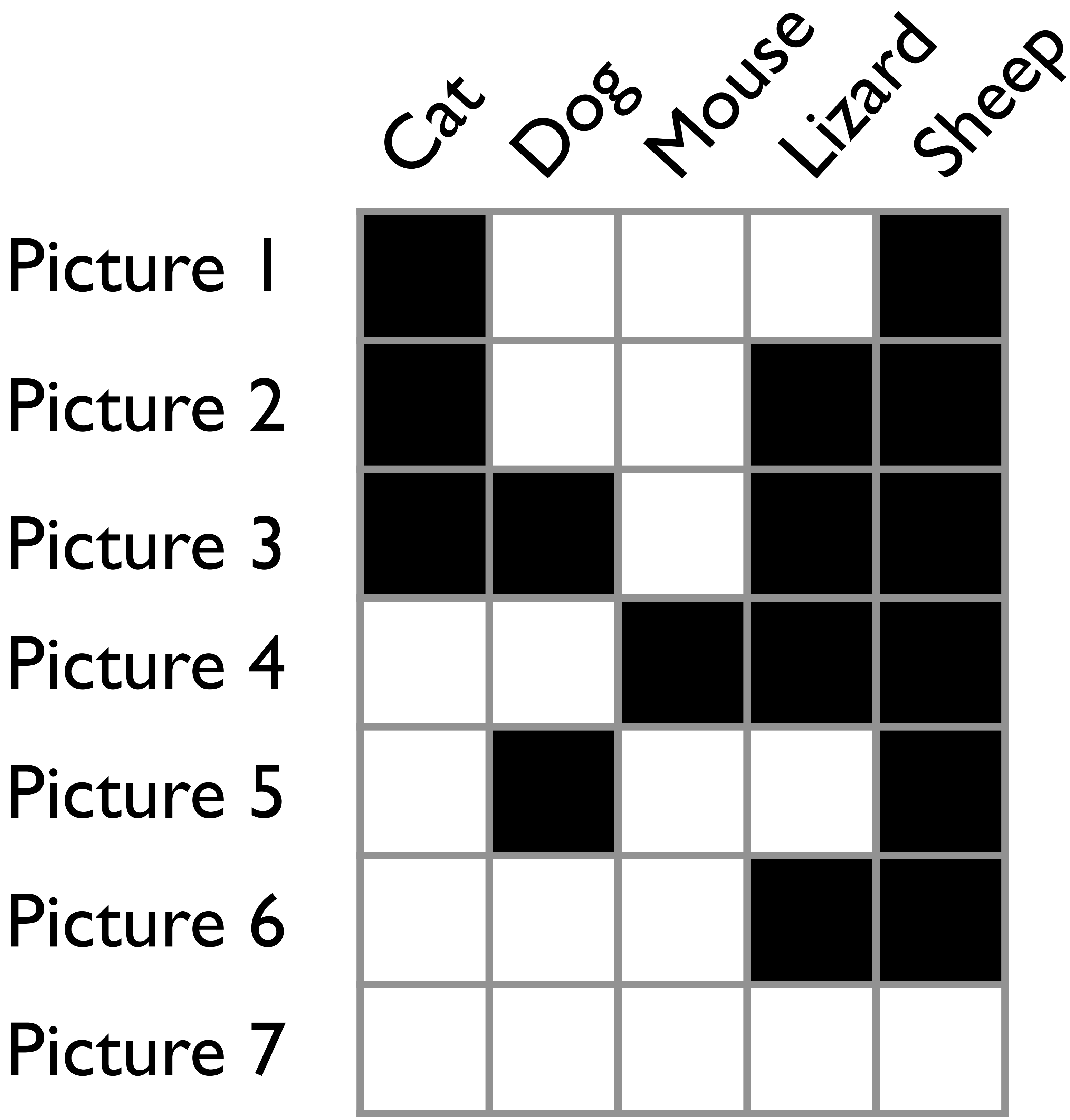


Latent feature allocation



"features",
"topics"

Latent feature allocation



"features",
"topics"

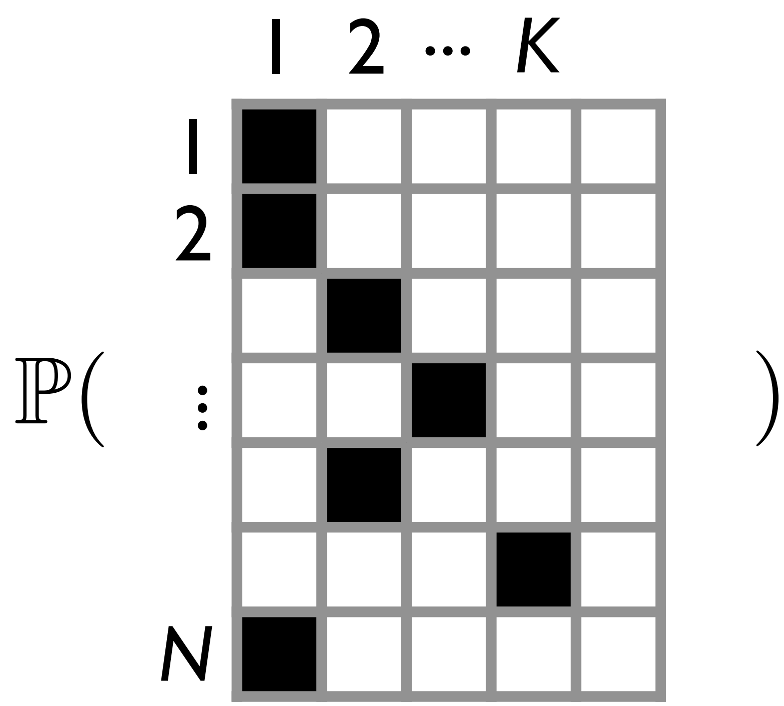
- Exchangeable
- Finite # of features per data point

Characterizations

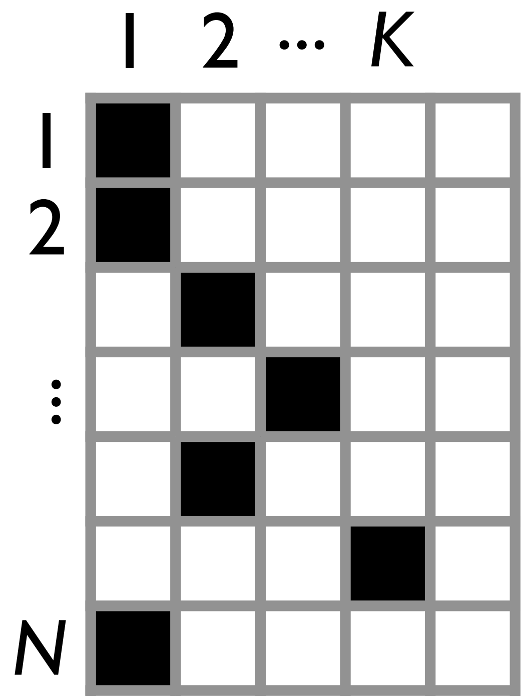
- Exchangeable cluster distributions are characterized

- What about exchangeable feature distributions?

Exchangeable probability functions

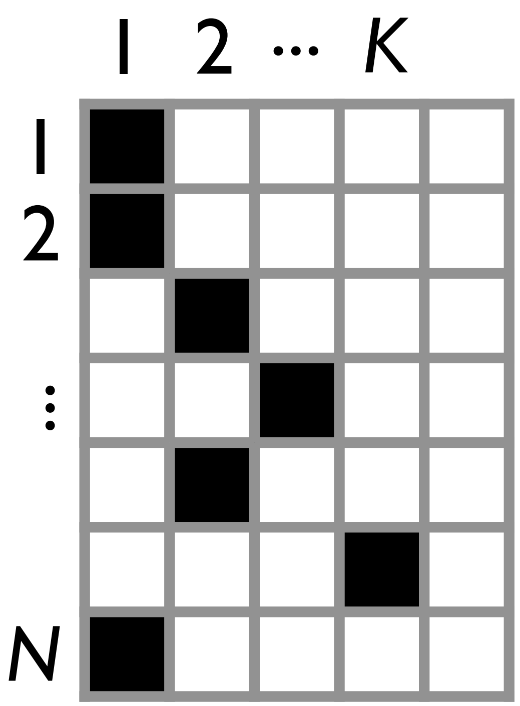


Exchangeable probability functions

$\mathbb{P}($  $) = p\,(S_{N,1},\ldots,S_{N,K})$

Exchangeable probability functions

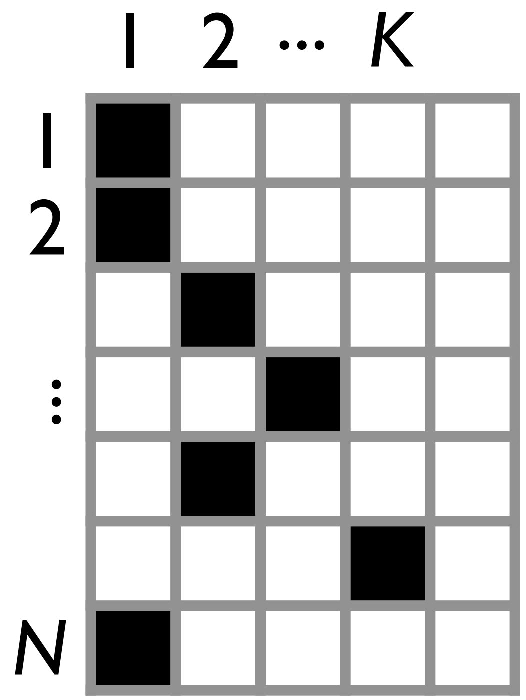
Size of *K*th cluster

$\mathbb{P}($ 

$) = p\left(S_{N,1}, \ldots, S_{N,K}\right)$

Exchangeable probability functions

Exchangeable partition probability function (EPPF)

$\mathbb{P}($ 

$) = p\left(S_{N,1}, \ldots, S_{N,K}\right)$

[Pitman 1995]

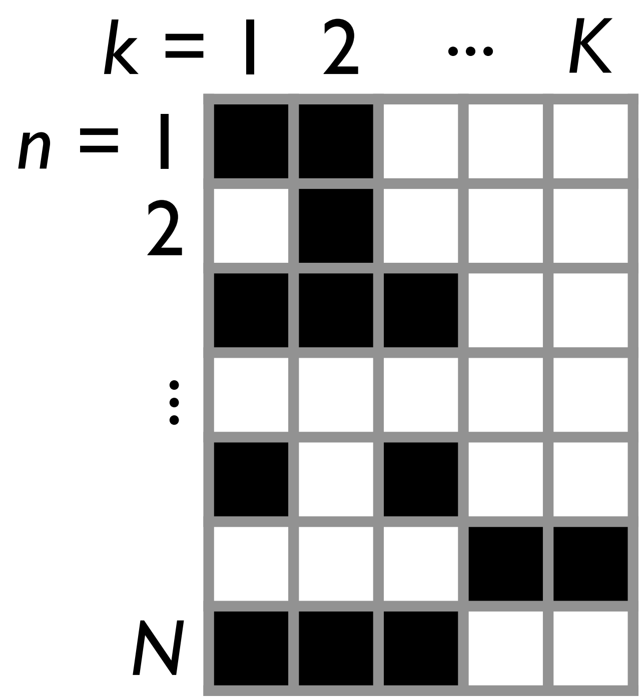
Exchangeable probability functions

“Exchangeable **feature** probability function” (EFPF)?

Example: Indian buffet process

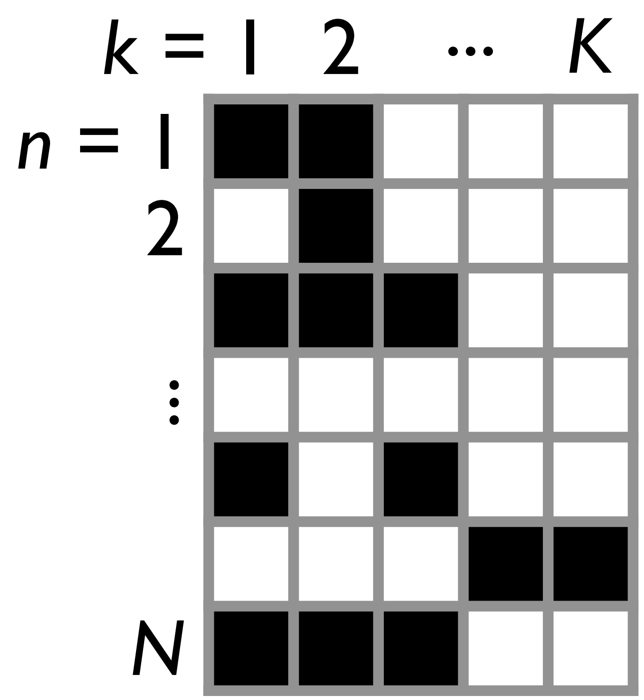
[Griffiths, Ghahramani 2006]

Example: Indian buffet process



[Griffiths, Ghahramani 2006]

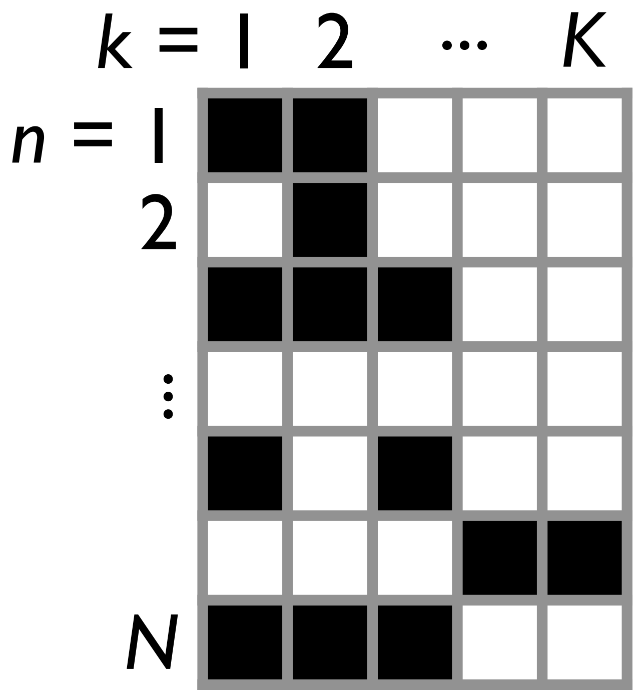
Example: Indian buffet process



For n = 1, 2, ..., N

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

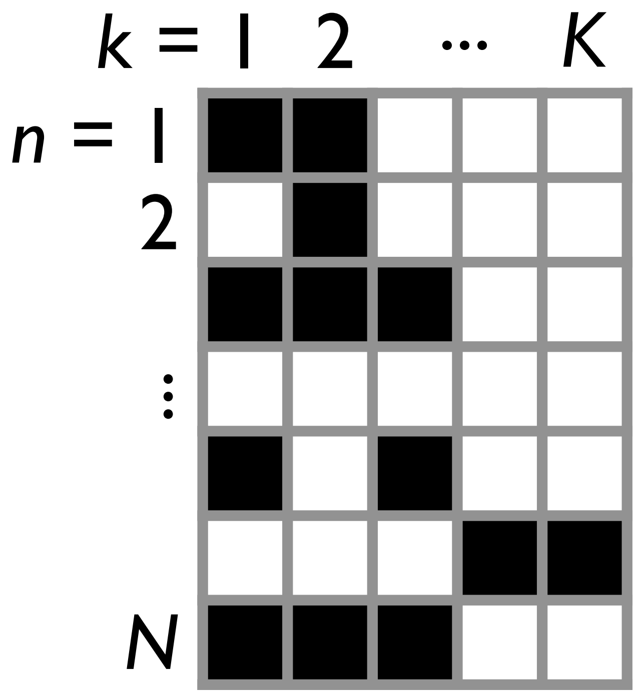


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\frac{S_{n-1,k}}{\theta + n - 1}$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

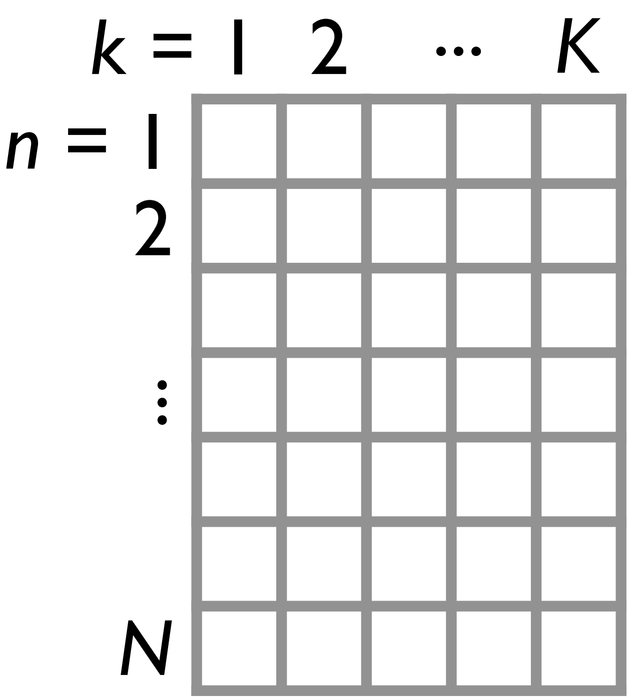


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

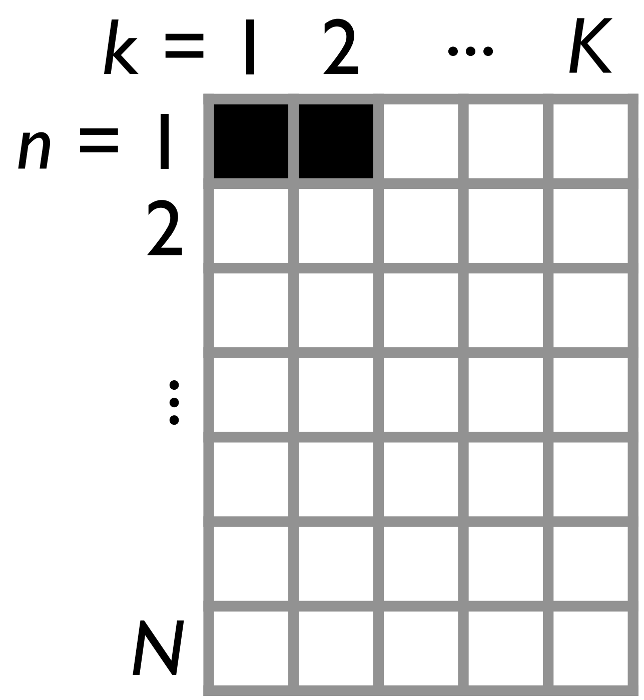


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

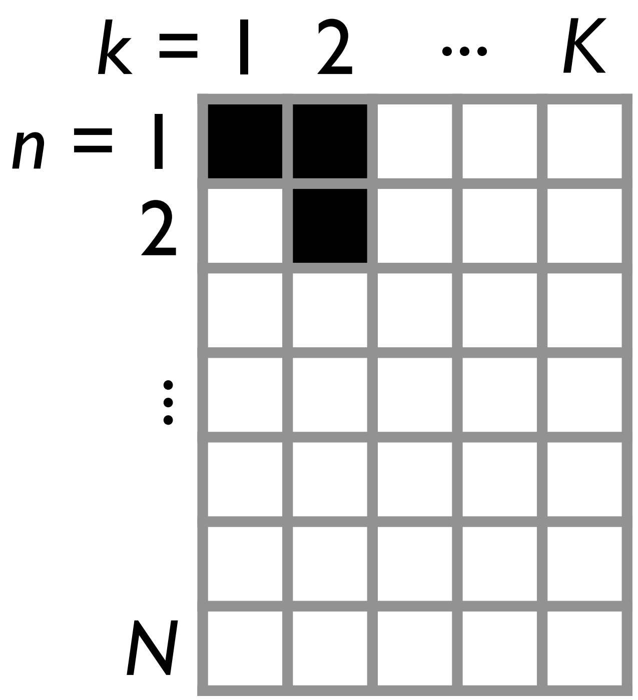


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

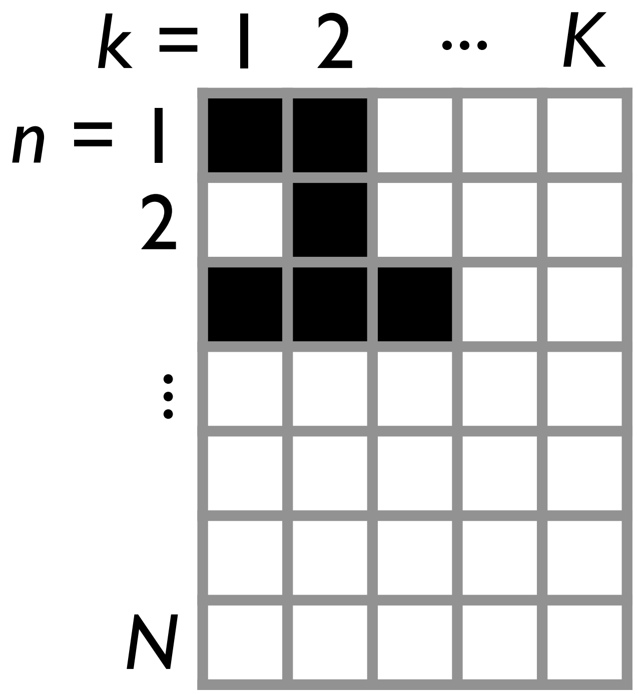


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

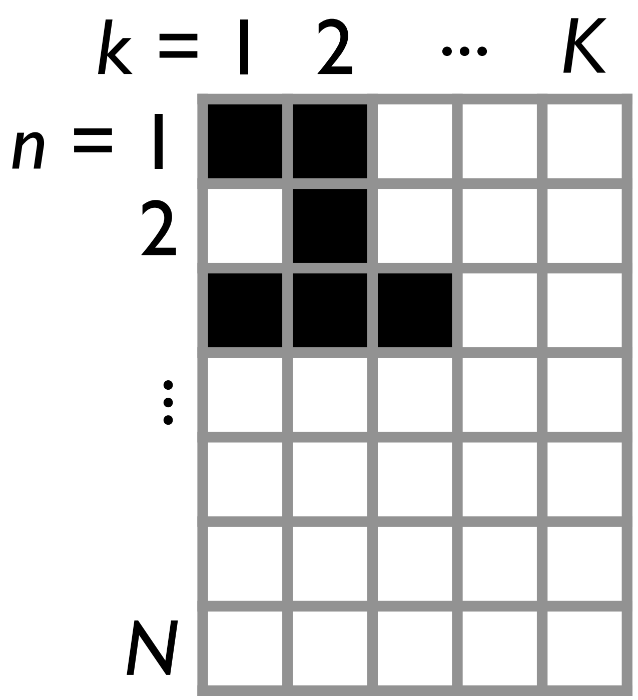


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



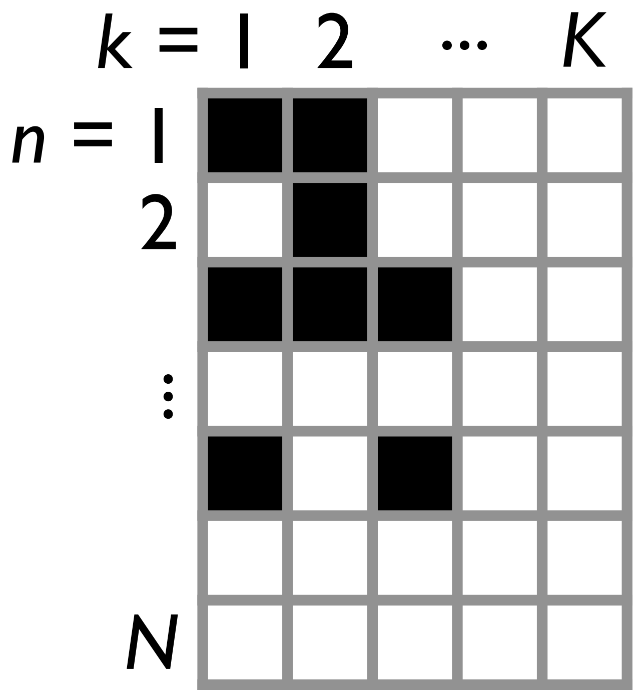
For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$

2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

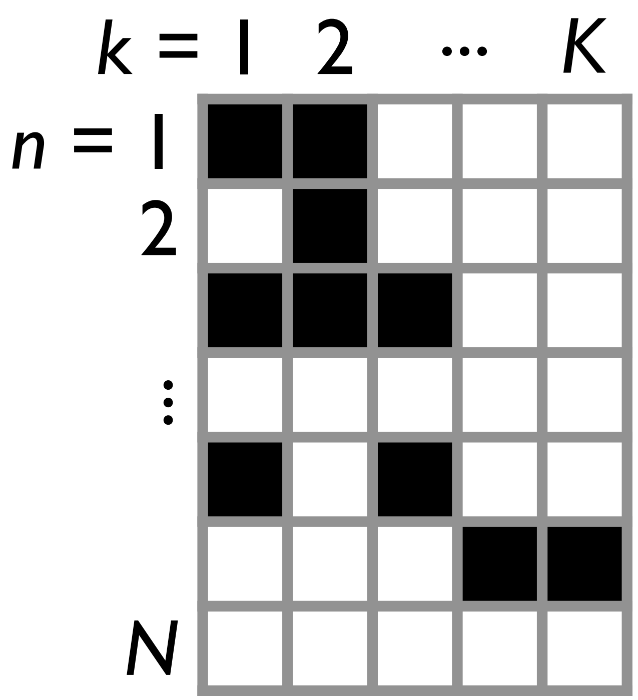


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process

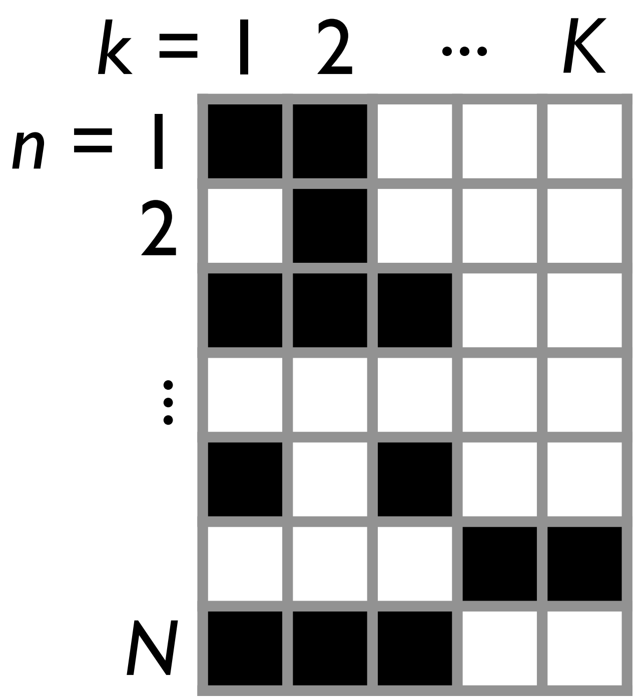


For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

Example: Indian buffet process



For n = 1, 2, ..., N

1. Data point n has an existing feature k that has already occurred $S_{n-1,k}$ times with probability $\dfrac{S_{n-1,k}}{\theta + n - 1}$
2. Number of new features for data point n: $K_n^+ = \text{Poisson}\left(\gamma \dfrac{\theta}{\theta + n - 1}\right)$

[Griffiths, Ghahramani 2006]

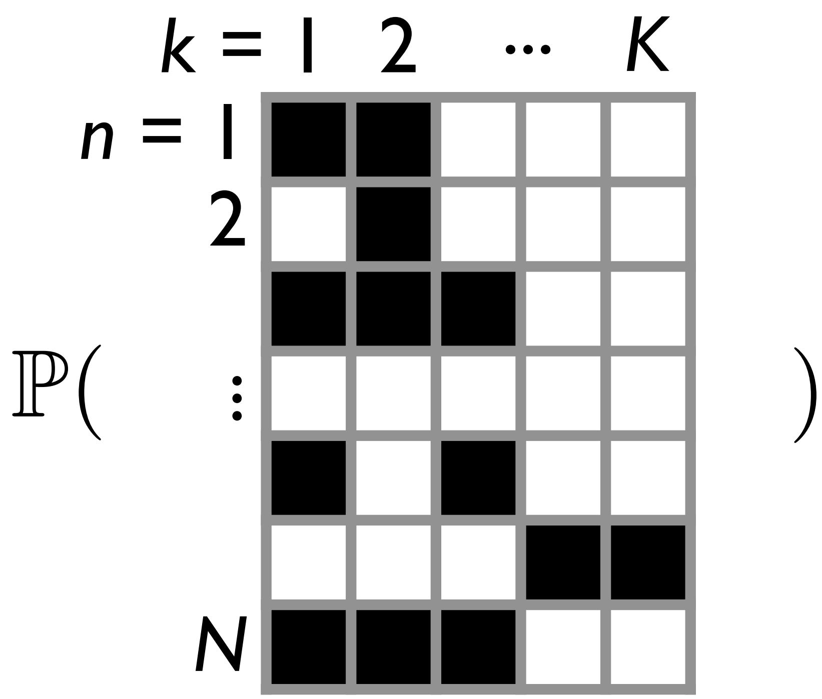
Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

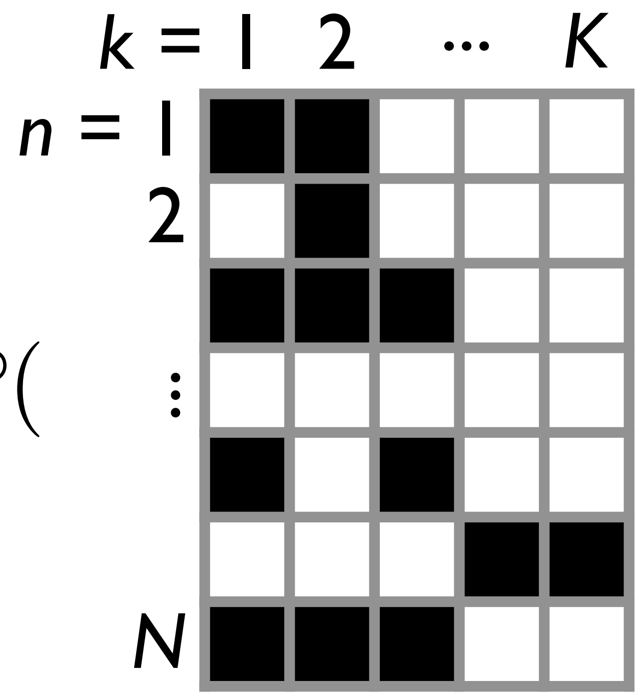
Example: Indian buffet process (IBP)



Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($ 

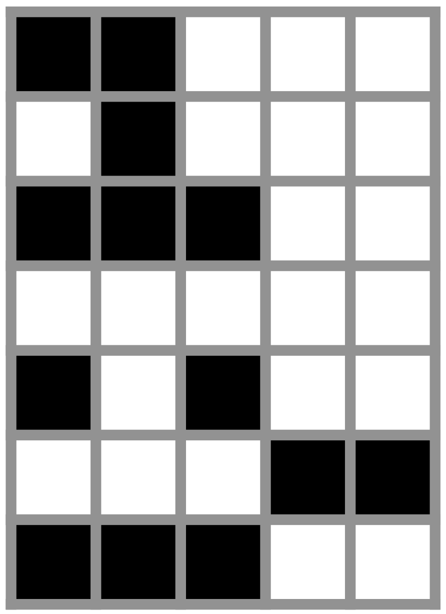
$)$

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1}\right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($  $)$

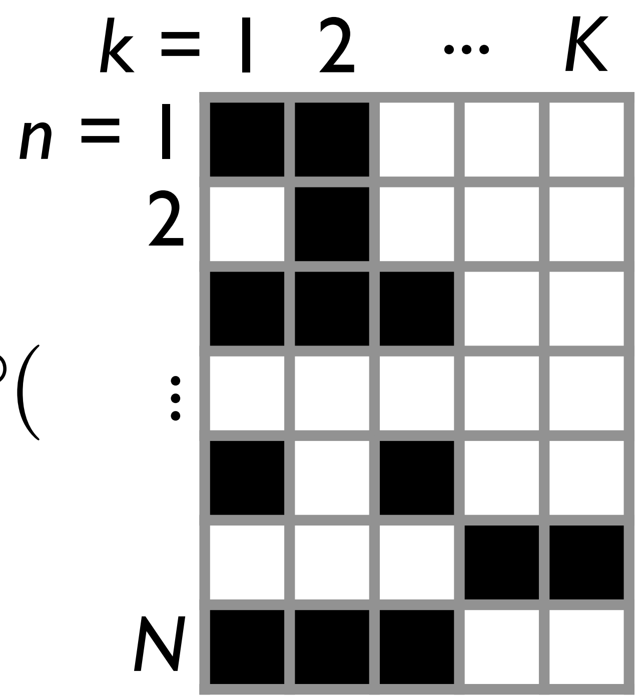
Size of kth feature

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta+n-1)^{-1}\right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N-S_{N,k}+\theta)}{\Gamma(N+\theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($  $)$

Size of *k*th feature

Number of features

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma\sum_{n=1}^{N}(\theta+n-1)^{-1}\right)\prod_{k=1}^{K_N}\frac{\Gamma(S_{N,k})\Gamma(N-S_{N,k}+\theta)}{\Gamma(N+\theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$$\mathbb{P}\left(\begin{array}{c|ccccc} & k=1 & 2 & & \cdots & K \\ \hline n=1 & 1 & 1 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 0 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ \vdots & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 1 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 1 \\ N & 1 & 1 & 1 & 0 & 0 \end{array} \right)$$

Number of data points

Size of kth feature

Number of features

$$= \frac{1}{K_N!} (\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$$
\mathbb{P}\left(
\begin{array}{c|ccccc}
 & k=1 & 2 & & \cdots & K \\
\hline
n=1 & 1 & 1 & 0 & 0 & 0 \\
2 & 0 & 1 & 0 & 0 & 0 \\
 & 1 & 1 & 1 & 0 & 0 \\
\vdots & 0 & 0 & 0 & 0 & 0 \\
 & 1 & 0 & 1 & 0 & 0 \\
 & 0 & 0 & 0 & 1 & 1 \\
N & 1 & 1 & 1 & 0 & 0
\end{array}
\right)
$$

Number of data points

Size of kth feature

Number of features

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma\sum_{n=1}^{N}(\theta+n-1)^{-1}\right)\prod_{k=1}^{K_N}\frac{\Gamma(S_{N,k})\Gamma(N-S_{N,k}+\theta)}{\Gamma(N+\theta)}$$

$$= p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})$$

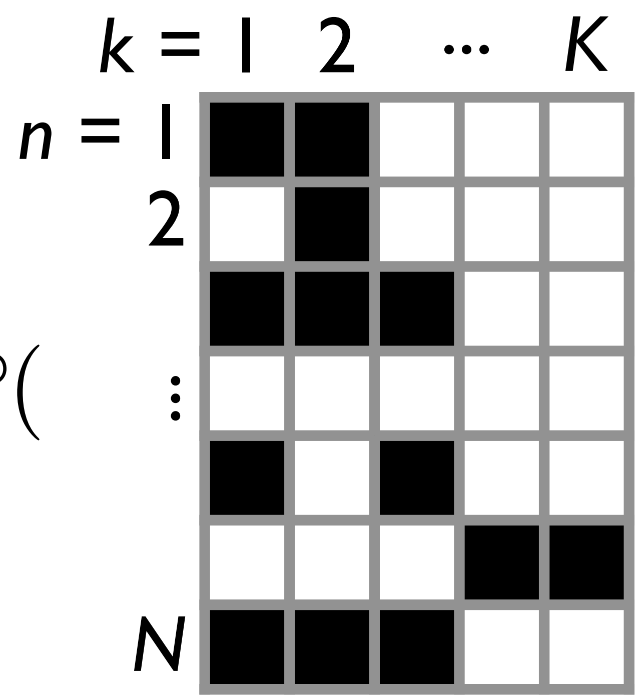
[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}\Big($



$\Big)$

Number of data points

Size of *k*th feature

Number of features

$$= \frac{1}{K_N!}(\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1}\right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k})\Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

$$= p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})$$

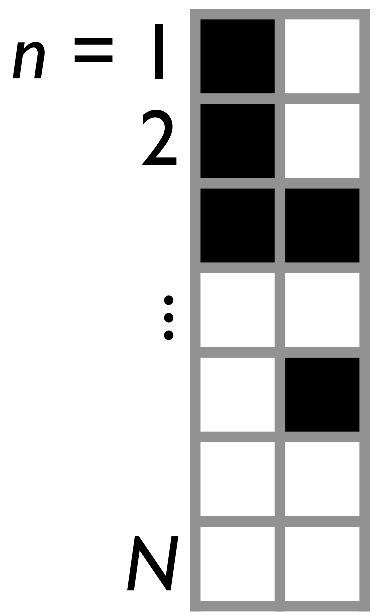
"EFPF"

[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

“Exchangeable feature probability function” (EFPF)?

Counterexample

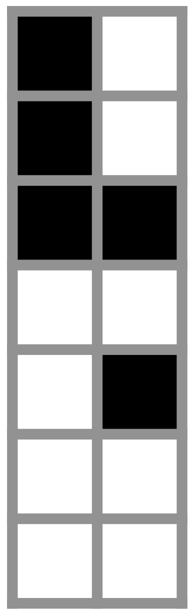


[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

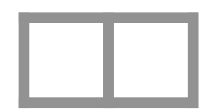
Counterexample



$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

$\mathbb{P}(\text{row} =$  $) = p_4$

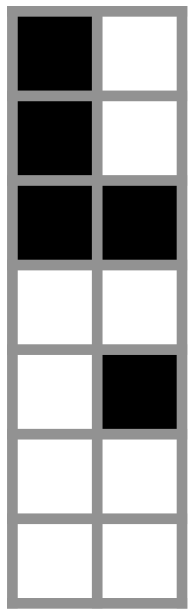
[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Counterexample

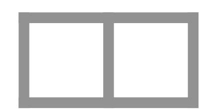
n = 1
2
⋮
N

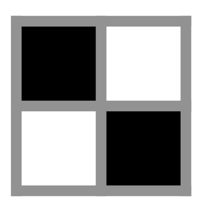


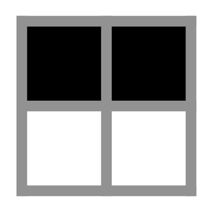
$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

$\mathbb{P}(\text{row} =$  $) = p_4$



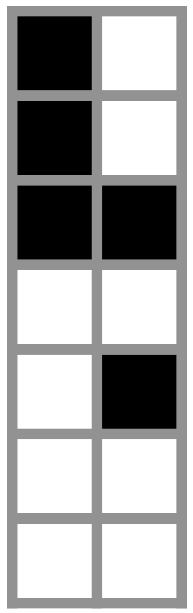


[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

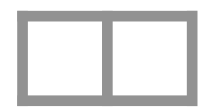
Counterexample

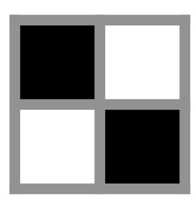
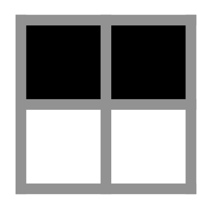


$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

$\mathbb{P}(\text{row} =$  $) = p_4$

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

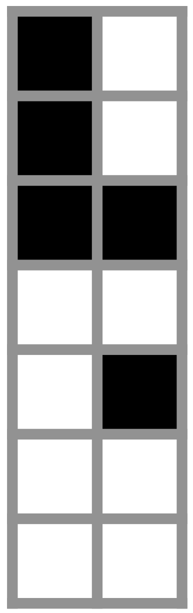
[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

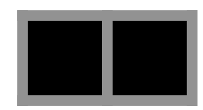
Counterexample

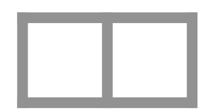
n = 1
2
⋮
N

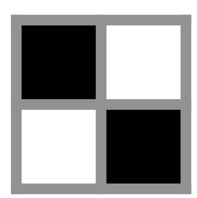
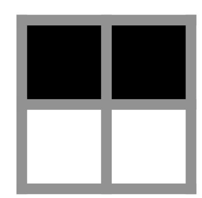


$\mathbb{P}(\text{row} =$  $) = p_1$

$\mathbb{P}(\text{row} =$  $) = p_2$

$\mathbb{P}(\text{row} =$  $) = p_3$

$\mathbb{P}(\text{row} =$  $) = p_4$

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

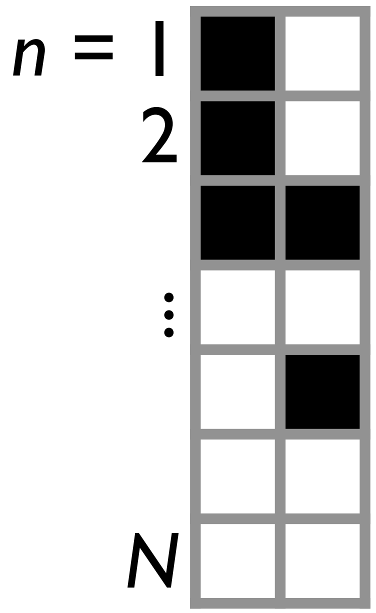
$p_1 p_2$ $p_3 p_4$

[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Counterexample



$$\mathbb{P}(\text{row} = \blacksquare\square) = p_1$$

$$\mathbb{P}(\text{row} = \square\blacksquare) = p_2$$

$$\mathbb{P}(\text{row} = \blacksquare\blacksquare) = p_3$$

$$\mathbb{P}(\text{row} = \square\square) = p_4$$

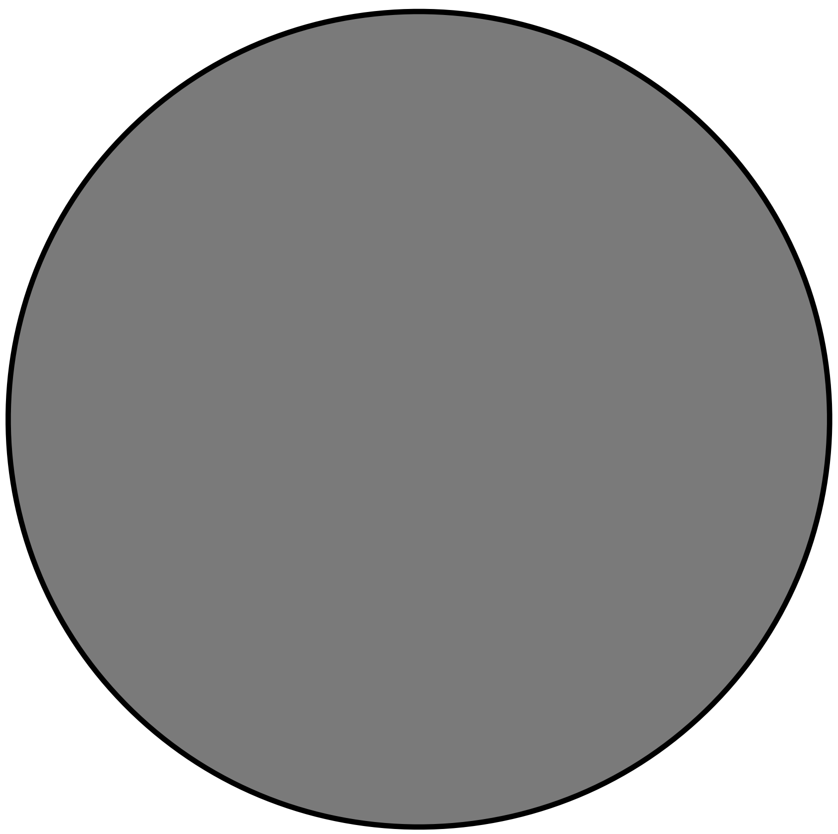
$$\mathbb{P}\left(\begin{matrix}\blacksquare & \square \\ \square & \blacksquare\end{matrix}\right) \neq \mathbb{P}\left(\begin{matrix}\blacksquare & \blacksquare \\ \square & \square\end{matrix}\right)$$

$$p_1 p_2 \neq p_3 p_4$$

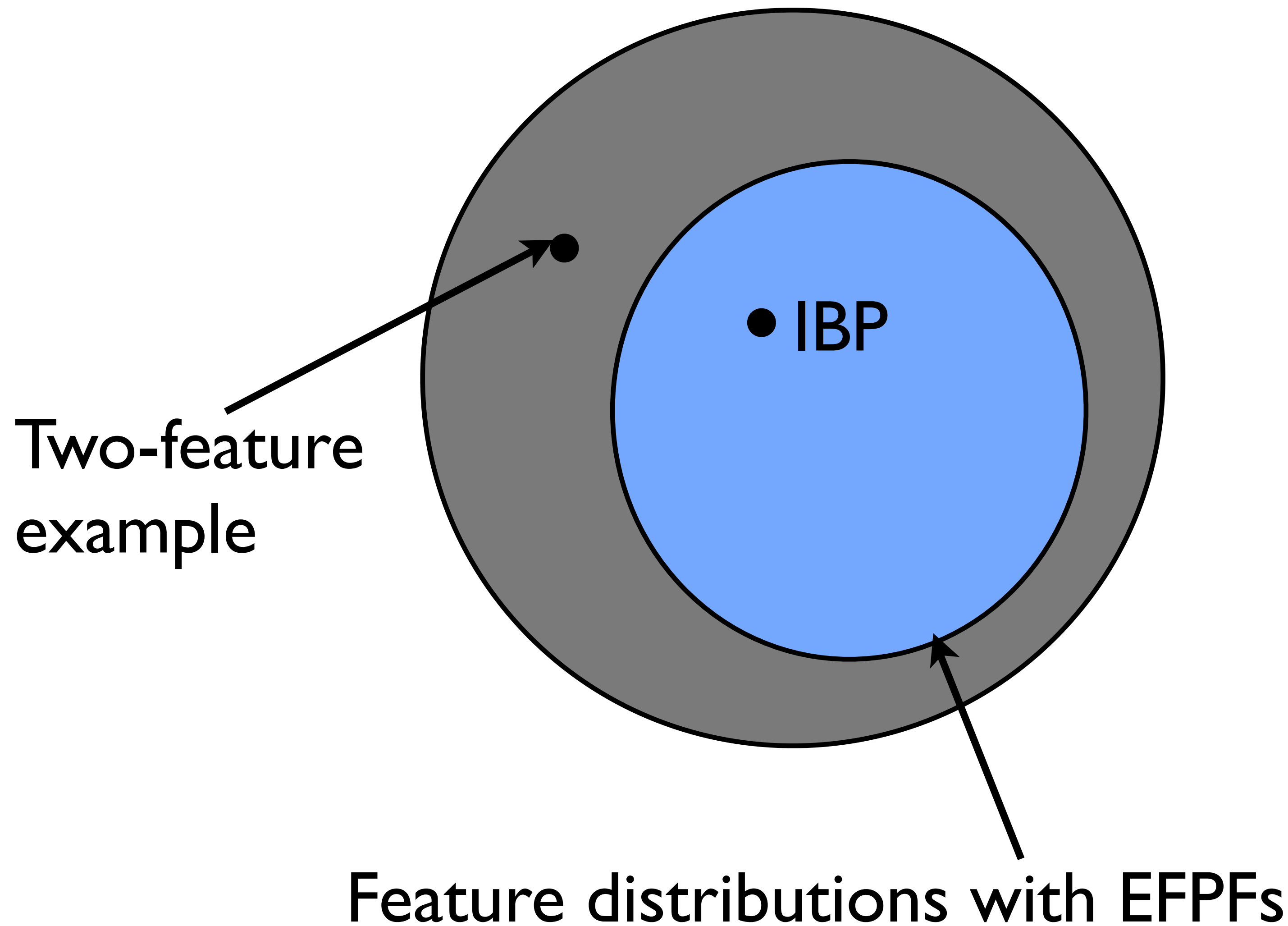
[Broderick, Jordan, Pitman 2013]

Exchangeable probability functions

Exchangeable cluster distributions
= Cluster distributions with EPPFs



Exchangeable feature distributions



[Broderick, Jordan, Pitman 2013]

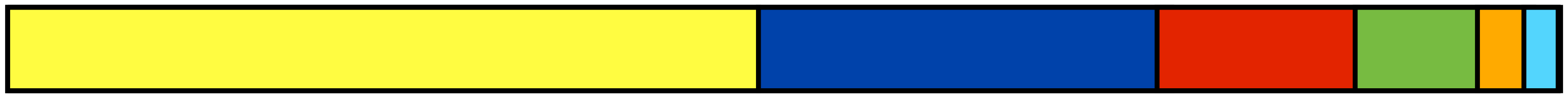
Paintboxes

Exchangeable partition: Kingman paintbox

[Kingman 1978]

Paintboxes

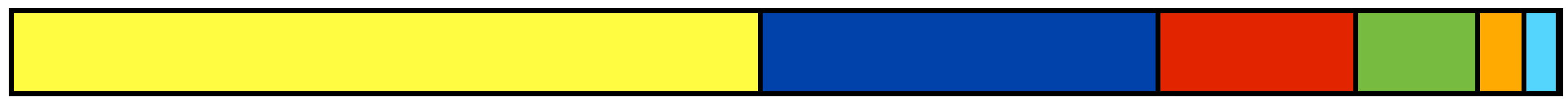
Exchangeable partition: Kingman paintbox

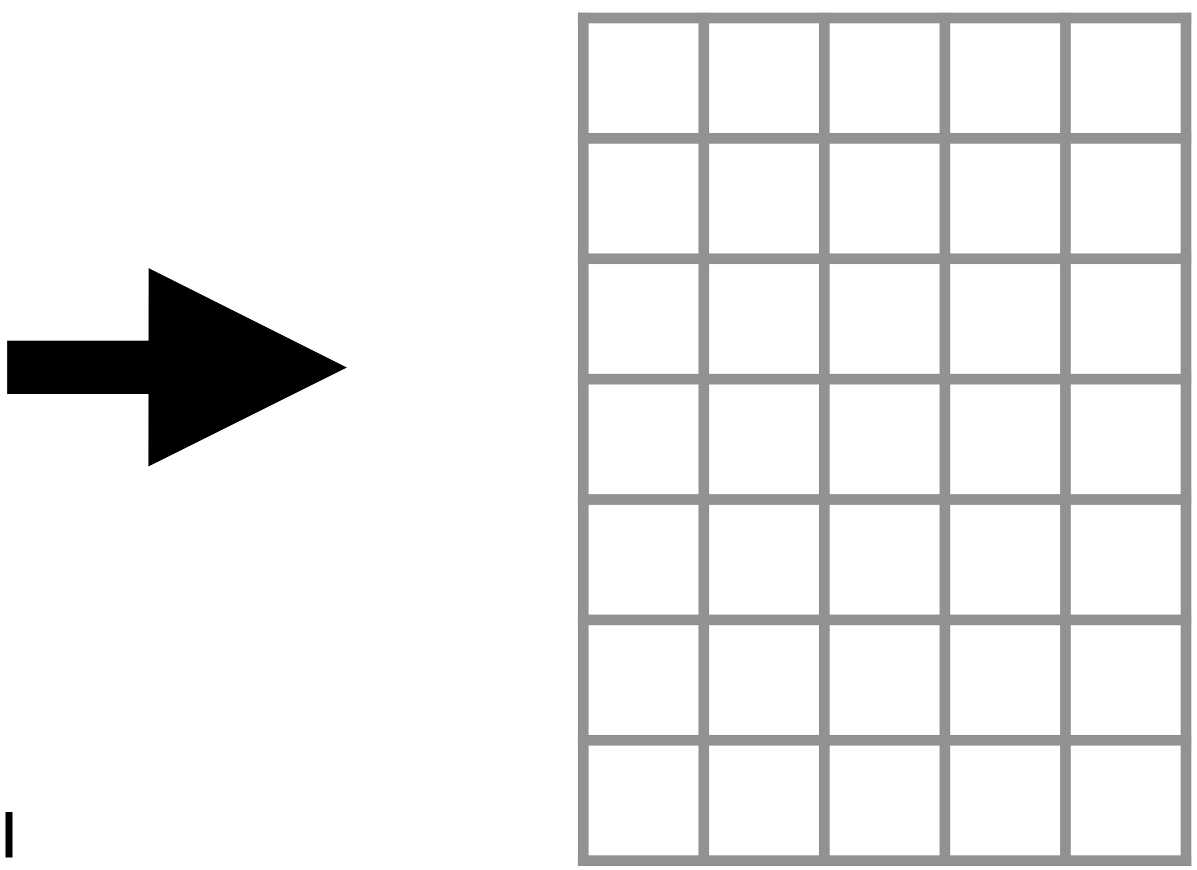


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

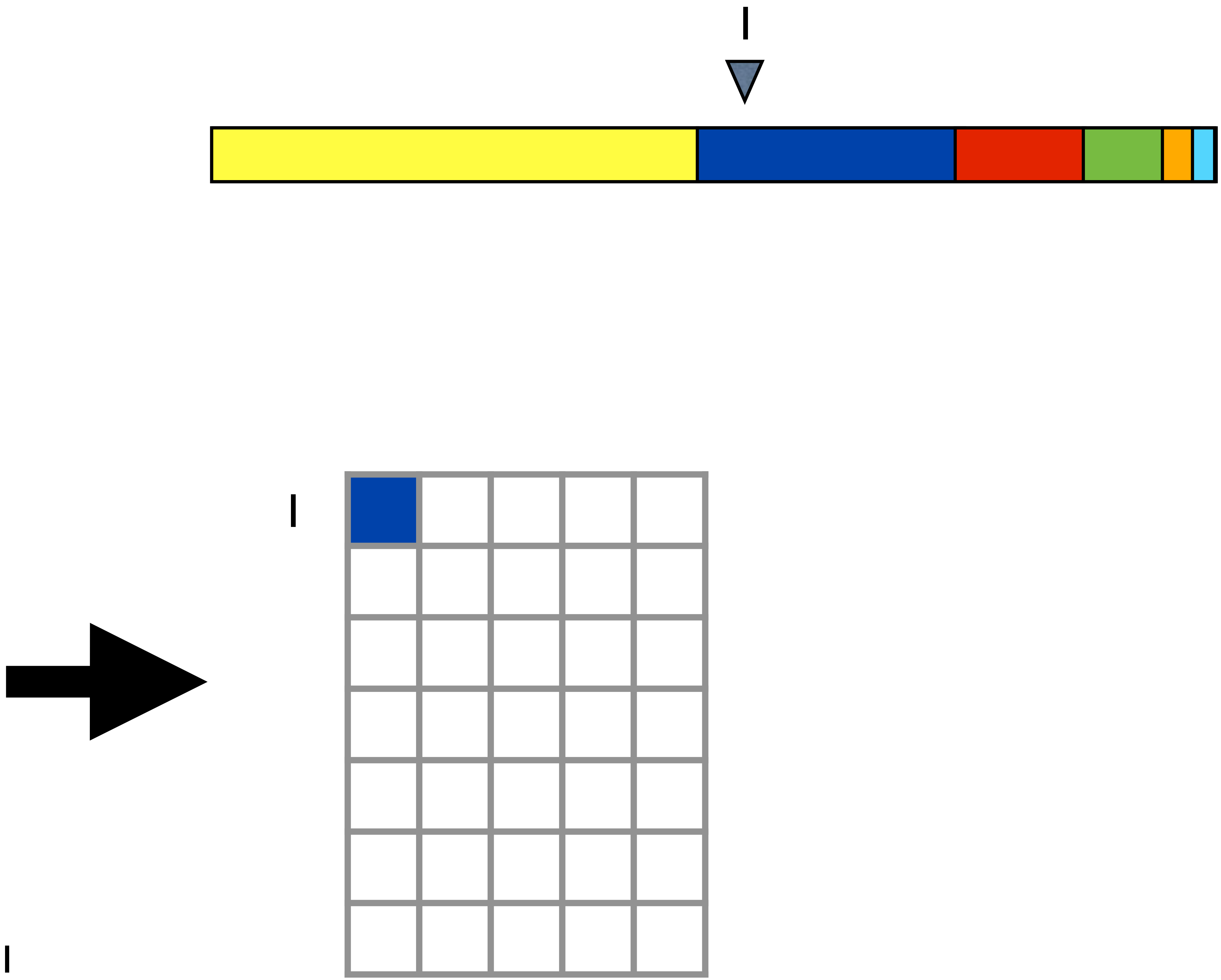




[Kingman 1978]

Paintboxes

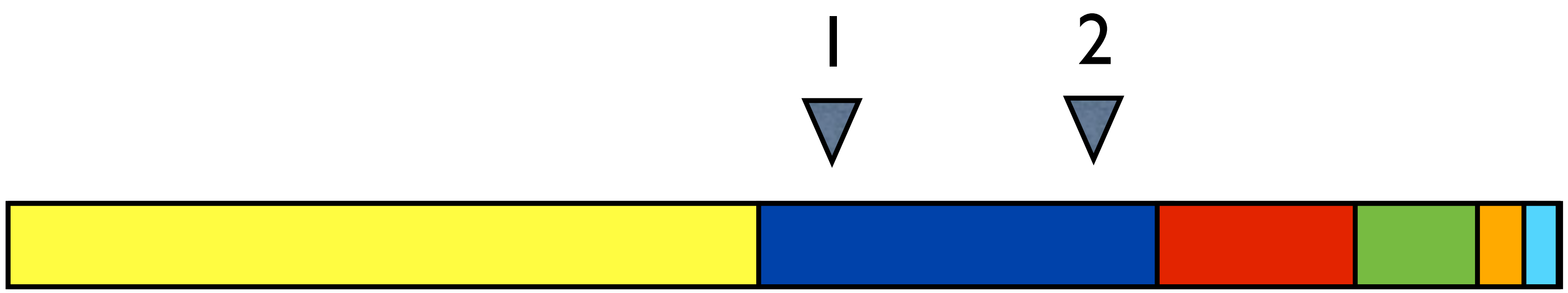
Exchangeable partition: Kingman paintbox

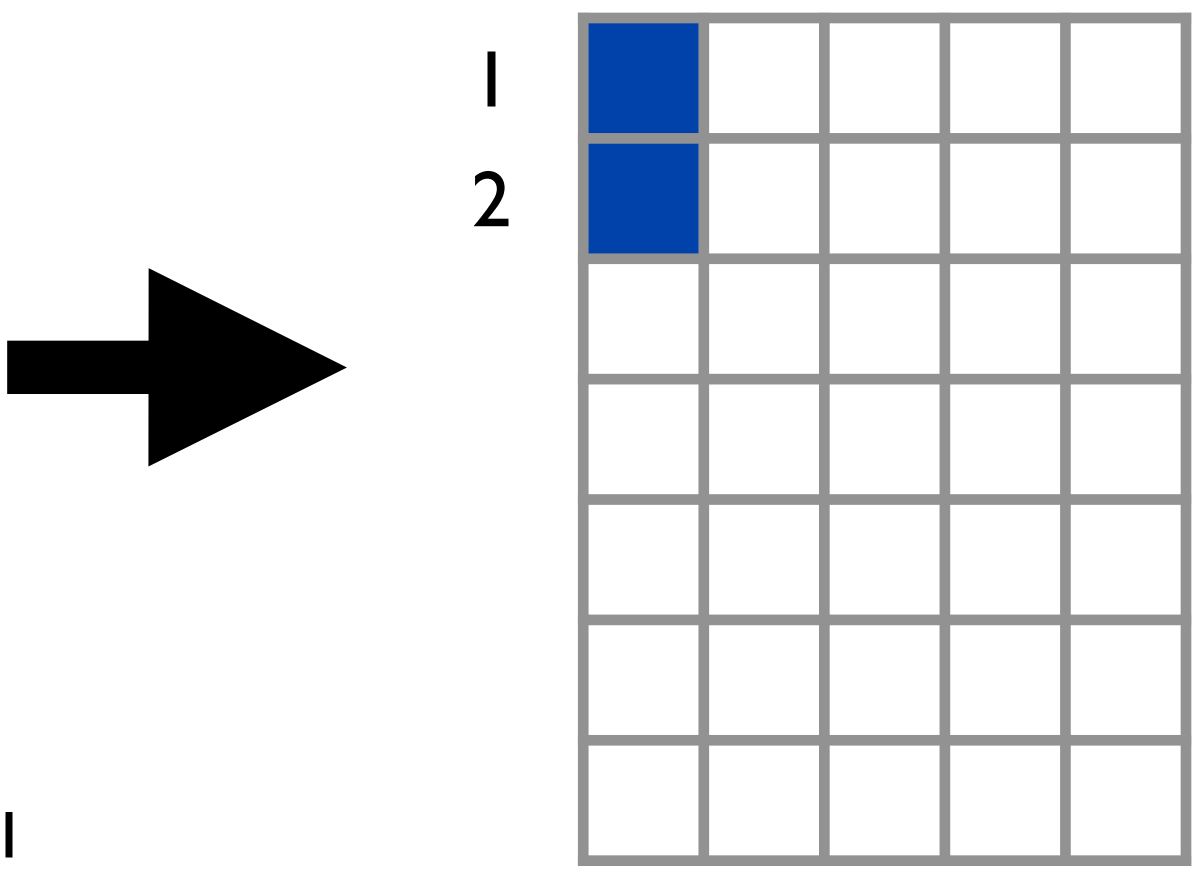


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

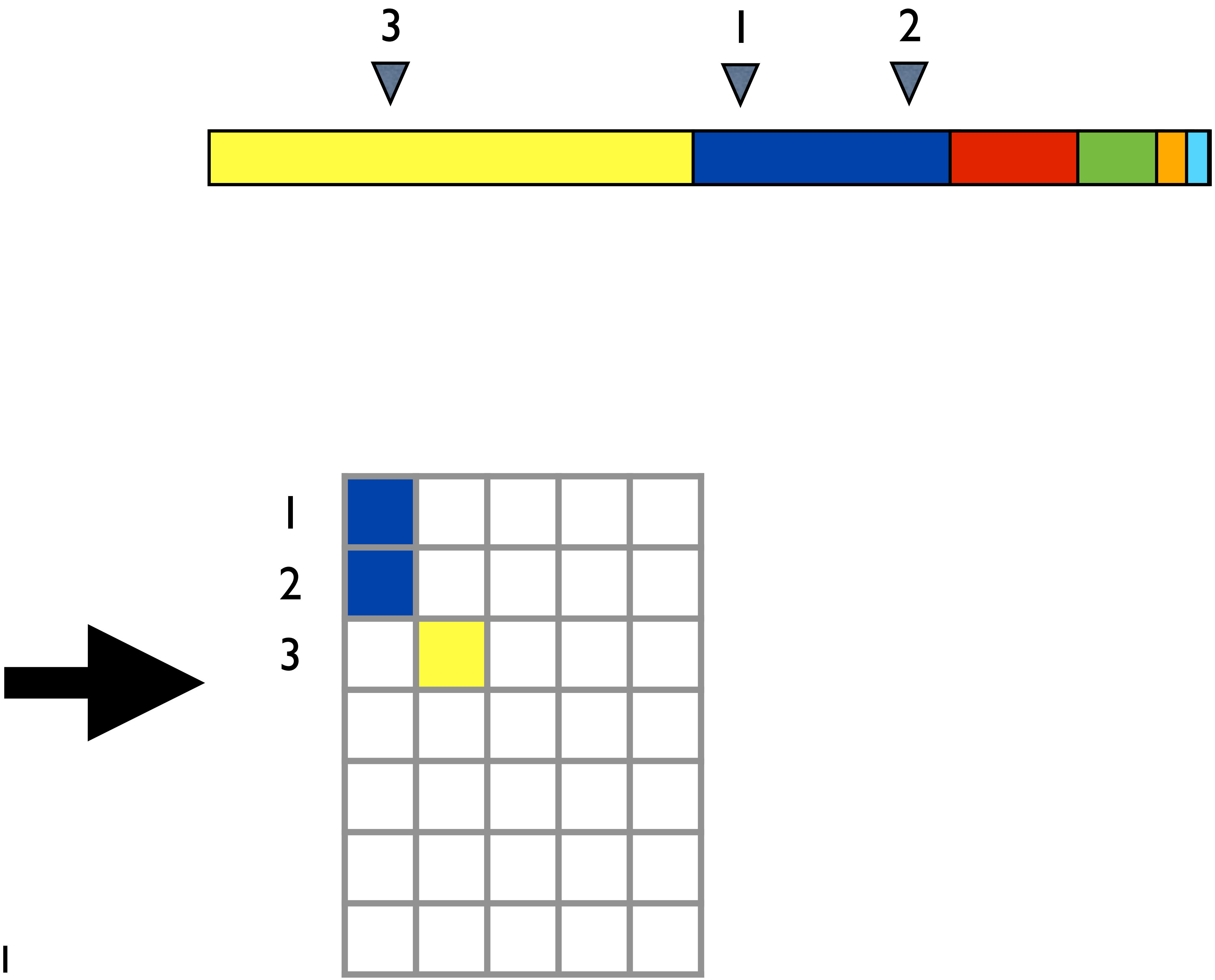




[Kingman 1978]

Paintboxes

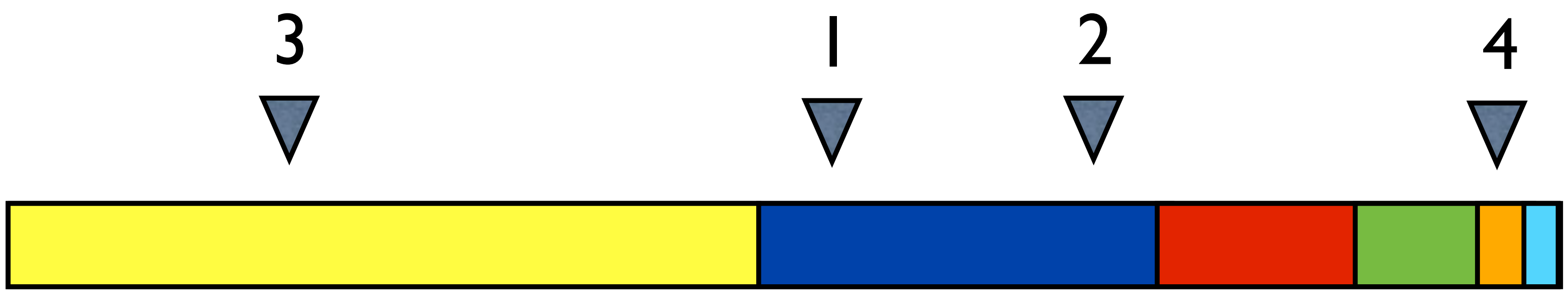
Exchangeable partition: Kingman paintbox

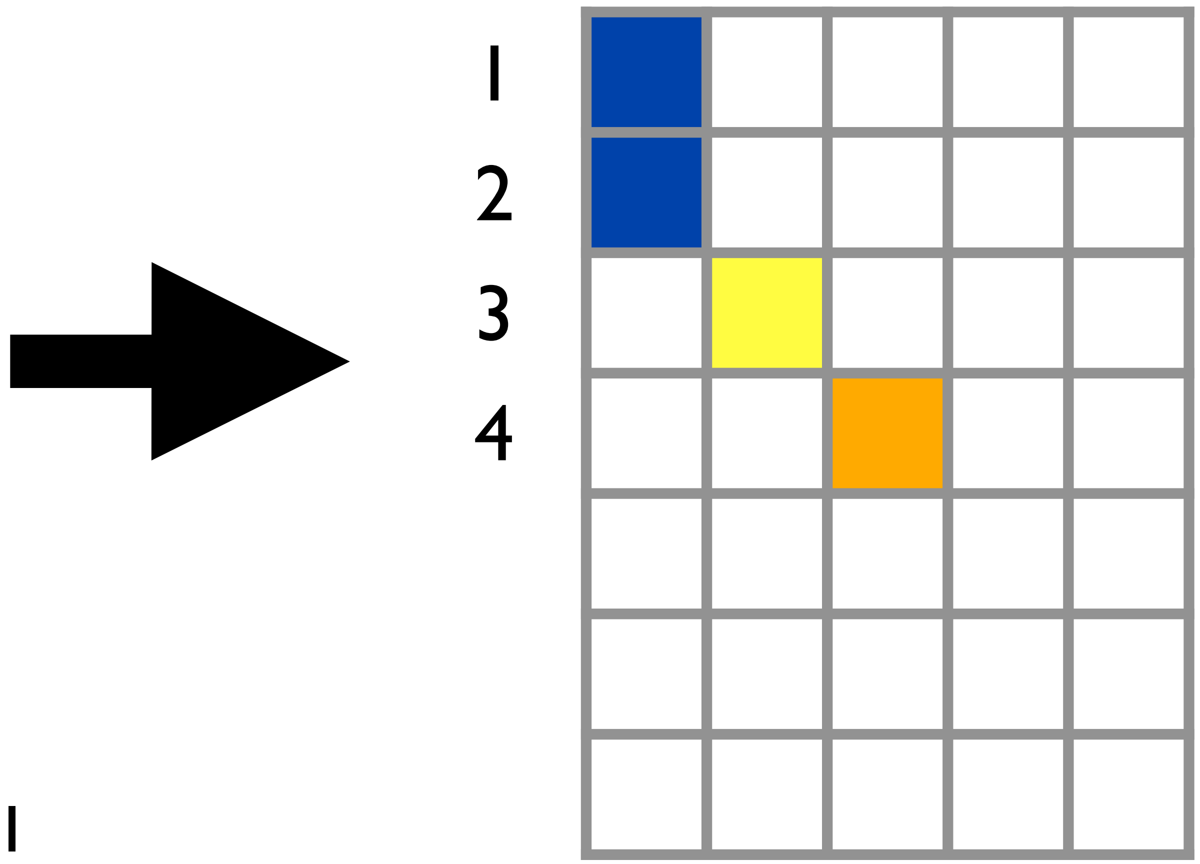


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

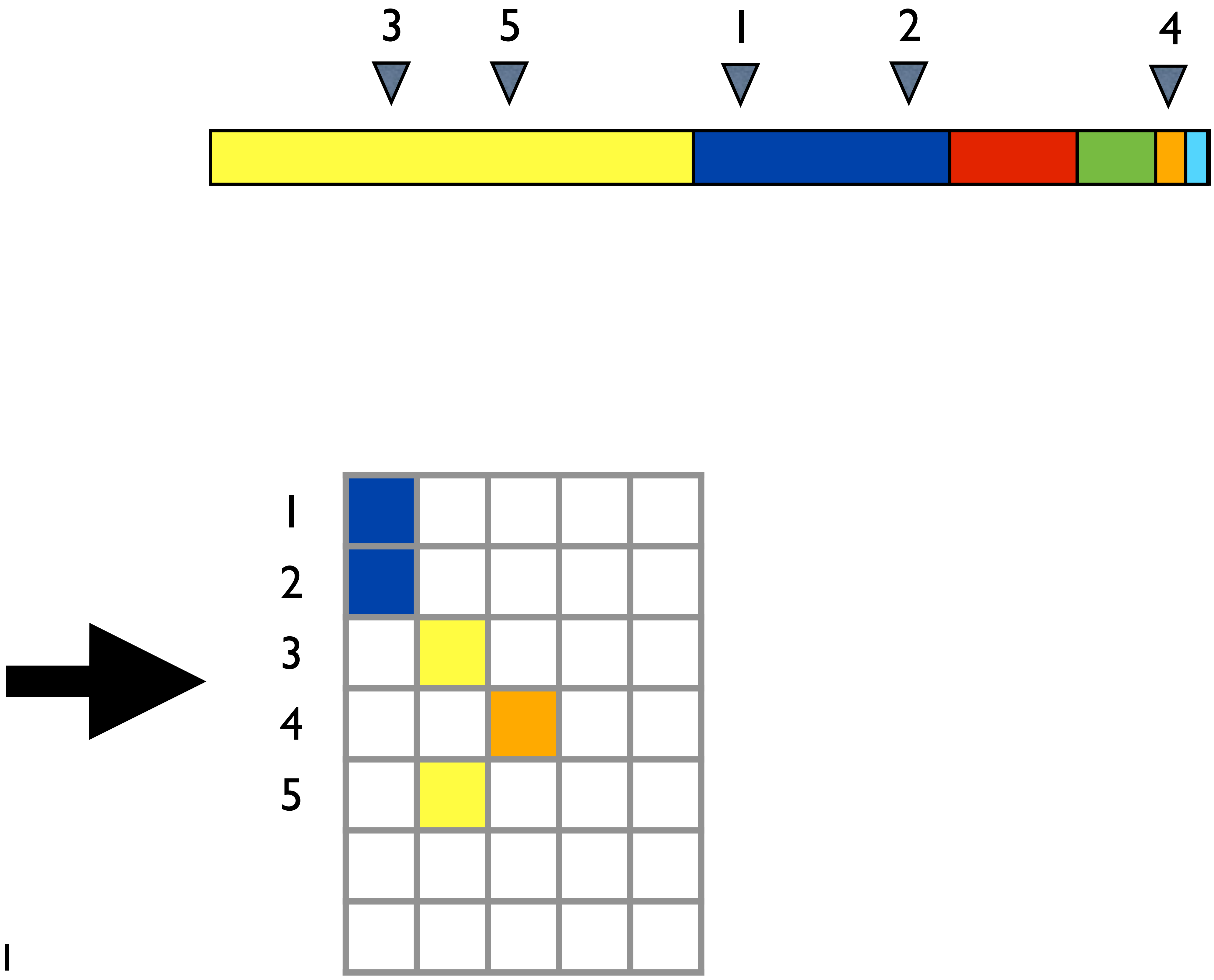




[Kingman 1978]

Paintboxes

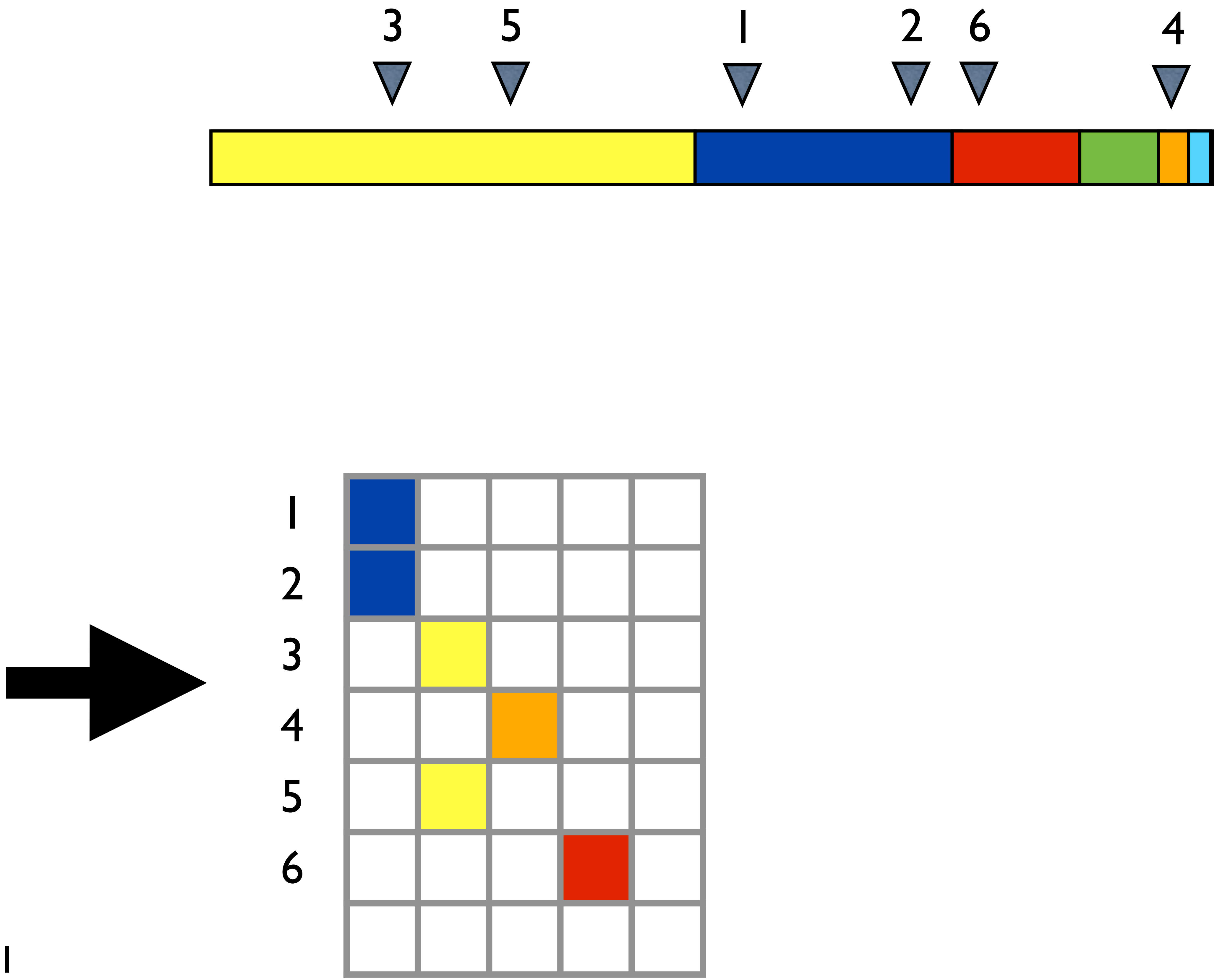
Exchangeable partition: Kingman paintbox



[Kingman 1978]

Paintboxes

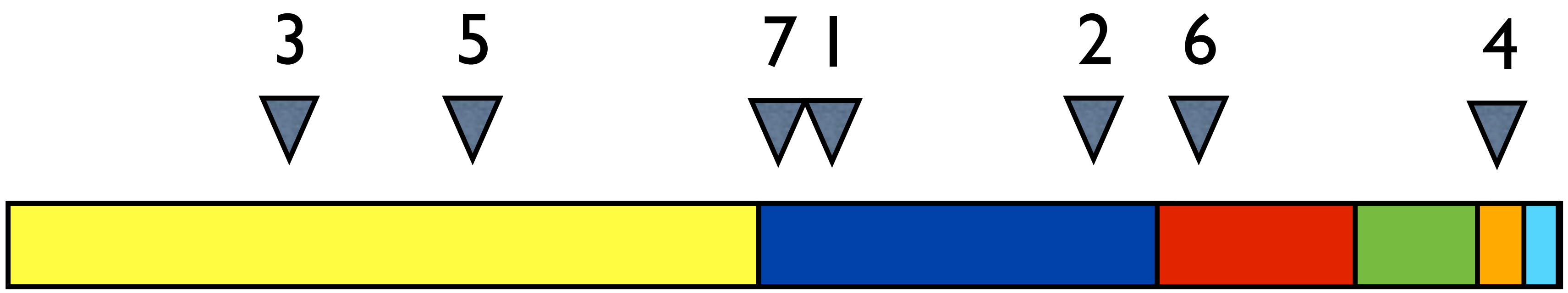
Exchangeable partition: Kingman paintbox

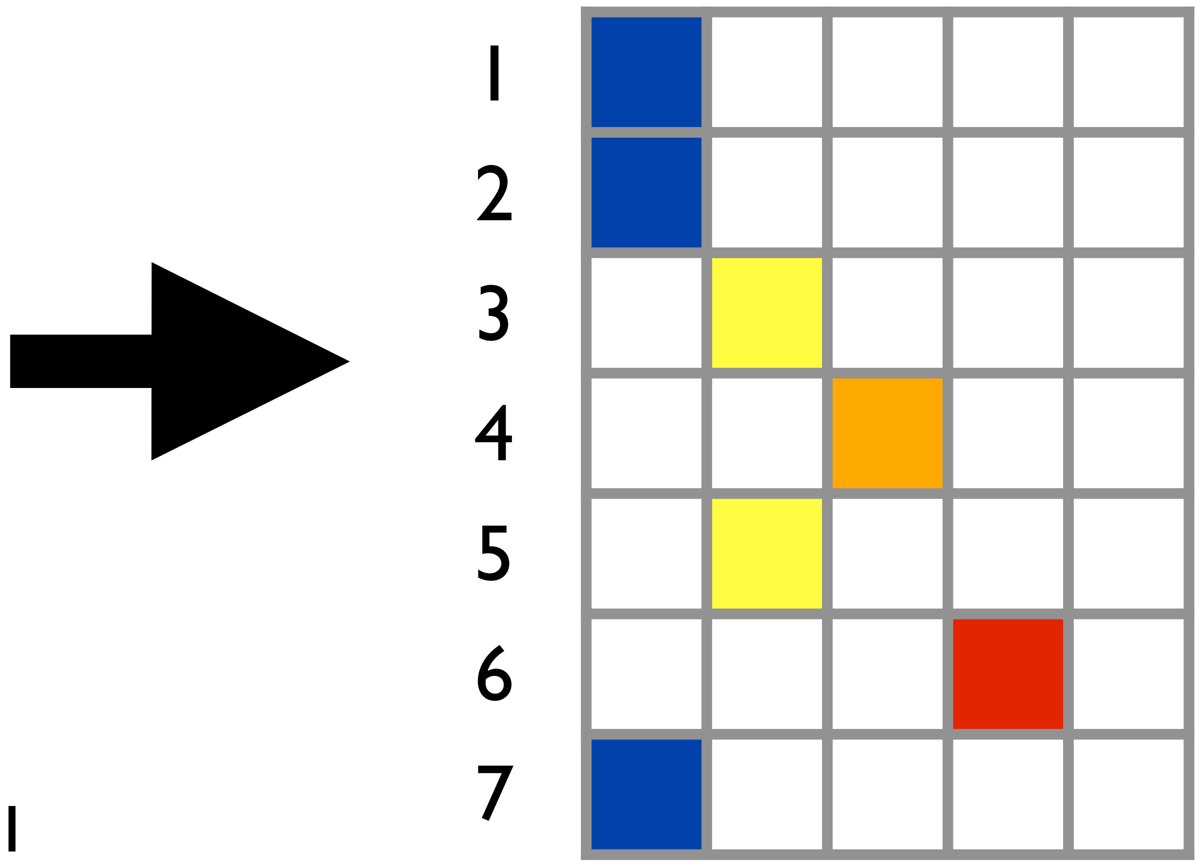


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

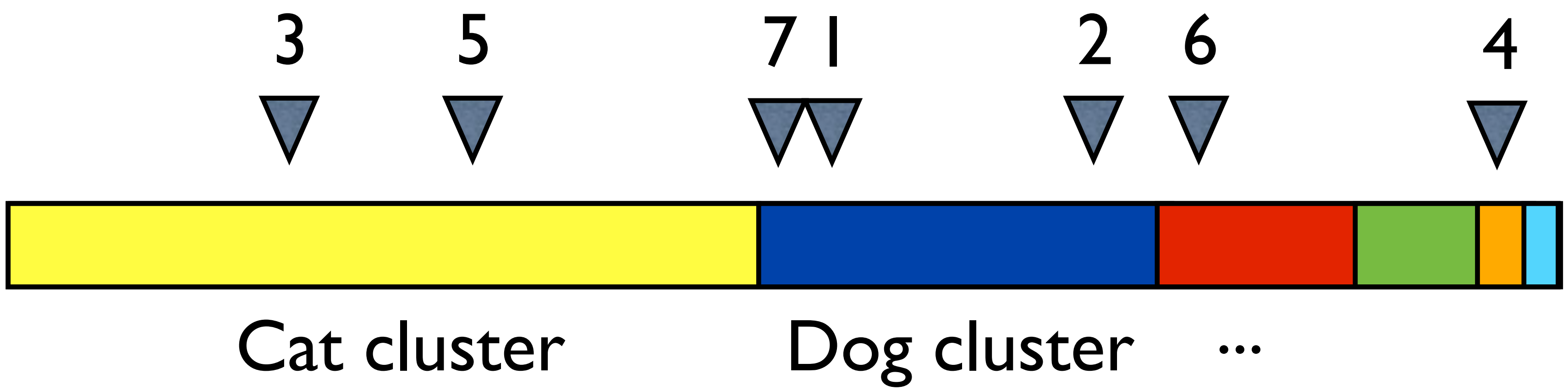


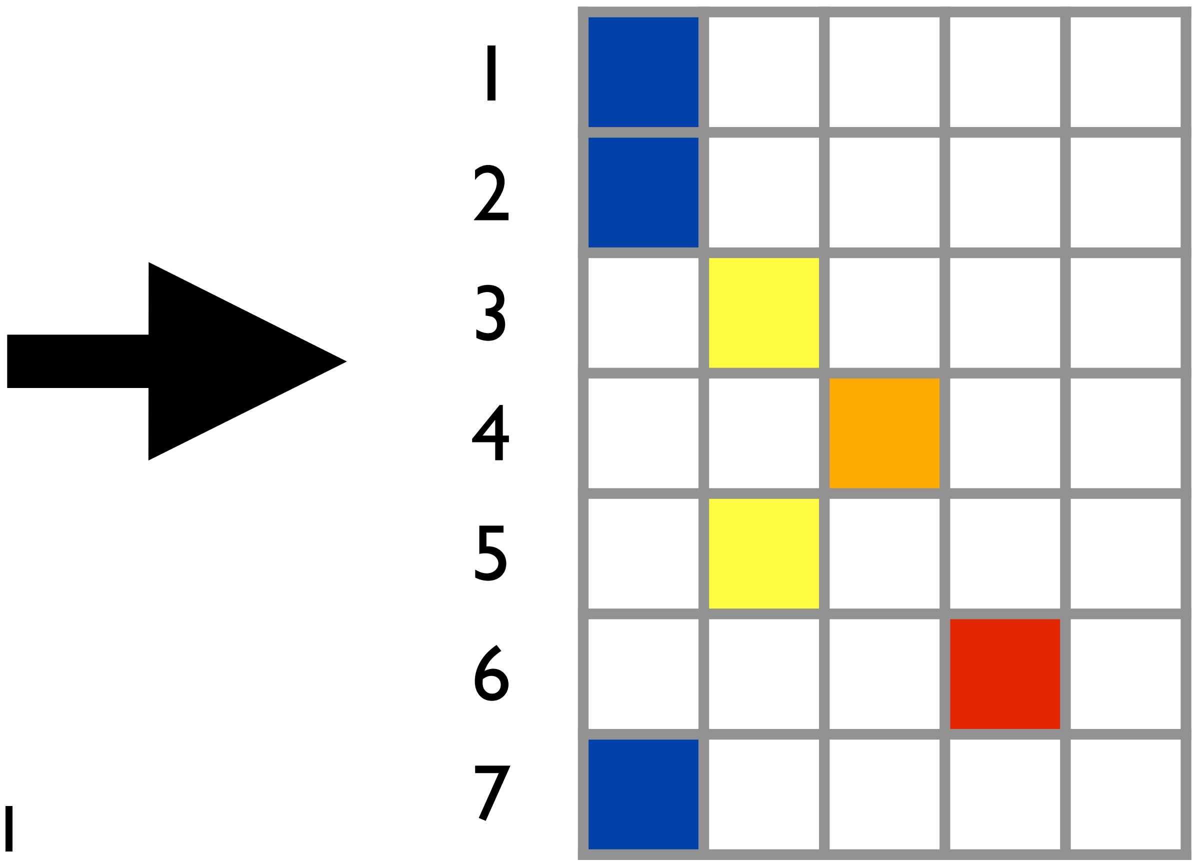


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

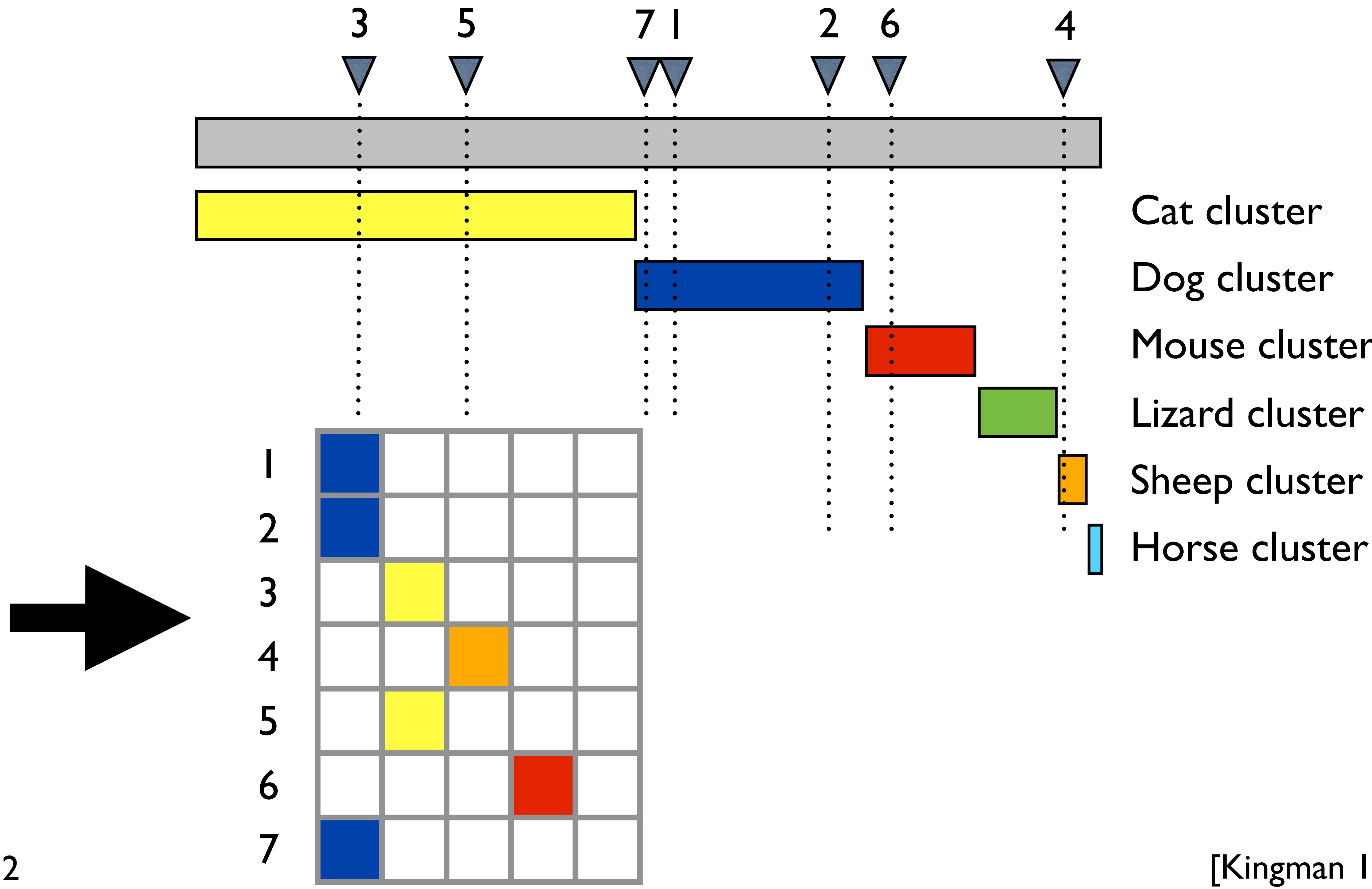




[Kingman 1978]

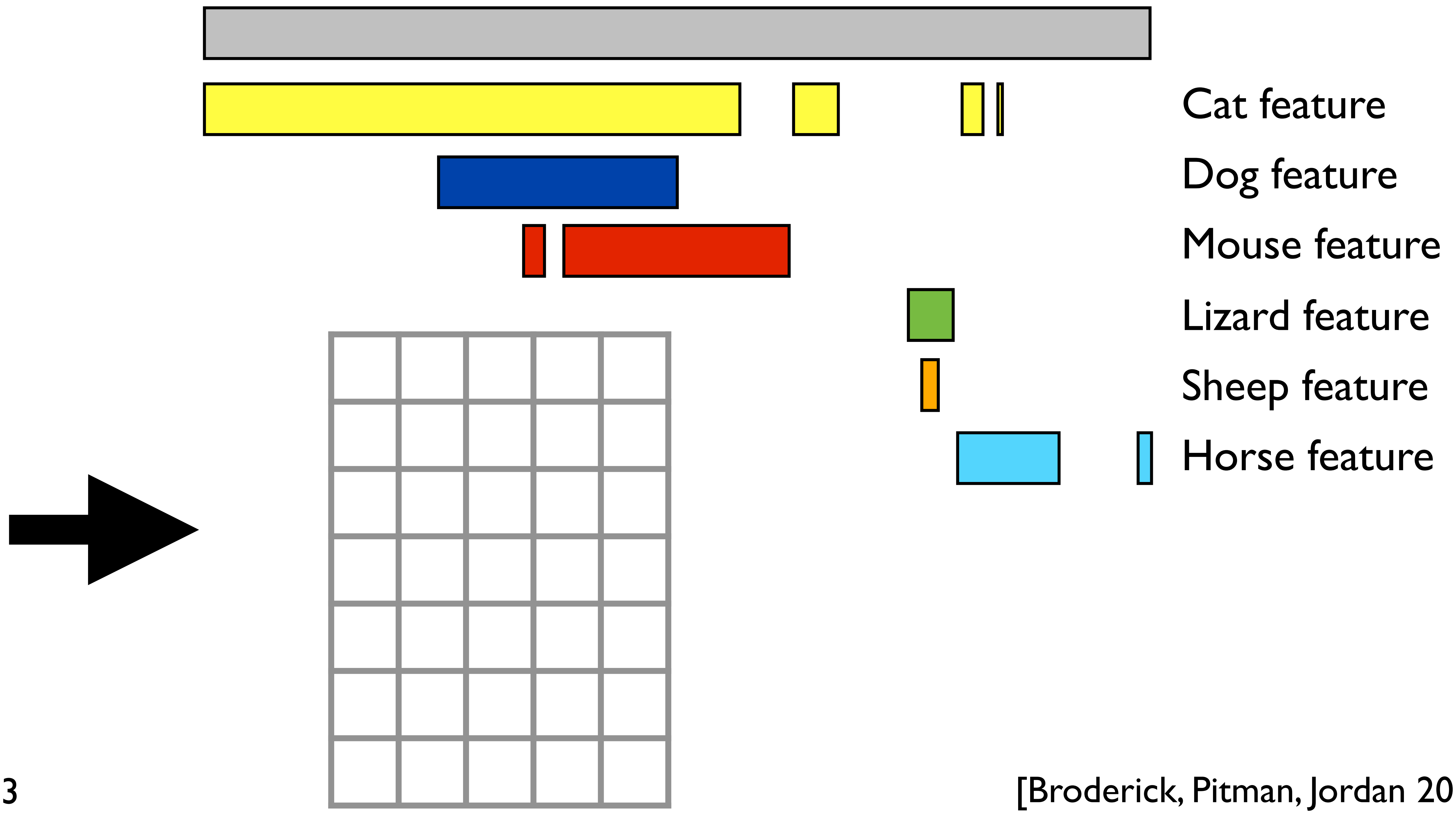
Paintboxes

Exchangeable partition: Kingman paintbox



[Kingman 1978]

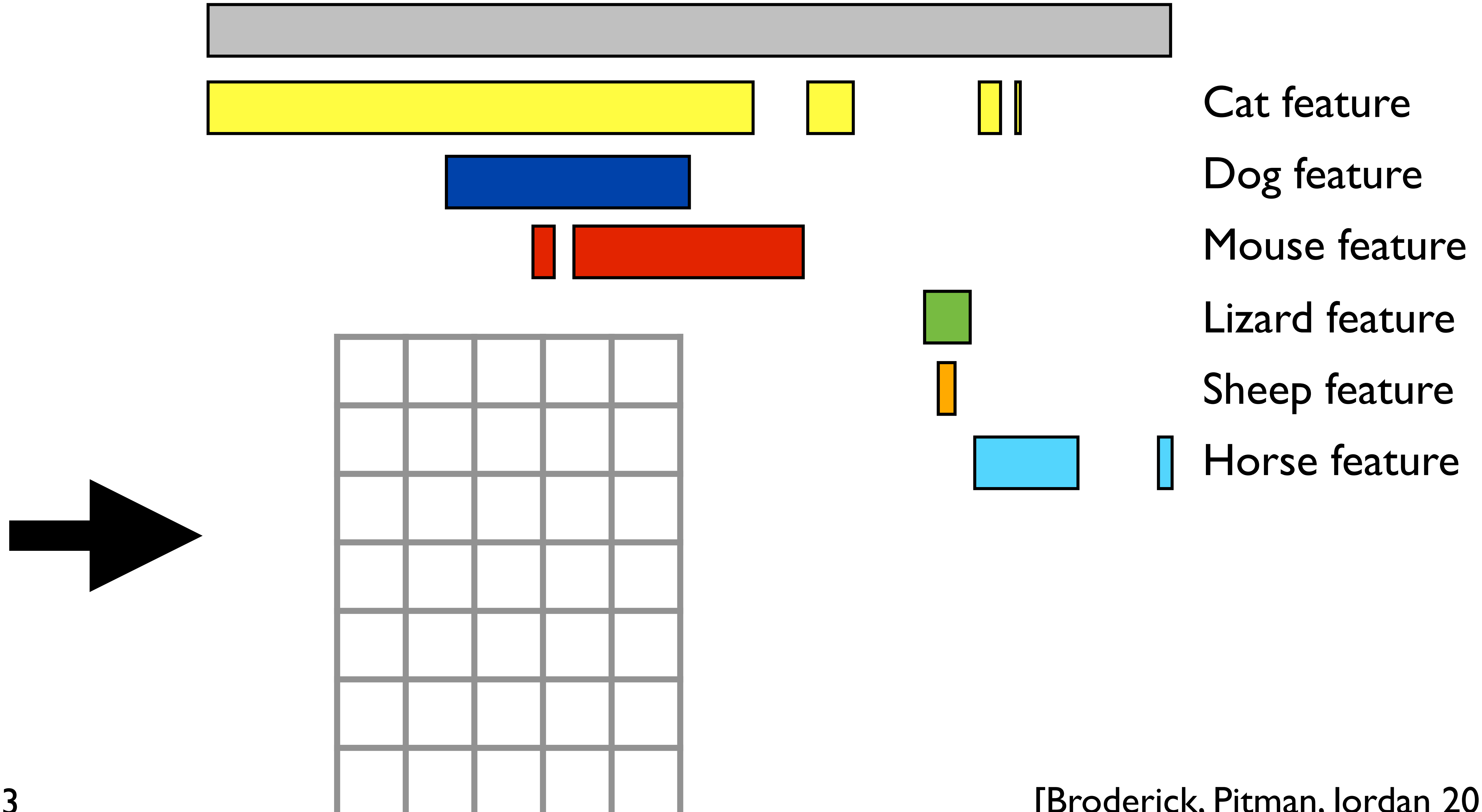
Paintboxes



[Broderick, Pitman, Jordan 2013]

Paintboxes

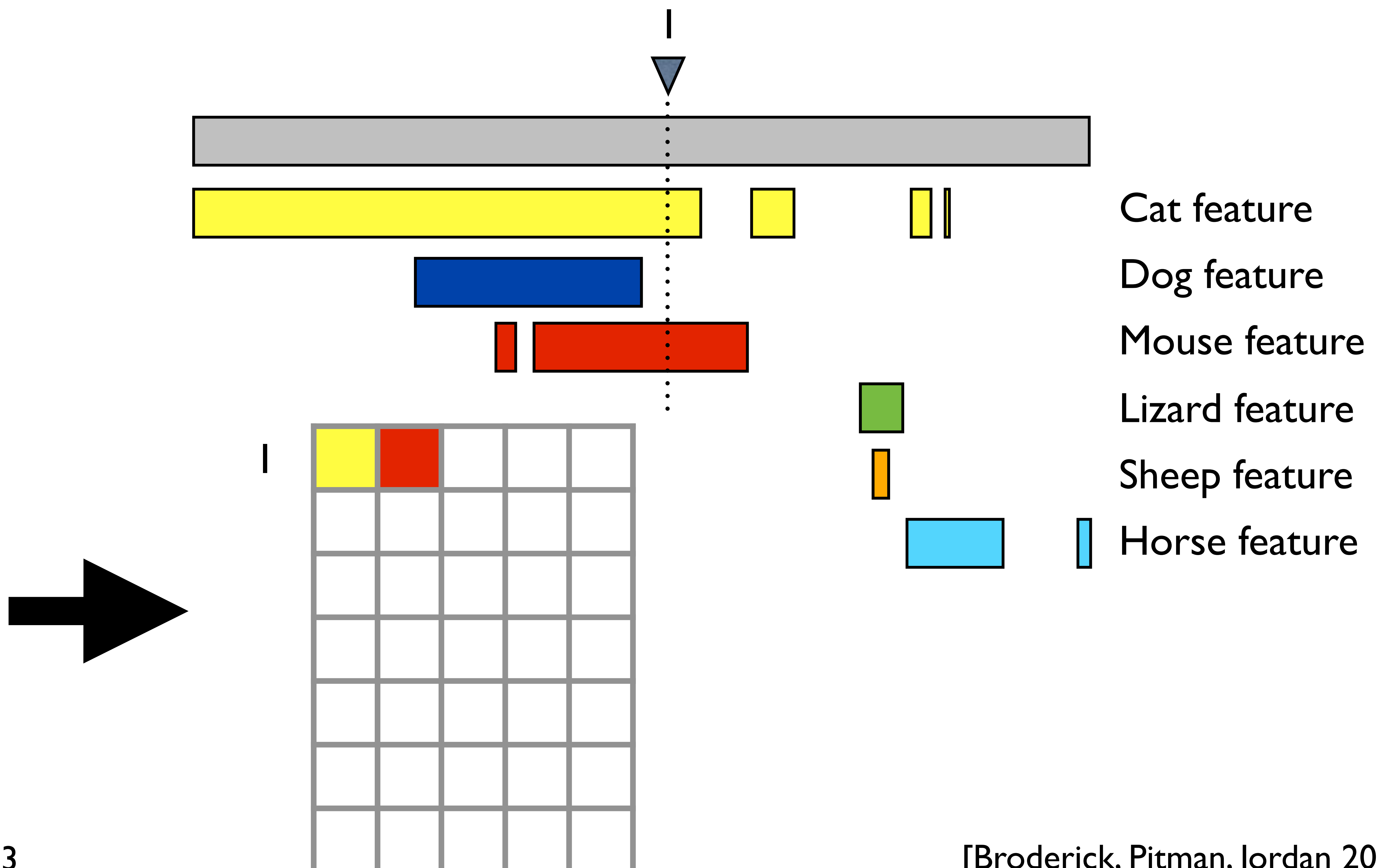
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

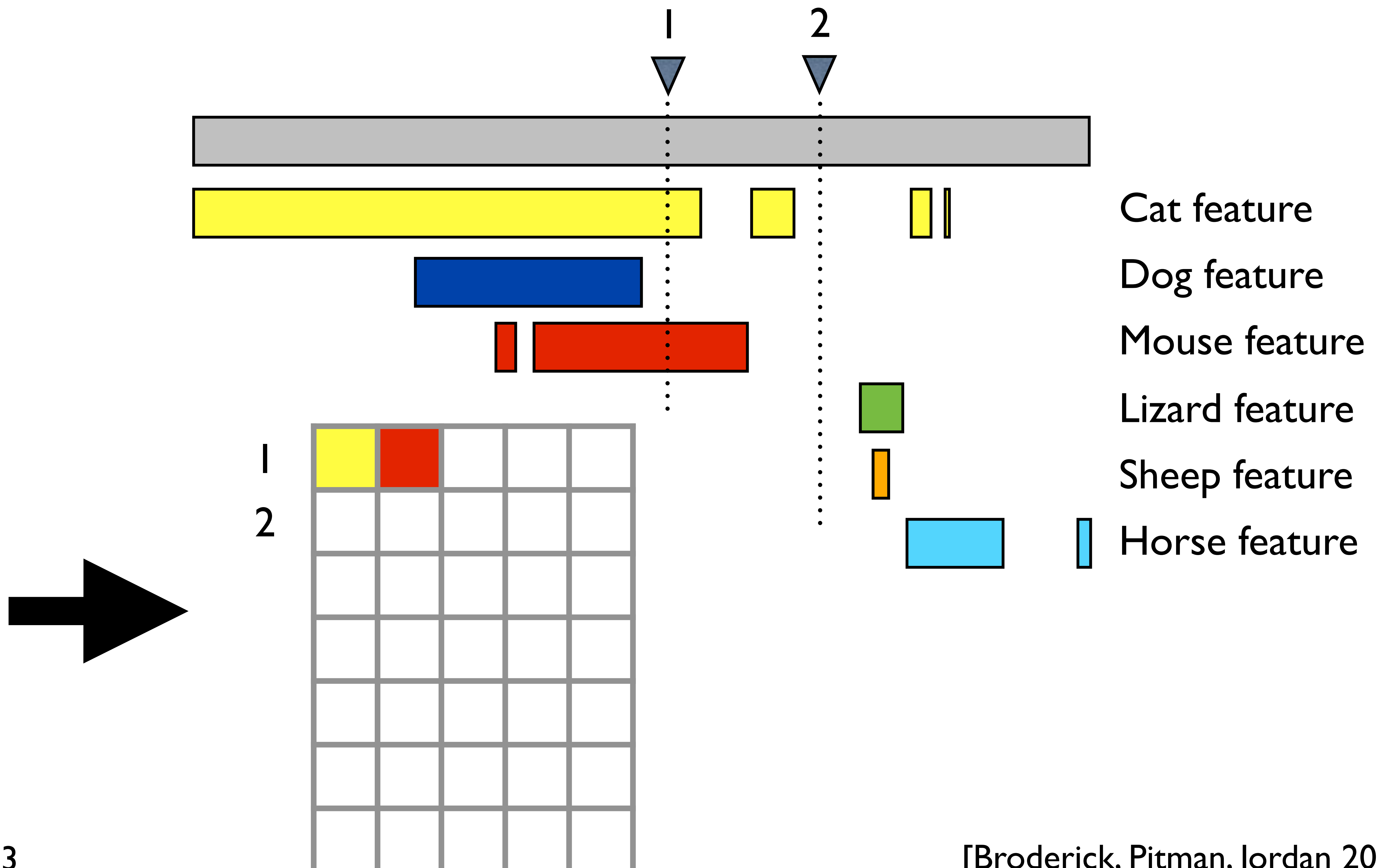
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

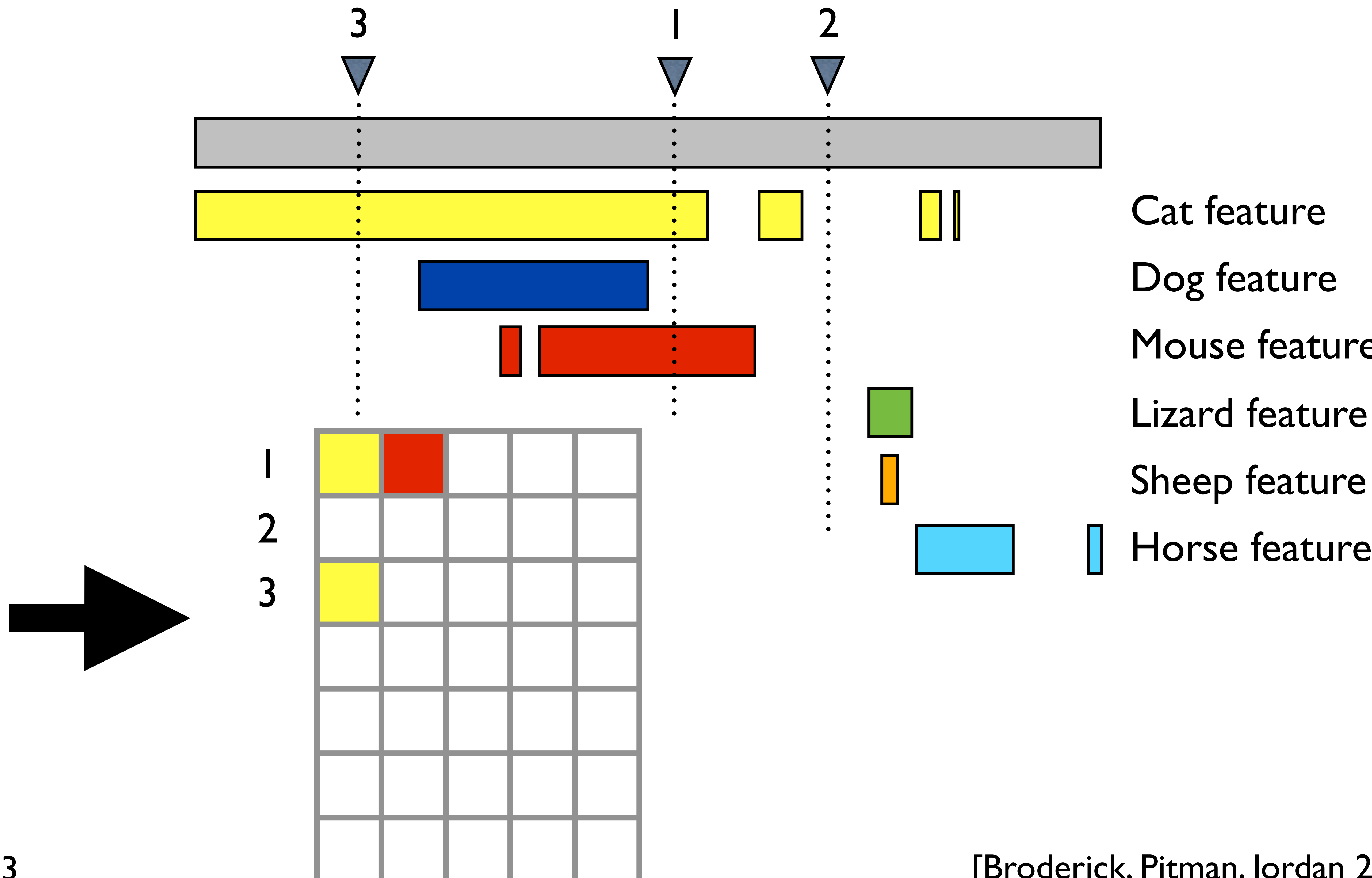
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

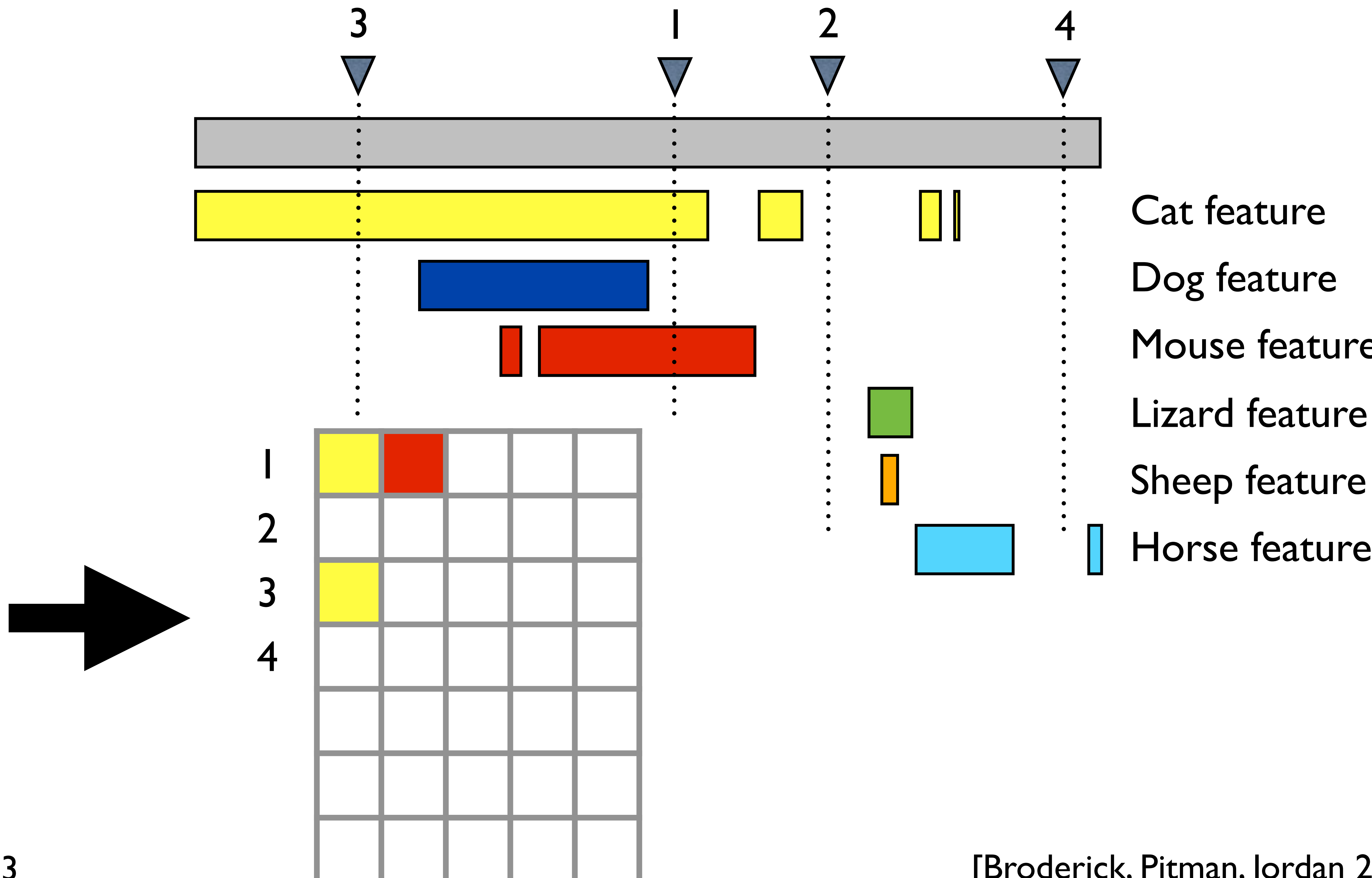
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

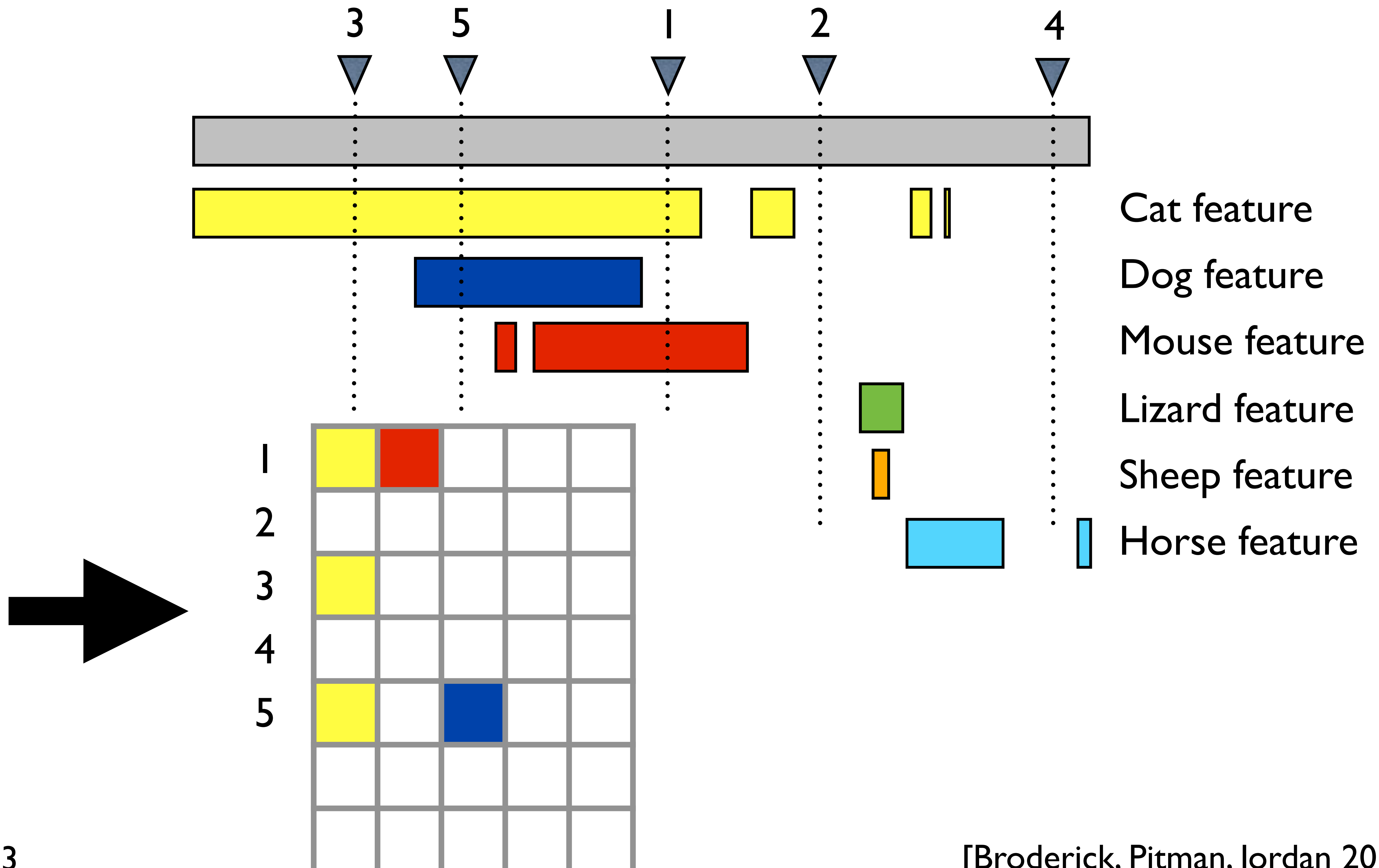
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

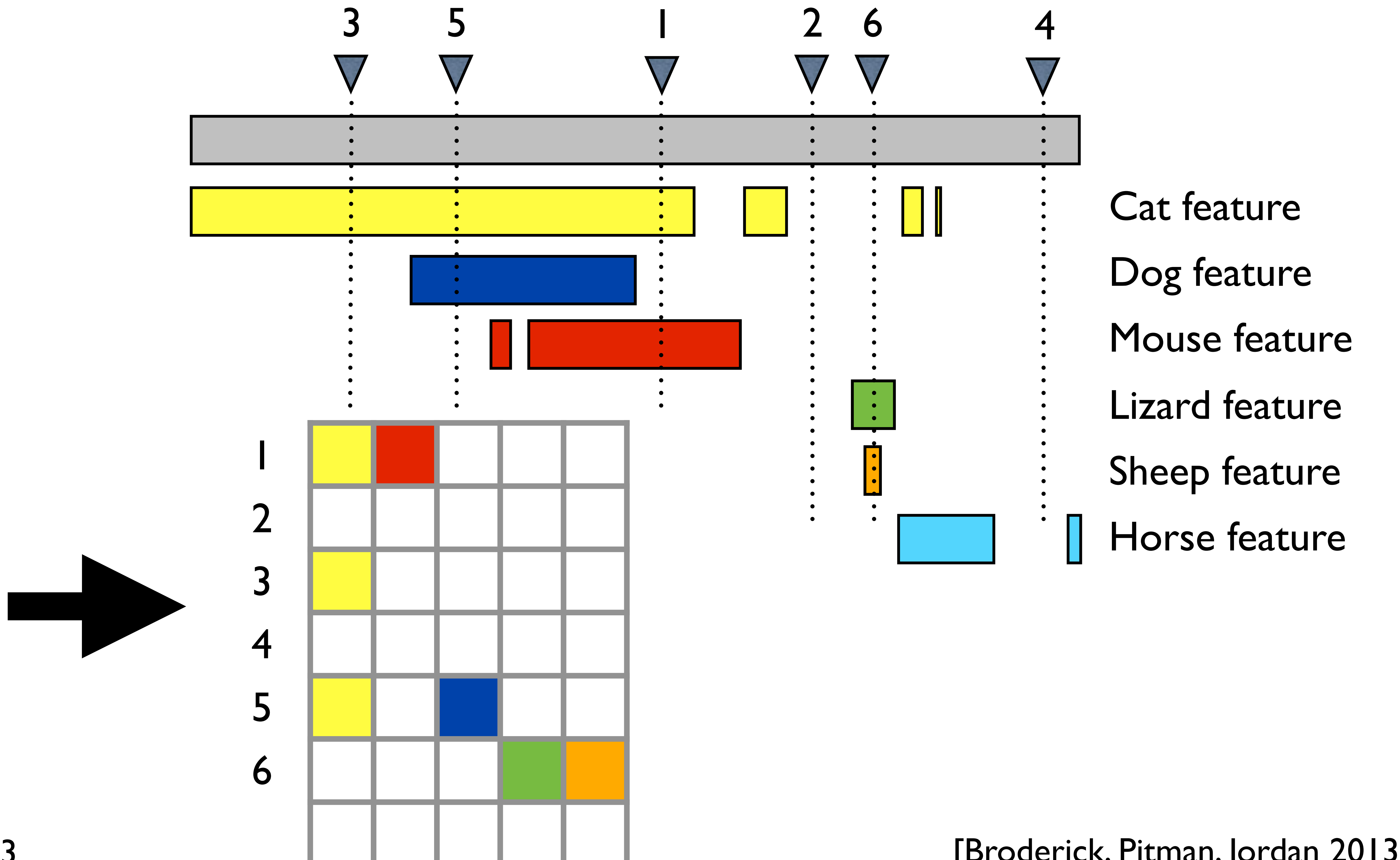
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

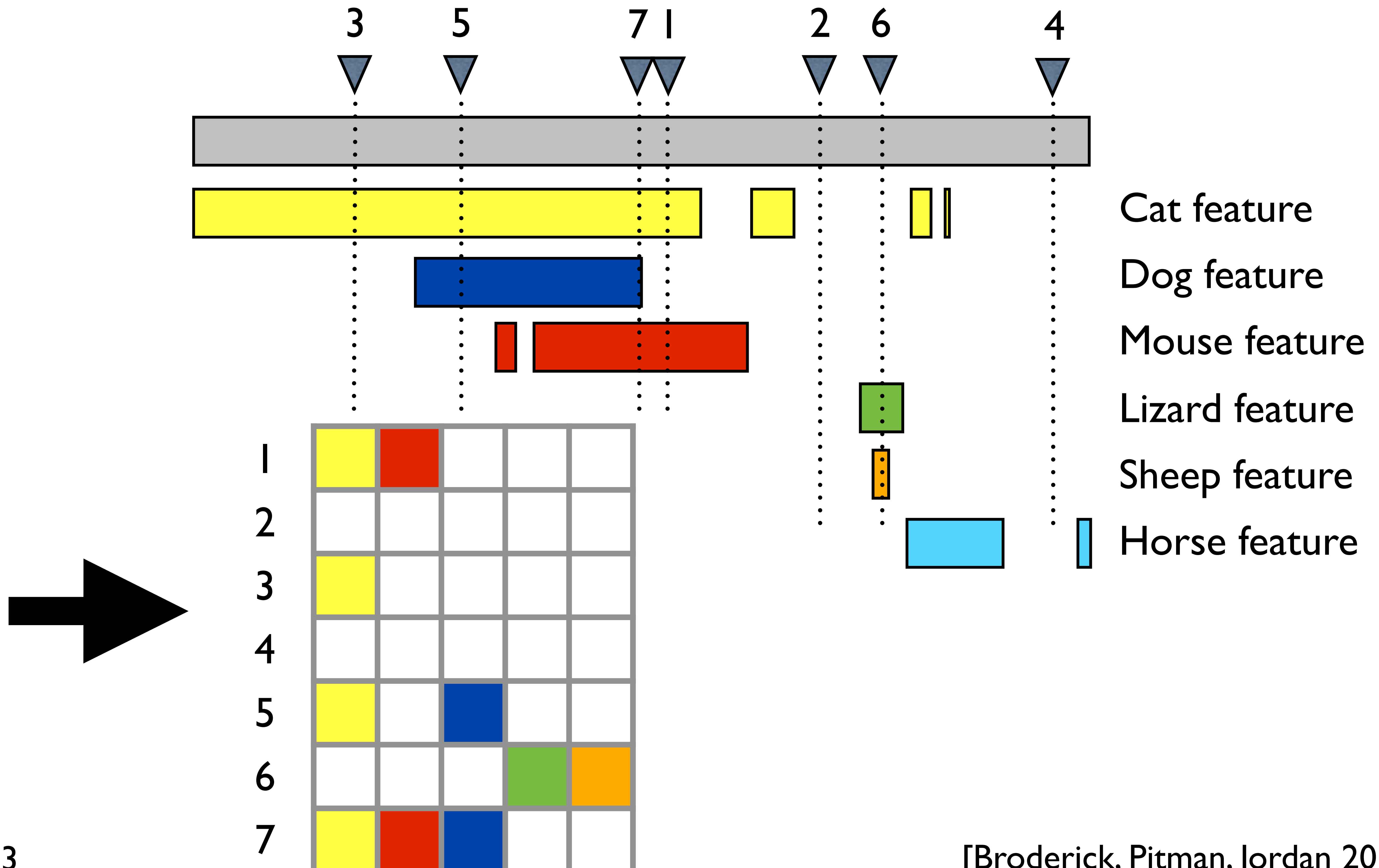
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

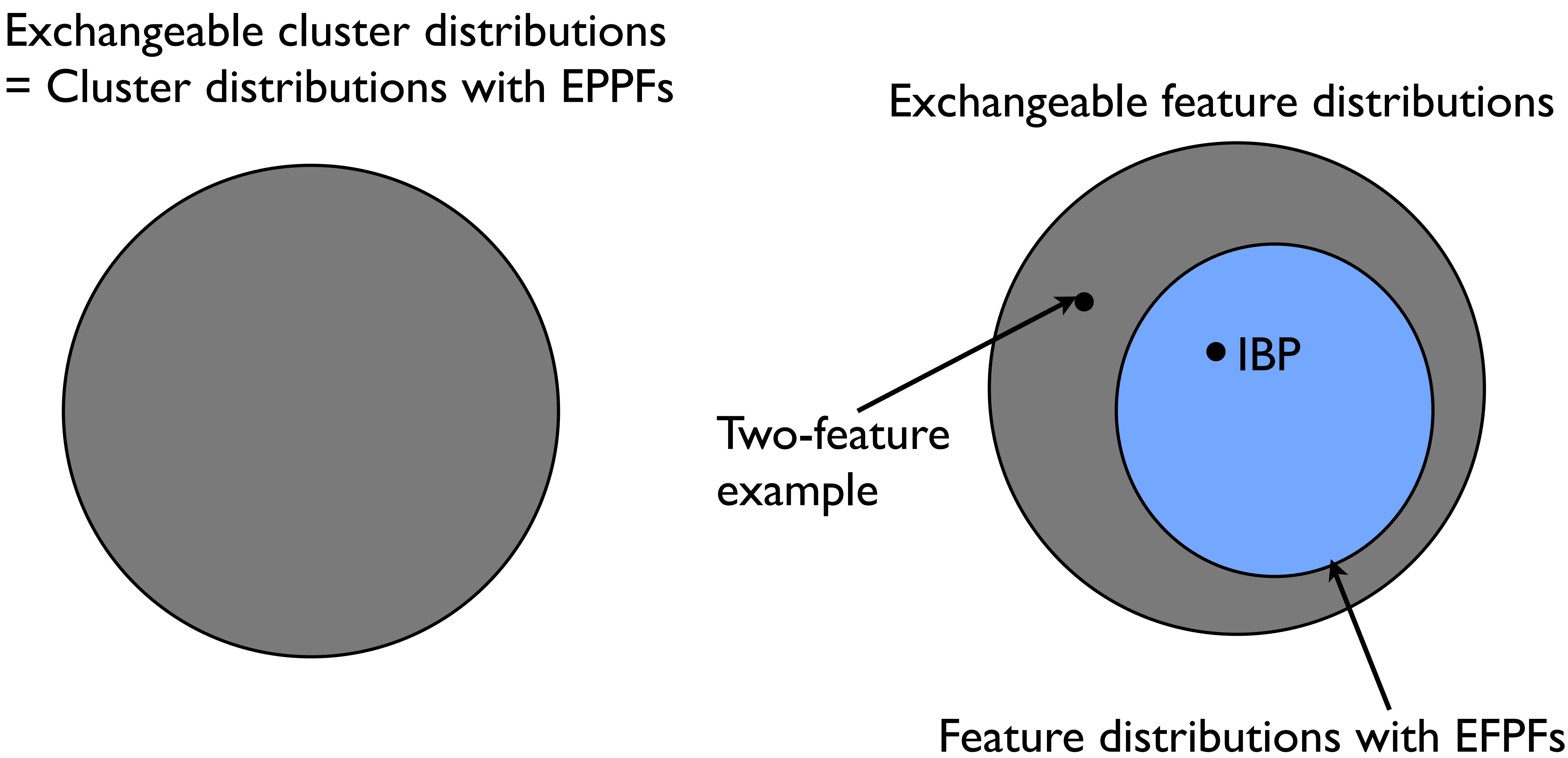
Paintboxes

Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan 2013]

Paintboxes

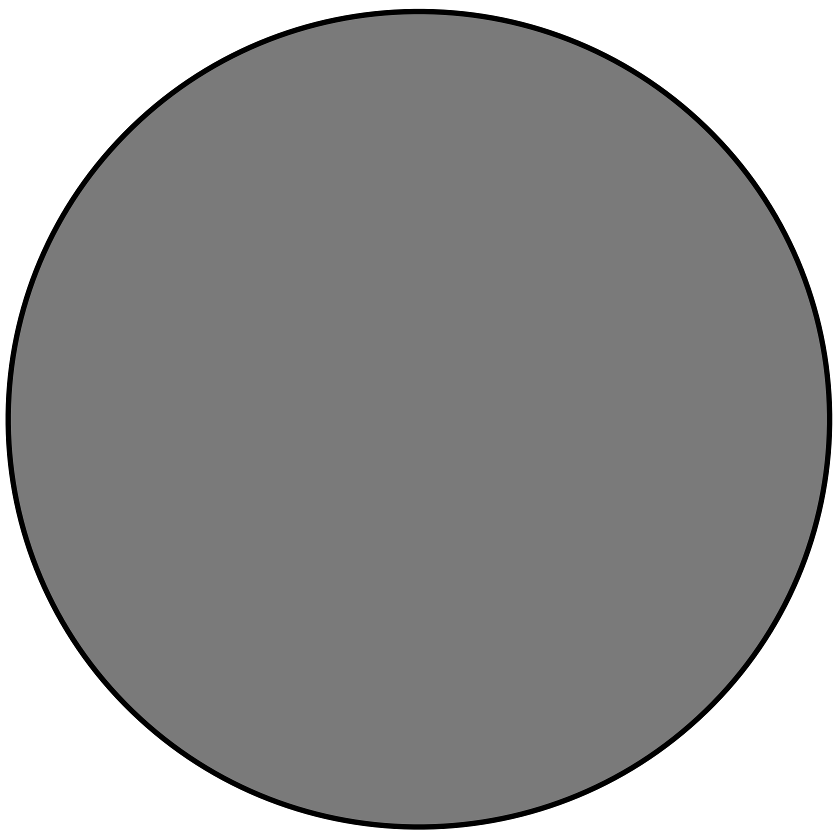


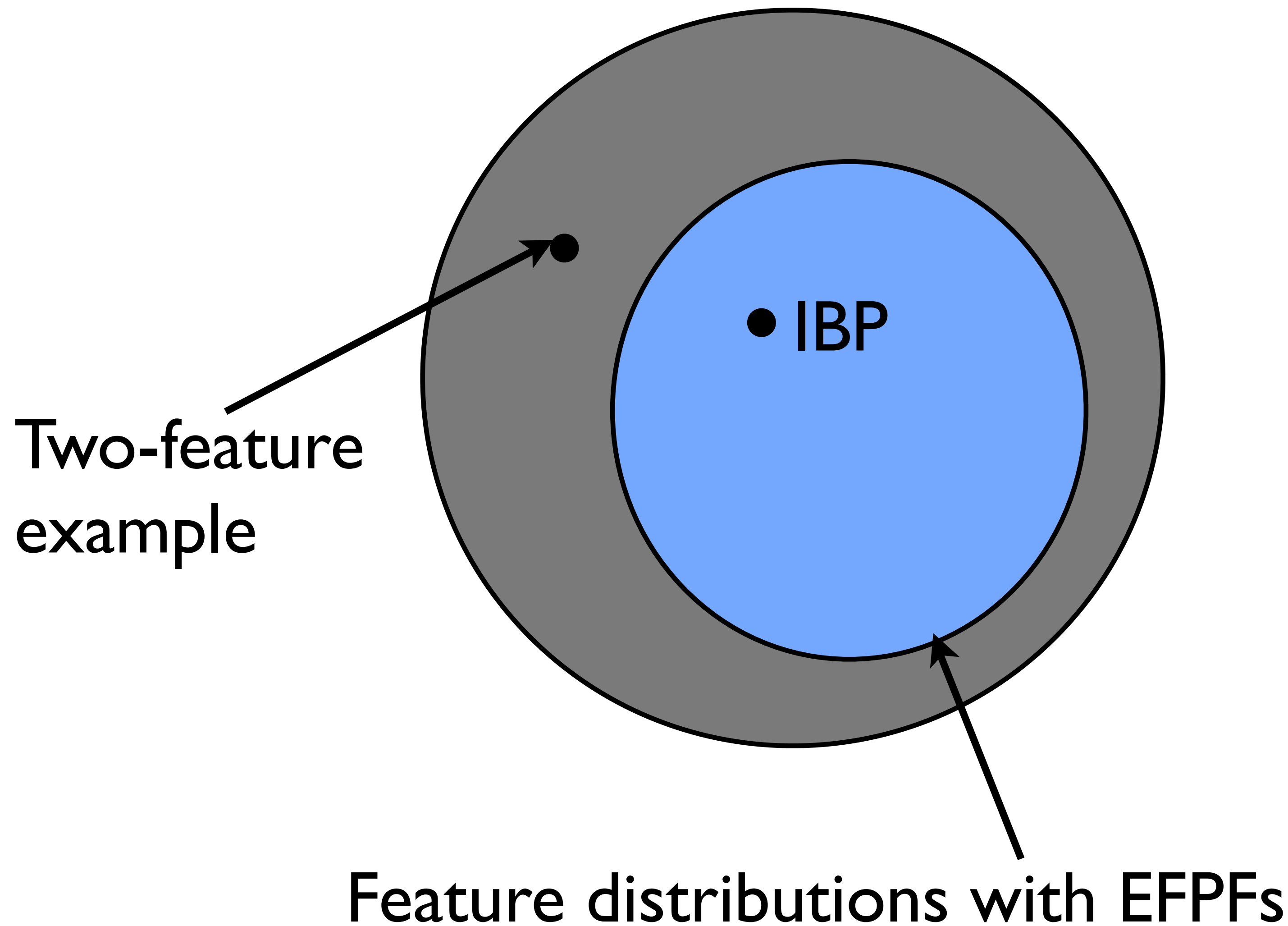
[Broderick, Pitman, Jordan 2013]

Paintboxes

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

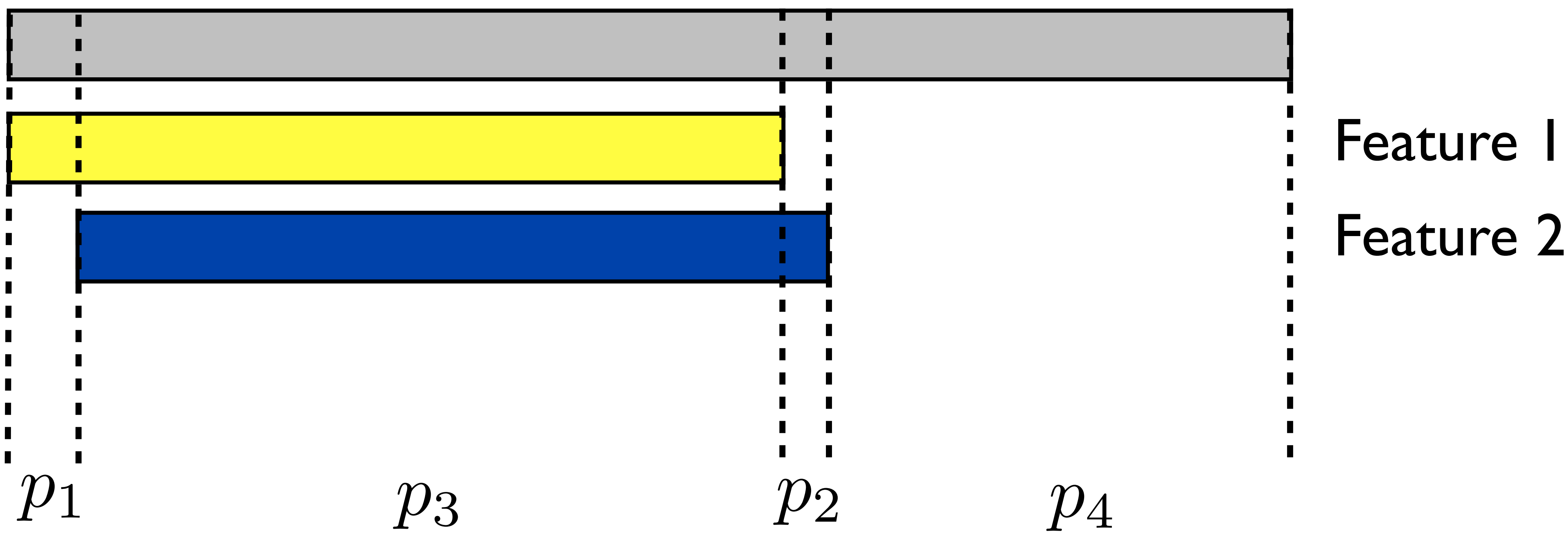




[Broderick, Pitman, Jordan 2013]

Paintboxes

Two feature example



$$\mathbb{P}(\text{row} = \blacksquare\square) = p_1$$

$$\mathbb{P}(\text{row} = \square\blacksquare) = p_2$$

$$\mathbb{P}(\text{row} = \blacksquare\blacksquare) = p_3$$

$$\mathbb{P}(\text{row} = \square\square) = p_4$$

Paintboxes

Indian buffet process: beta feature frequencies

[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For *m* = 1, 2, ...

1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + m - 1}\right)$

[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For *m* = 1, 2, ...

1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$

[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...

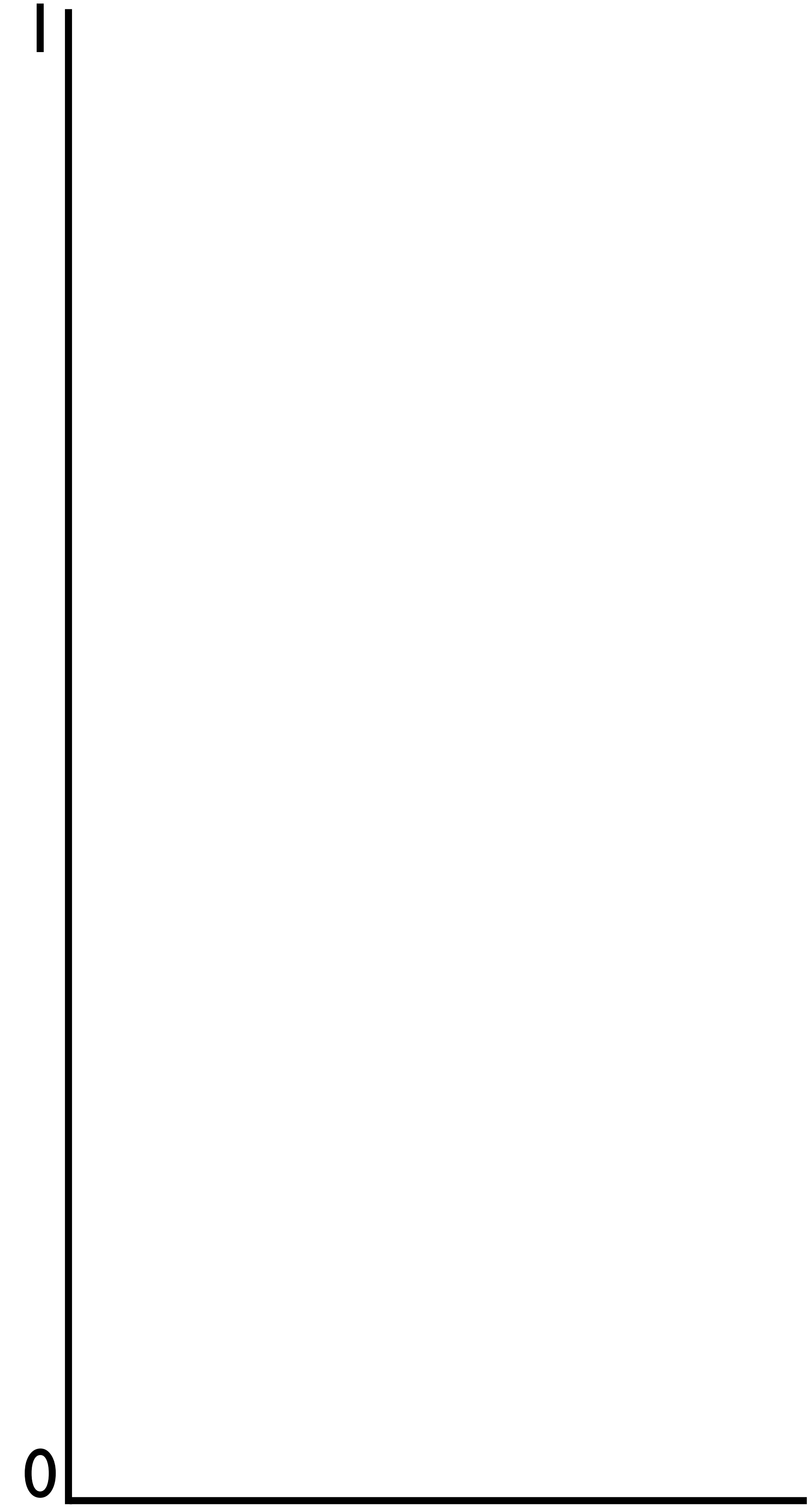
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...

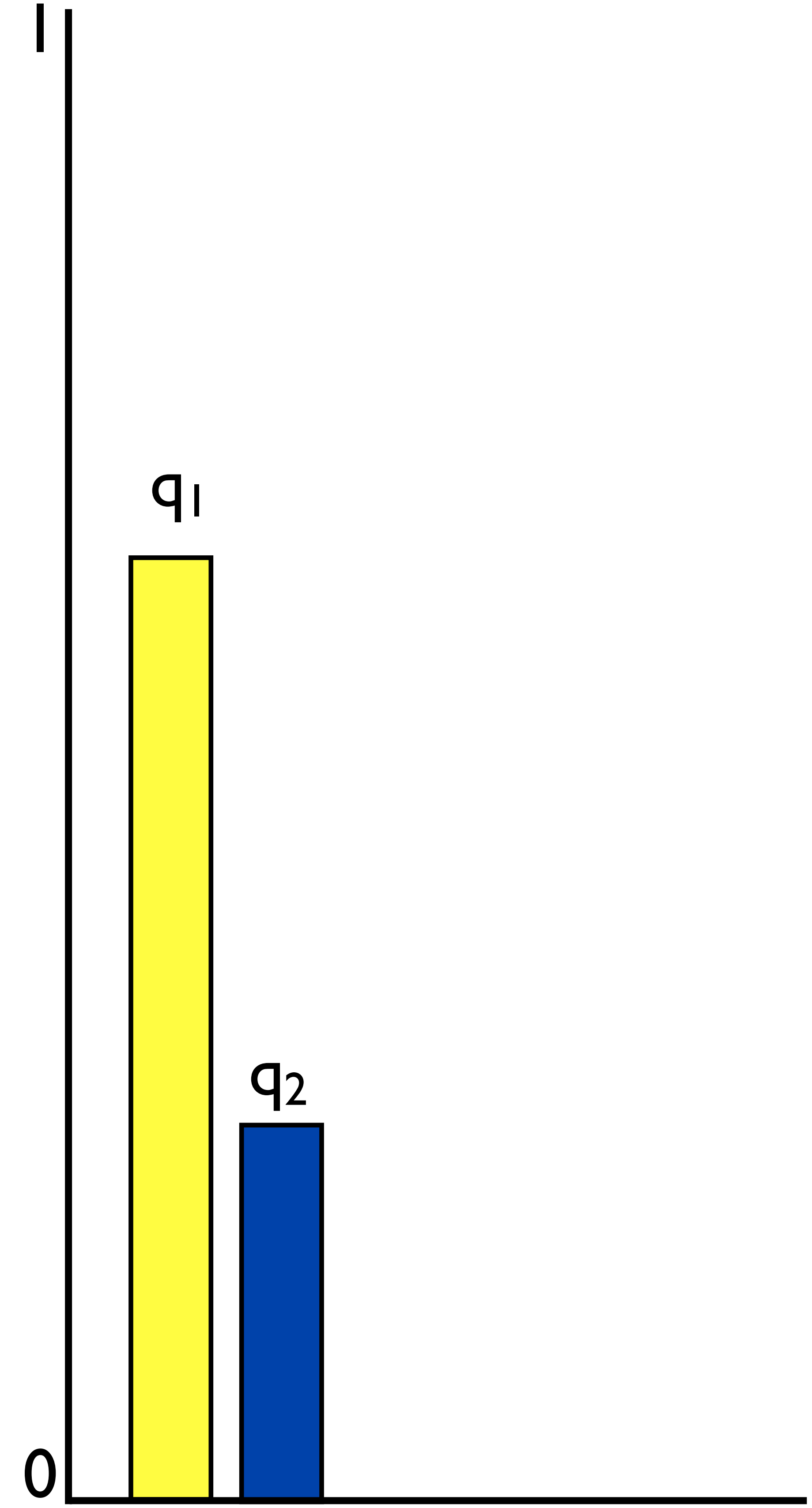
1. Draw $K_m^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \text{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...

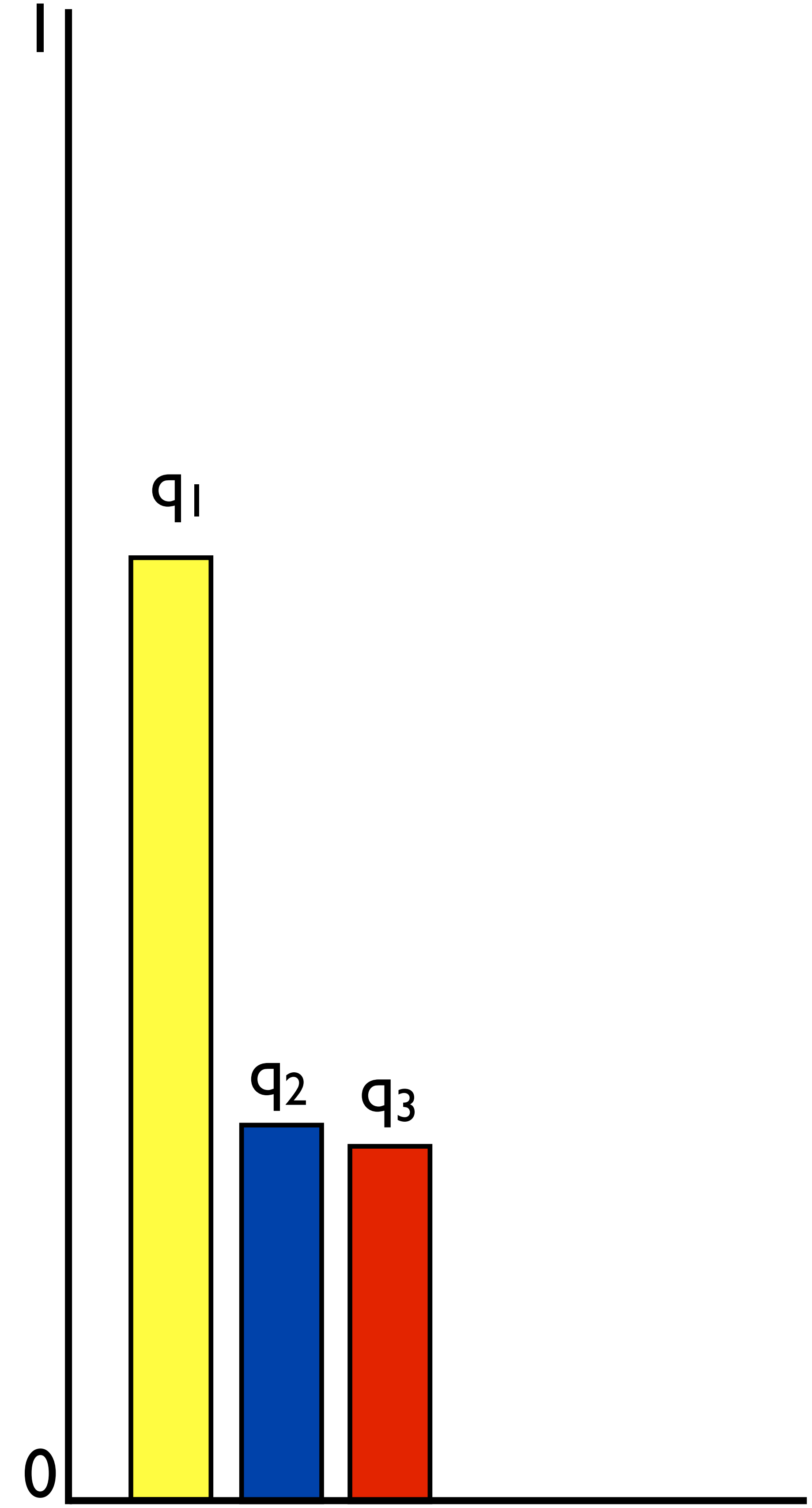
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For *m* = 1, 2, ...

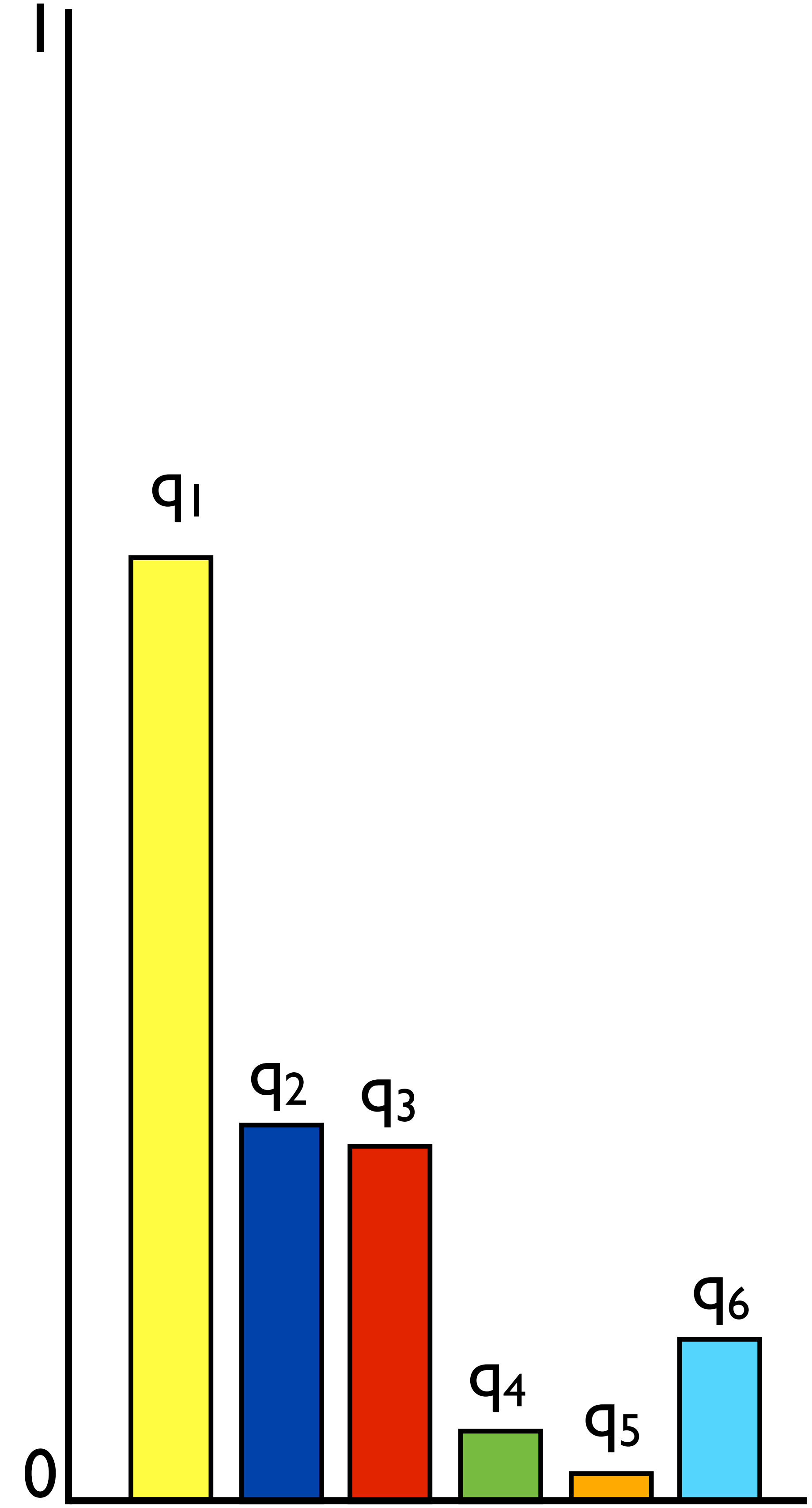
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...

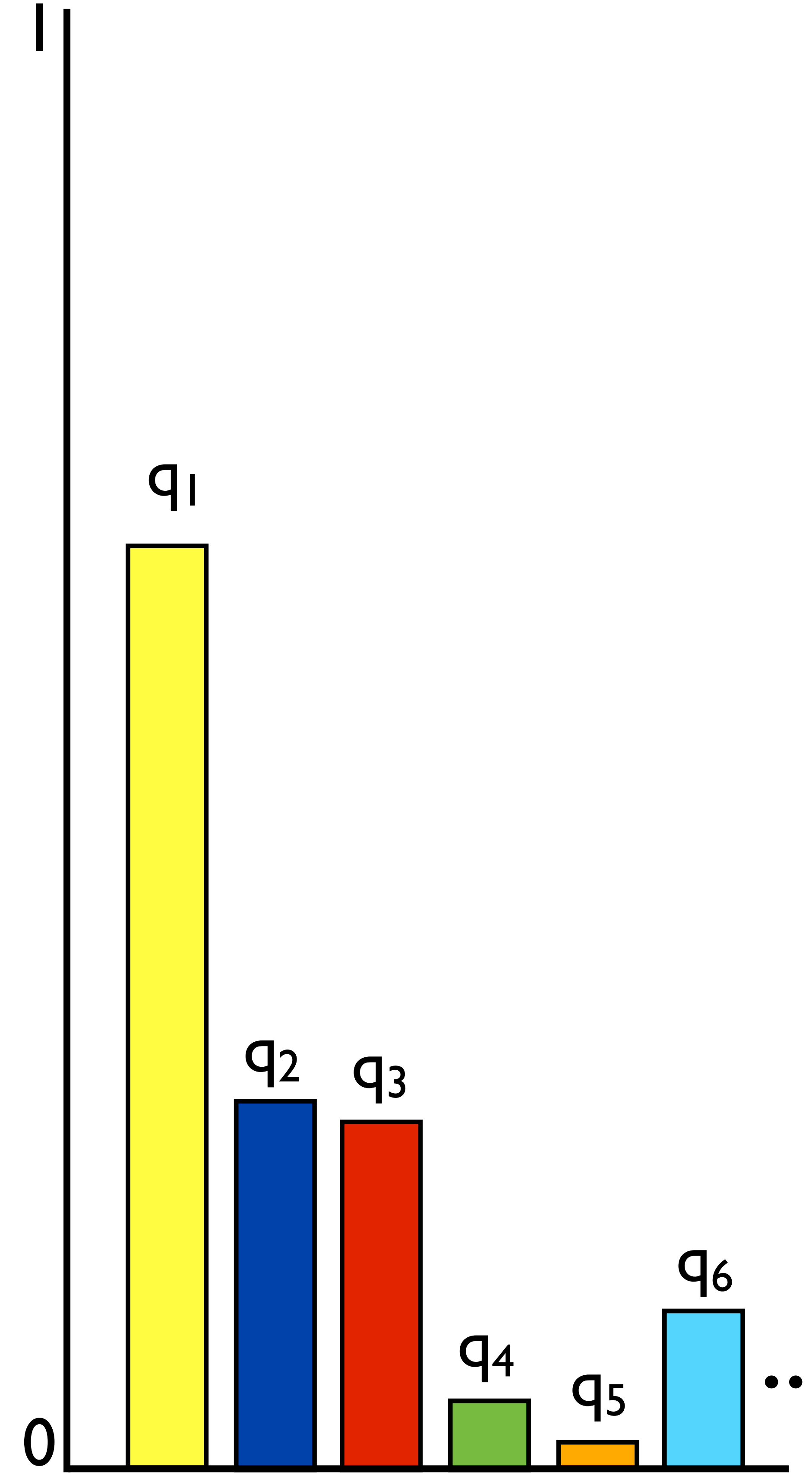
1. Draw $K_m^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

   $$q_k \sim \text{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...

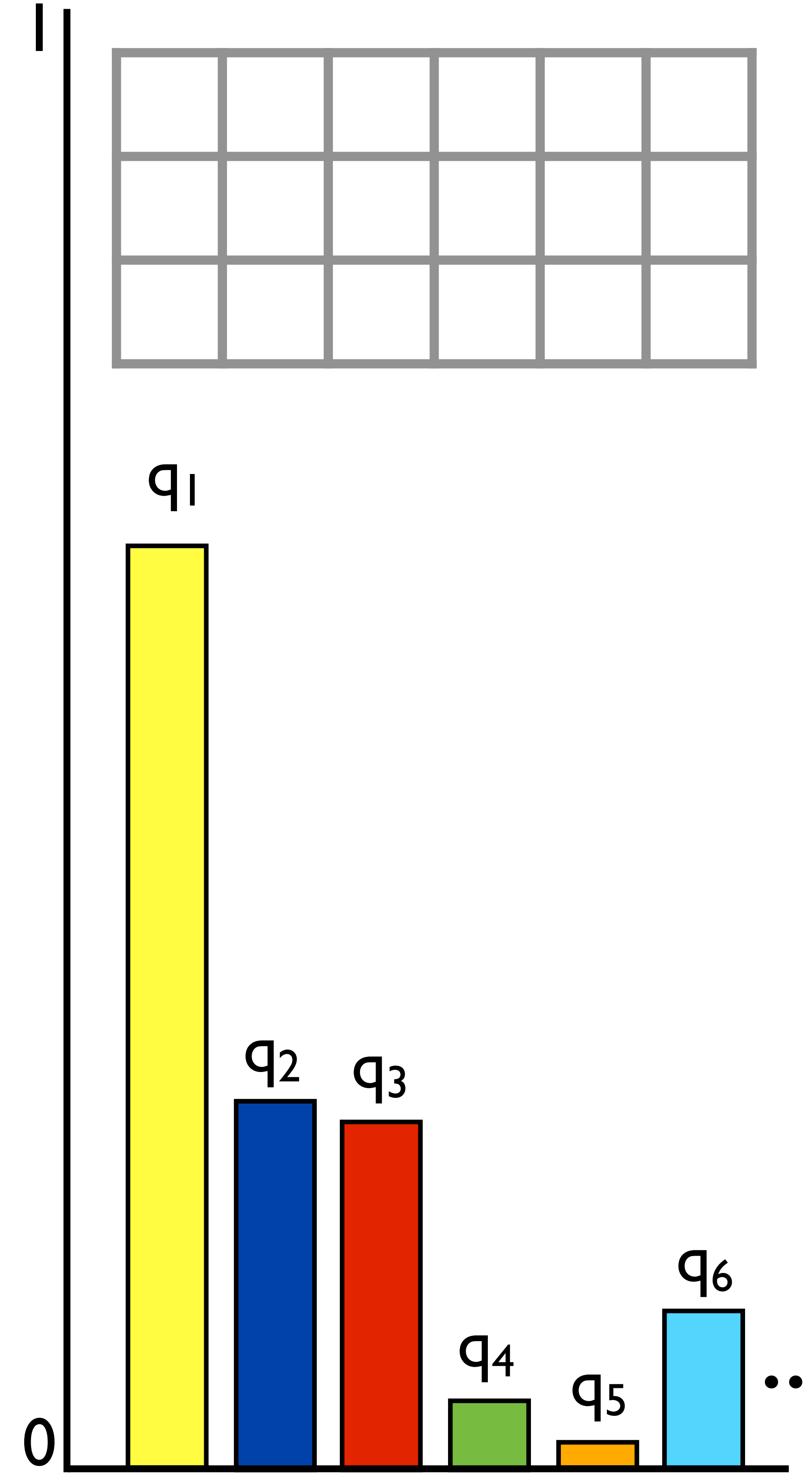
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...

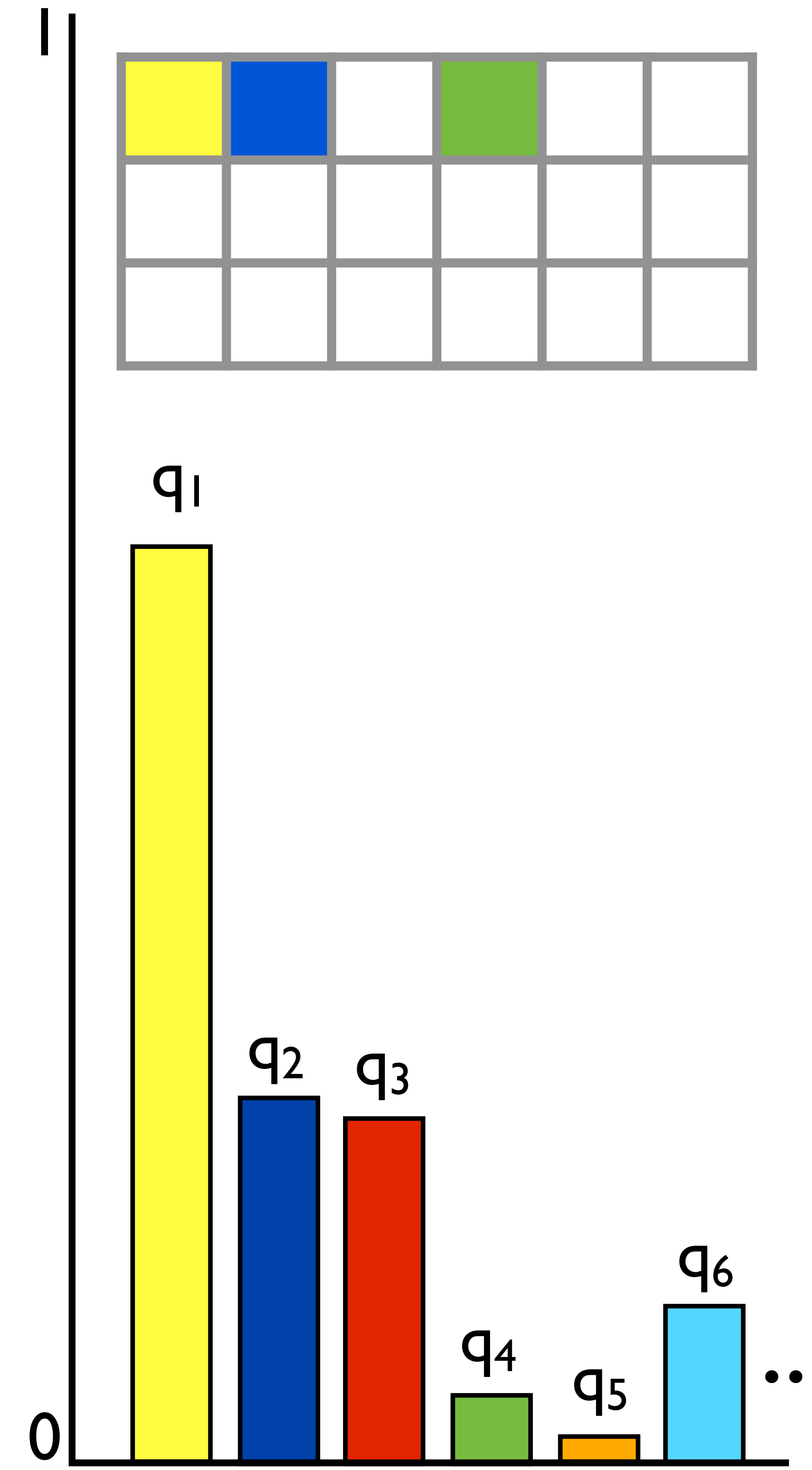
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \displaystyle\sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For *m* = 1, 2, ...

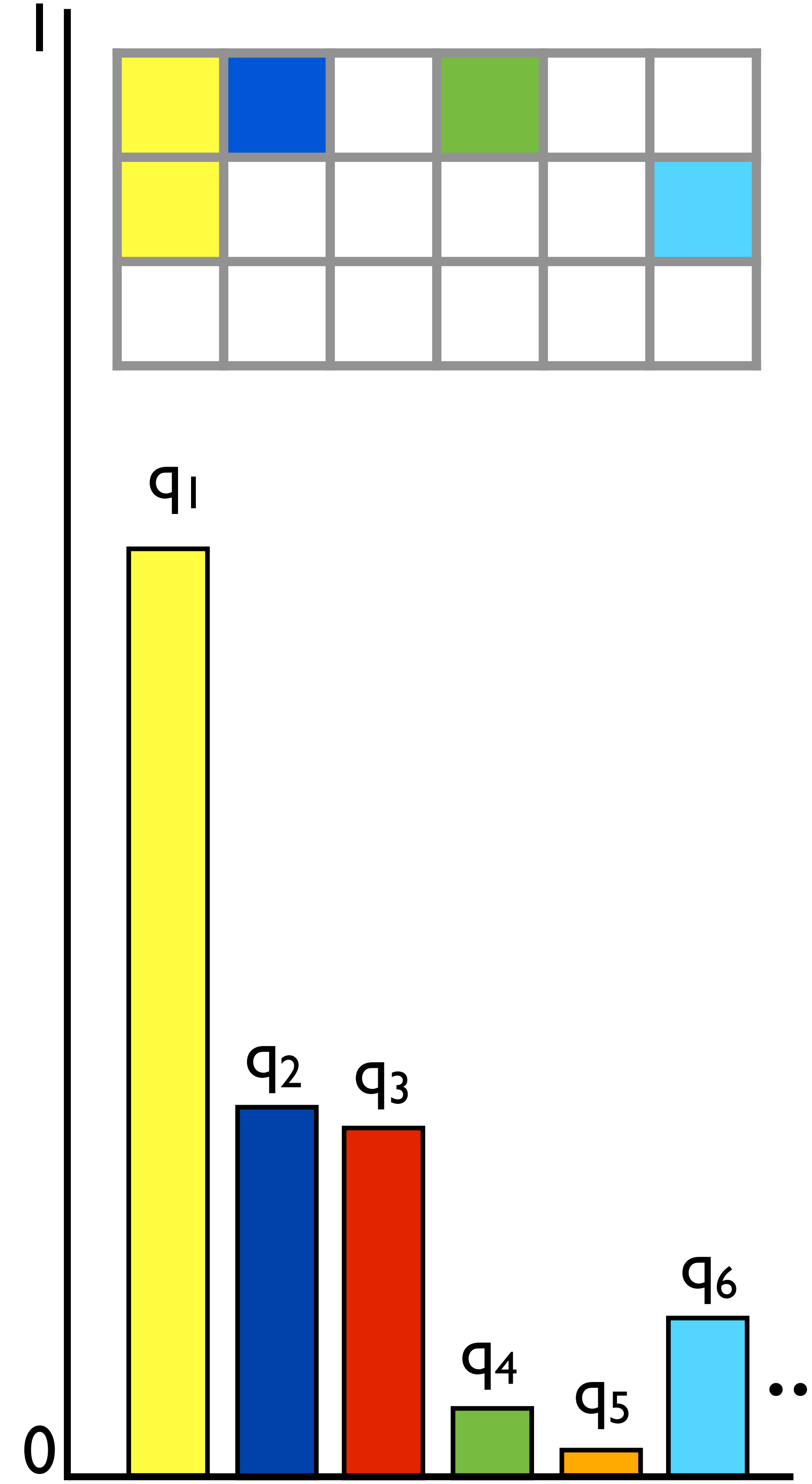
1. Draw $K_m^+ = \text{Poisson}\left(\gamma \dfrac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

$$q_k \sim \text{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...

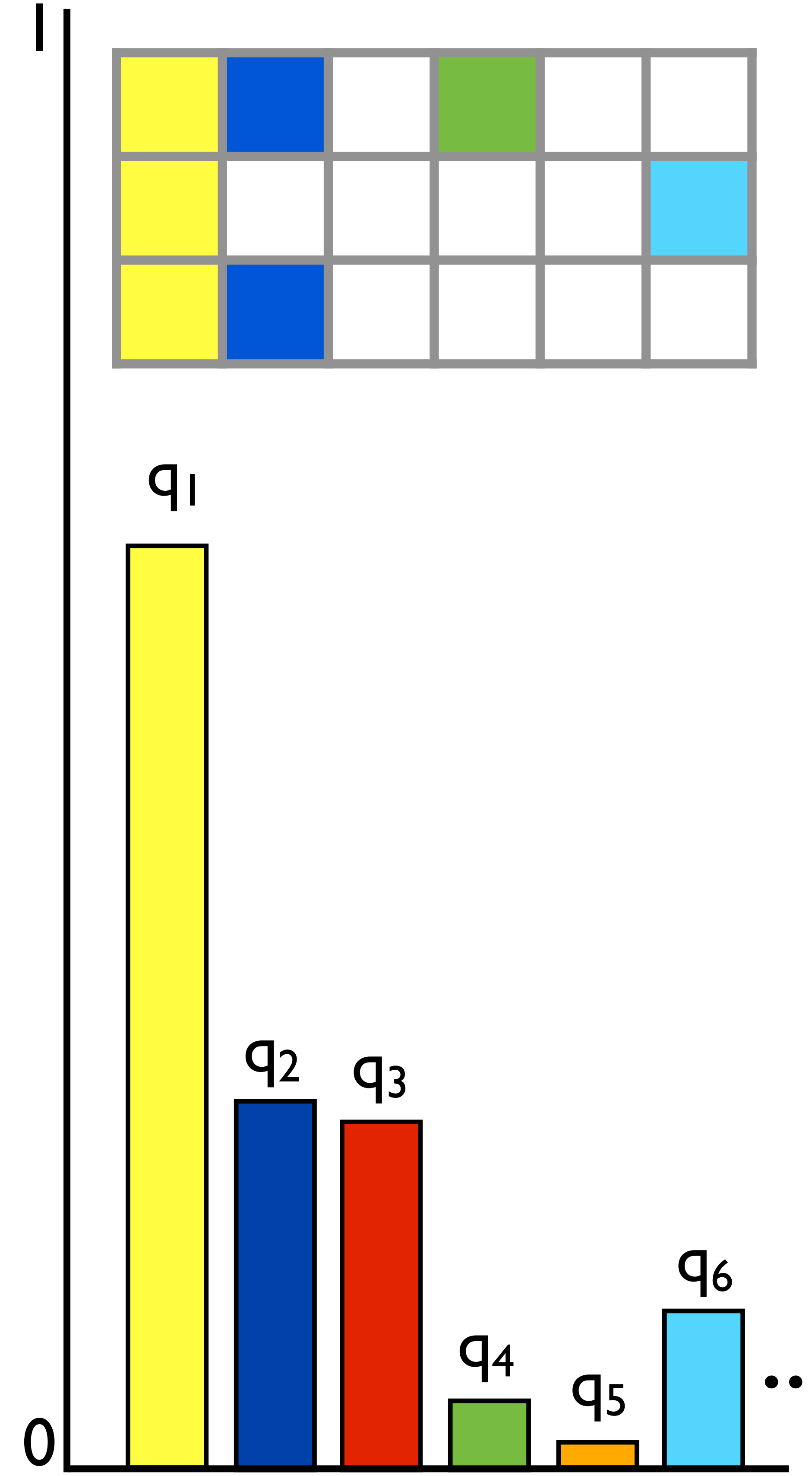
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

   Set $K_m = \sum_{j=1}^{m} K_j^+$

2. For k = $K_{m-1} + 1, \ldots, K_m$

   Draw a frequency of size

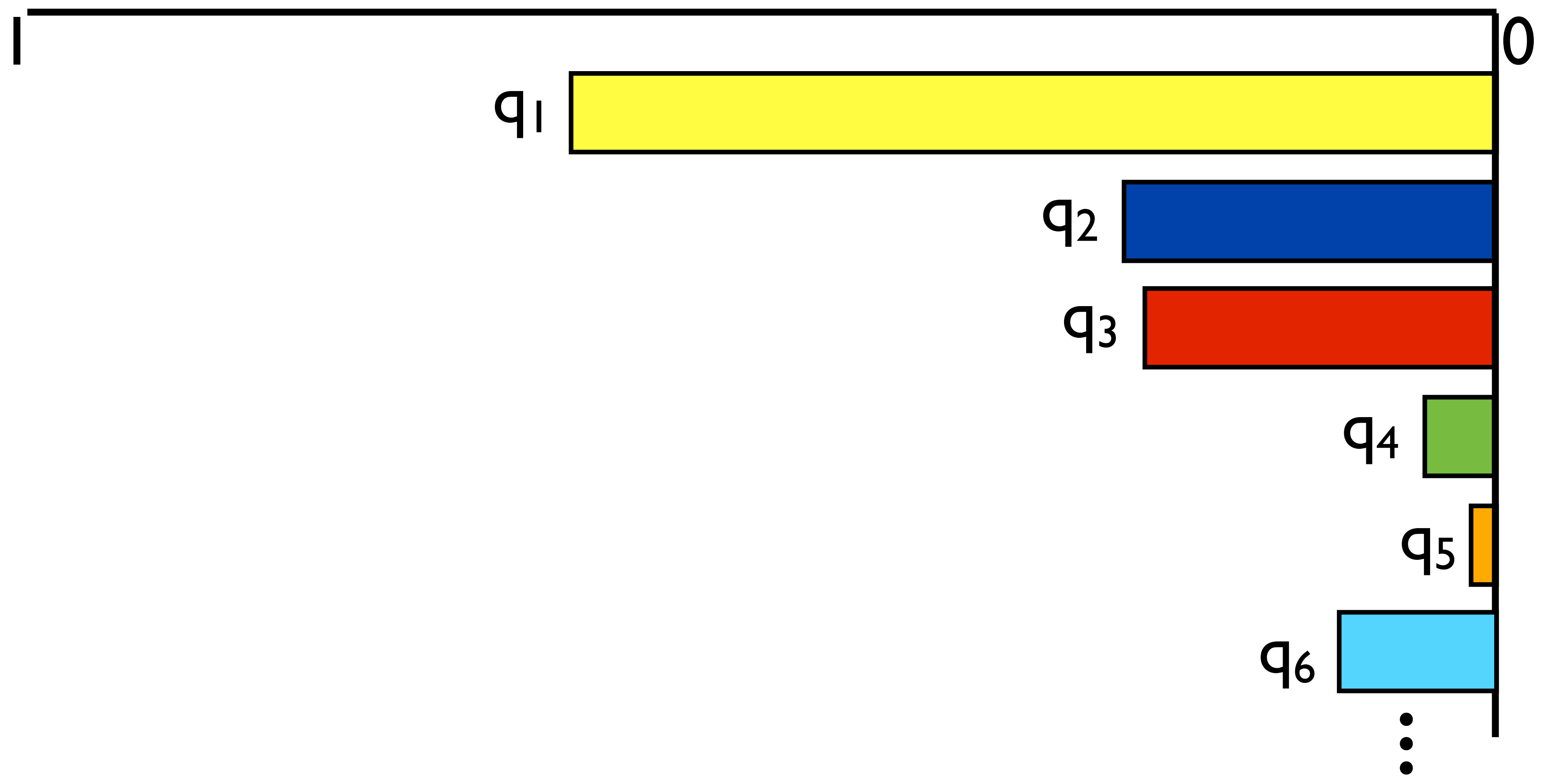
$$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

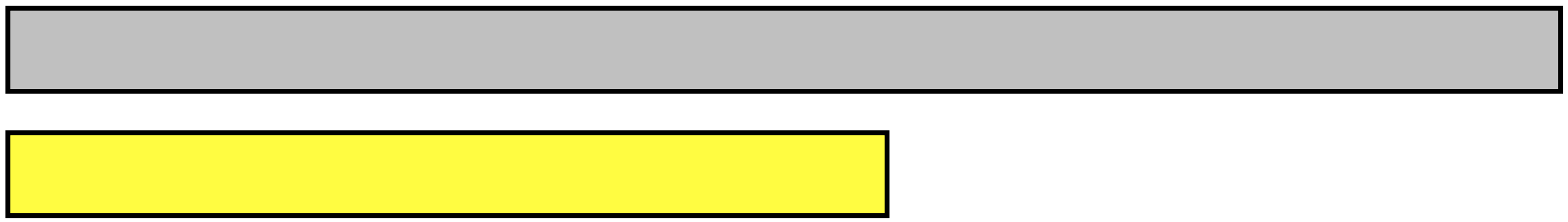


Paintboxes

Indian buffet process: beta feature frequencies

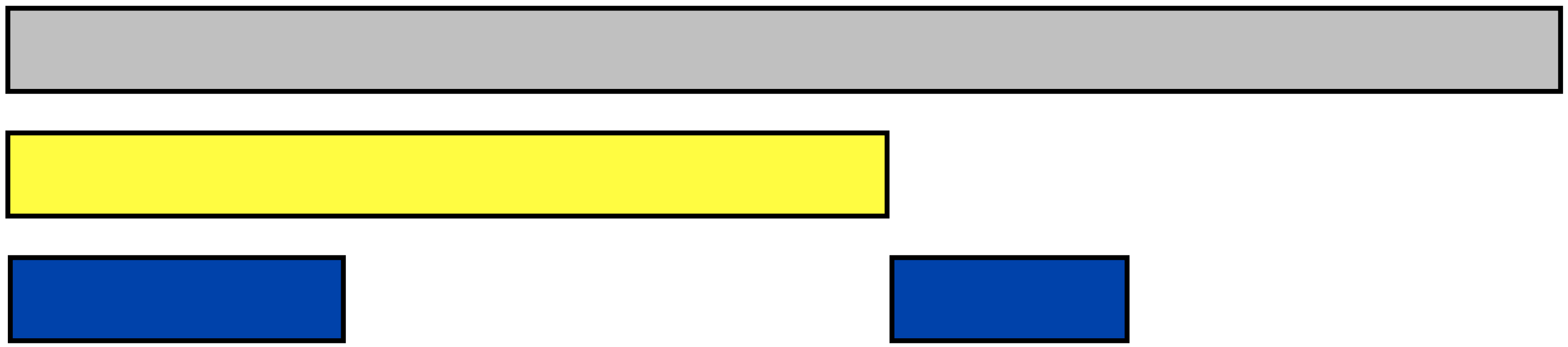
Paintboxes

Indian buffet process: beta feature frequencies



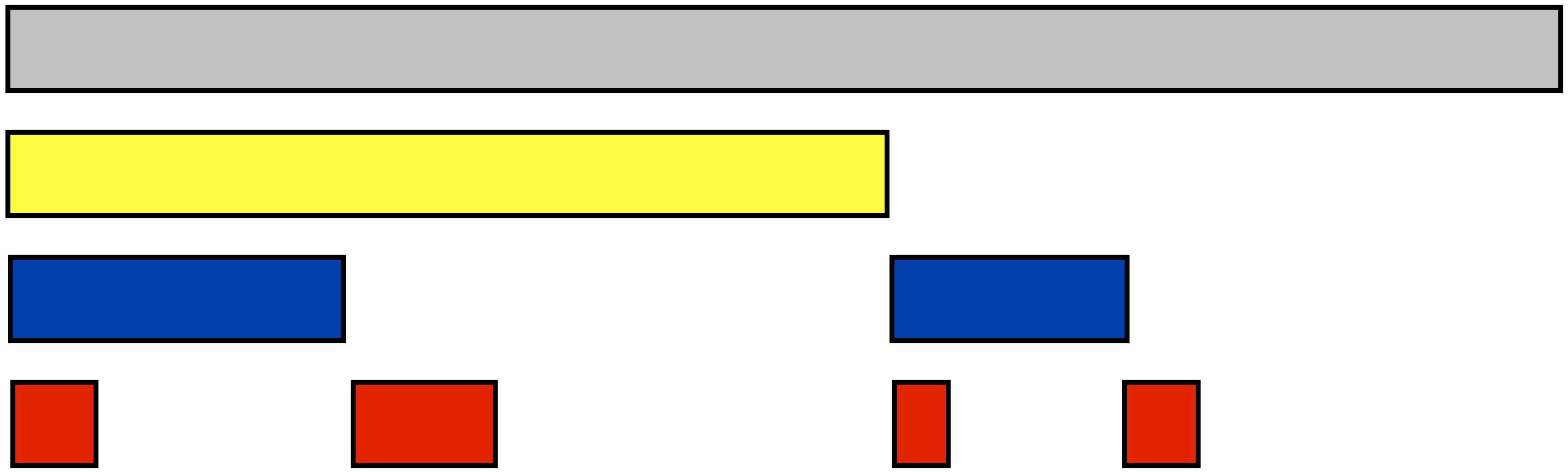
Paintboxes

Indian buffet process: beta feature frequencies



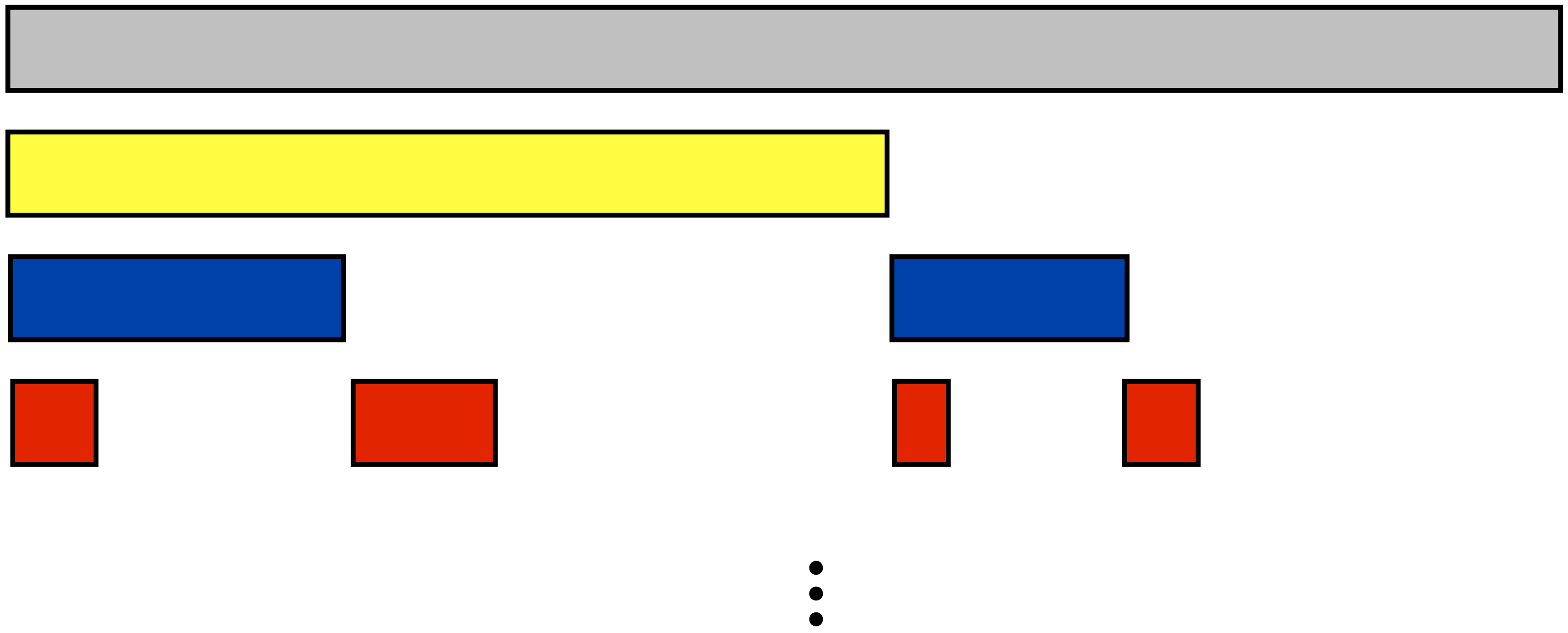
Paintboxes

Indian buffet process: beta feature frequencies

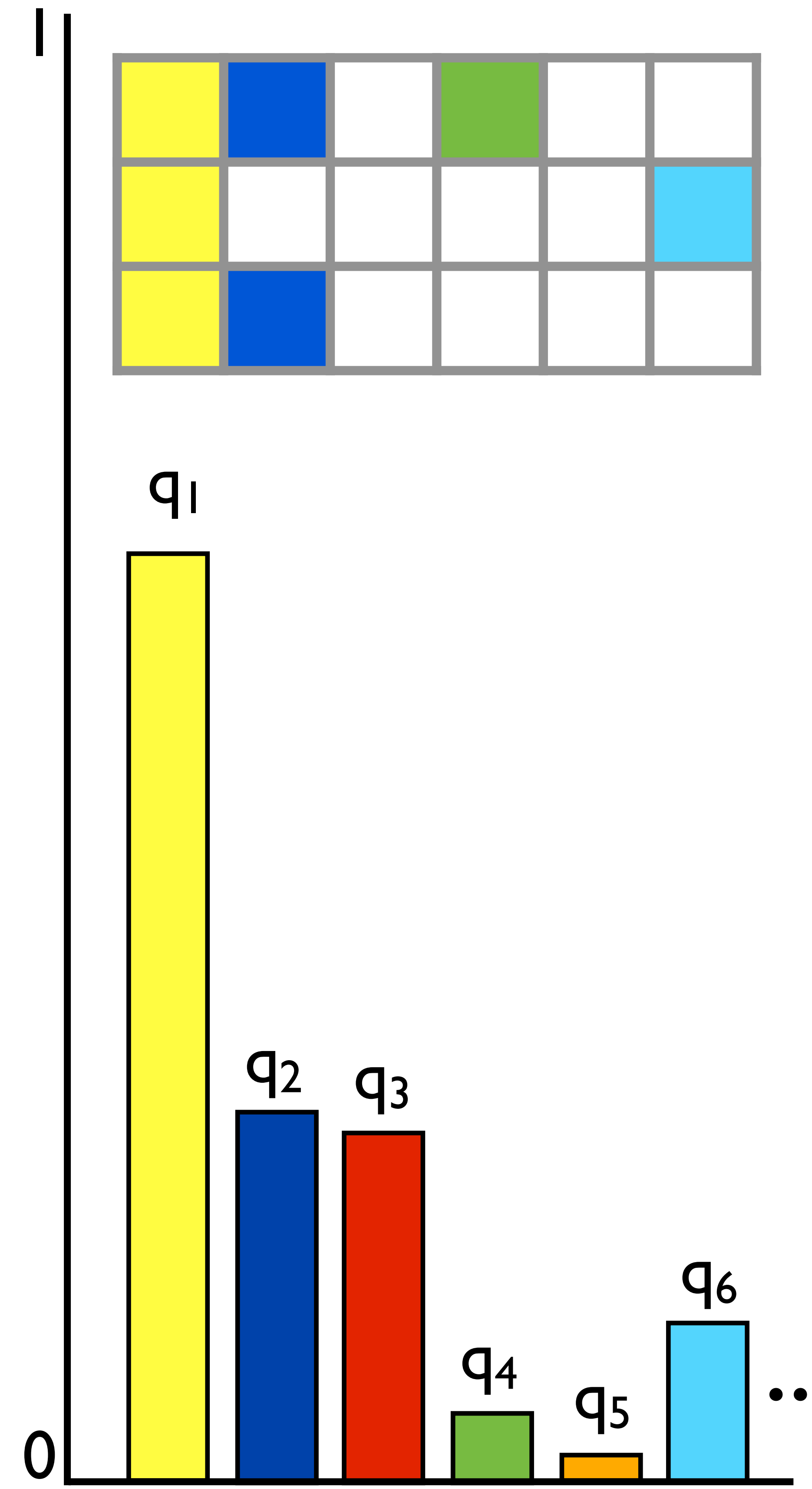


Paintboxes

Indian buffet process: beta feature frequencies

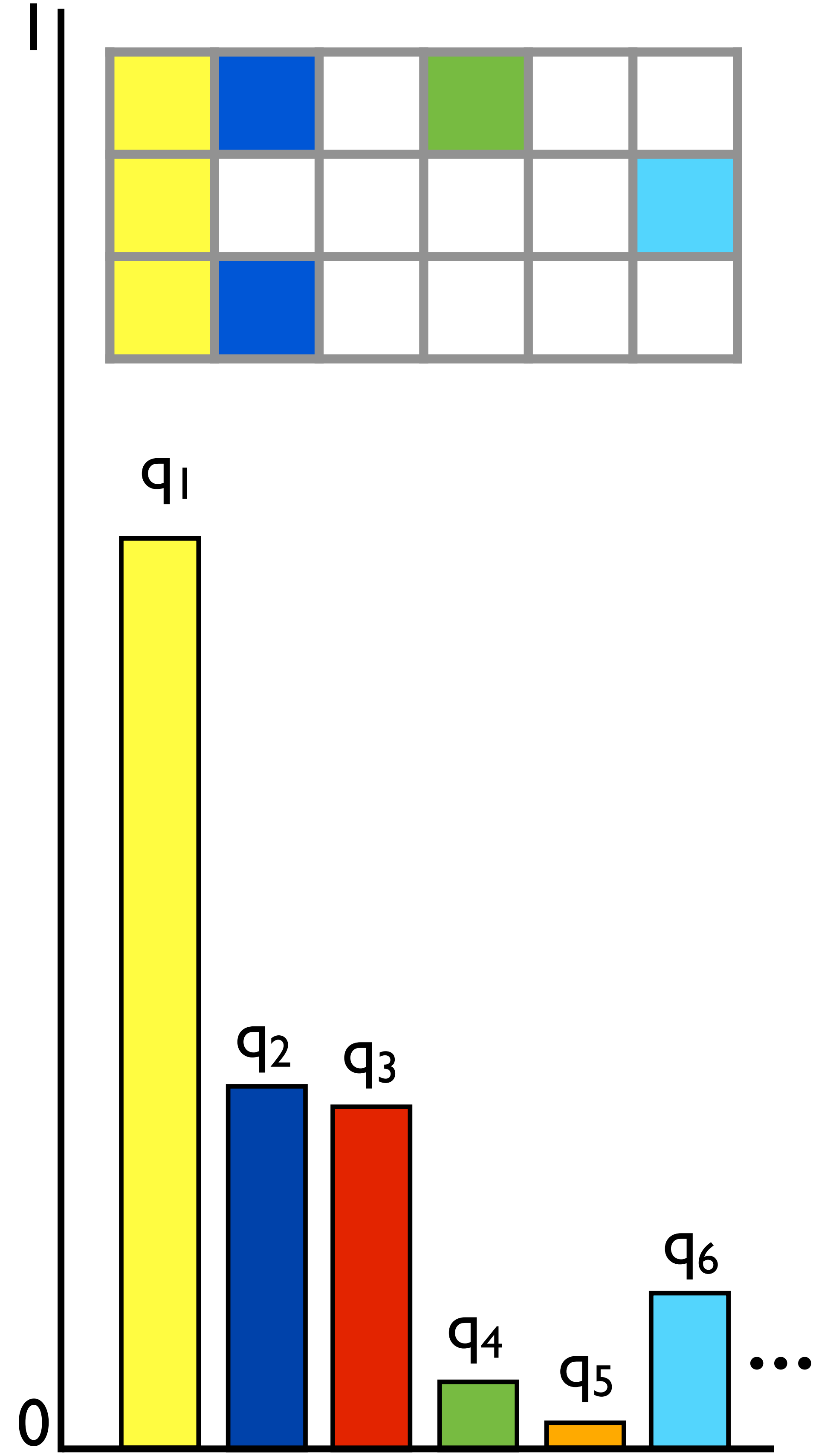


Paintboxes



Paintboxes

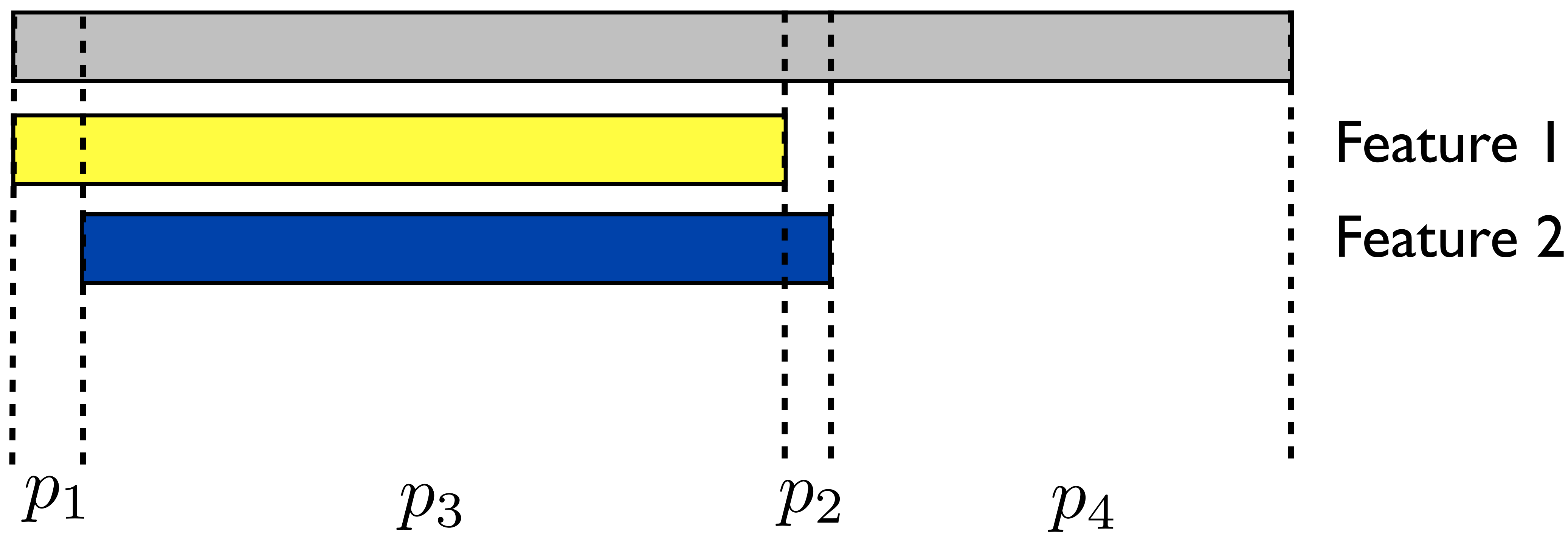
"Feature frequency models"



[Broderick, Pitman, Jordan 2013]

Paintboxes

Two feature example



$$\mathbb{P}(\mathrm{row} = \blacksquare\square) = p_1$$

$$\mathbb{P}(\mathrm{row} = \square\blacksquare) = p_2$$

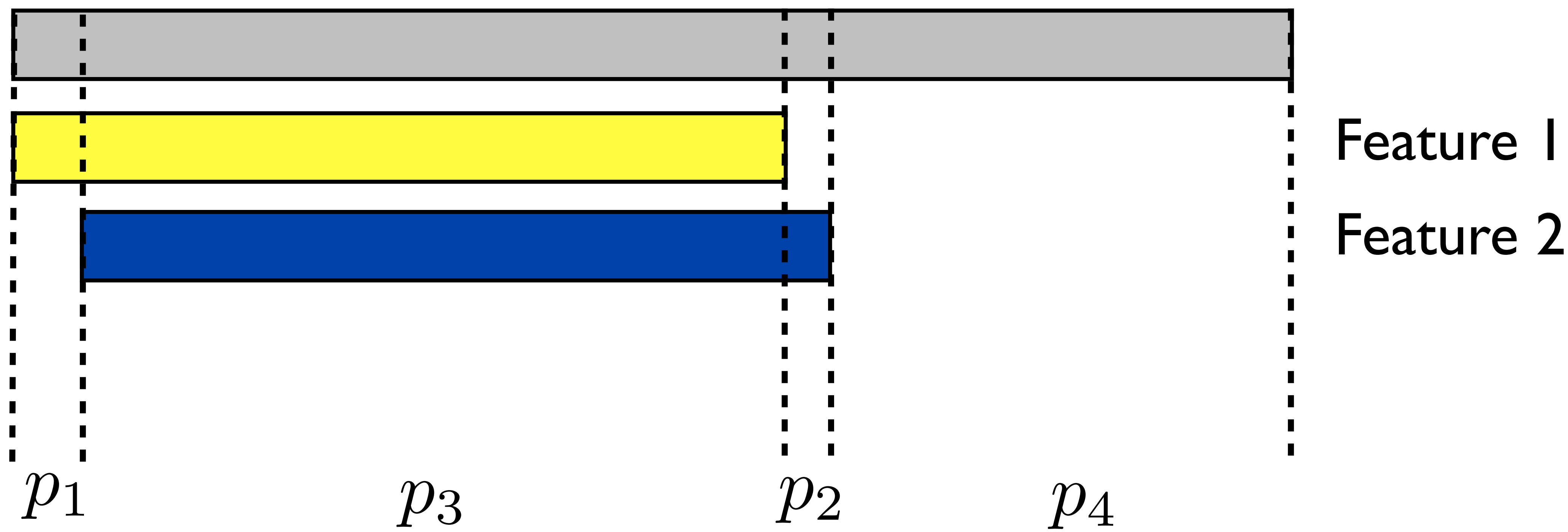
$$\mathbb{P}(\mathrm{row} = \blacksquare\blacksquare) = p_3$$

$$\mathbb{P}(\mathrm{row} = \square\square) = p_4$$

Paintboxes

Two feature example

Not a feature frequency model



$$\mathbb{P}(\text{row} = \blacksquare\square) = p_1$$

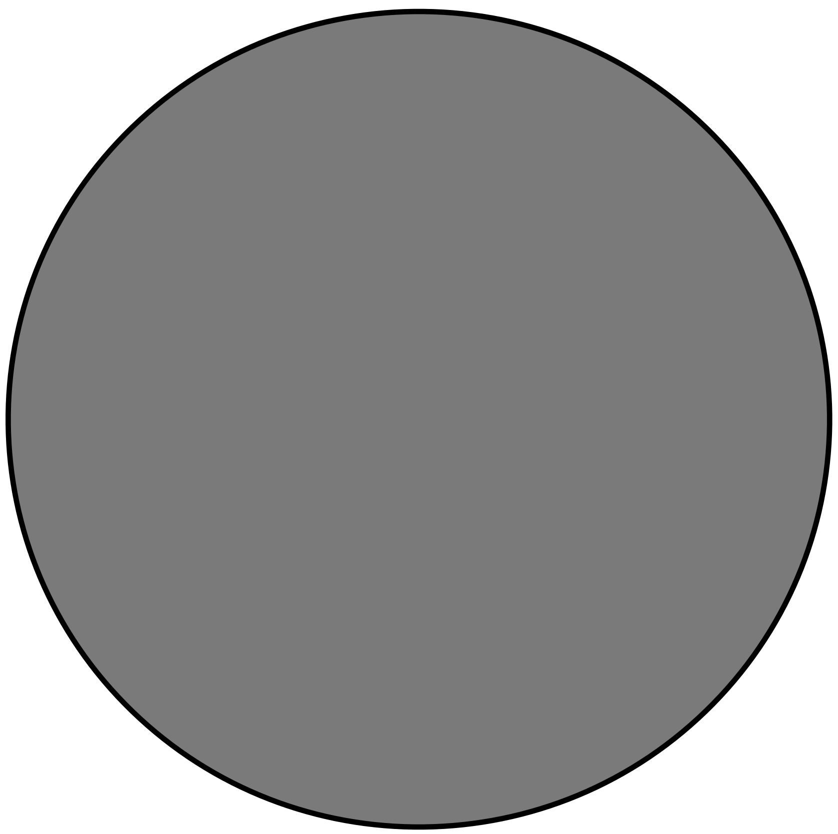
$$\mathbb{P}(\text{row} = \square\blacksquare) = p_2$$

$$\mathbb{P}(\text{row} = \blacksquare\blacksquare) = p_3$$

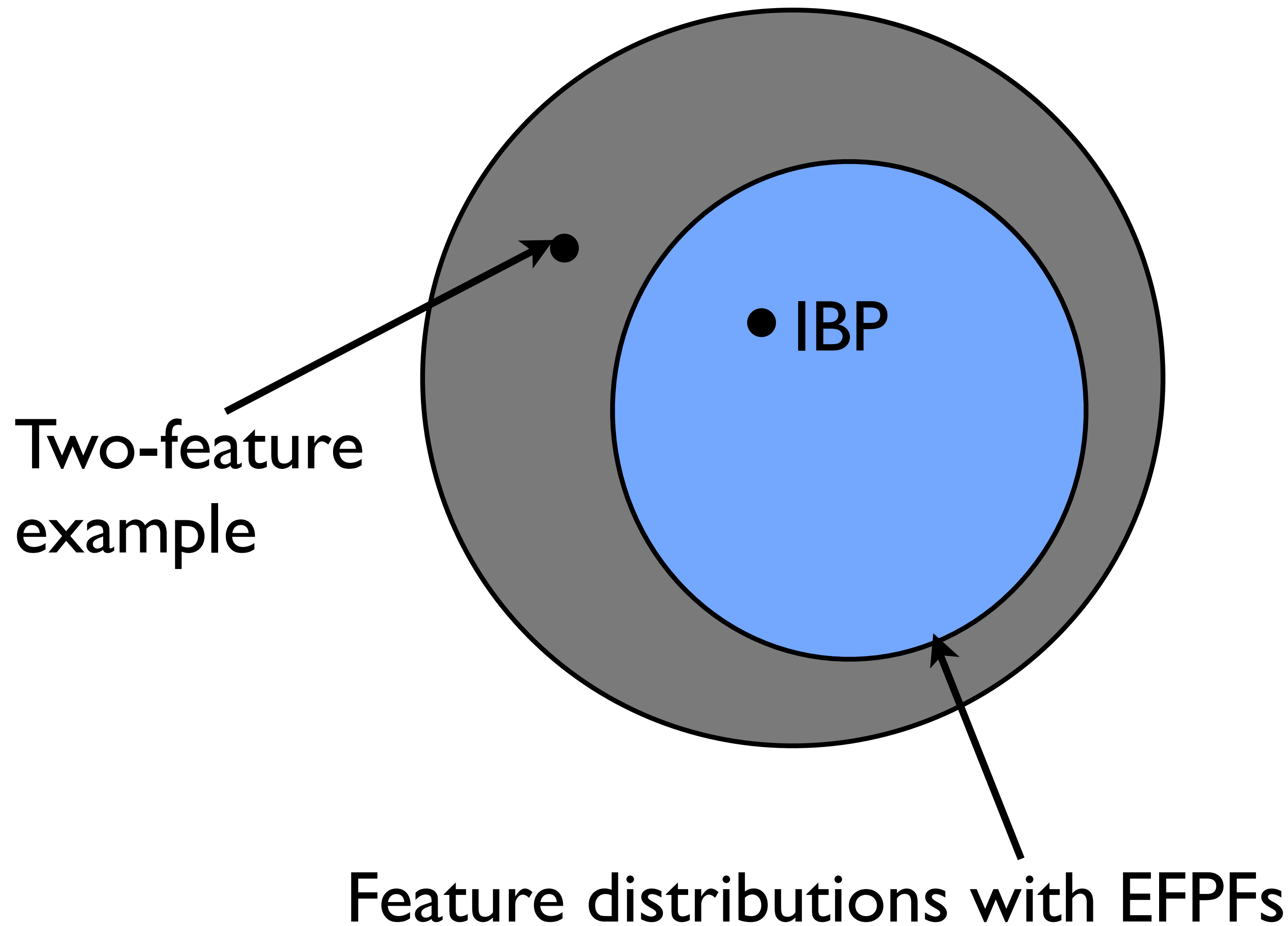
$$\mathbb{P}(\text{row} = \square\square) = p_4$$

Paintboxes

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions



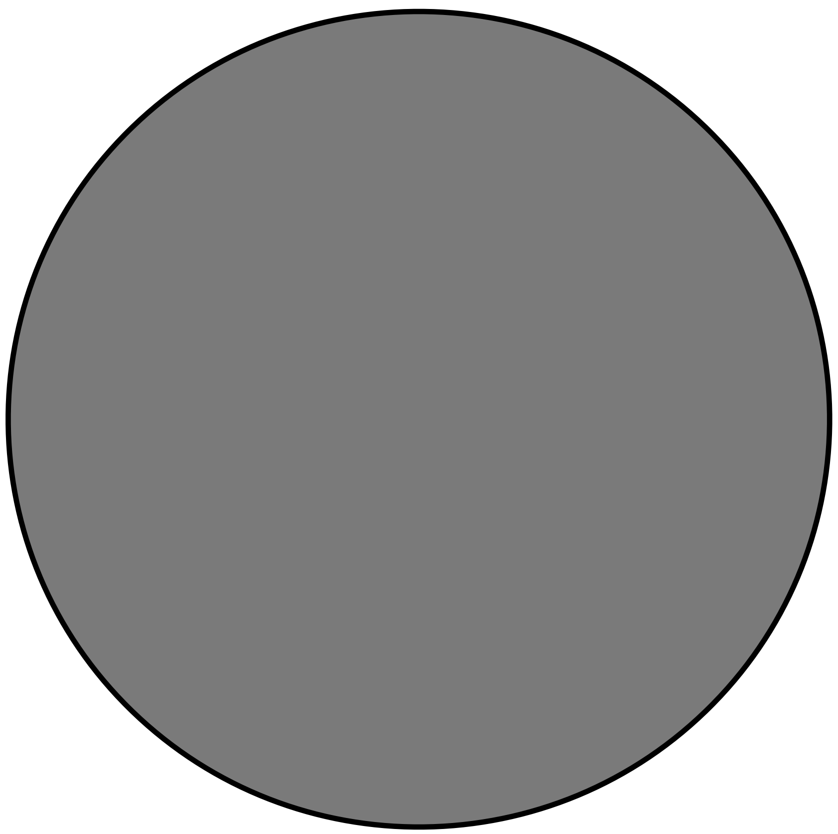
Exchangeable feature distributions
= Feature paintbox allocations



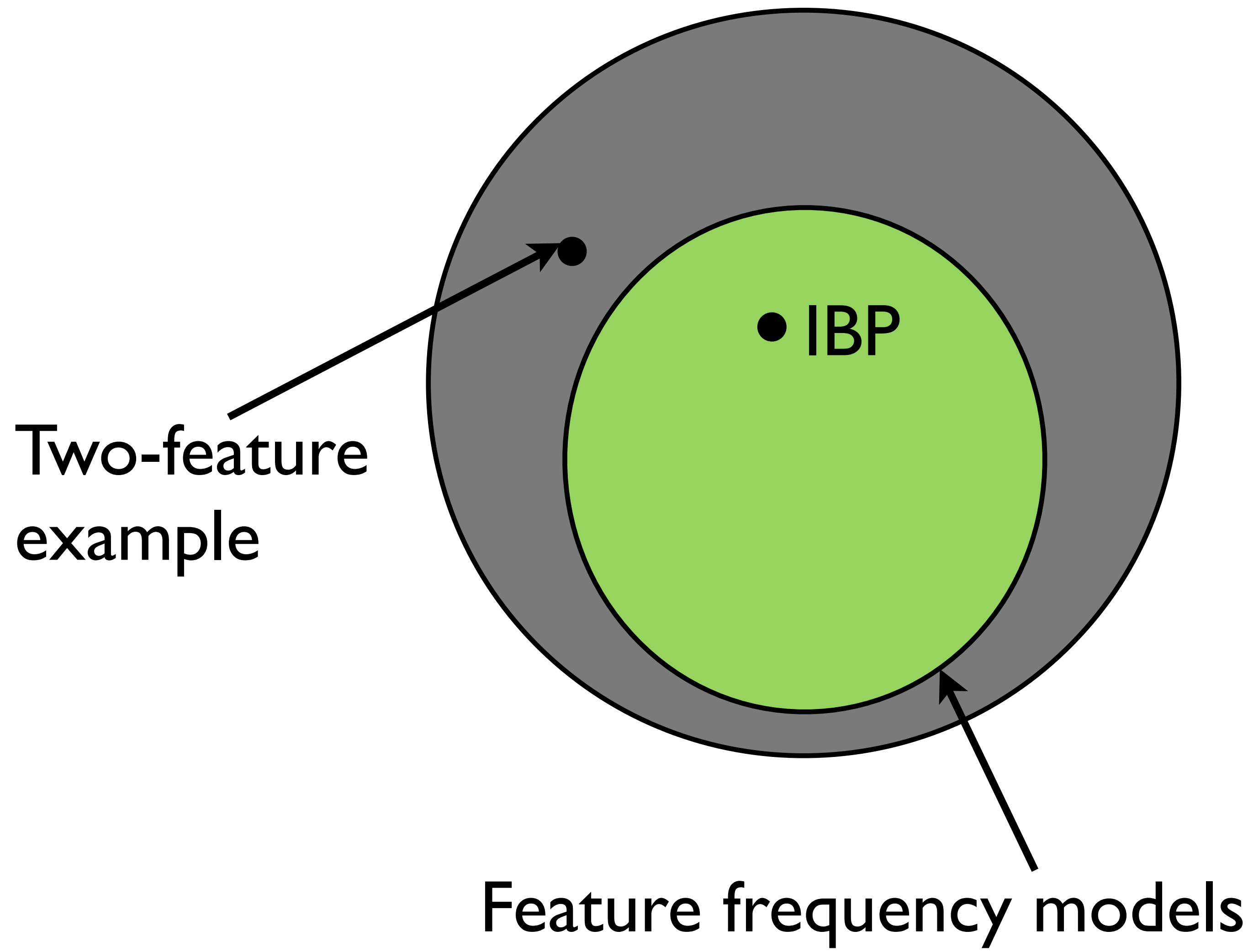
[Broderick, Pitman, Jordan 2013]

Paintboxes

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

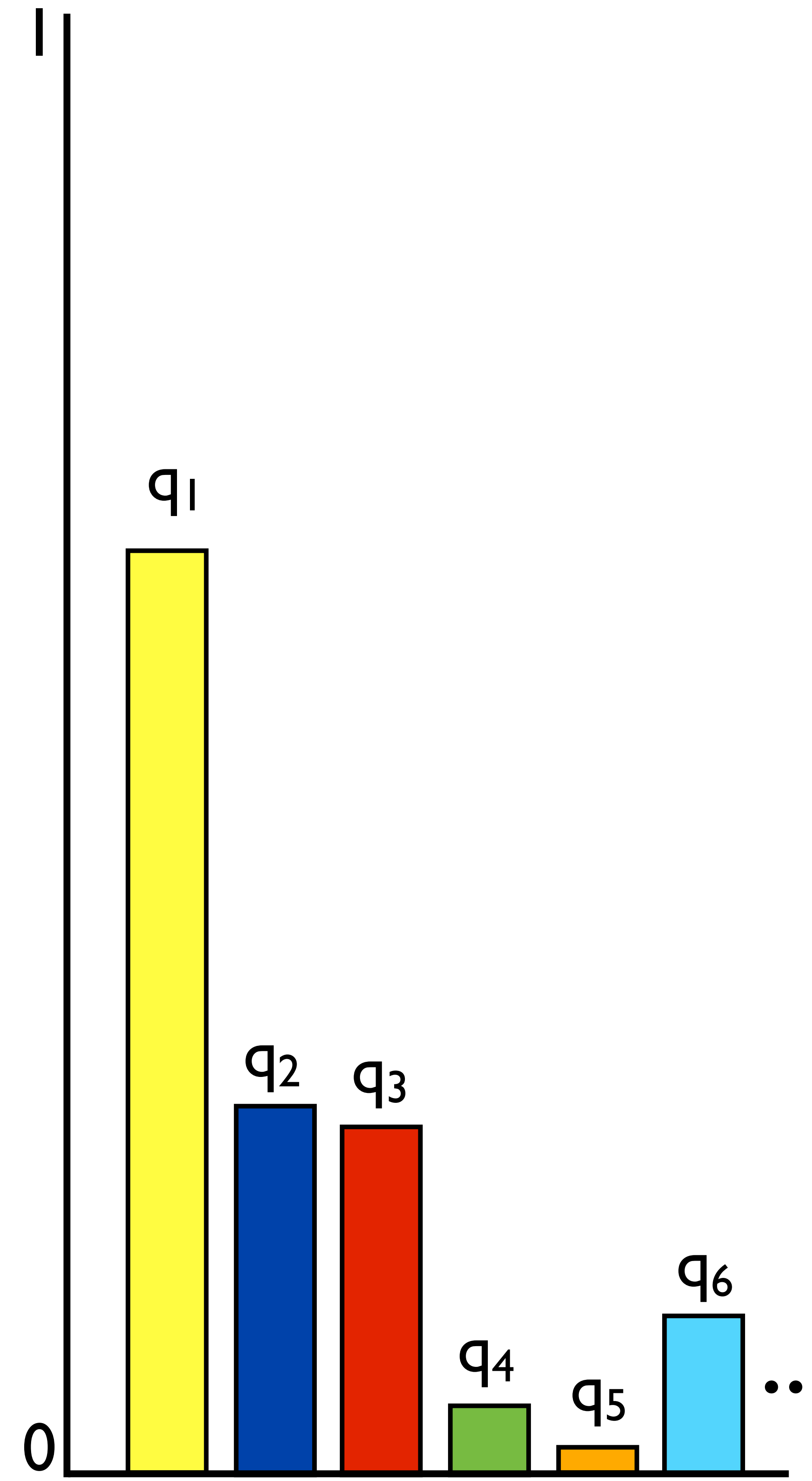


Exchangeable feature distributions
= Feature paintbox allocations

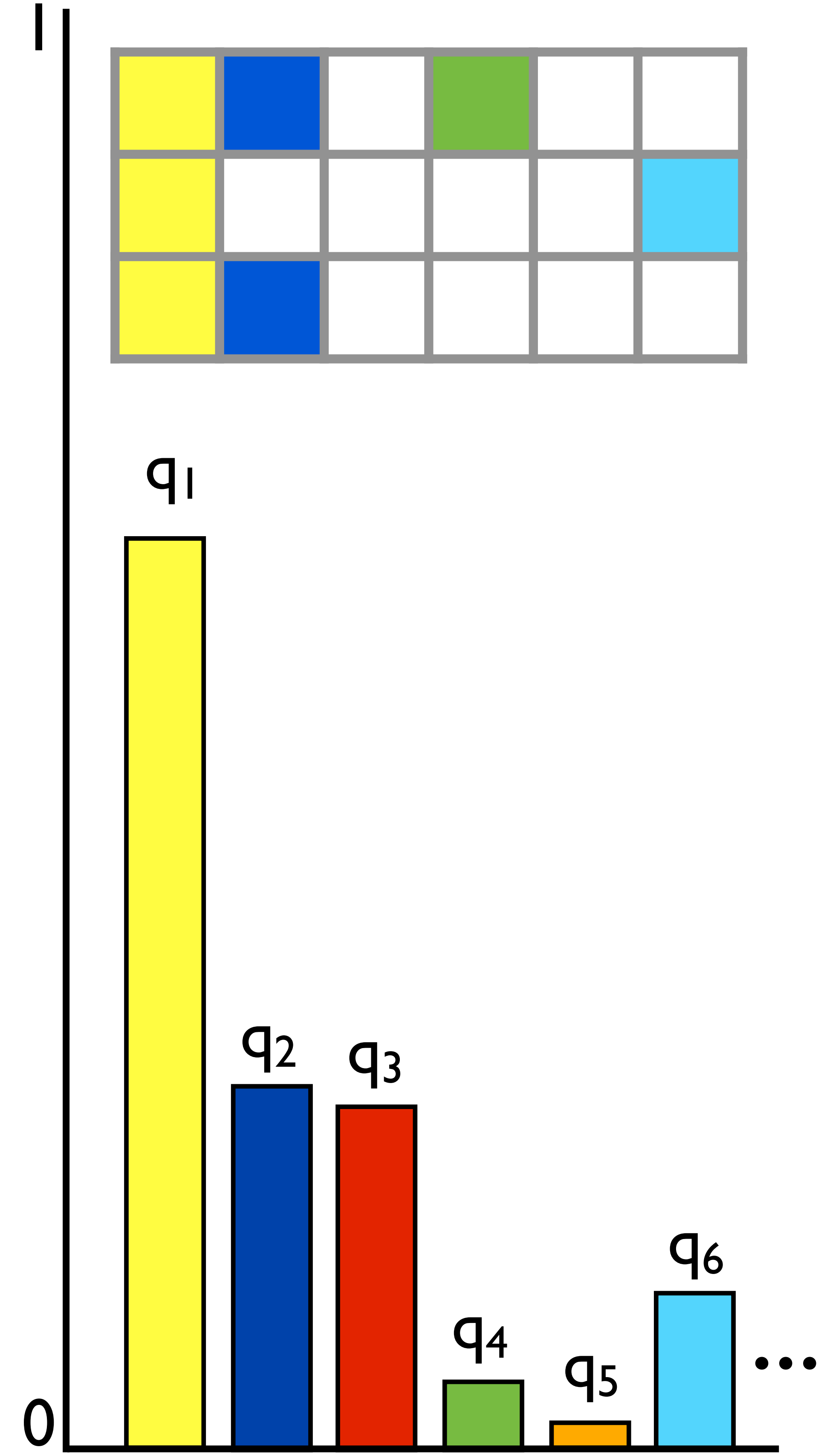


[Broderick, Pitman, Jordan 2013]

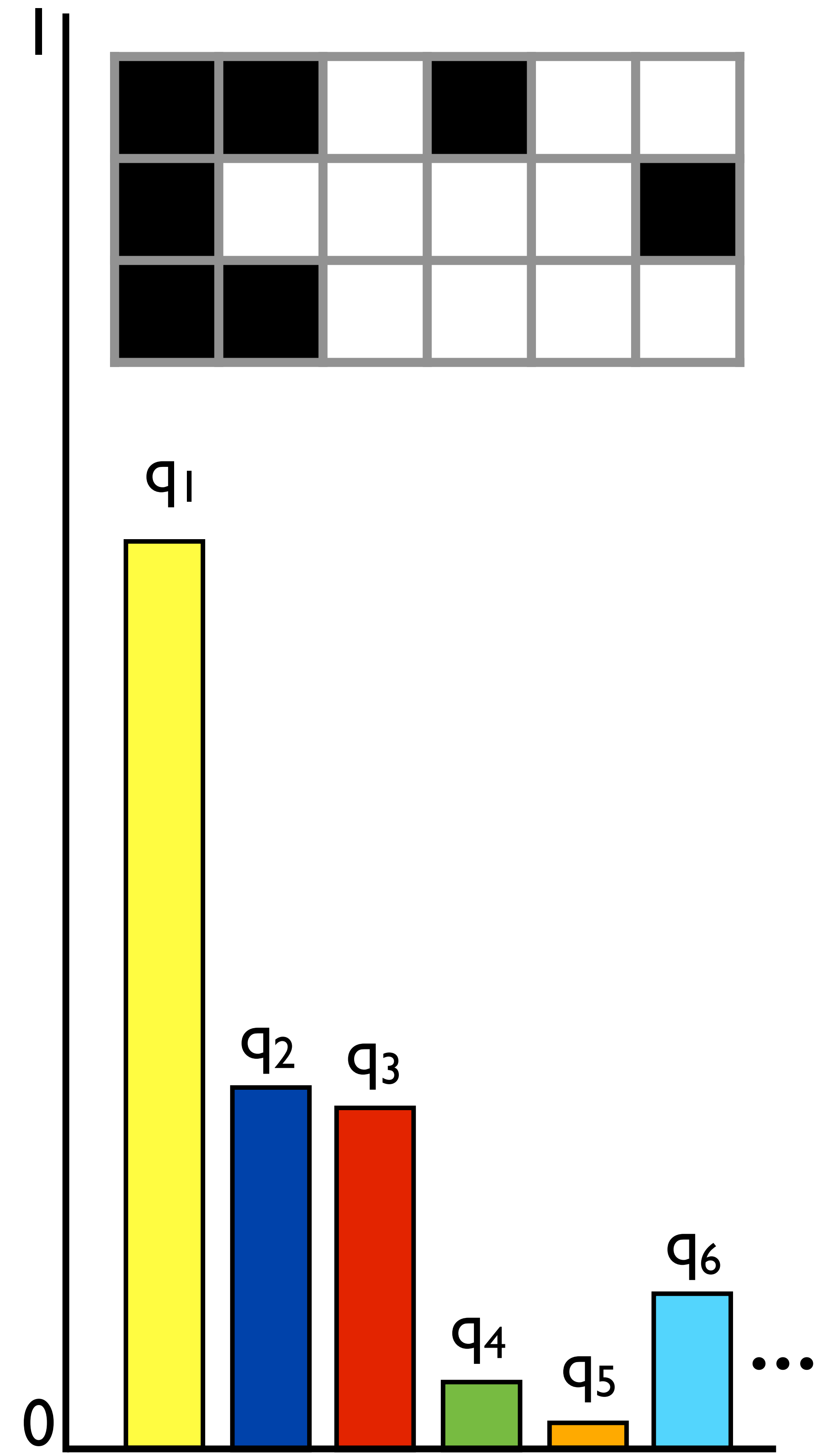
Feature frequency models: EFPFs?



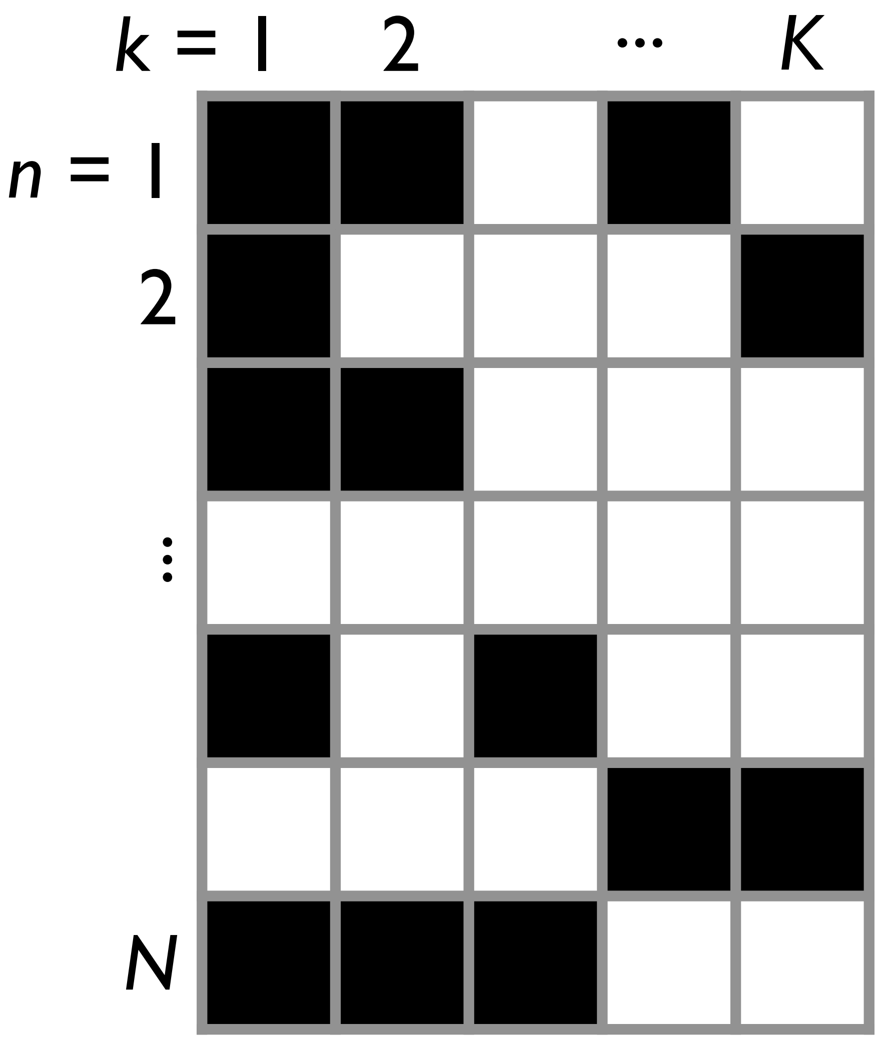
Feature frequency models: EFPFs?

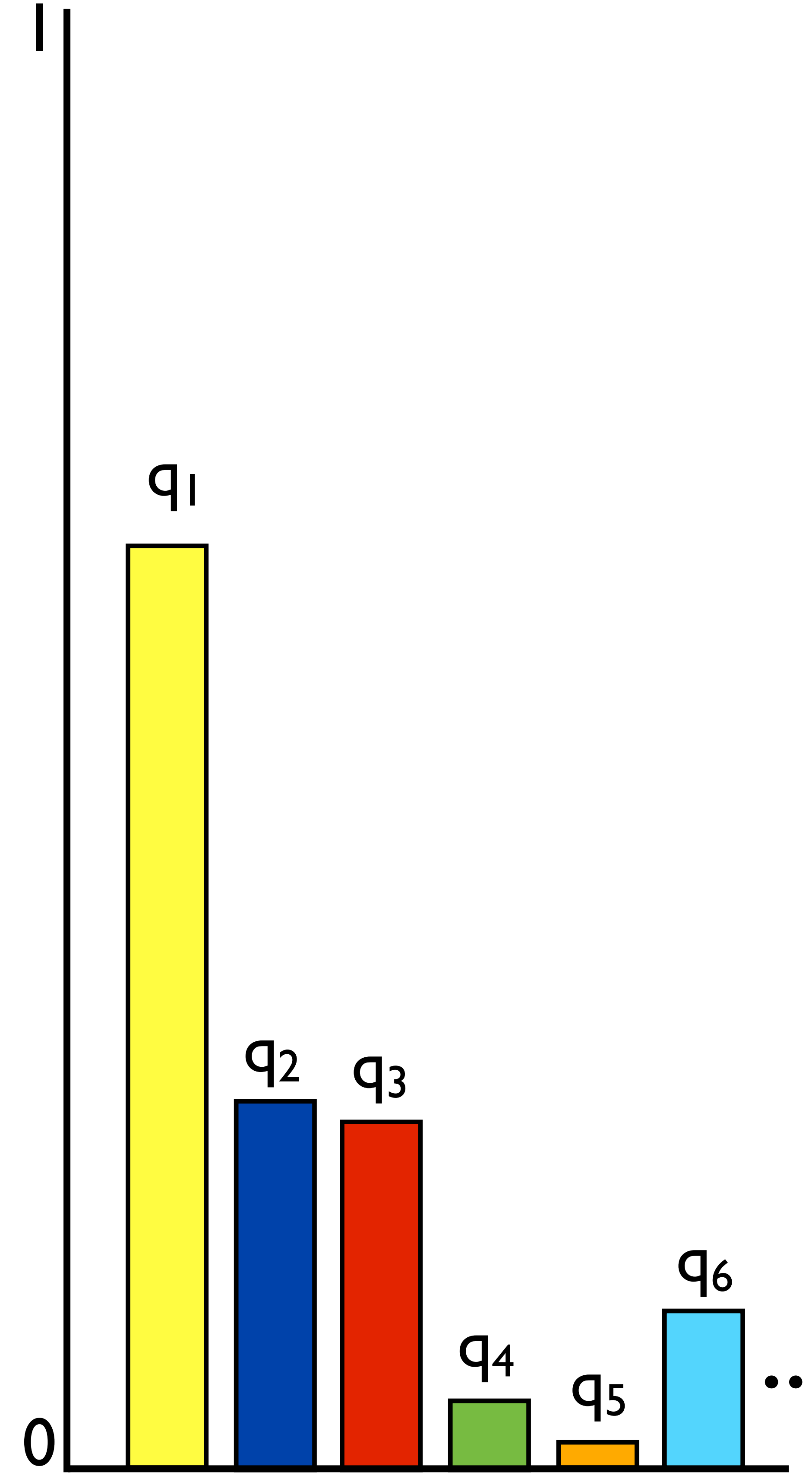


Feature frequency models: EFPFs?

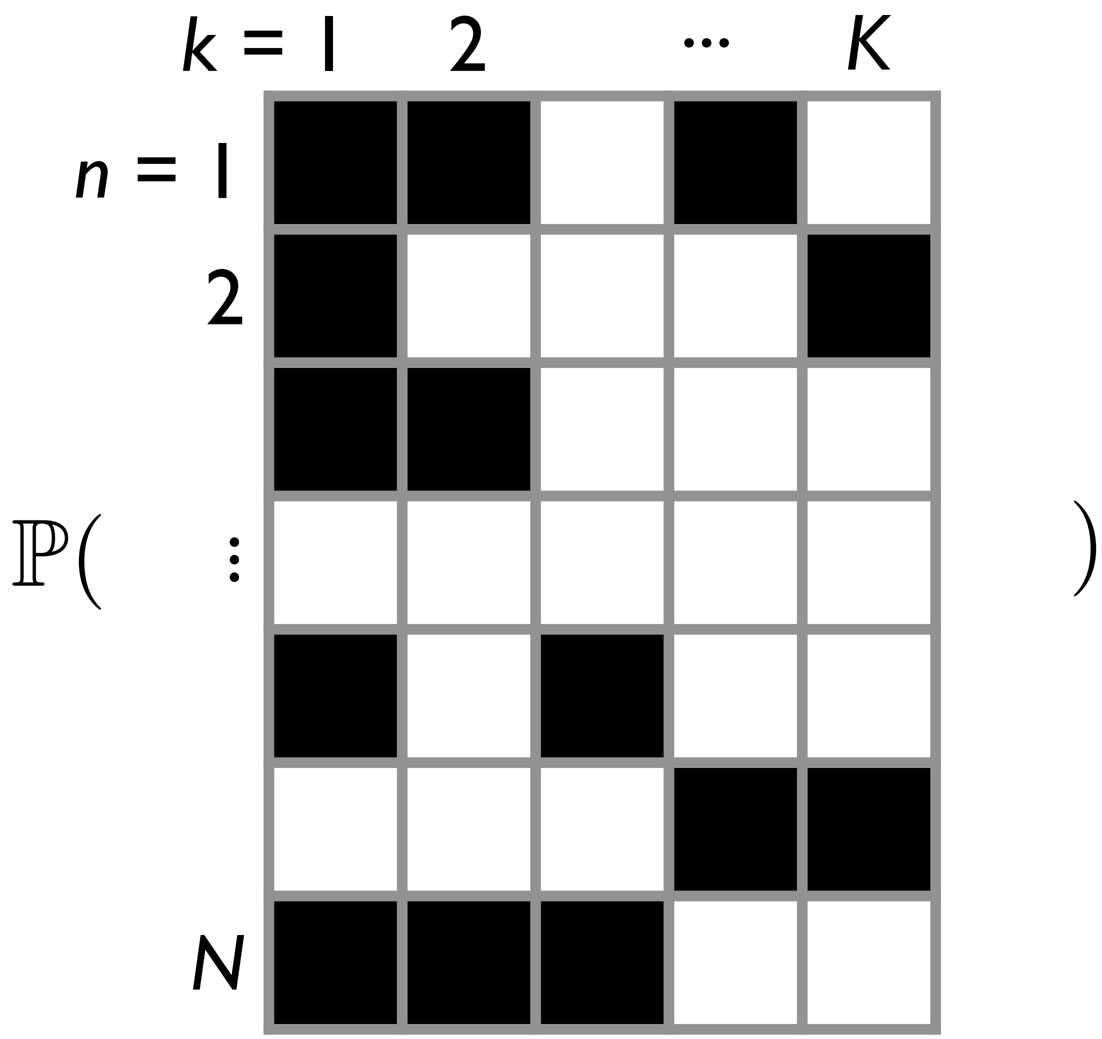


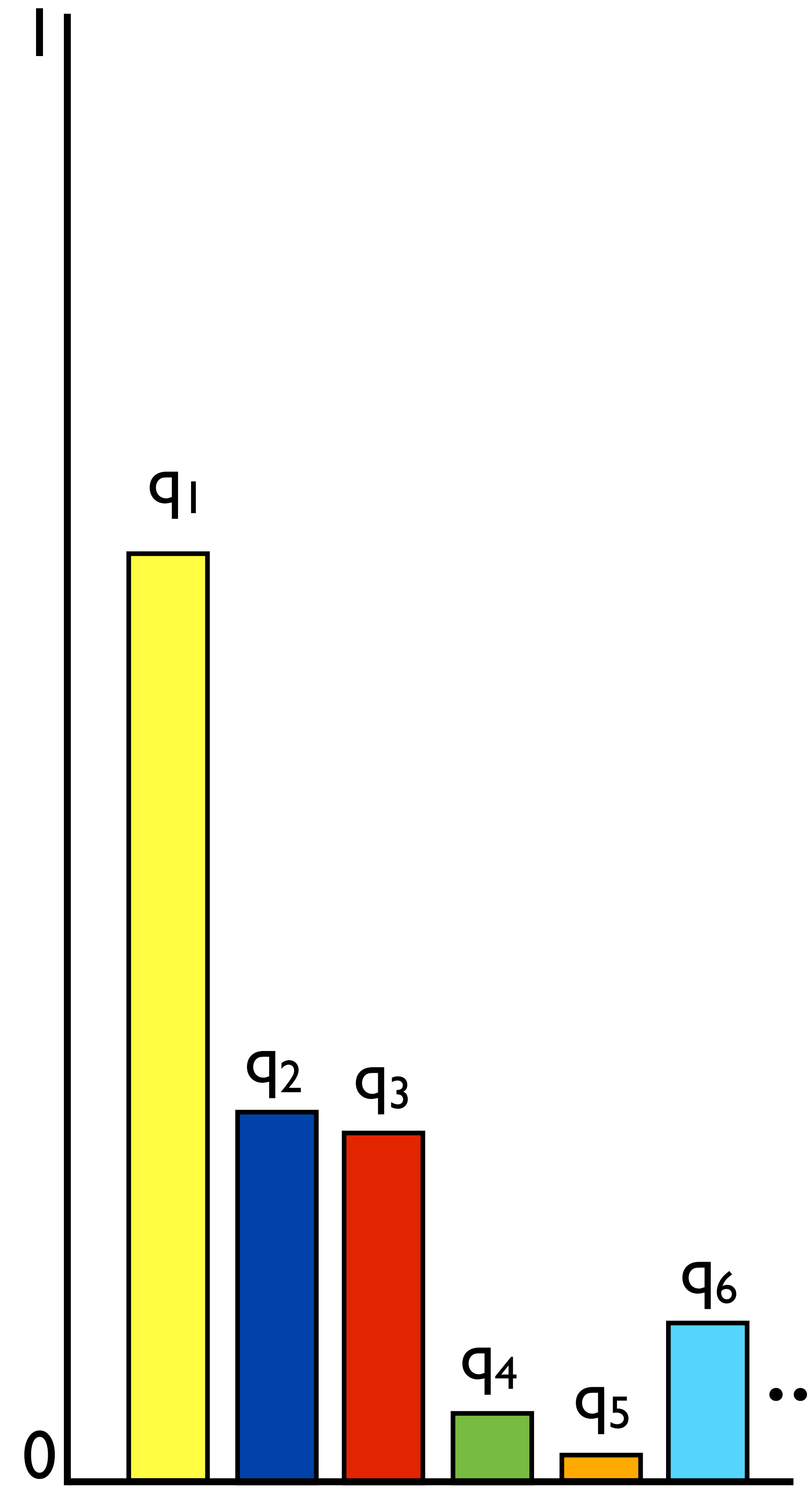
Feature frequency models: EFPFs?



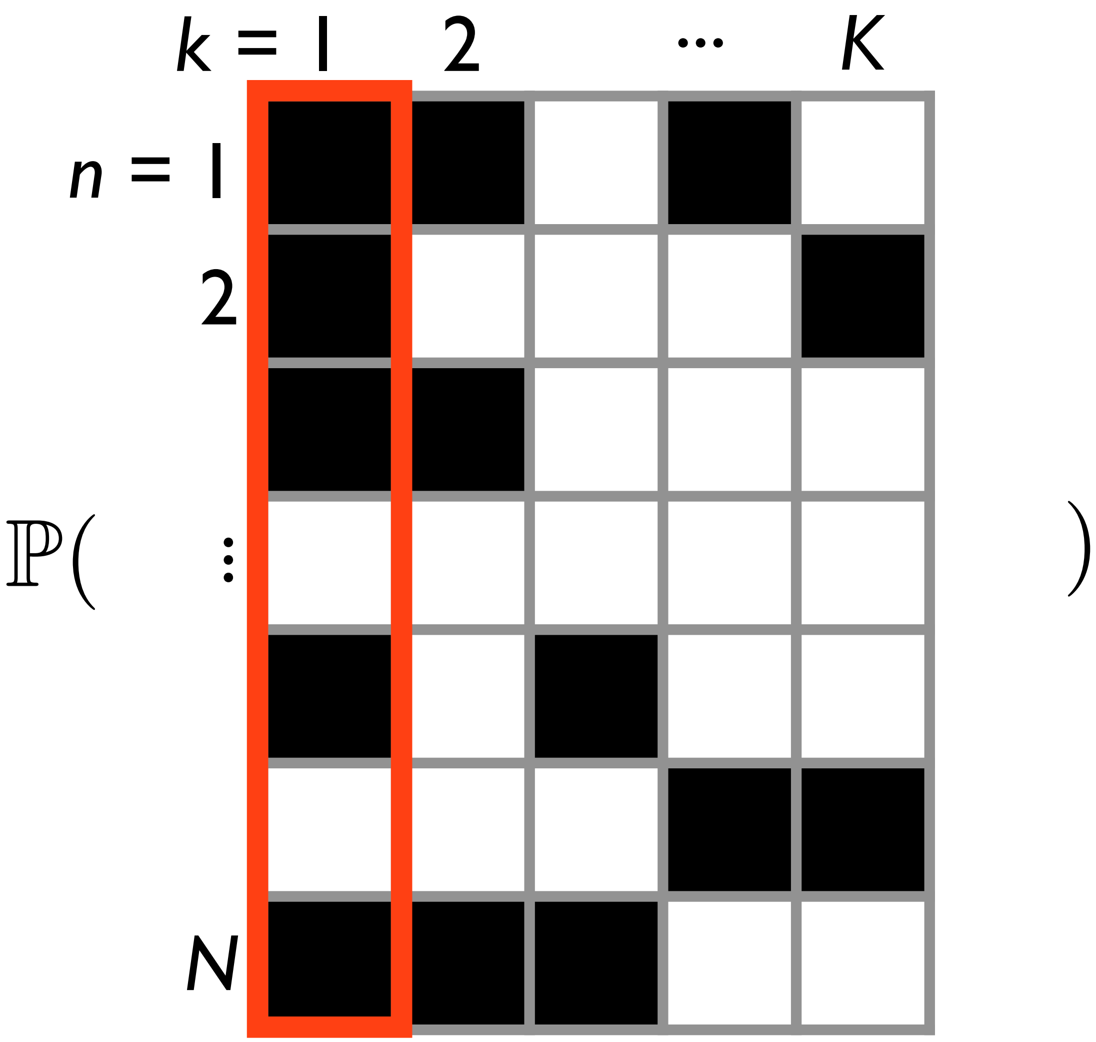


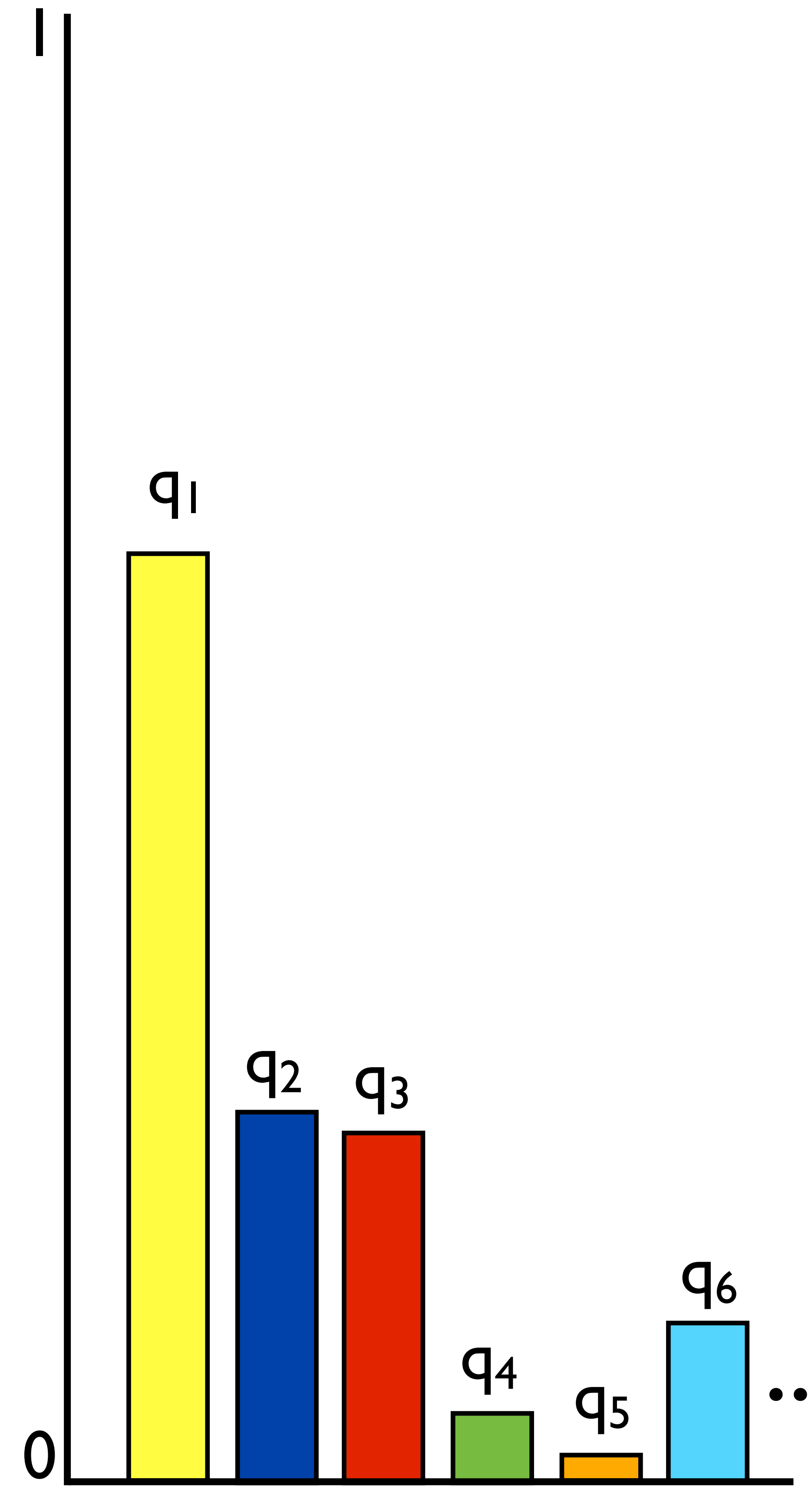
Feature frequency models: EFPFs?



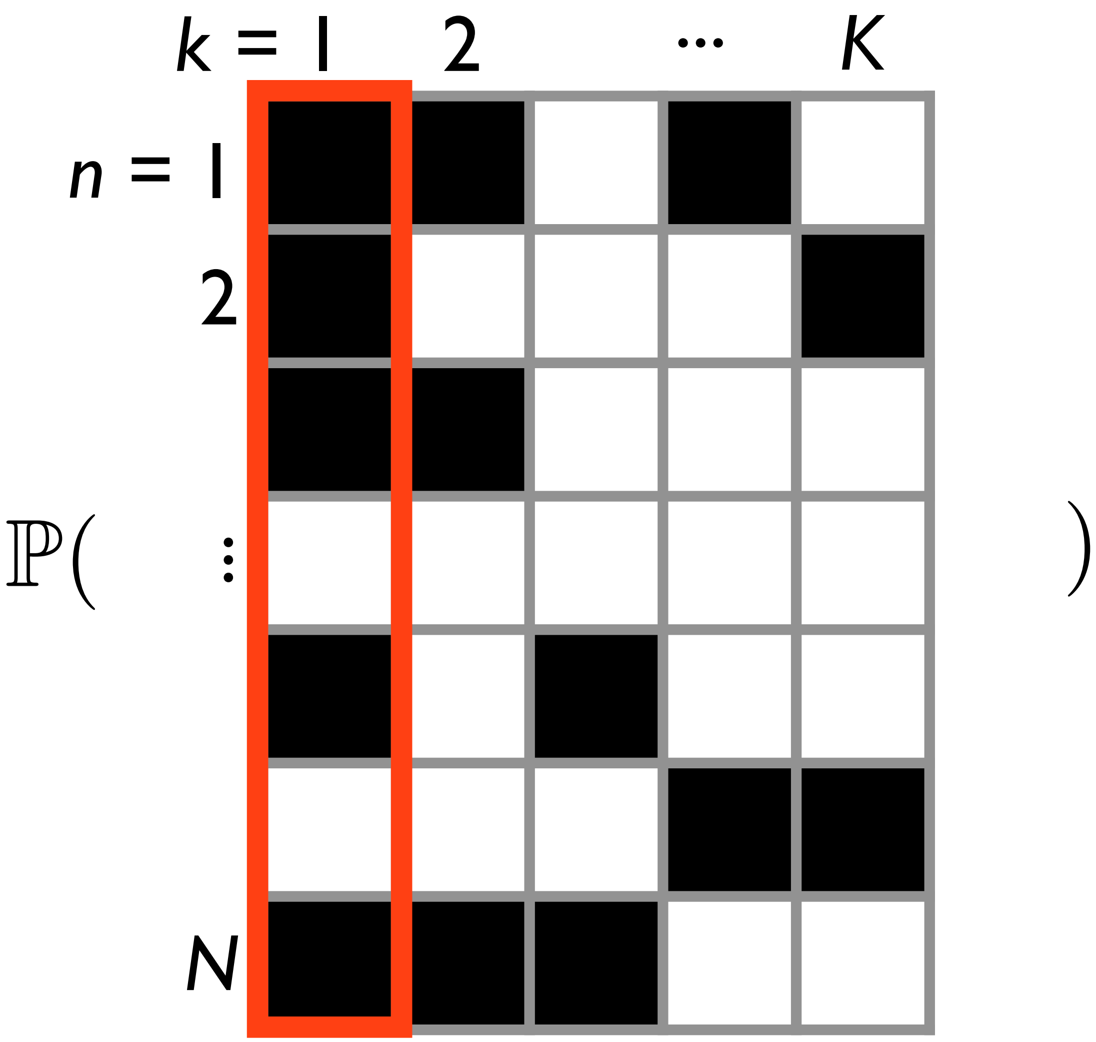


Feature frequency models: EFPFs?

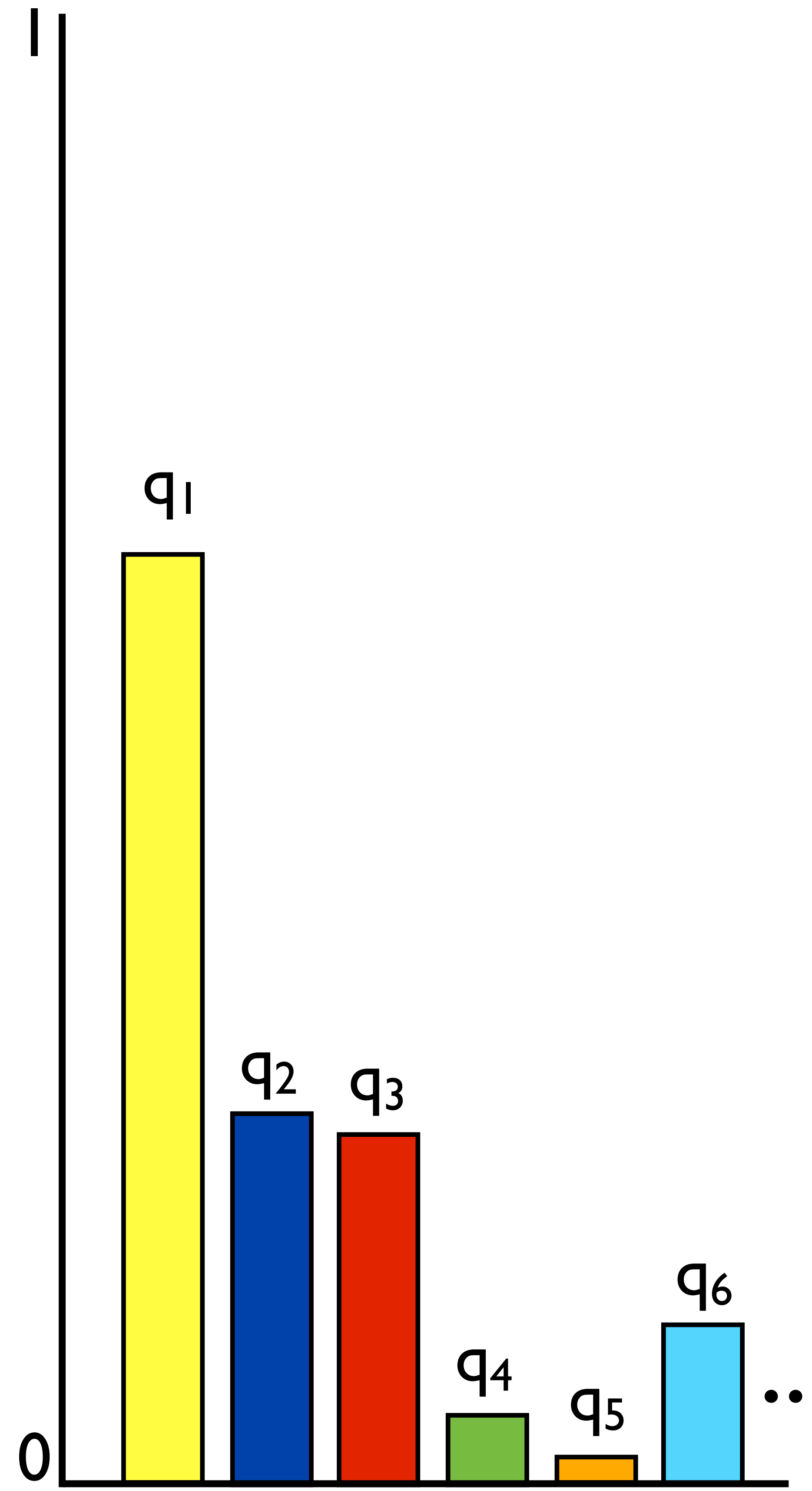




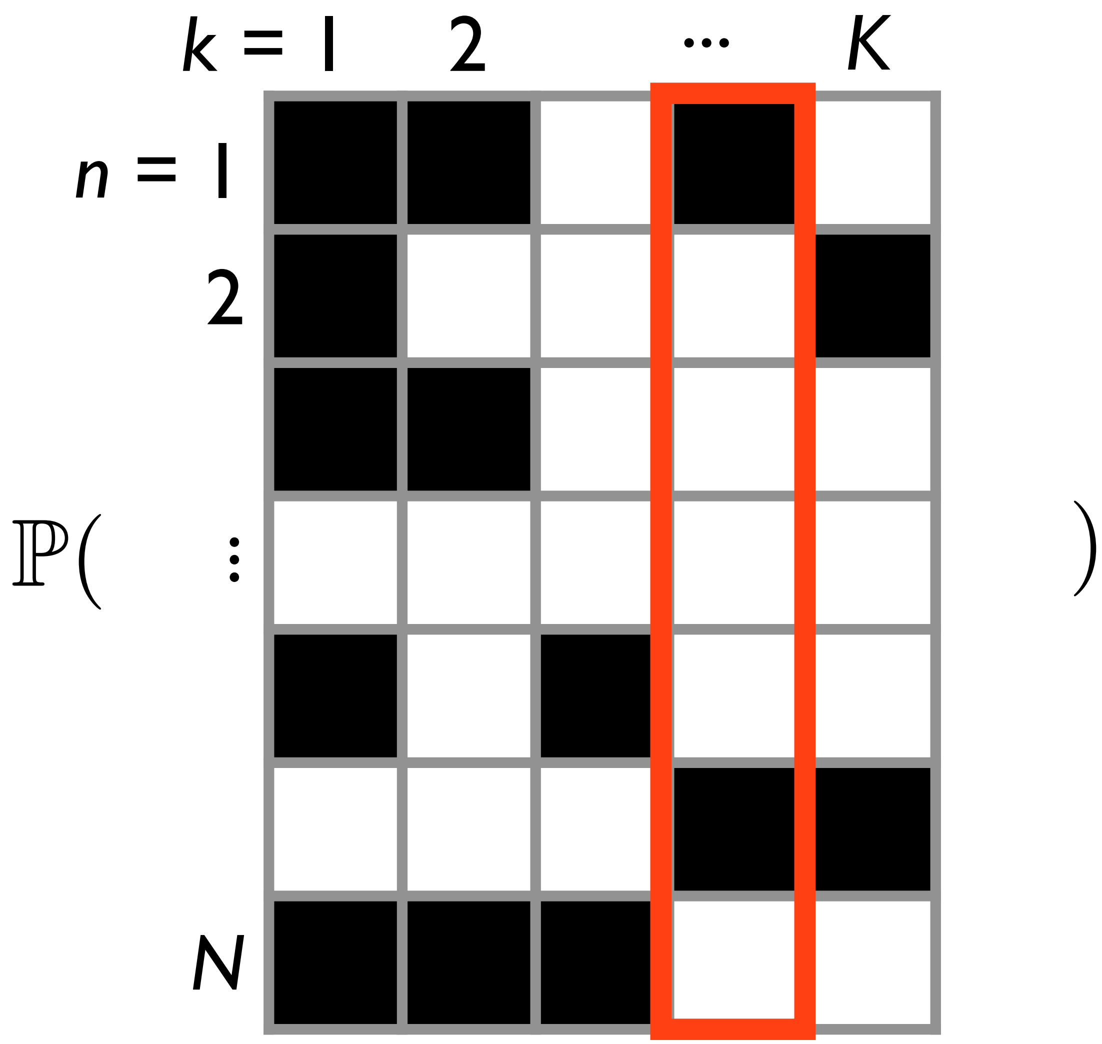
Feature frequency models: EFPFs?



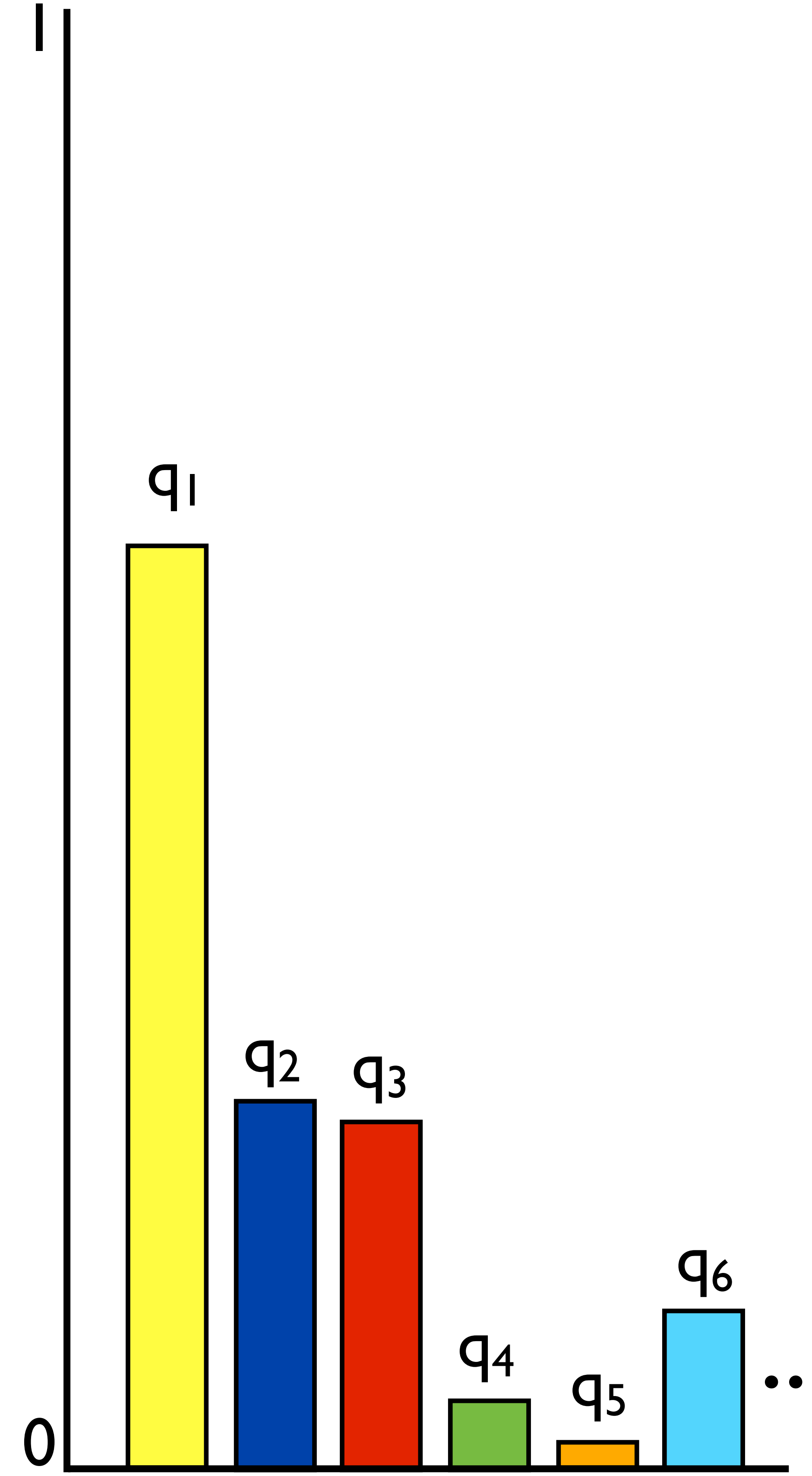
$$q_1^{S_{N,1}}(1 - q_1)^{N - S_{N,1}}$$



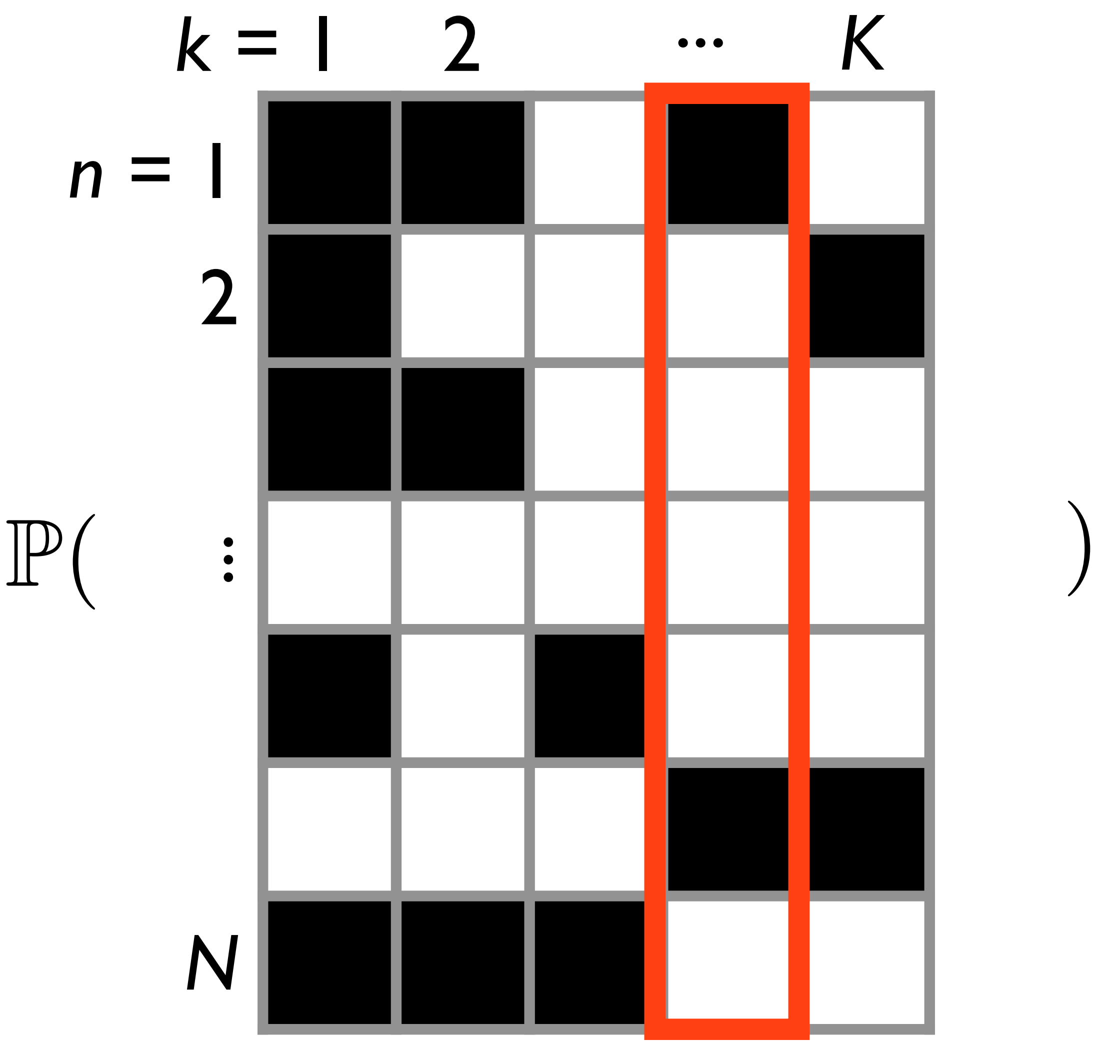
Feature frequency models: EFPFs?



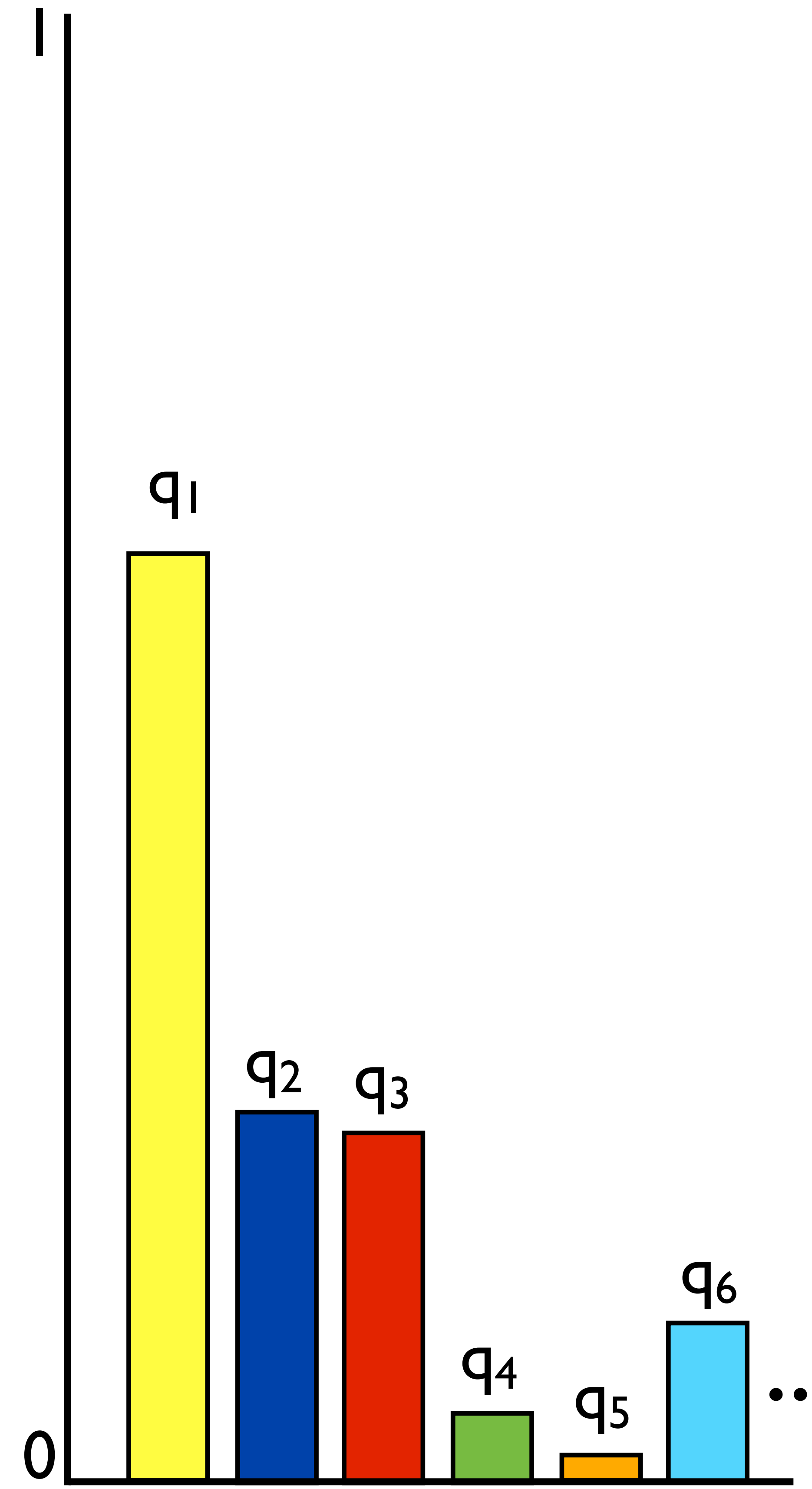
$$q_k^{S_{N,k}}(1-q_k)^{N-S_{N,k}}$$



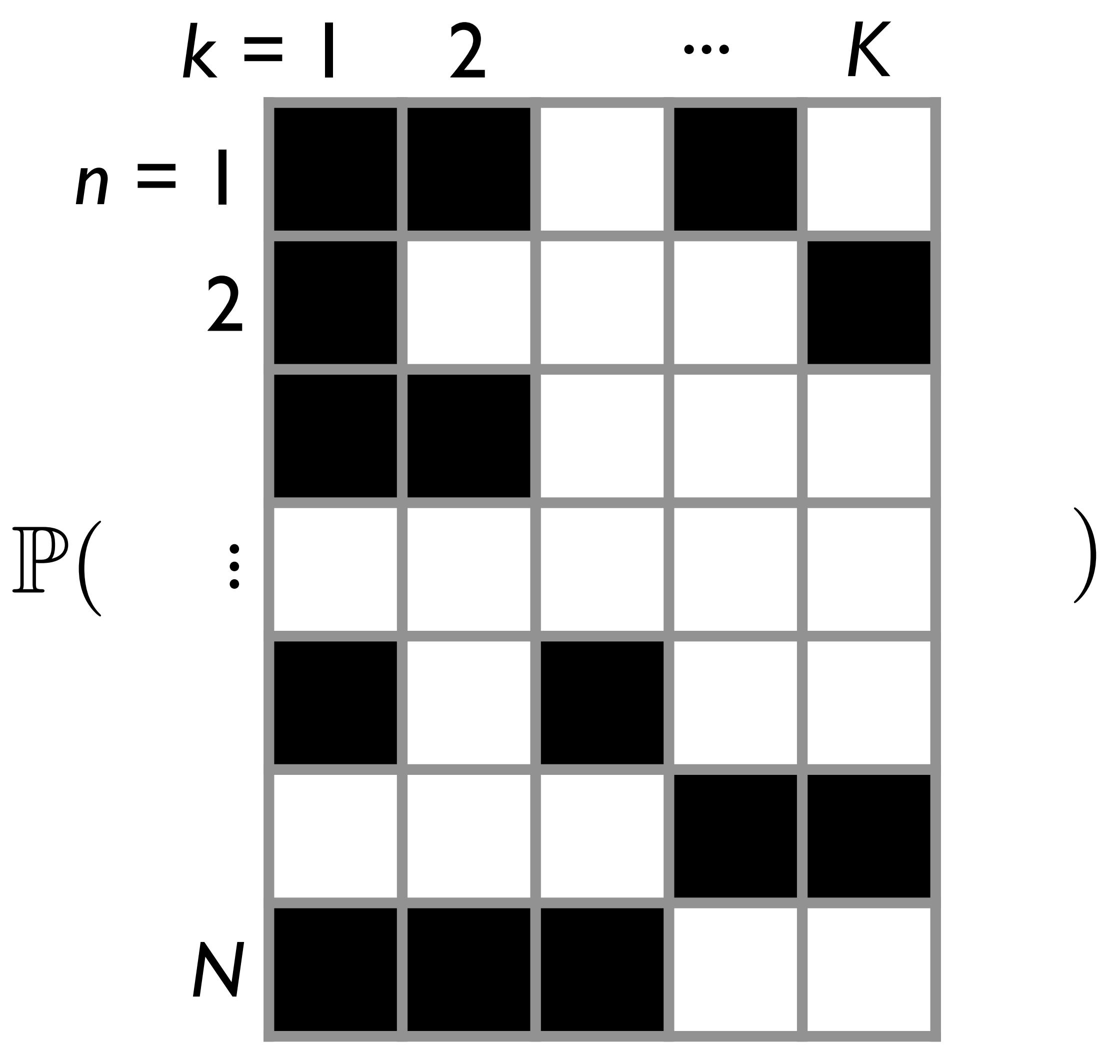
Feature frequency models: EFPFs?



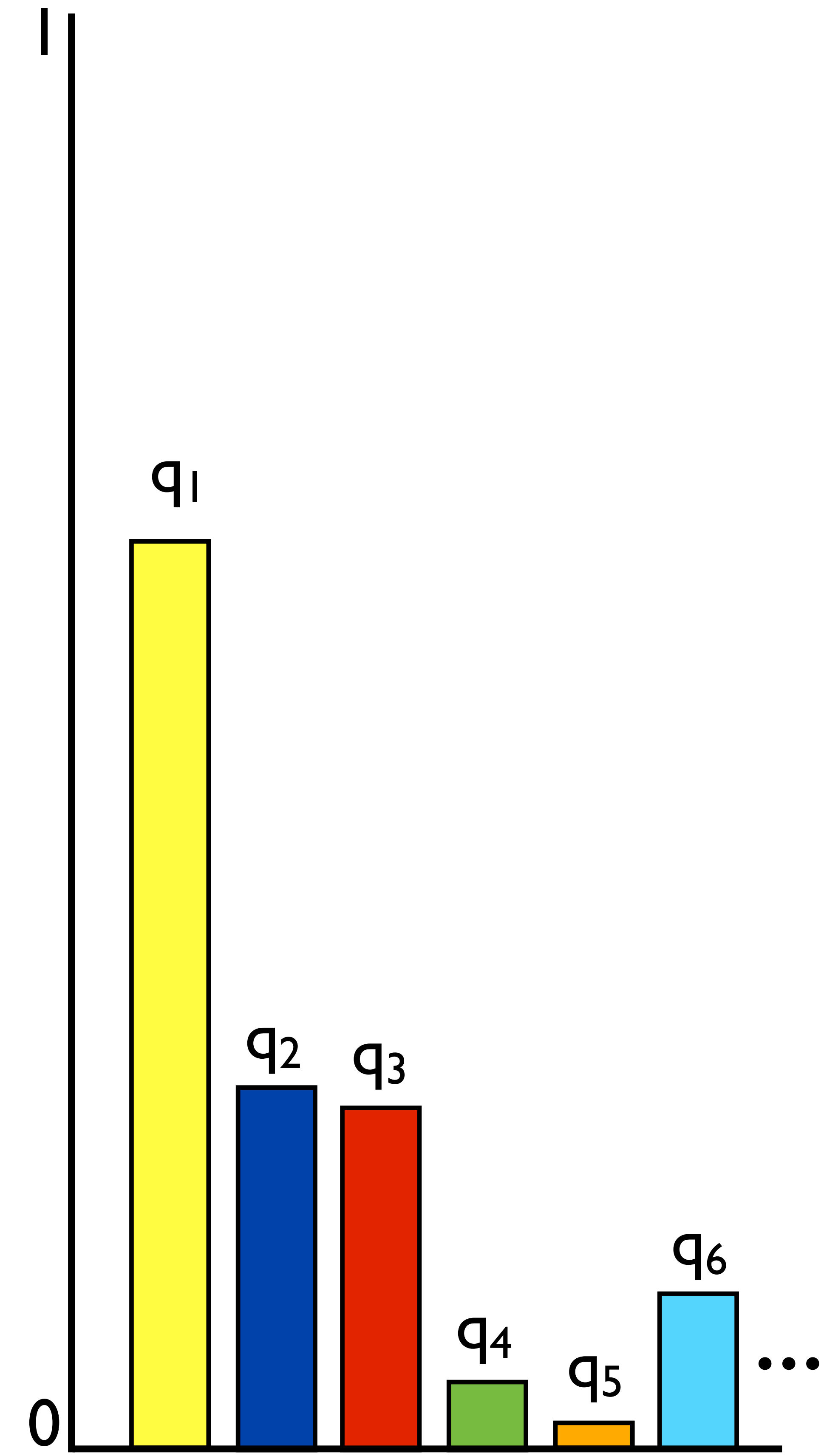
$$q_{i_k}^{S_{N,k}}(1 - q_{i_k})^{N - S_{N,k}}$$



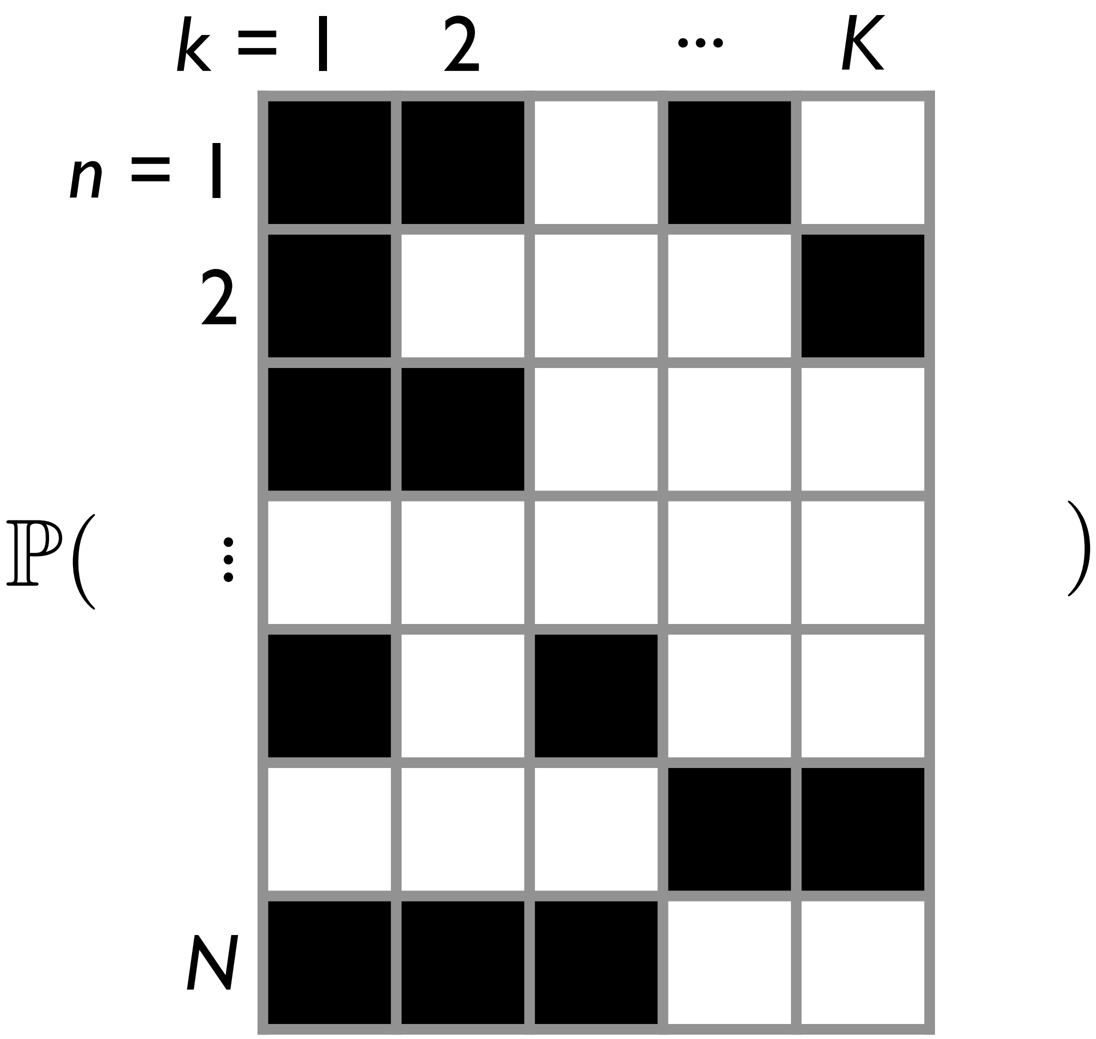
Feature frequency models: EFPFs?



$$\prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}}$$

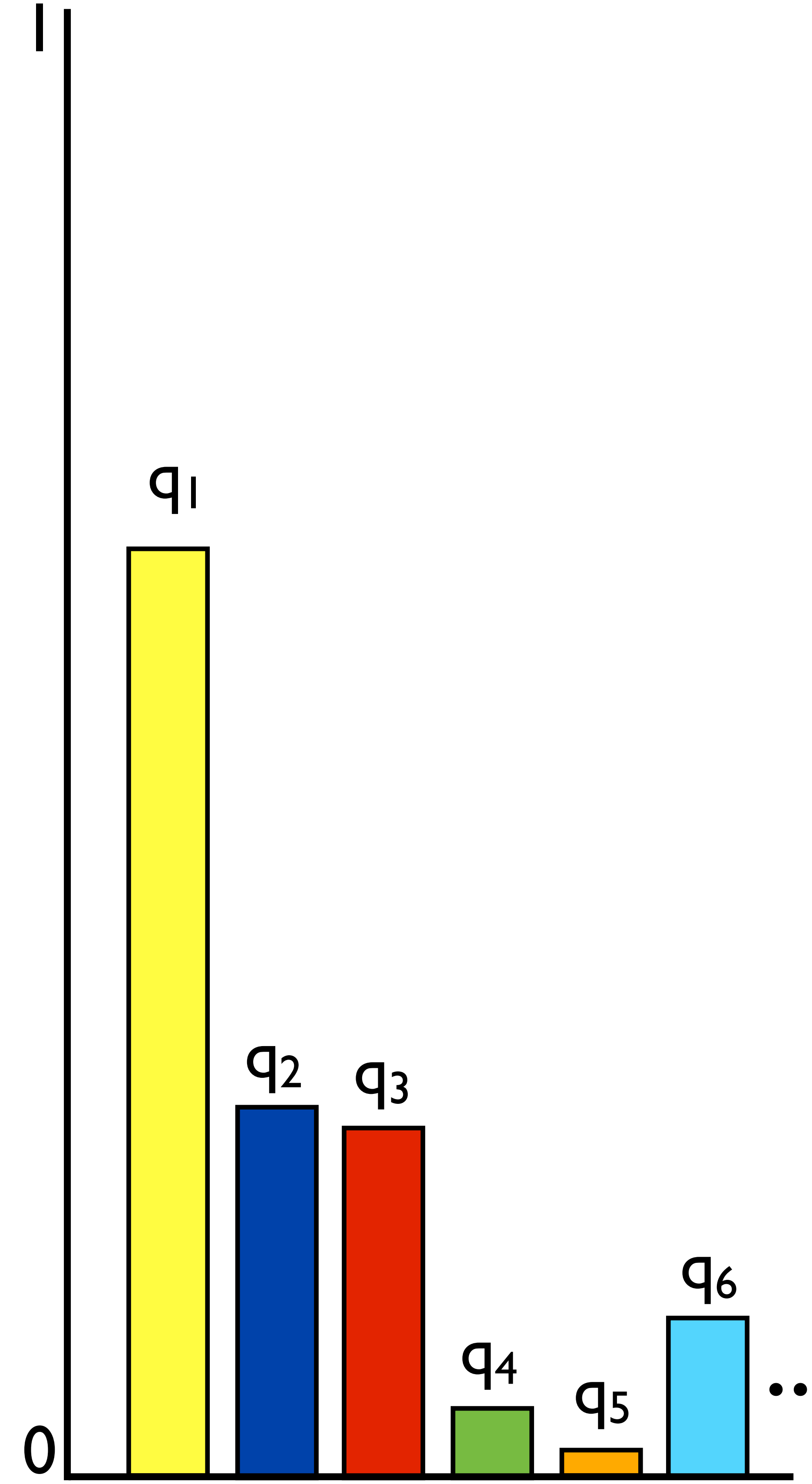


Feature frequency models: EFPFs?

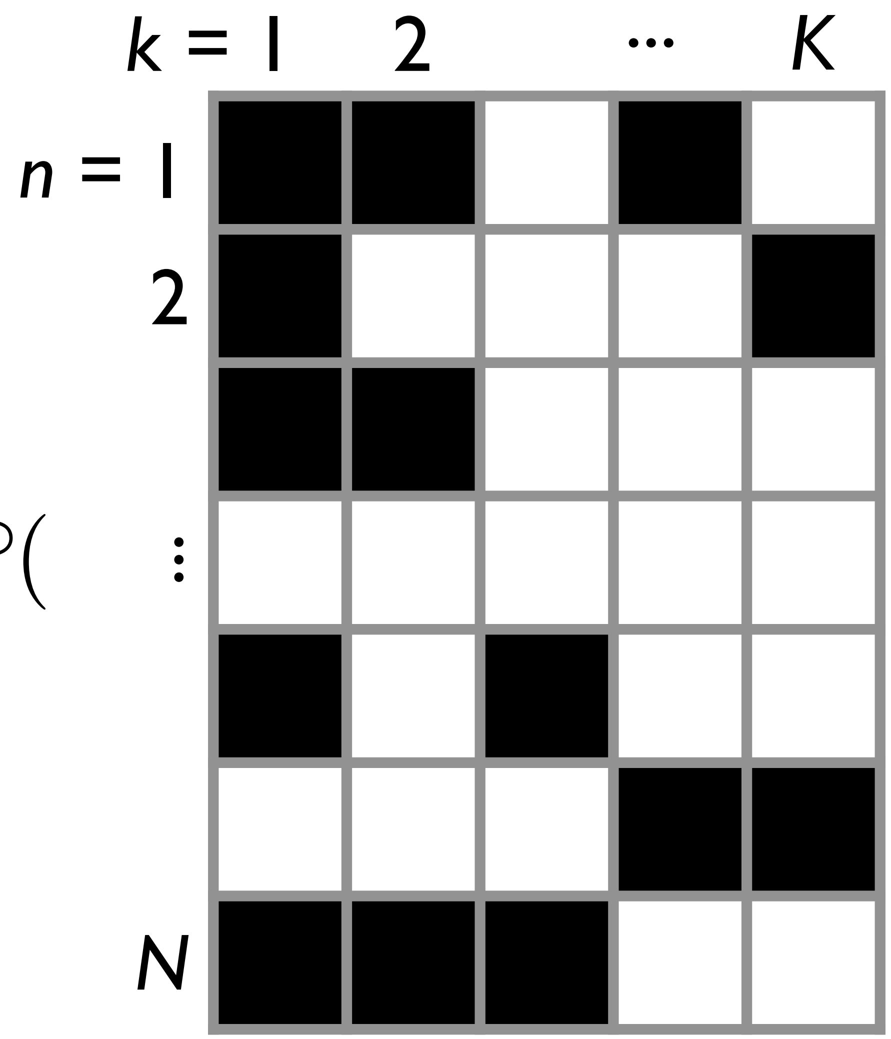


$$\prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}}$$

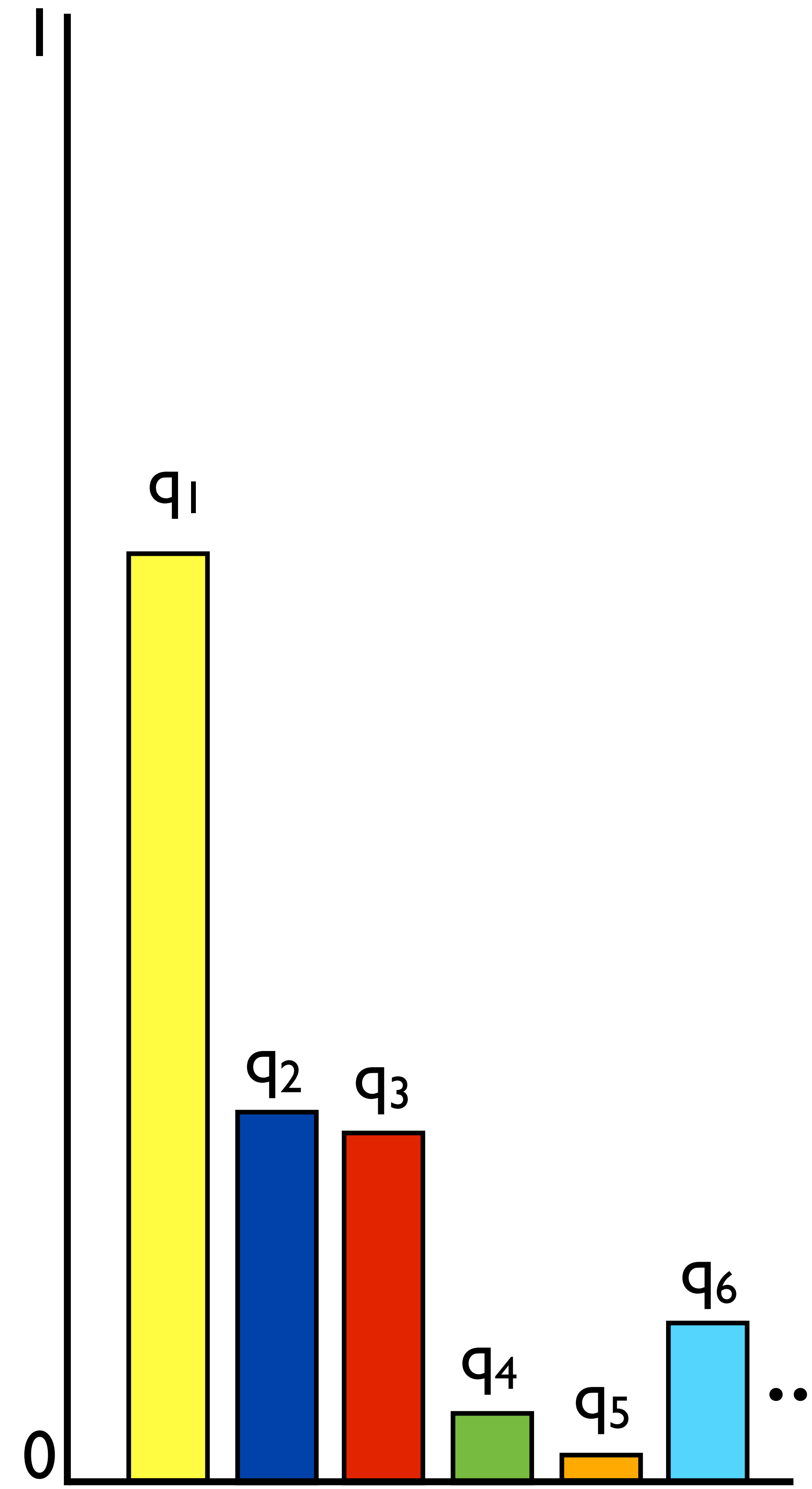
$$\cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N$$



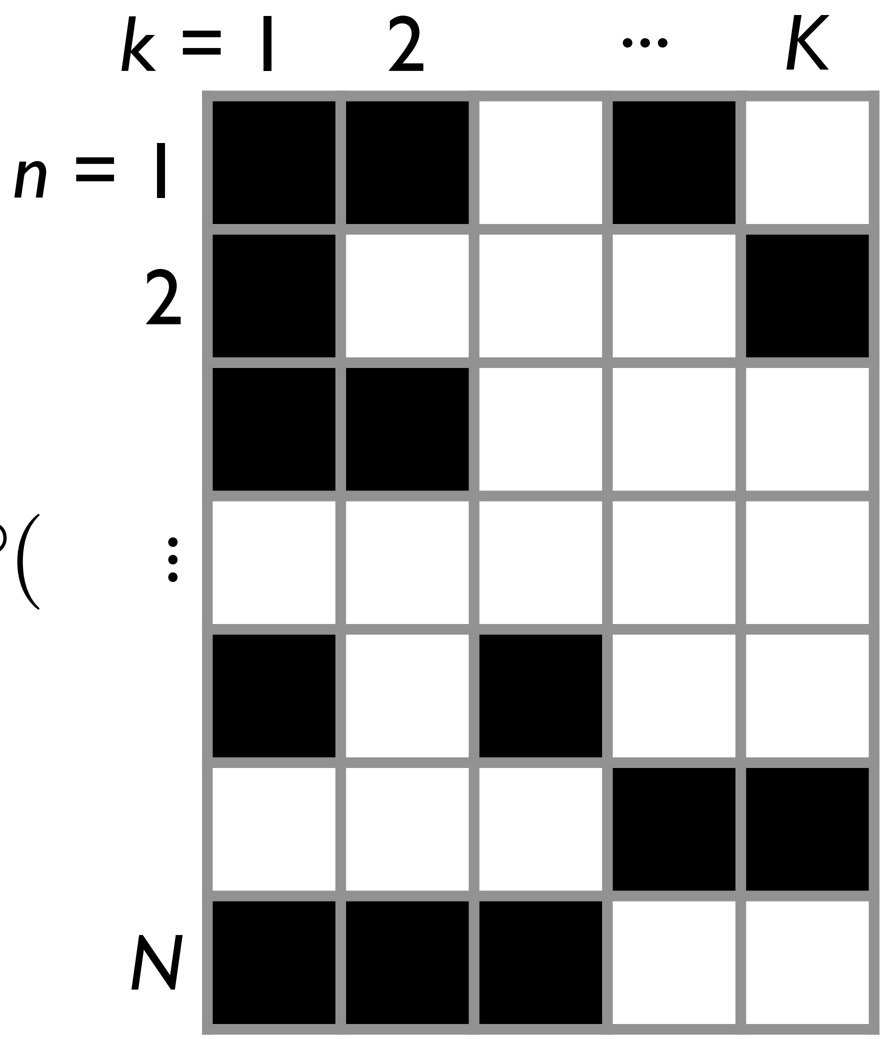
Feature frequency models: EFPFs?

$\mathbb{P}($  $)$

$$= \mathbb{E}\Big[\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N \Big]$$



Feature frequency models: EFPFs?

$\mathbb{P}($  $)$

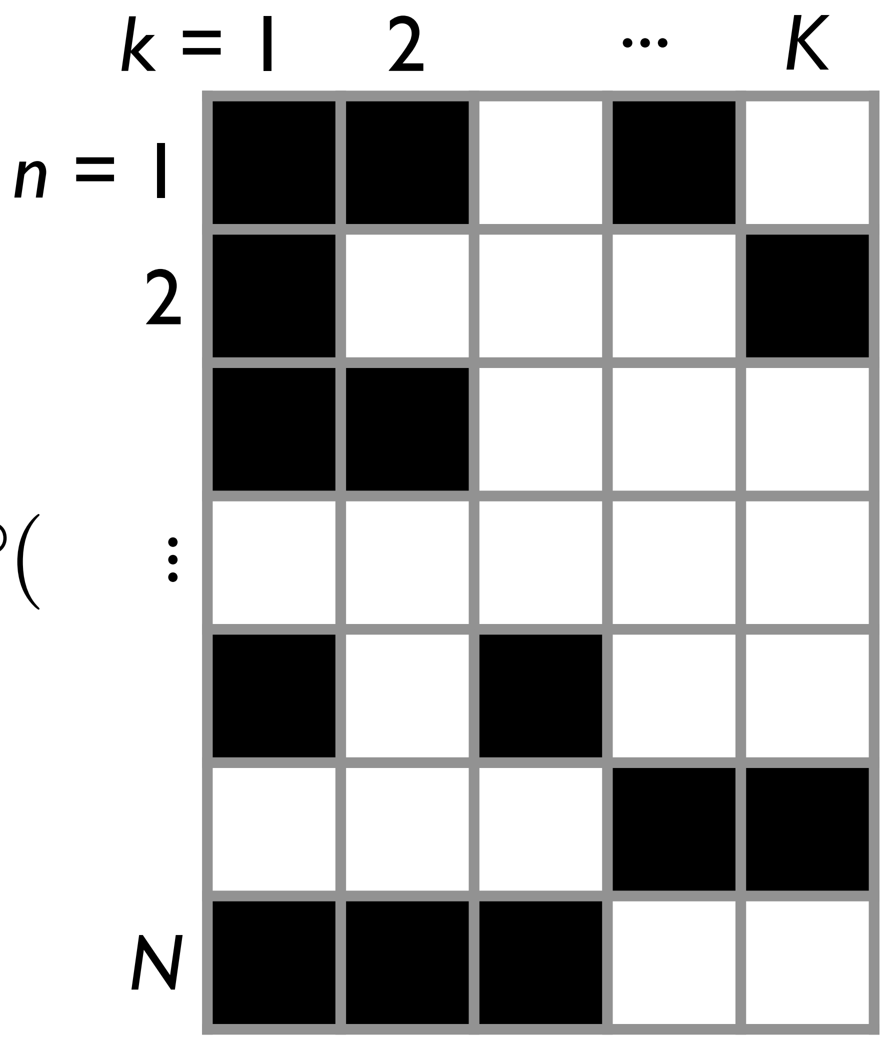
$$
= \mathbb{E}[\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}}
$$

$$
\cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N]
$$

[Broderick, Pitman, Jordan 2013]

Feature frequency models: EFPFs?

$\mathbb{P}($



$)$

$$= \mathbb{E}\Big[\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}}$$

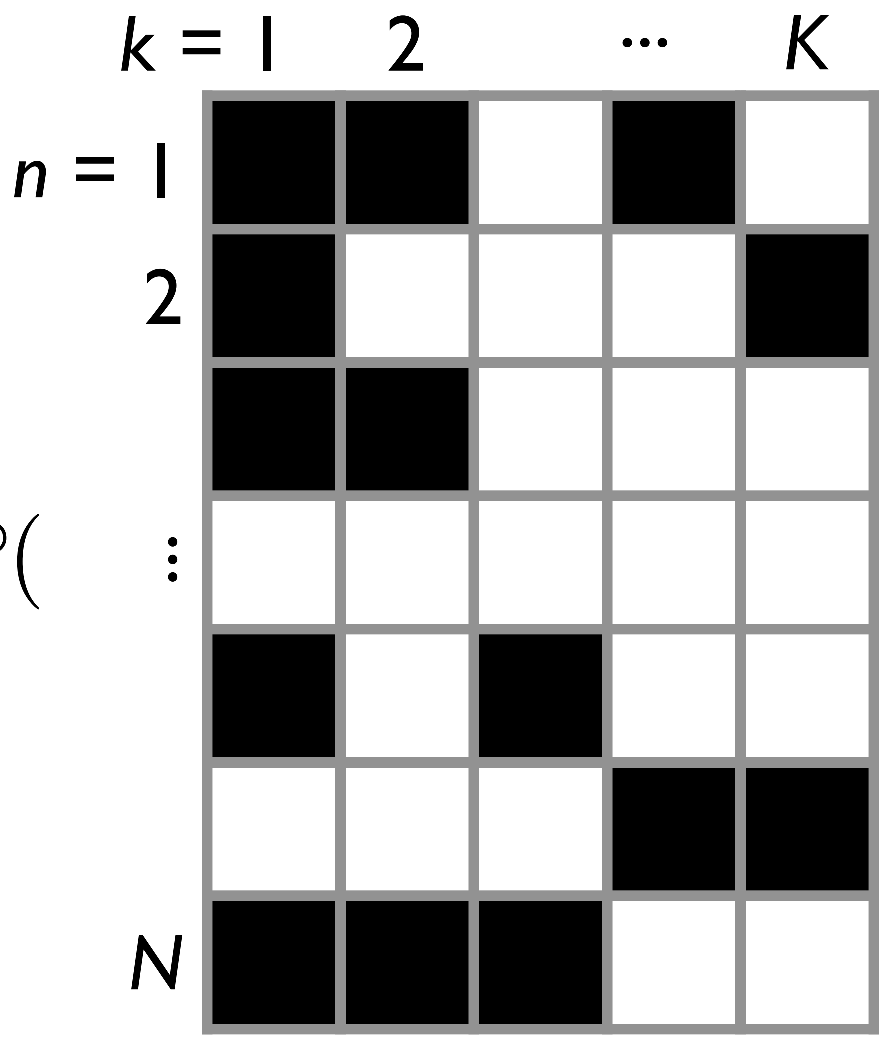
$$\cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^{N} \Big]$$

Size of *k*th feature

[Broderick, Pitman, Jordan 2013]

Feature frequency models: EFPFs?

$\mathbb{P}($



$)$

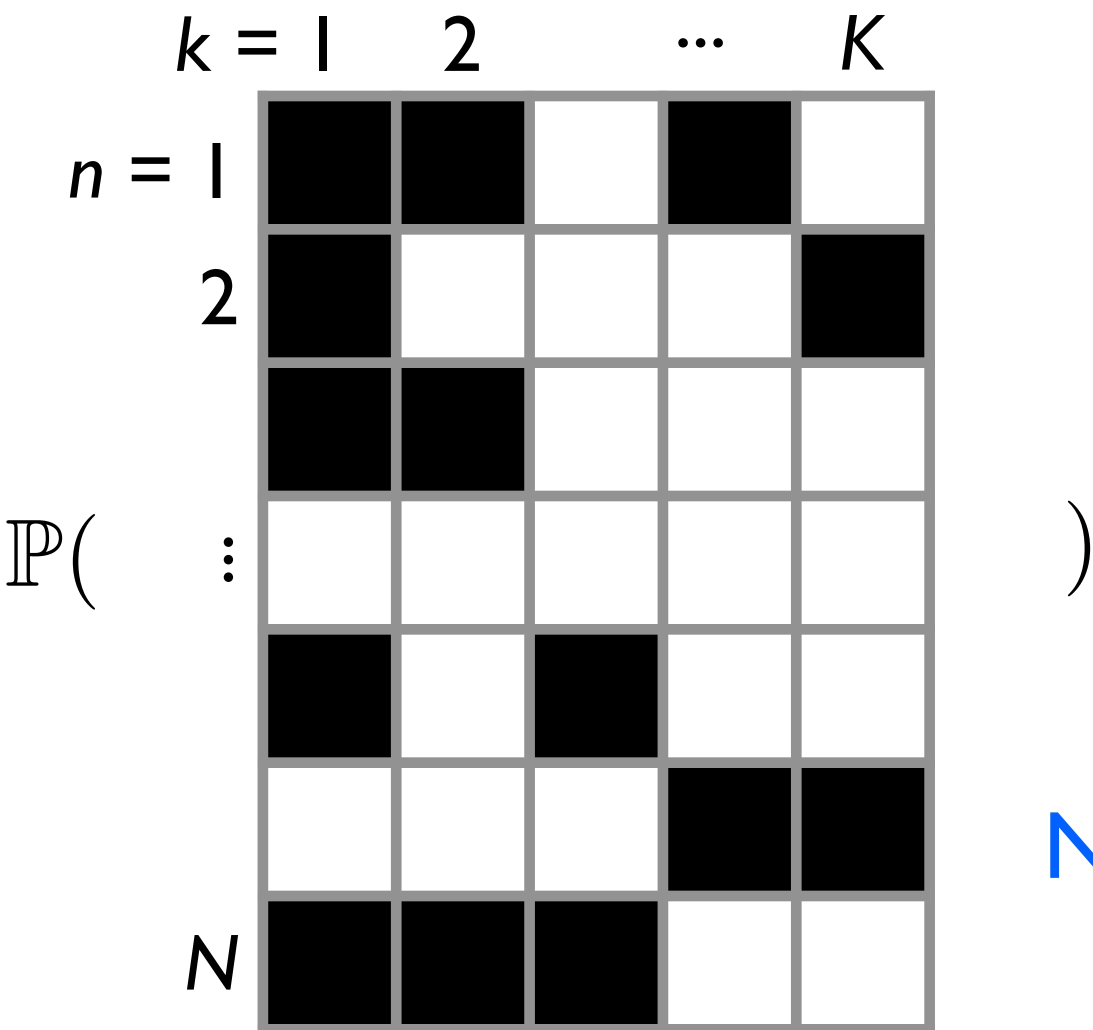
Number of features

Size of kth feature

$$= \mathbb{E}\Big[\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N \Big]$$

[Broderick, Pitman, Jordan 2013]

Feature frequency models: EFPFs?



Number of features

Number of data points

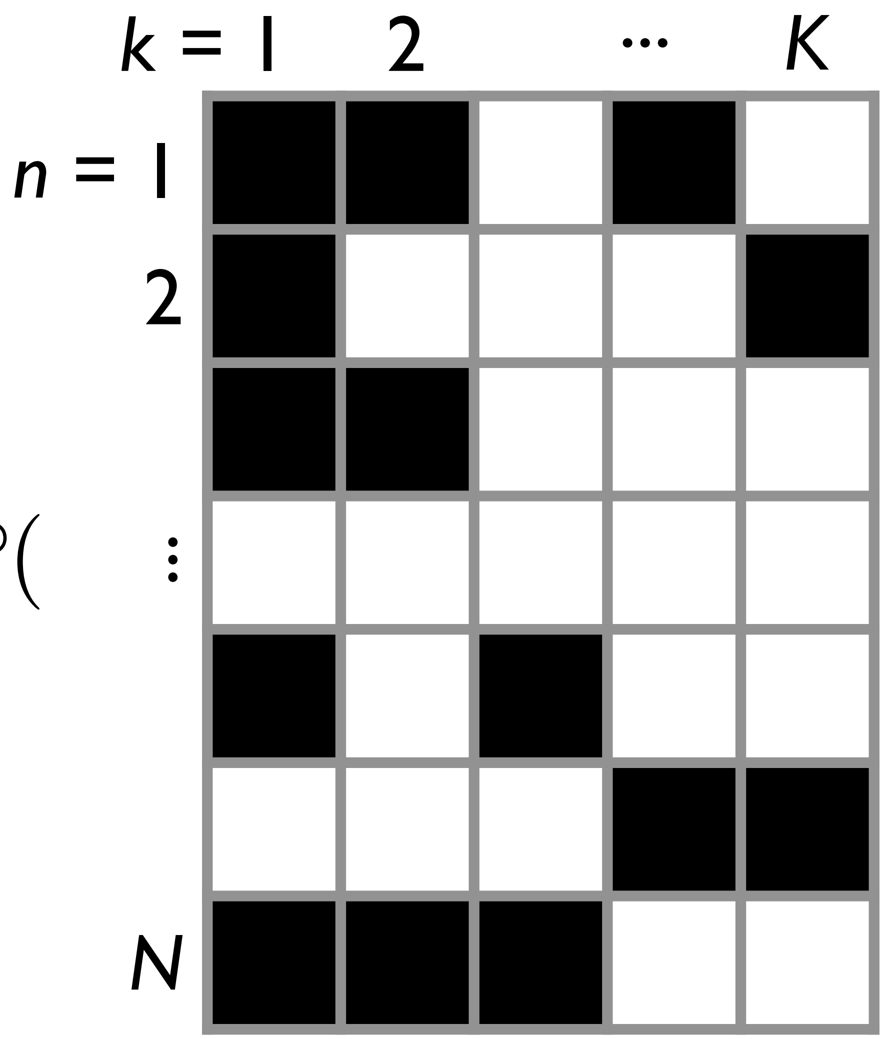
Size of kth feature

$$= \mathbb{E}\Big[\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N \Big]$$

[Broderick, Pitman, Jordan 2013]

Feature frequency models: EFPFs?

$\mathbb{P}($



$)$

Number of features

Number of data points

Size of kth feature

$$= \mathbb{E}\Big[\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}}$$

$$\cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N \Big] = p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})$$

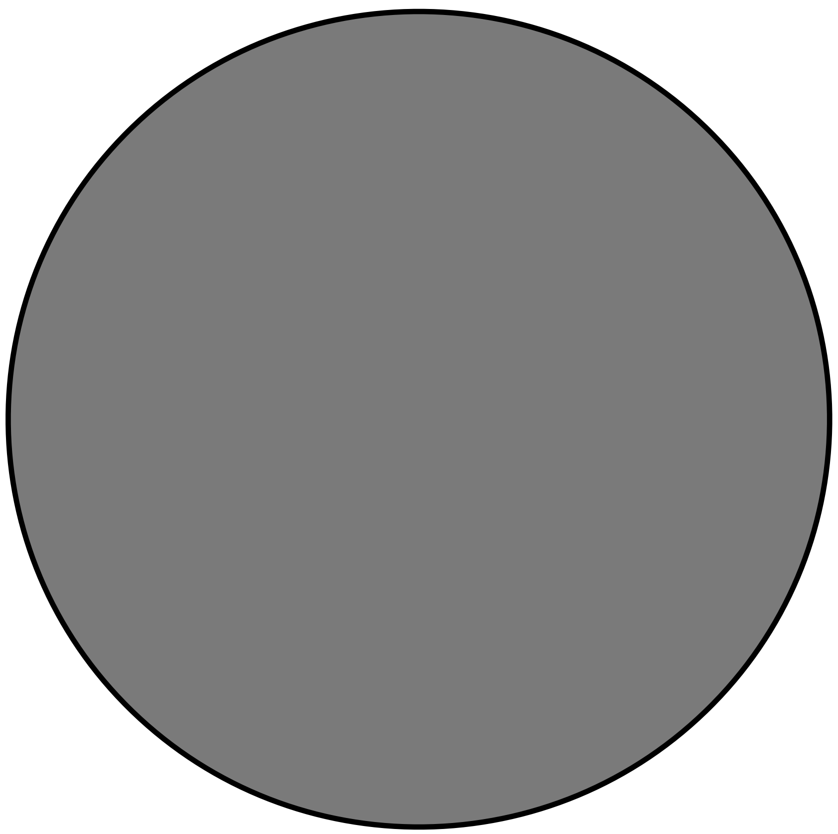
EFPF

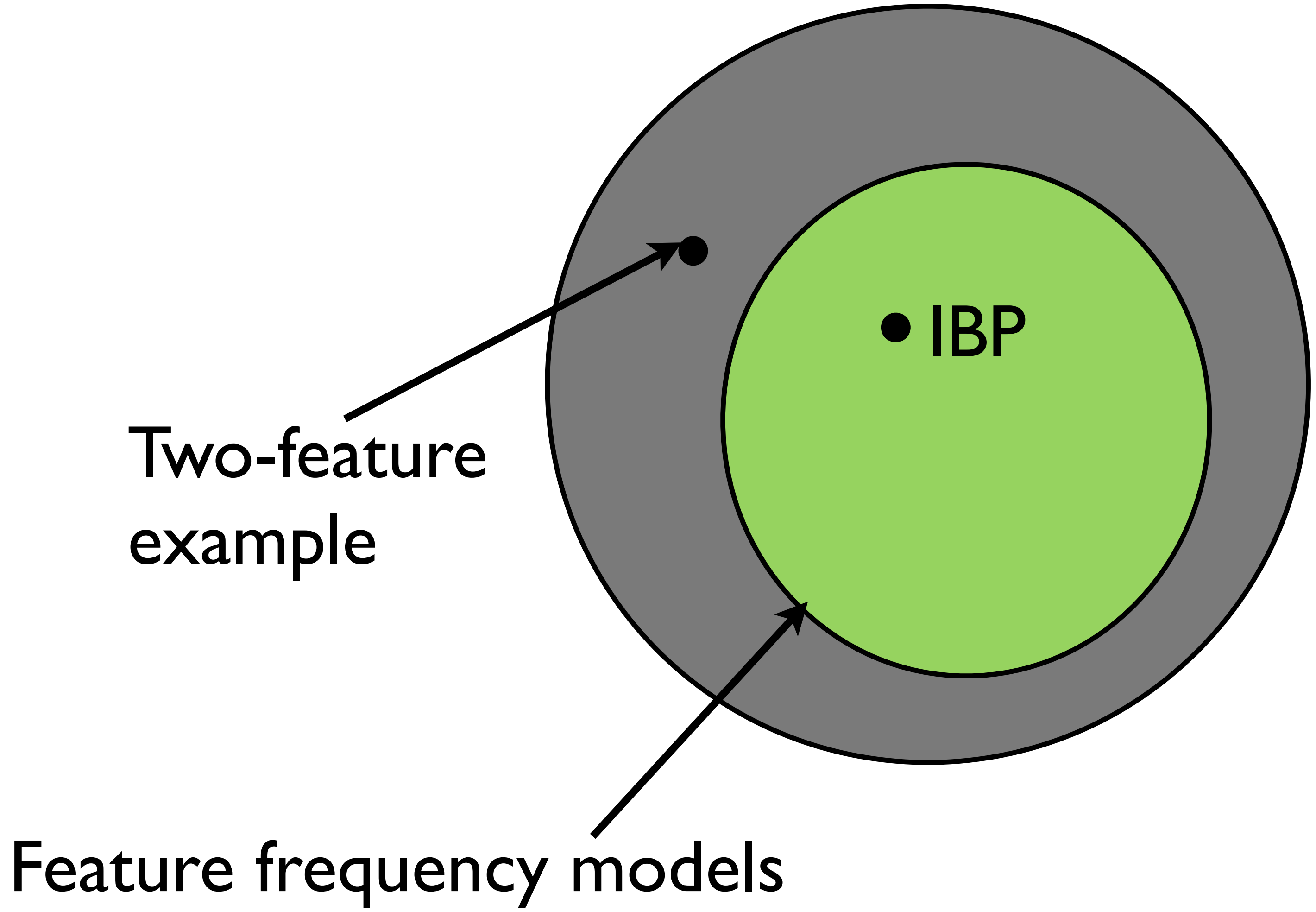
[Broderick, Pitman, Jordan 2013]

Feature frequency models: EFPFs?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions

Exchangeable feature distributions
= Feature paintbox allocations

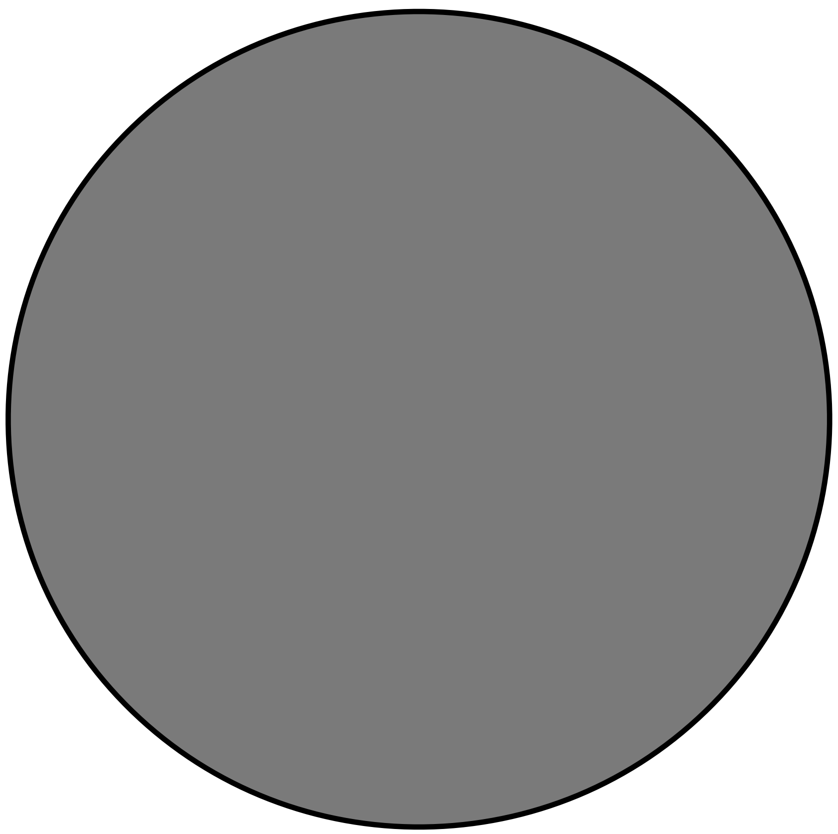




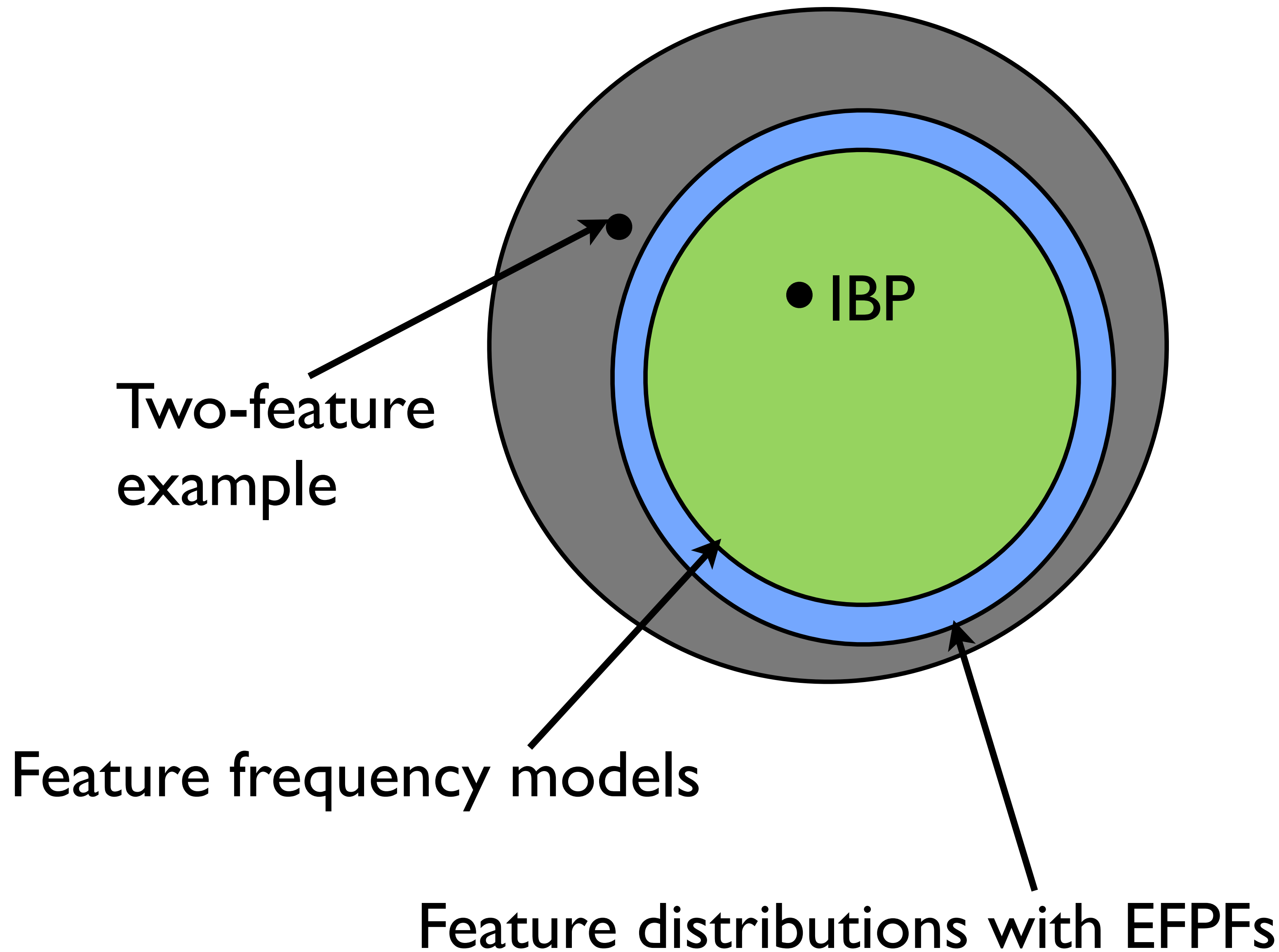
[Broderick, Pitman, Jordan 2013]

Feature frequency models: EFPFs?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions



Exchangeable feature distributions
= Feature paintbox allocations



[Broderick, Pitman, Jordan 2013]

Distributions with EFPFs: frequencies?

[Broderick, Pitman, Jordan 2013]

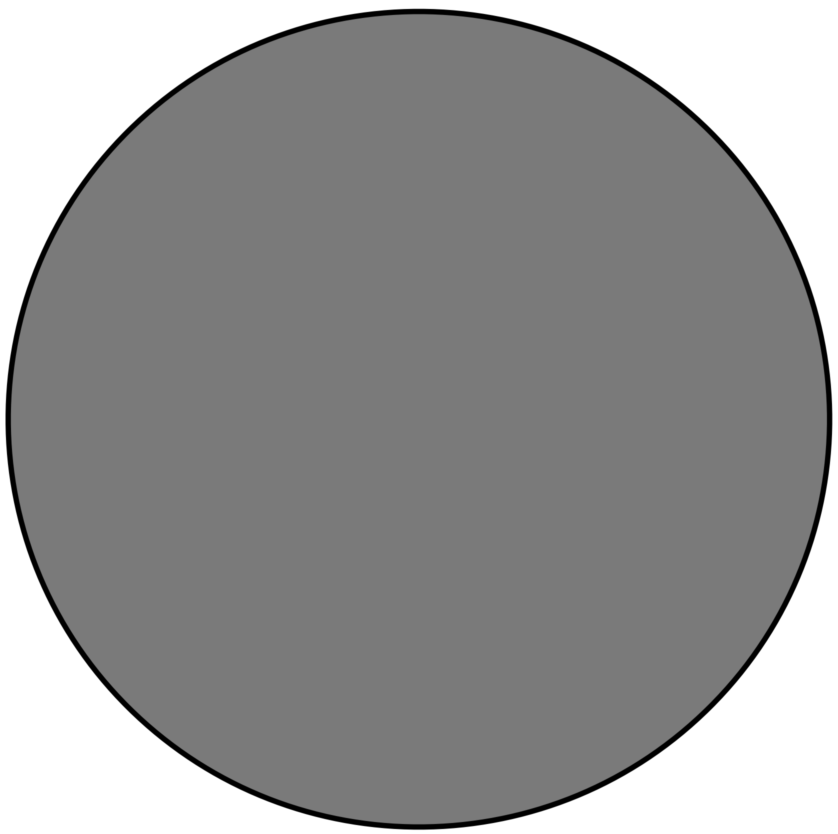
Distributions with EFPFs: frequencies?

✔ • Any number (+unbounded case) of features

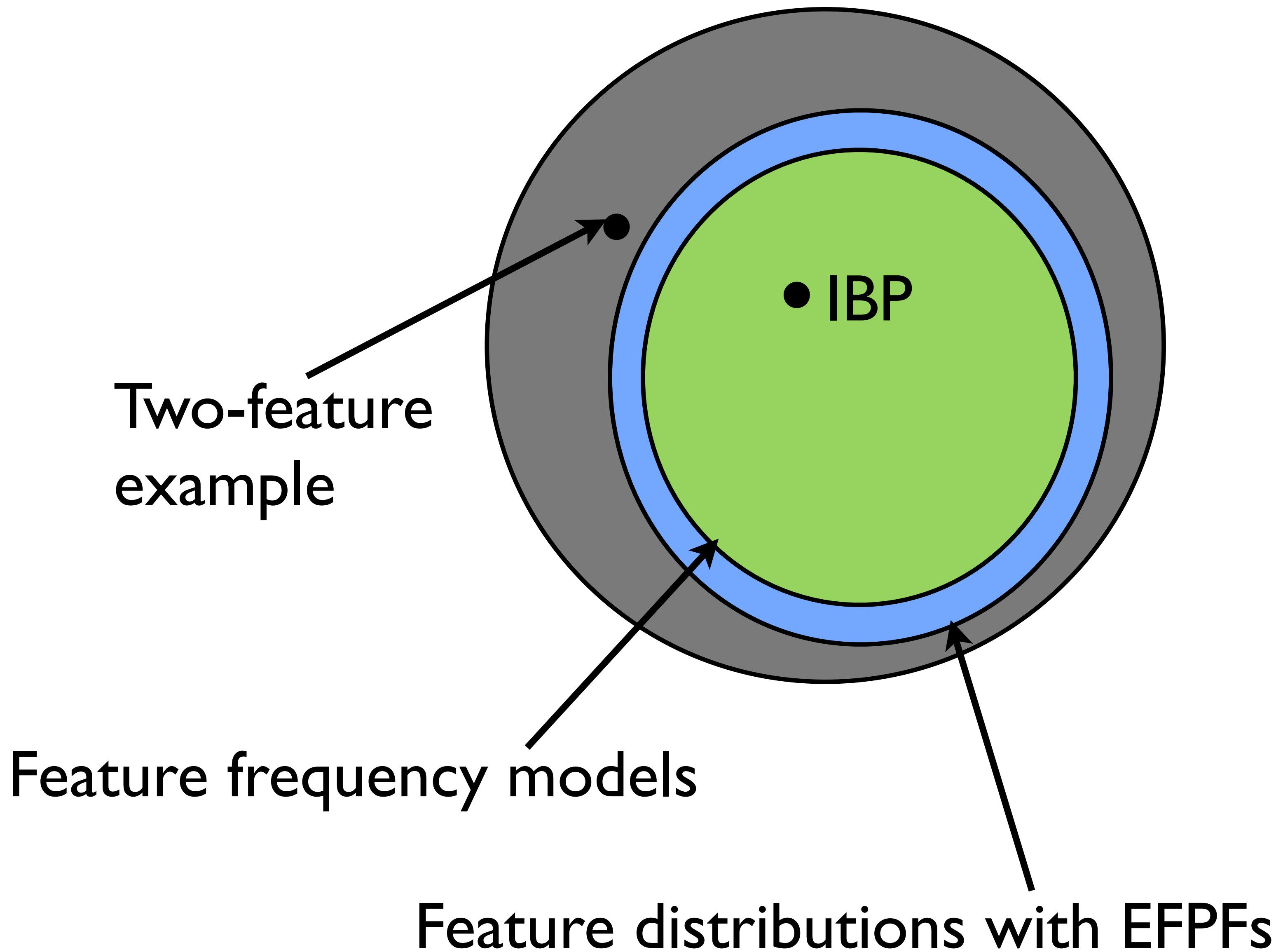
[Broderick, Pitman, Jordan 2013]

Distributions with EFPFs: frequencies?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions



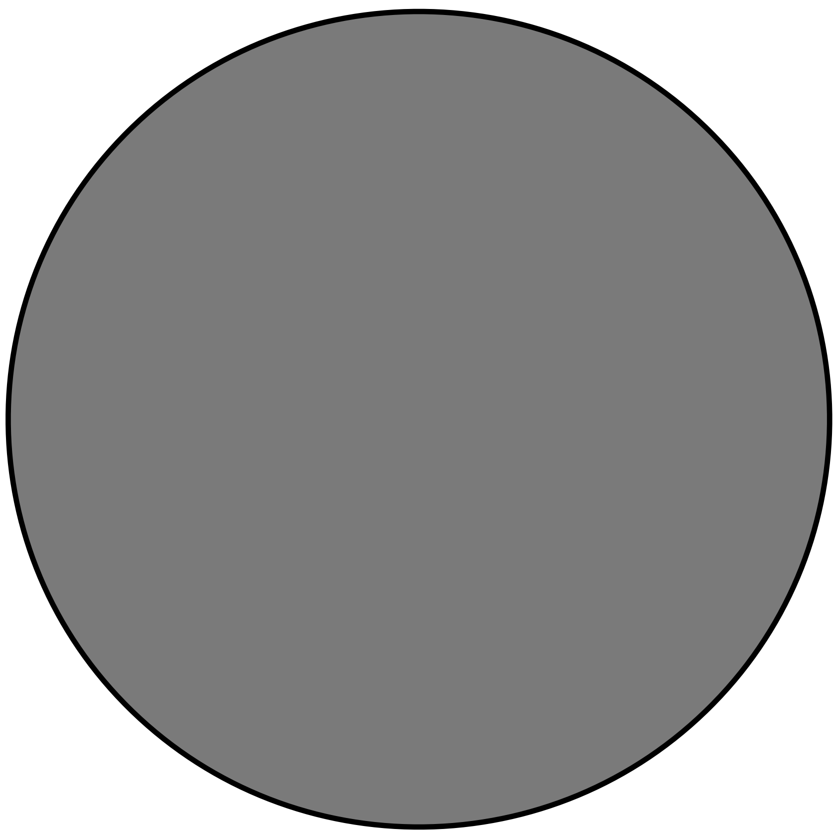
Exchangeable feature distributions
= Feature paintbox allocations



[Broderick, Pitman, Jordan 2013]

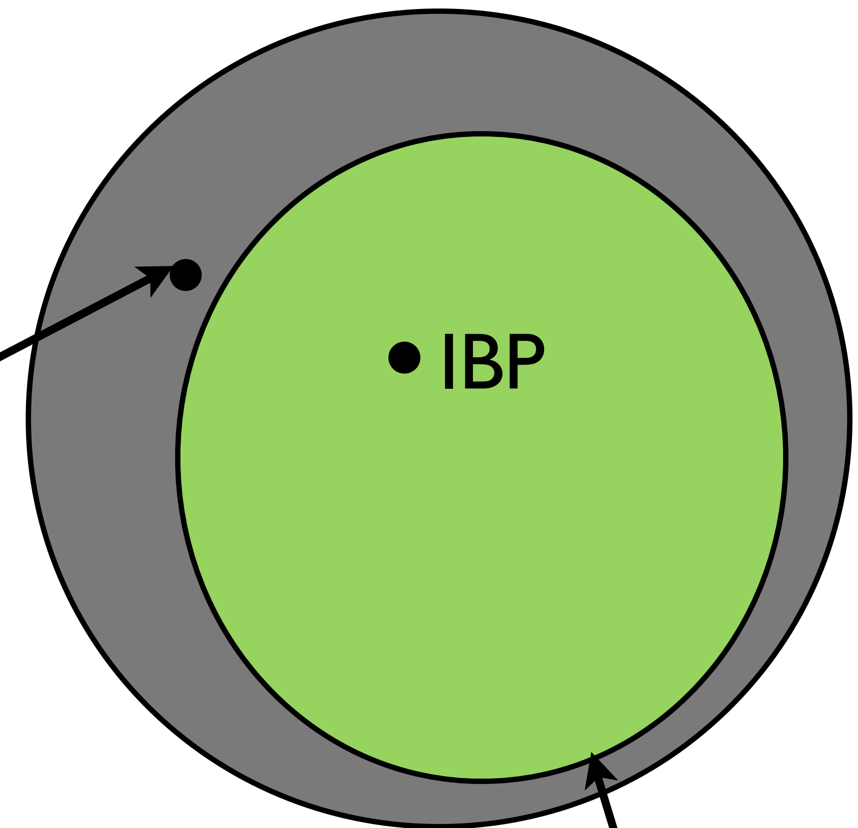
Distributions with EFPFs: frequencies?

Exchangeable cluster distributions
= Cluster distributions with EPPFs
= Kingman paintbox partitions



Exchangeable feature distributions
= Feature paintbox allocations

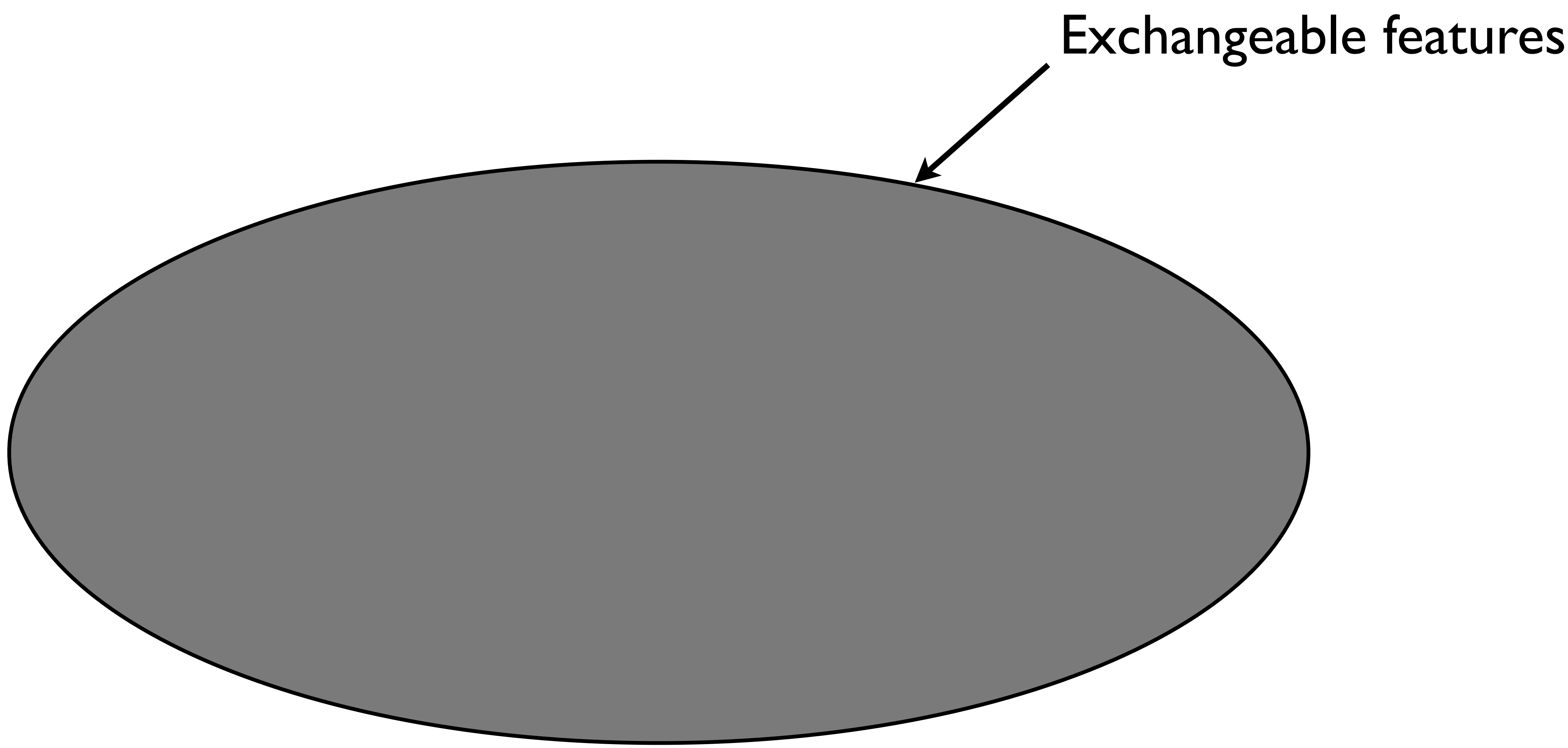
Two-feature example



Feature distributions with EFPFs
= Feature frequency models

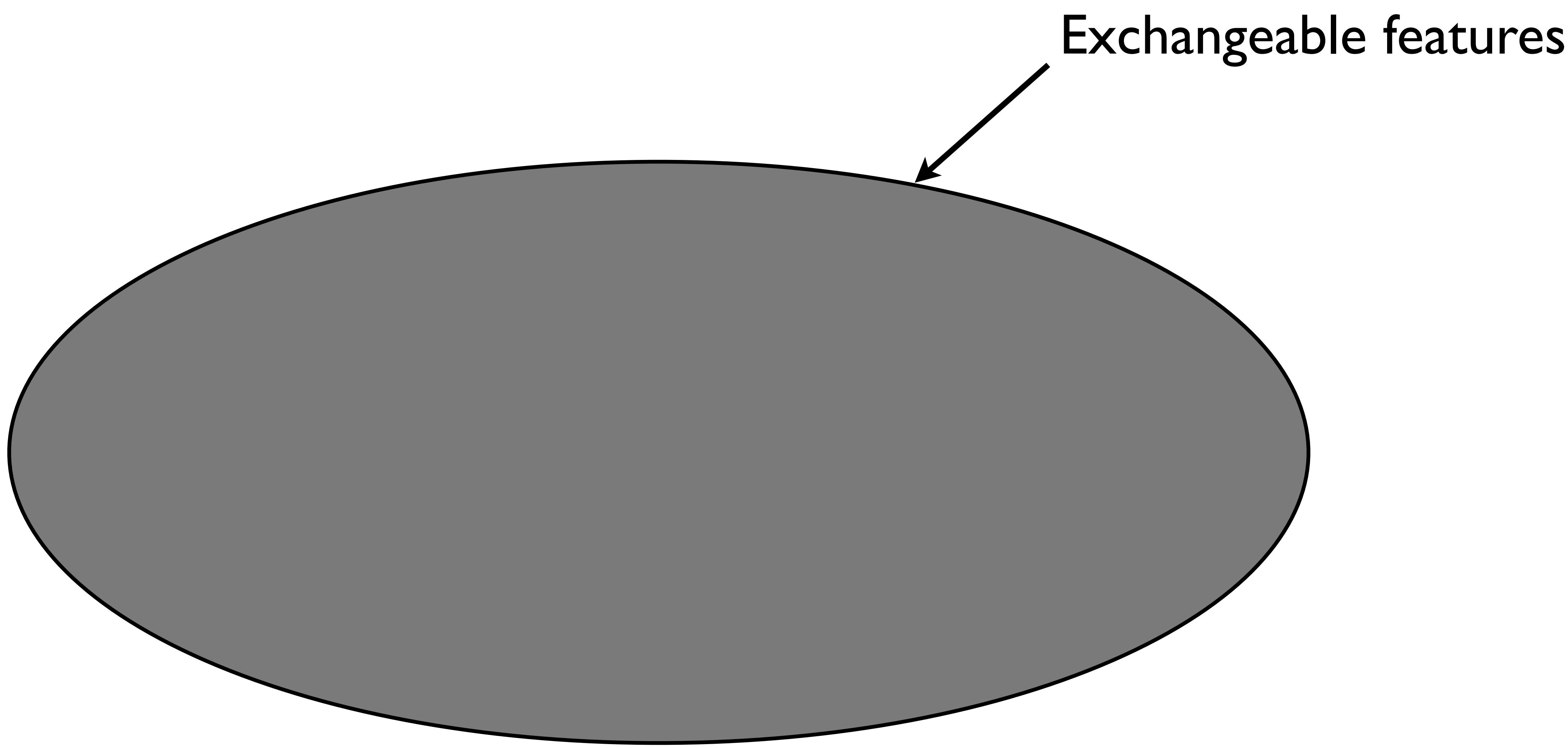
[Broderick, Pitman, Jordan 2013]

Theory conclusions



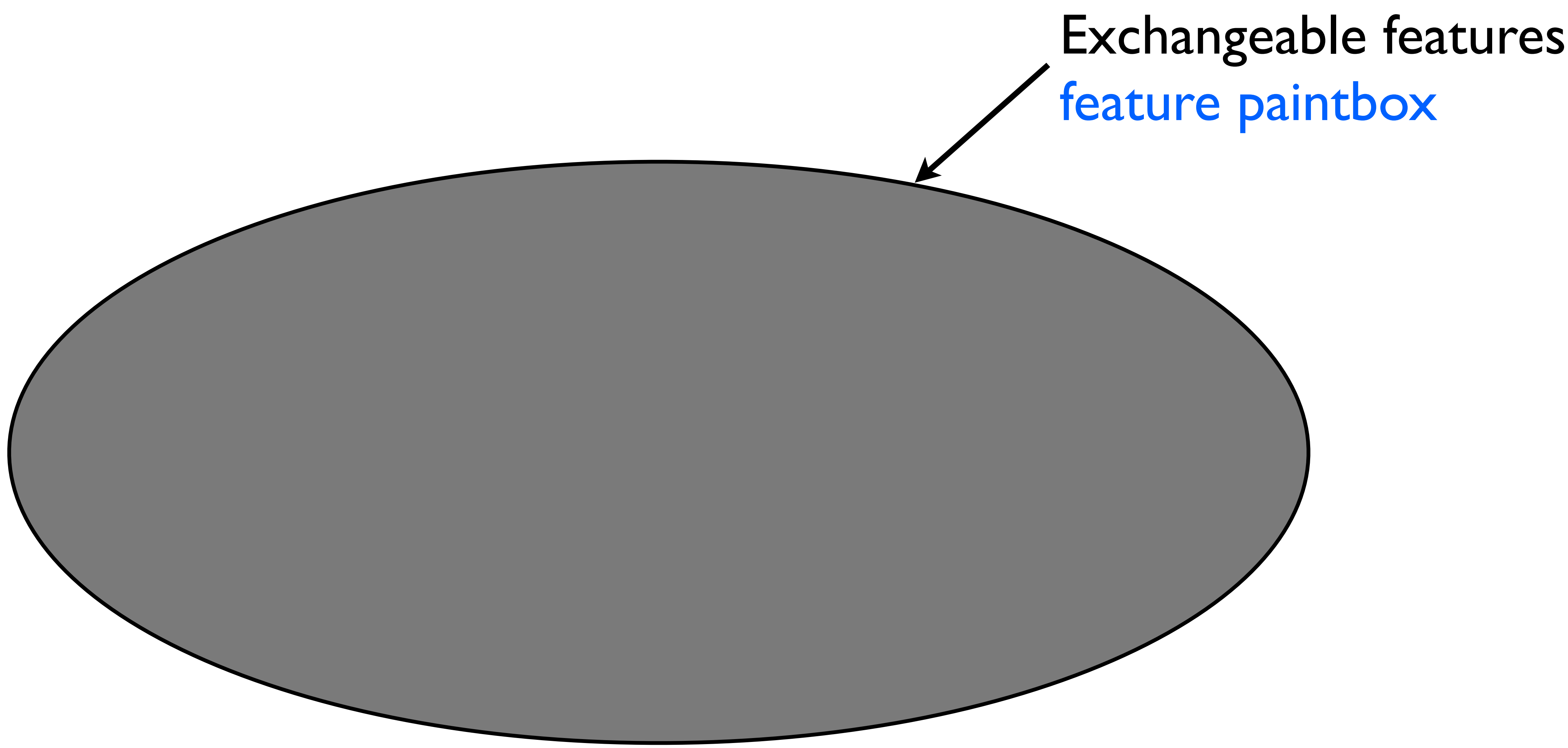
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models



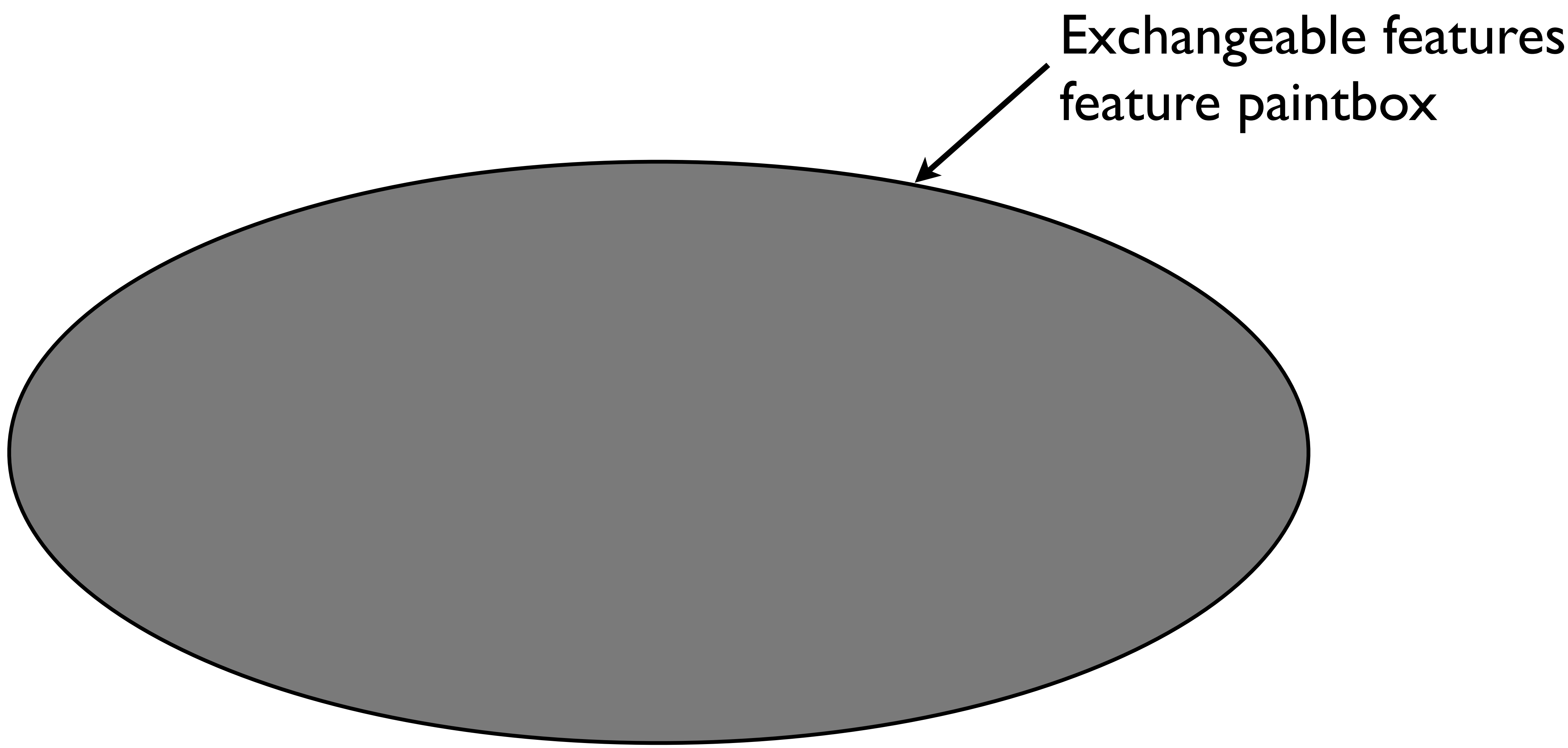
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models



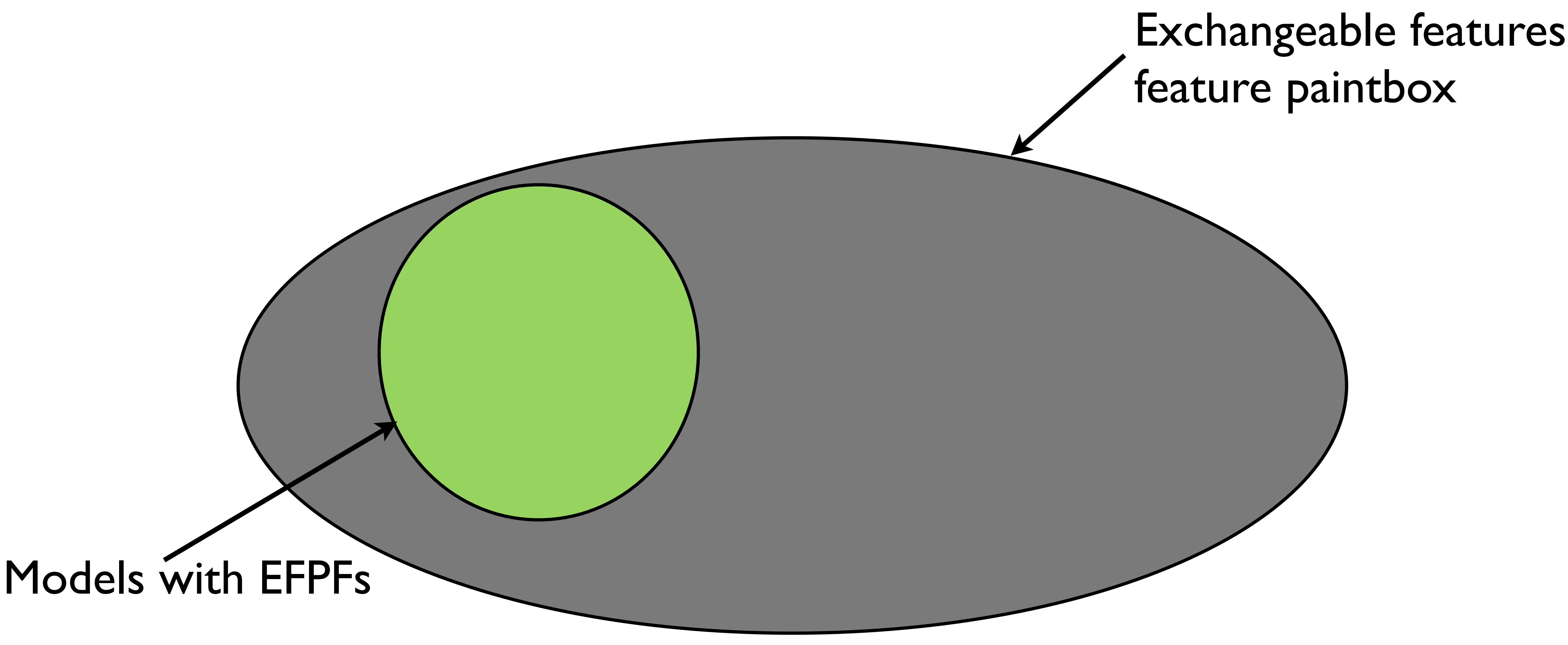
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Limits of clustering characterizations in feature case?



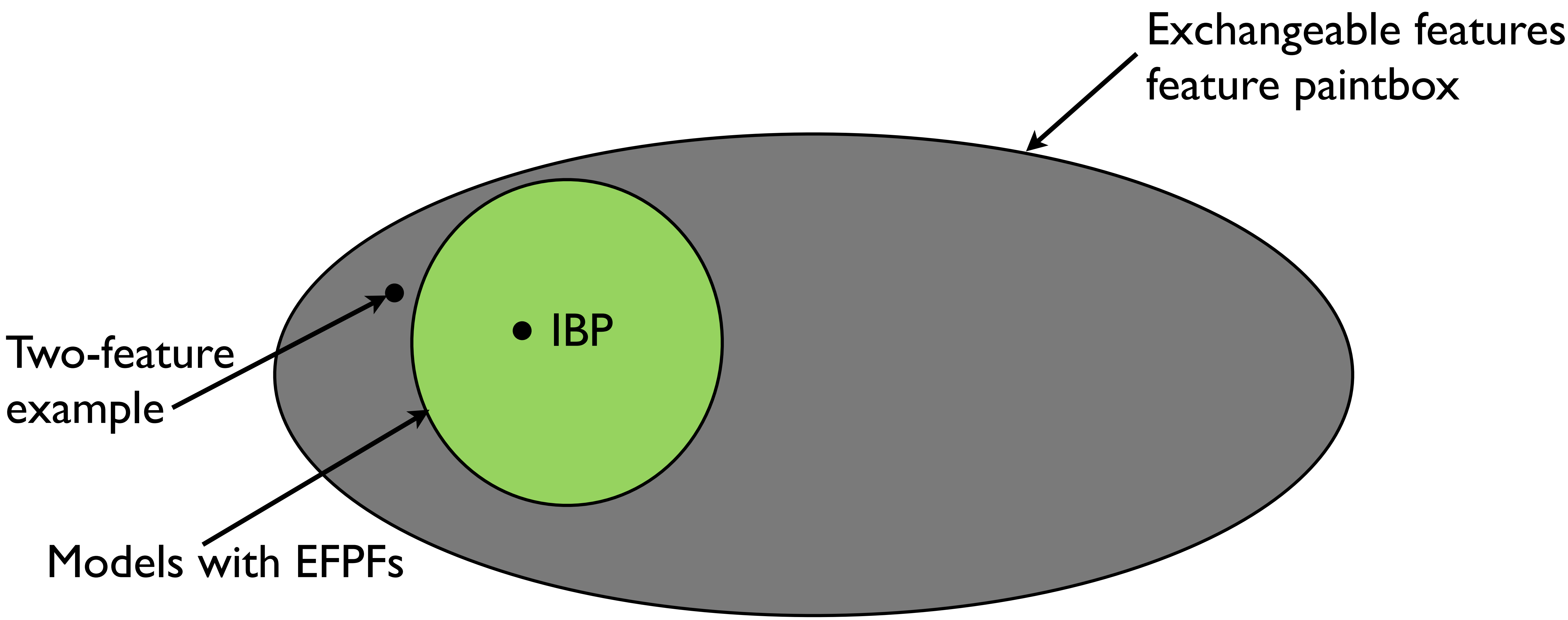
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Limits of clustering characterizations in feature case?



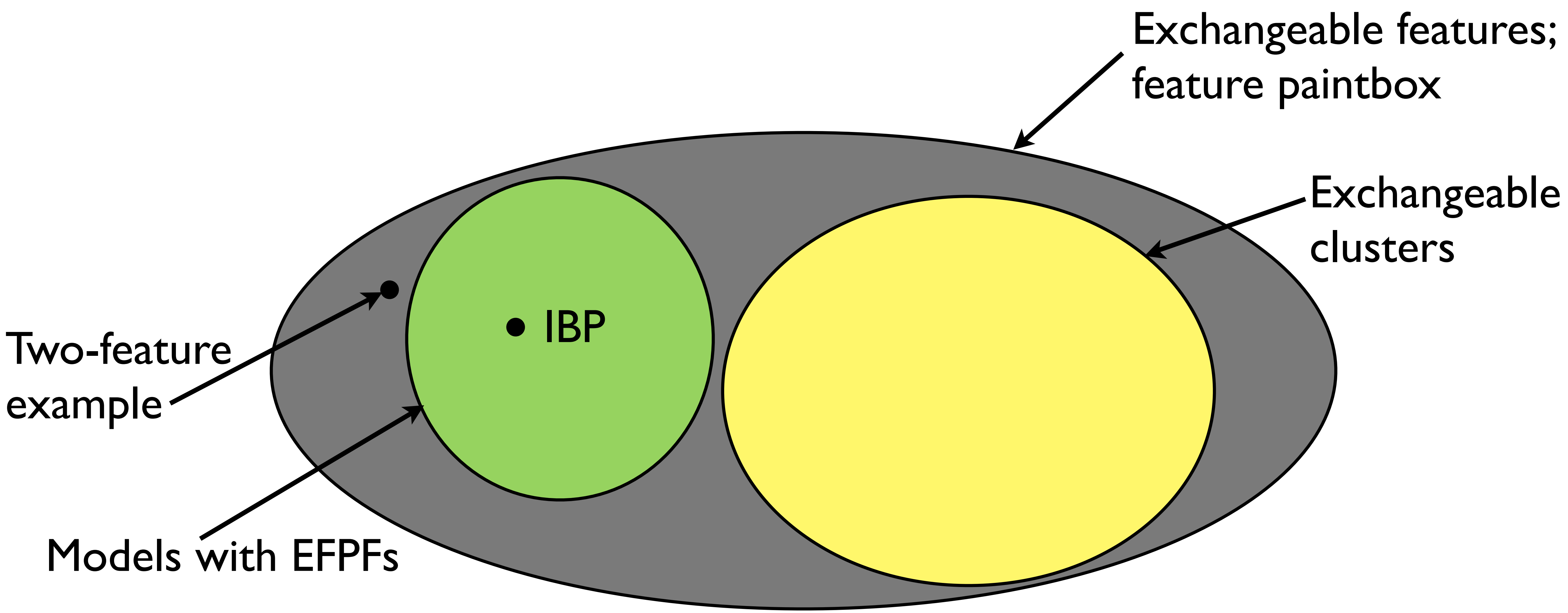
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Limits of clustering characterizations in feature case?



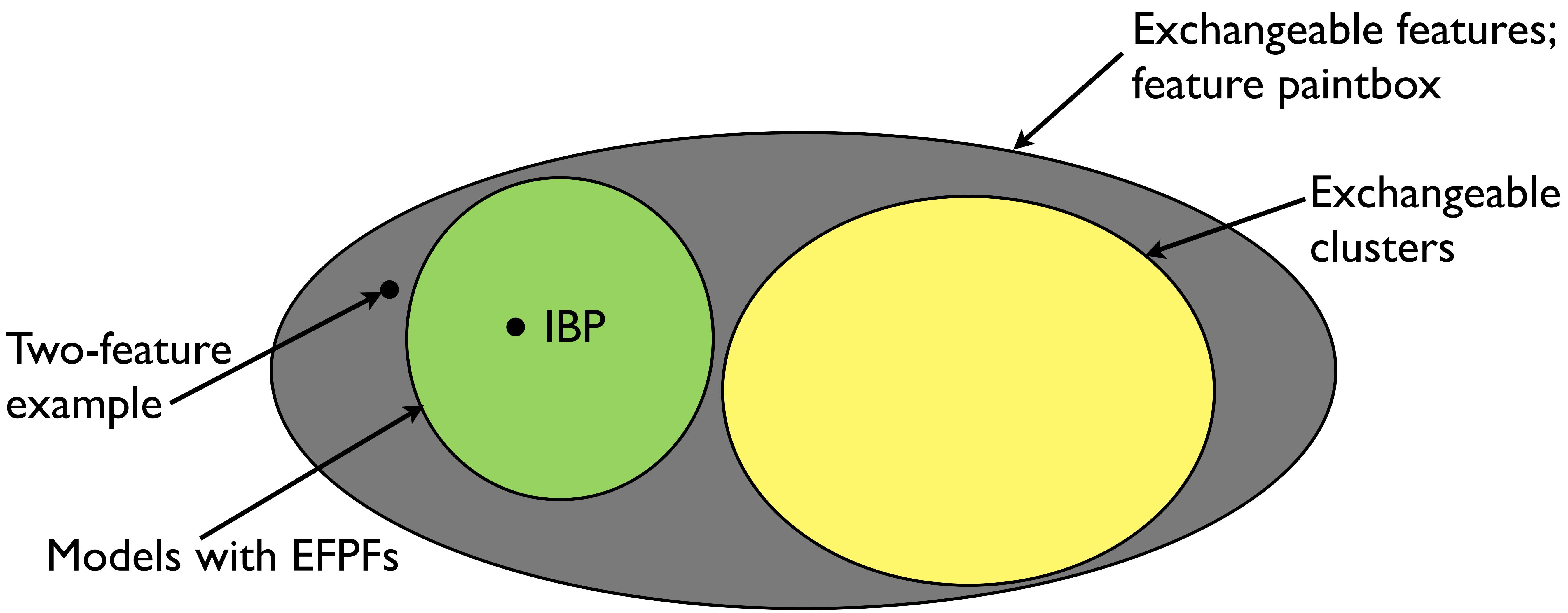
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Limits of clustering characterizations in feature case?



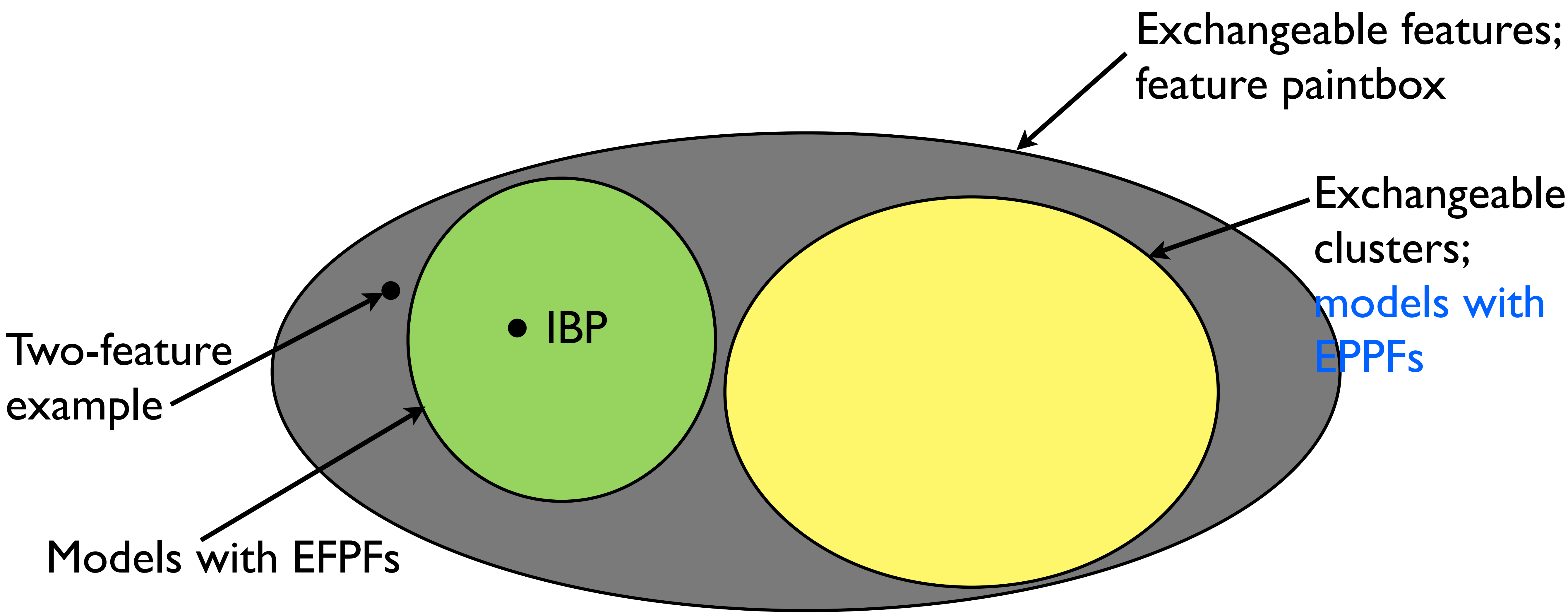
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure



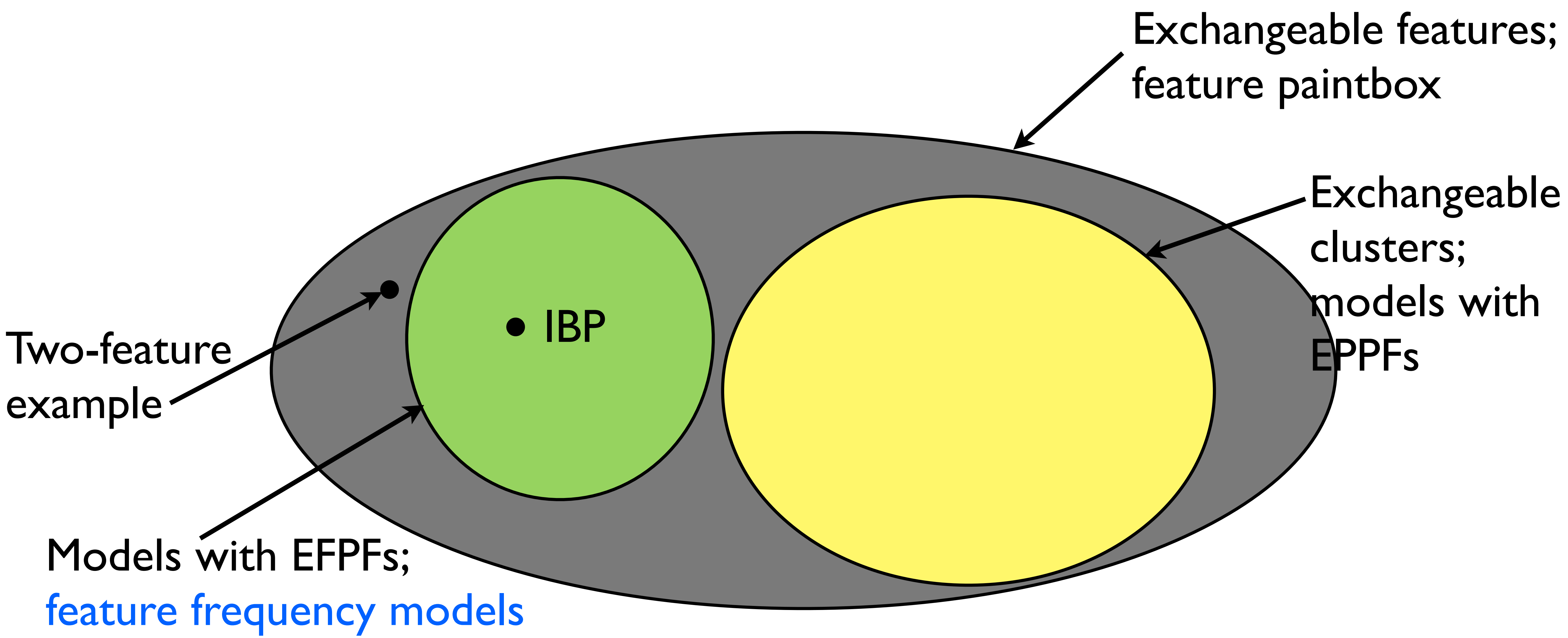
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure



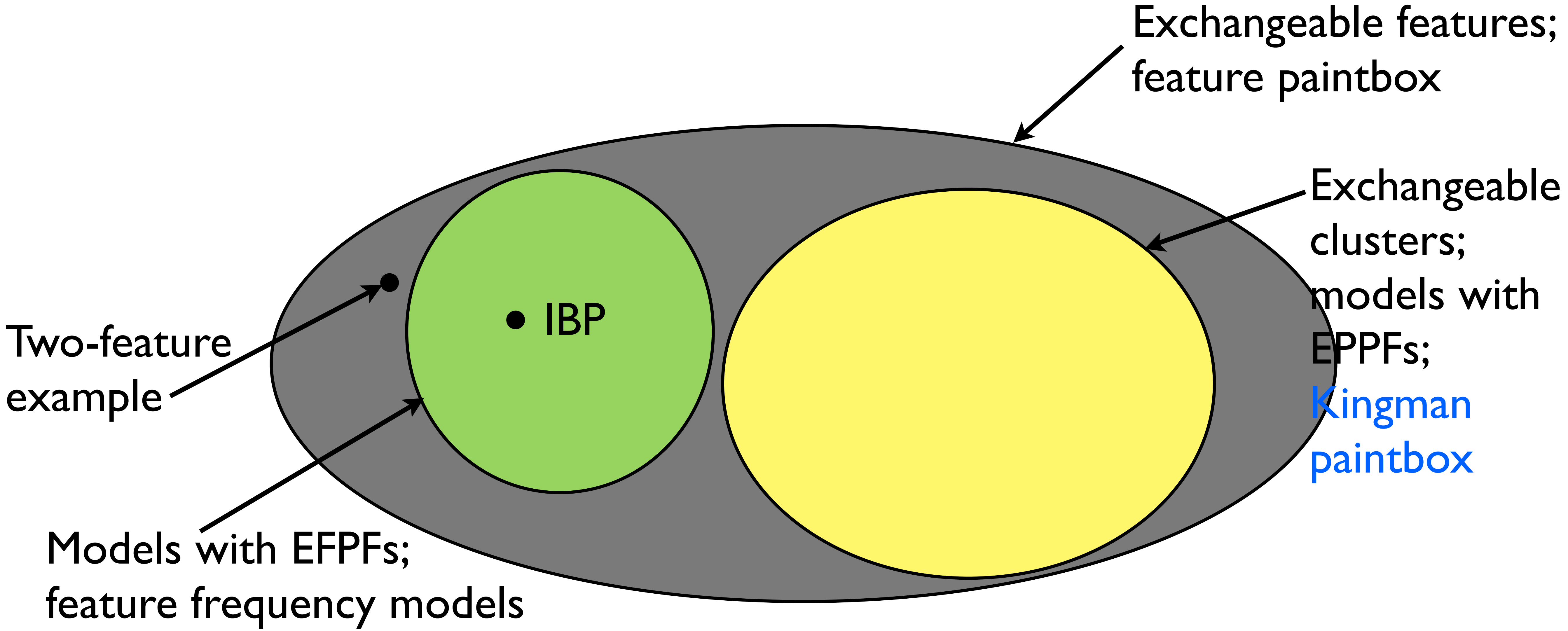
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure



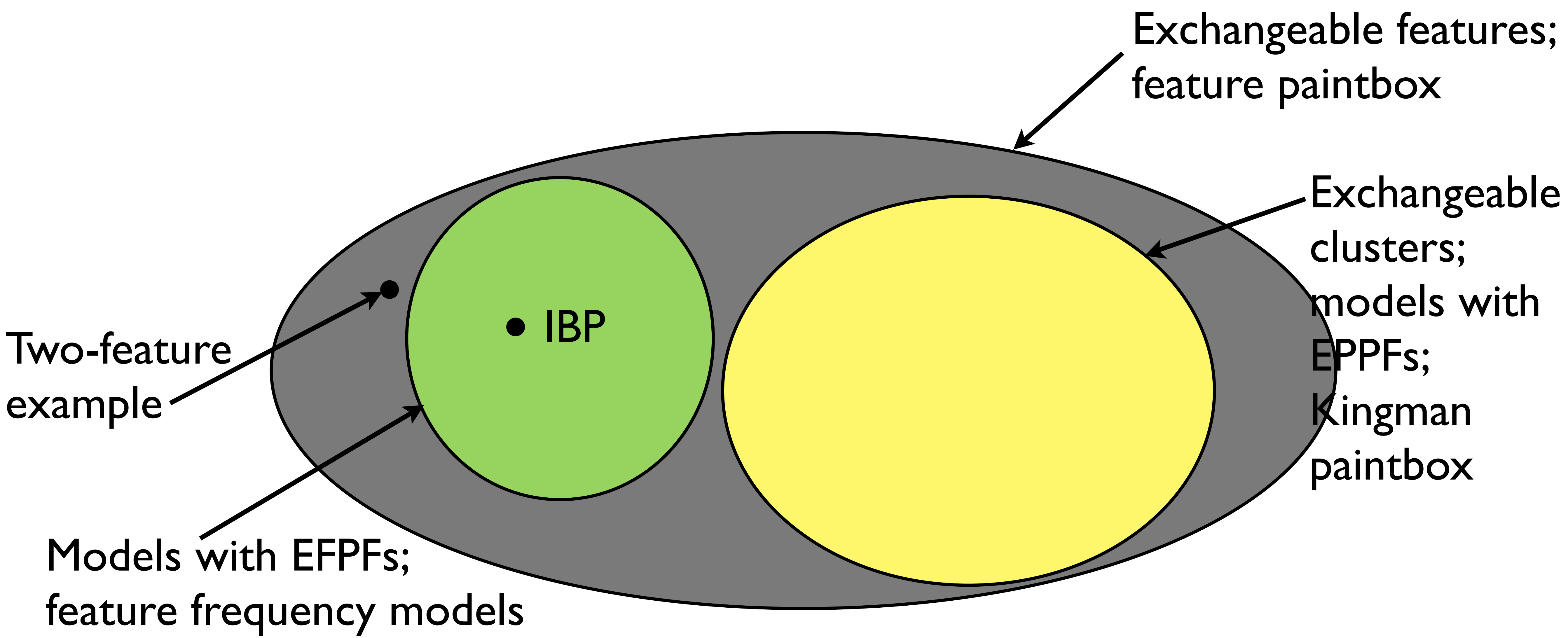
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure



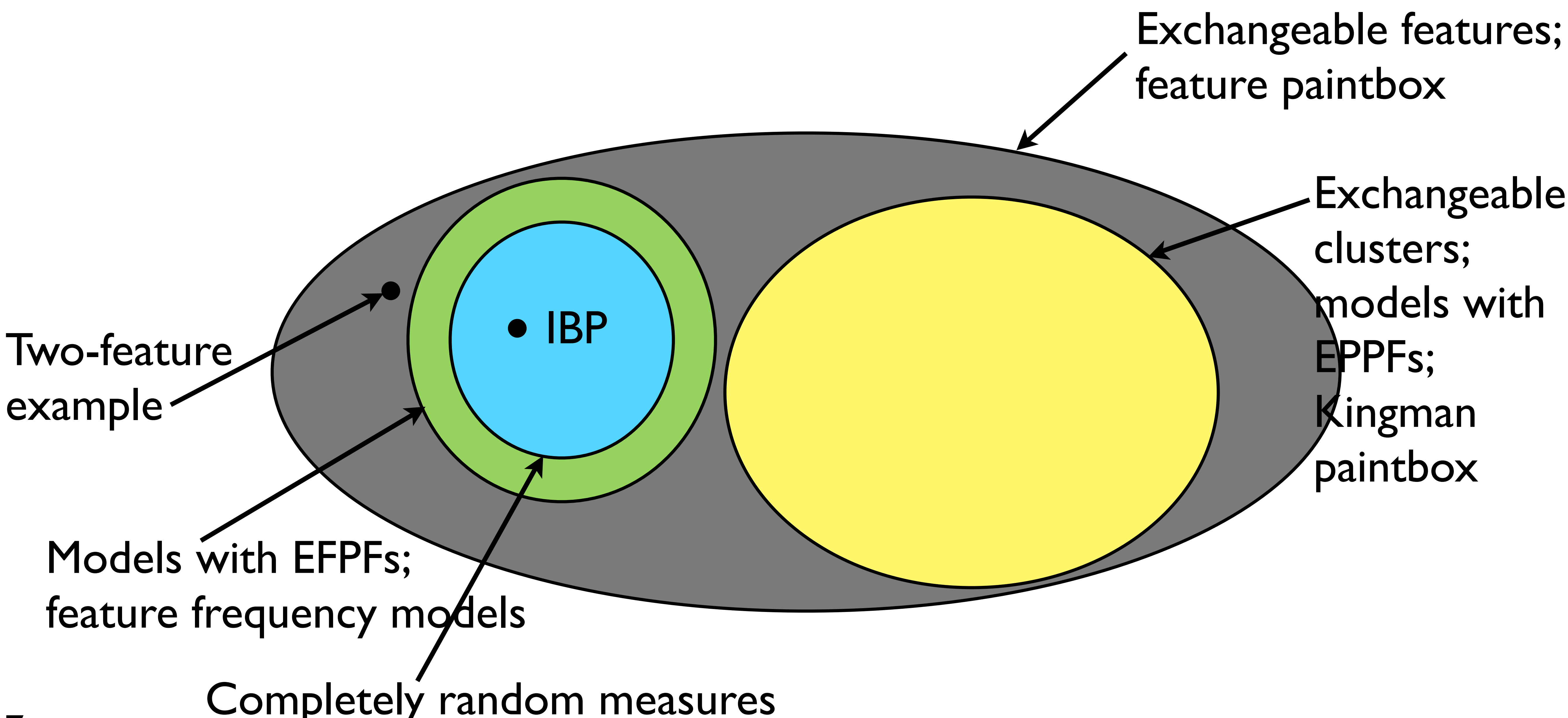
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections



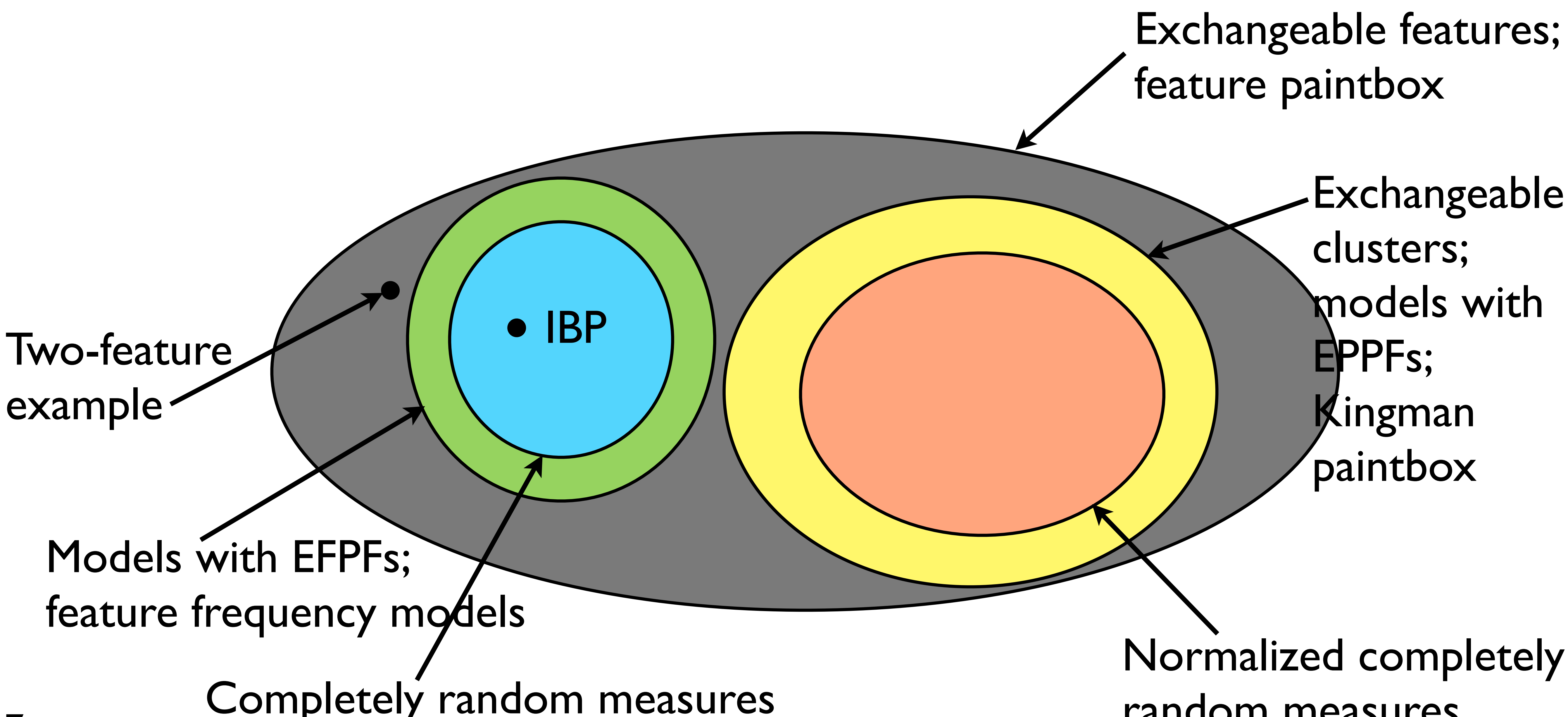
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs)



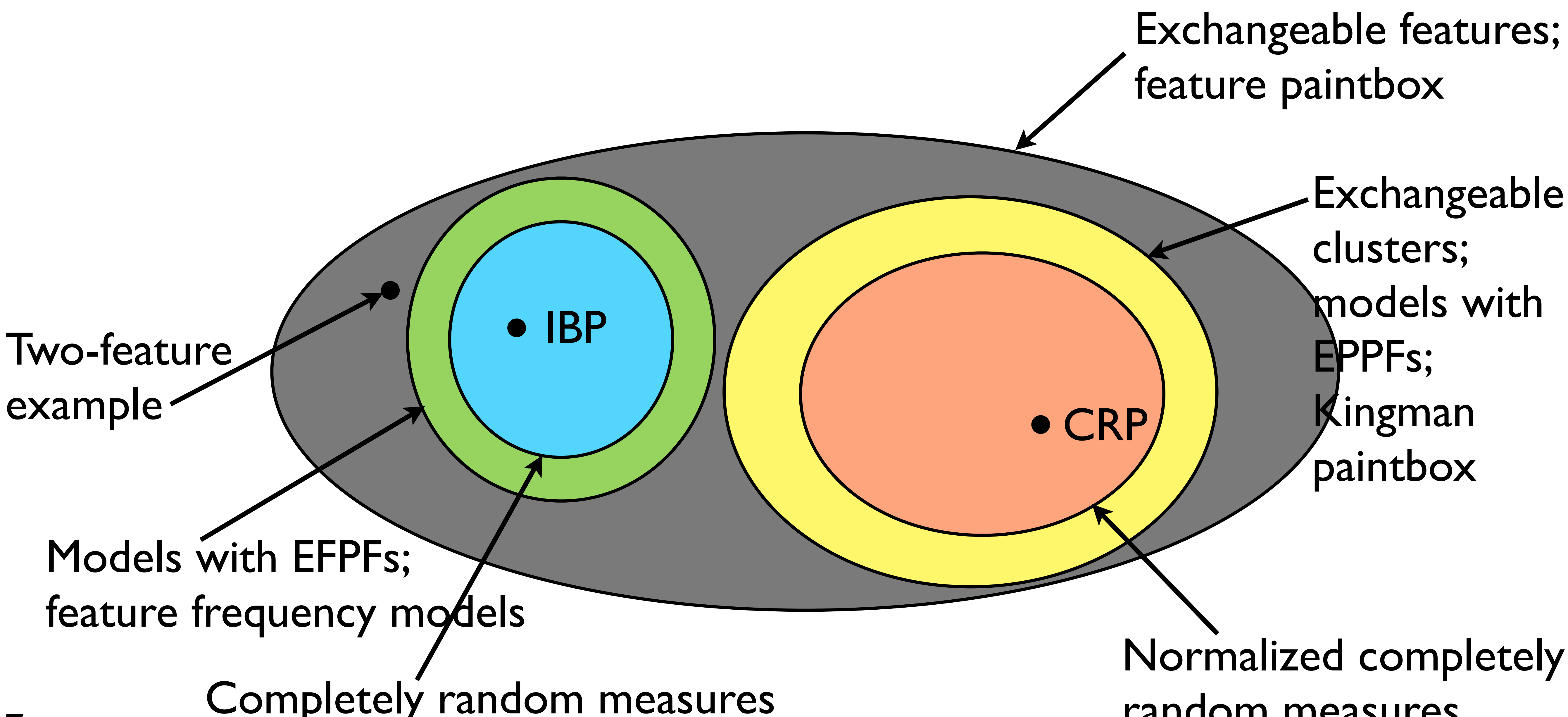
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs)



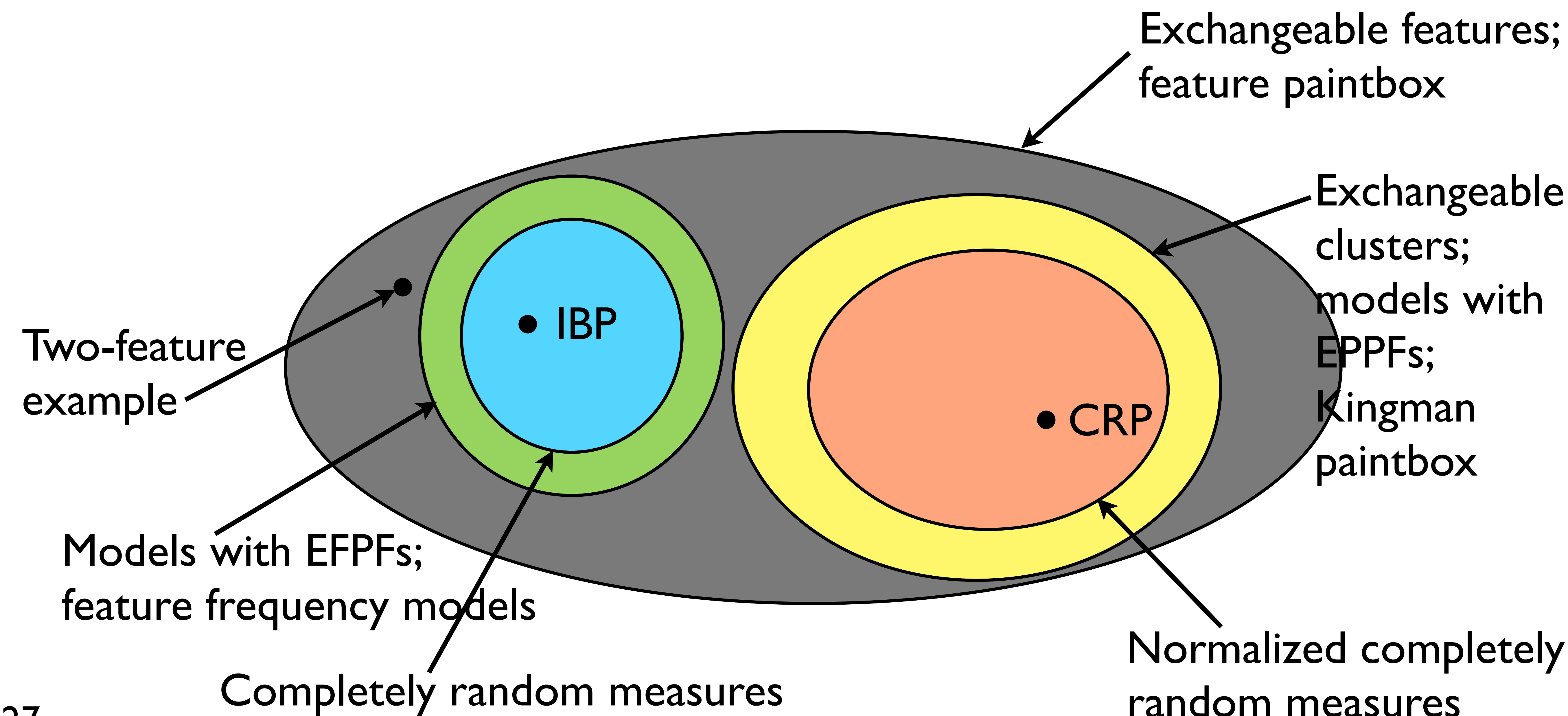
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs)



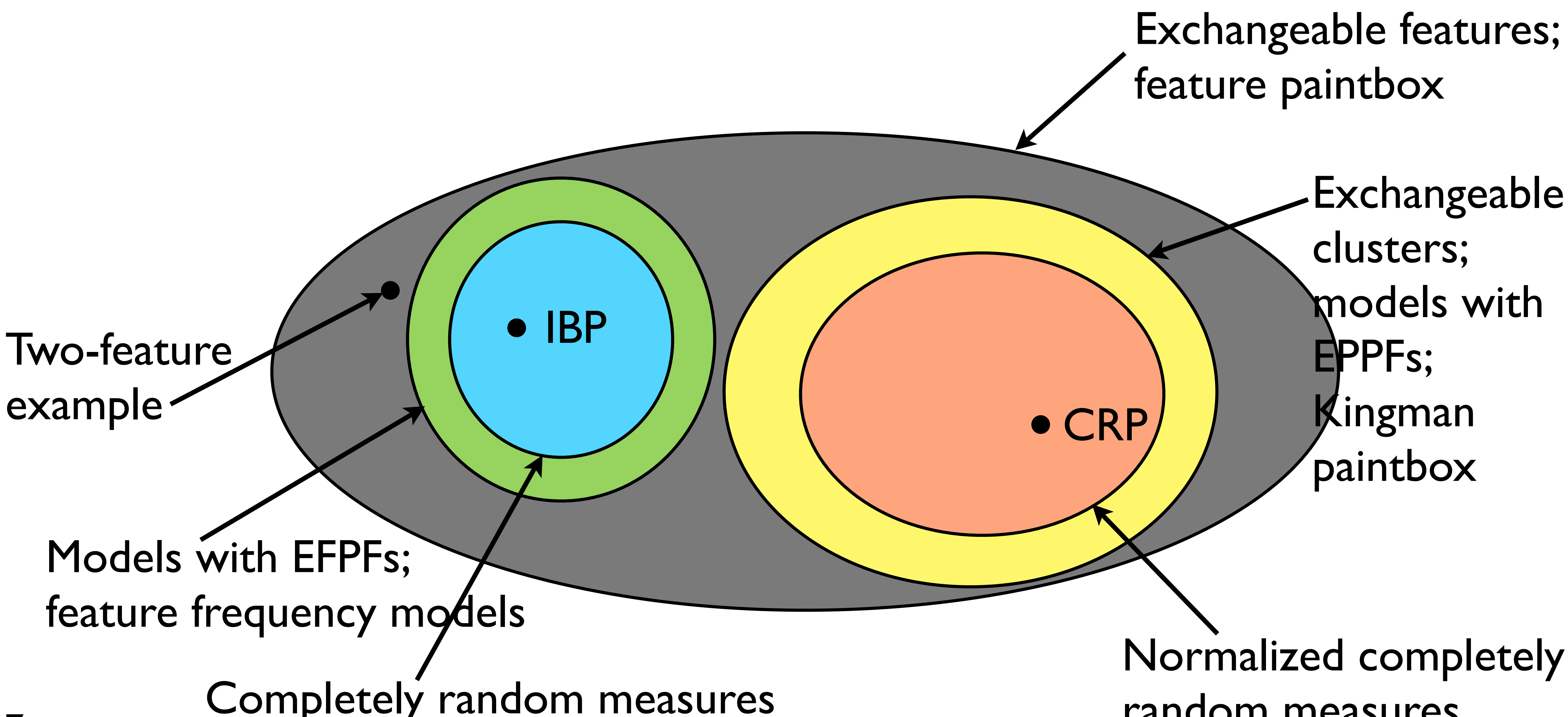
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs, dust)



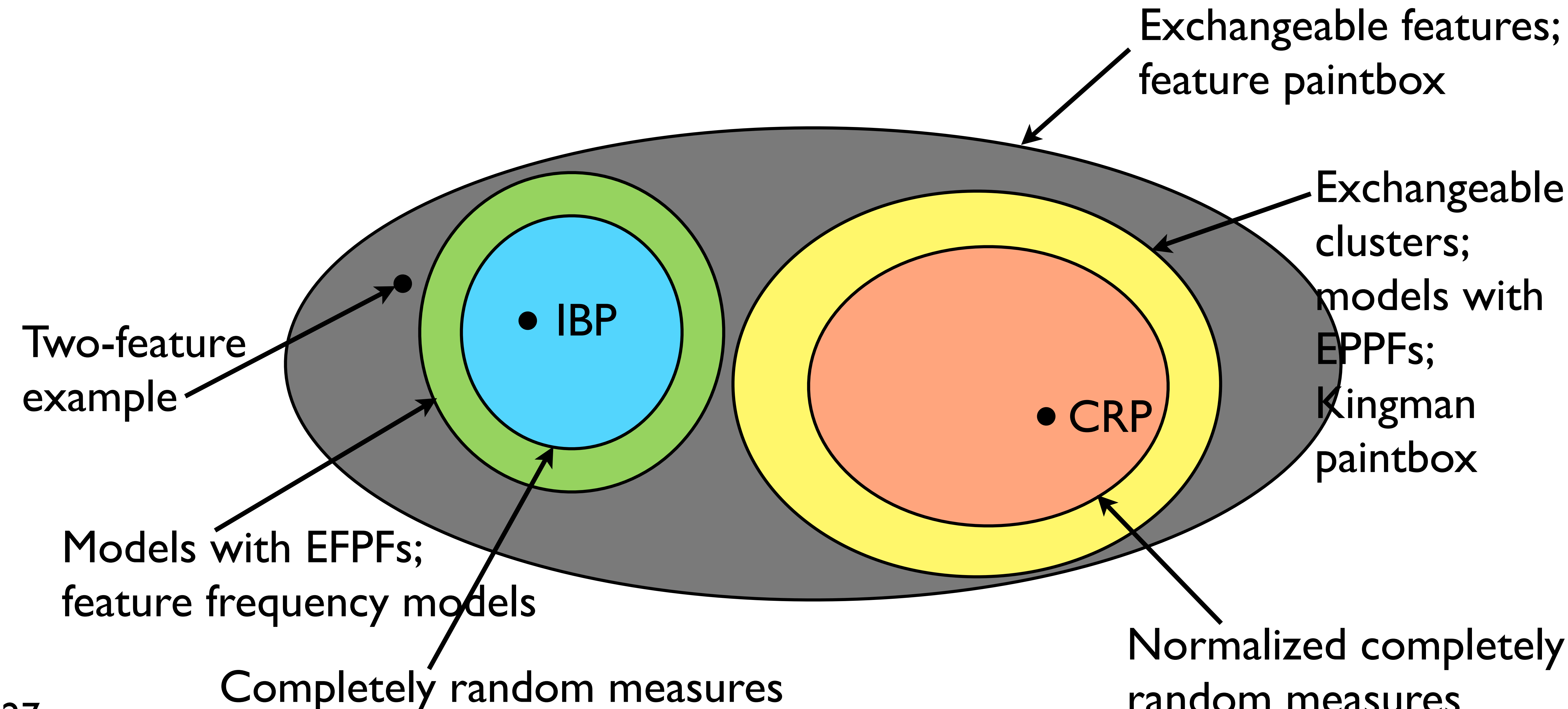
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs, dust, etc)



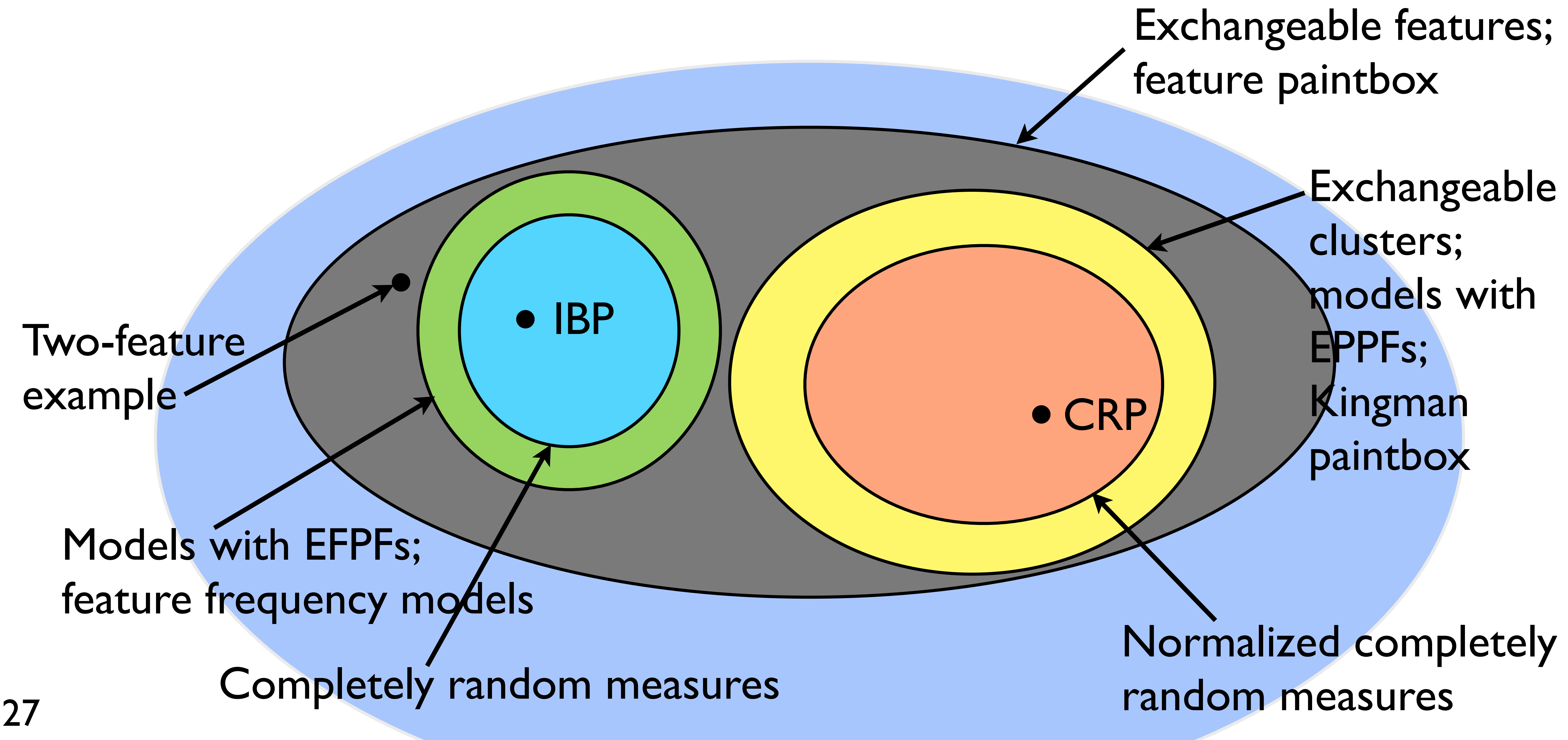
Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs, dust, etc)
- Other combinatorial structures



Theory conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections (CRMs, dust, etc)
- Other combinatorial structures



References

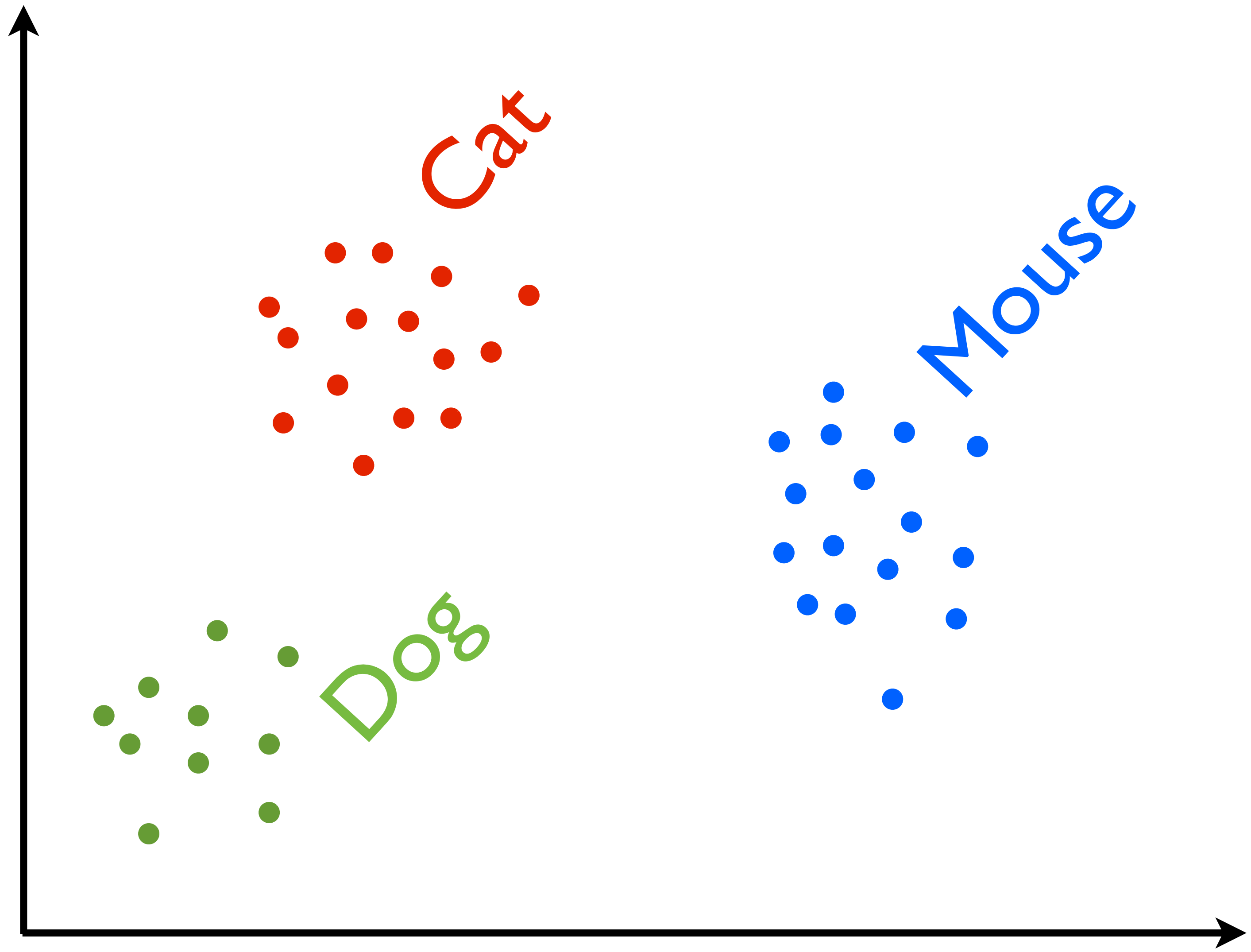
T. Broderick, J. Pitman, and M. I. Jordan. Feature allocations, probability functions, and paintboxes. *Bayesian Analysis*, 8(4):801-836, 2013.

T. Broderick, M. I. Jordan, and J. Pitman. Cluster and feature modeling from combinatorial stochastic processes. *Statistical Science*, 28(3):289-312, 2013.

T. Broderick, L. Mackey, J. Paisley, and M. I. Jordan. Combinatorial clustering and the beta negative binomial process. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

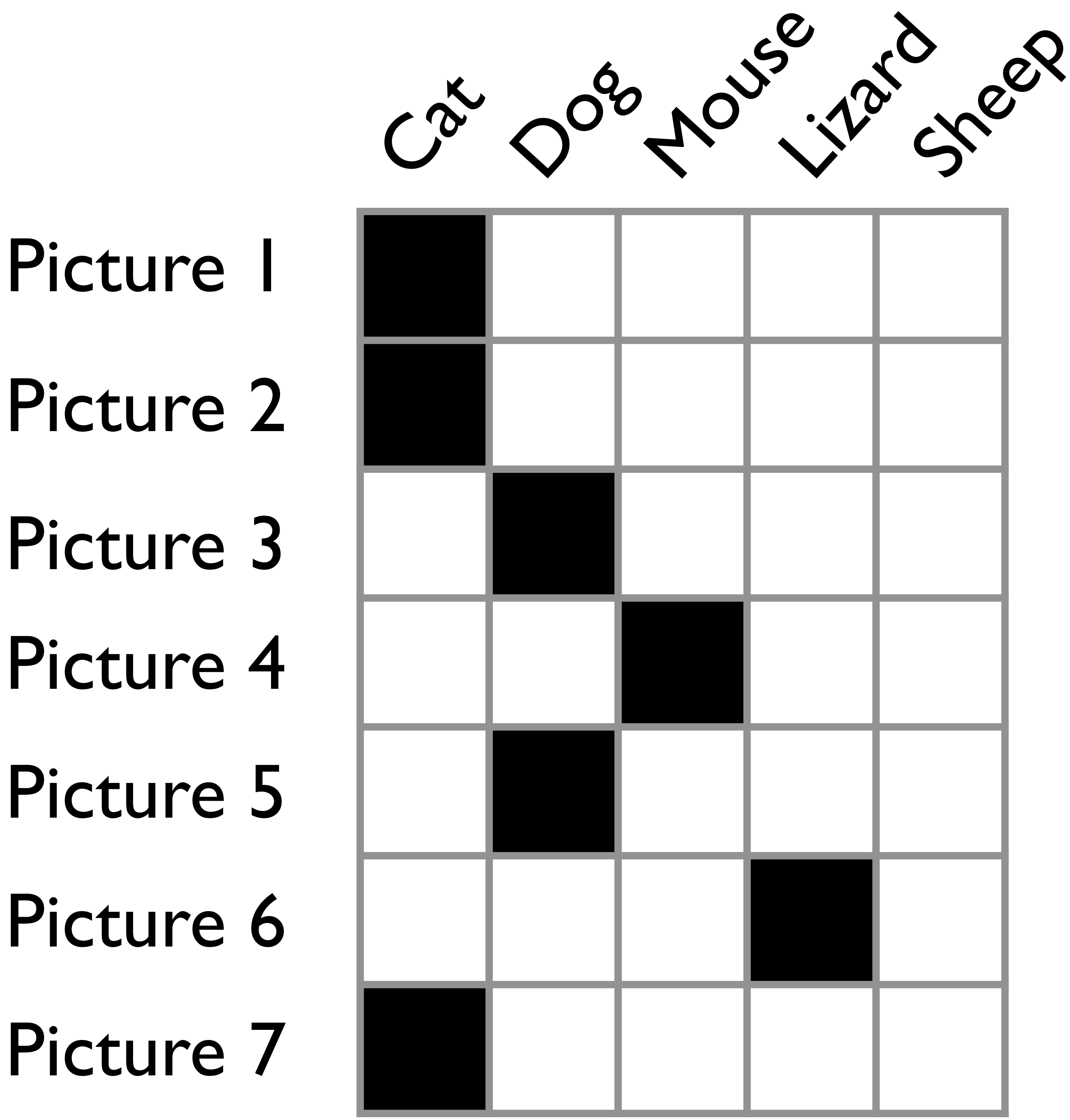
T. Broderick, A. C. Wilson, and M. I. Jordan. Posteriors, conjugacy, and exponential families for completely random measures. Submitted.

Clusters

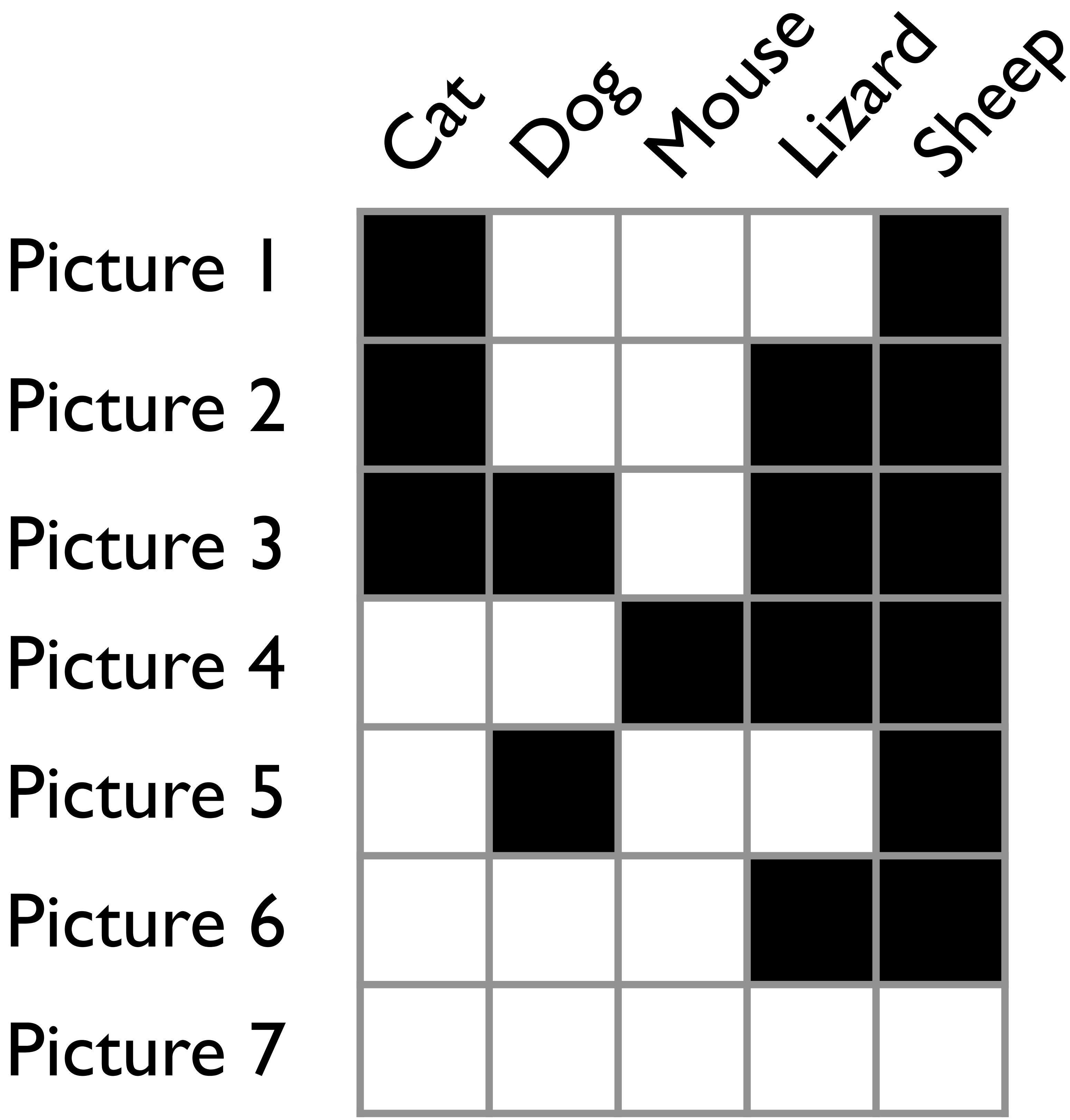


"clusters"

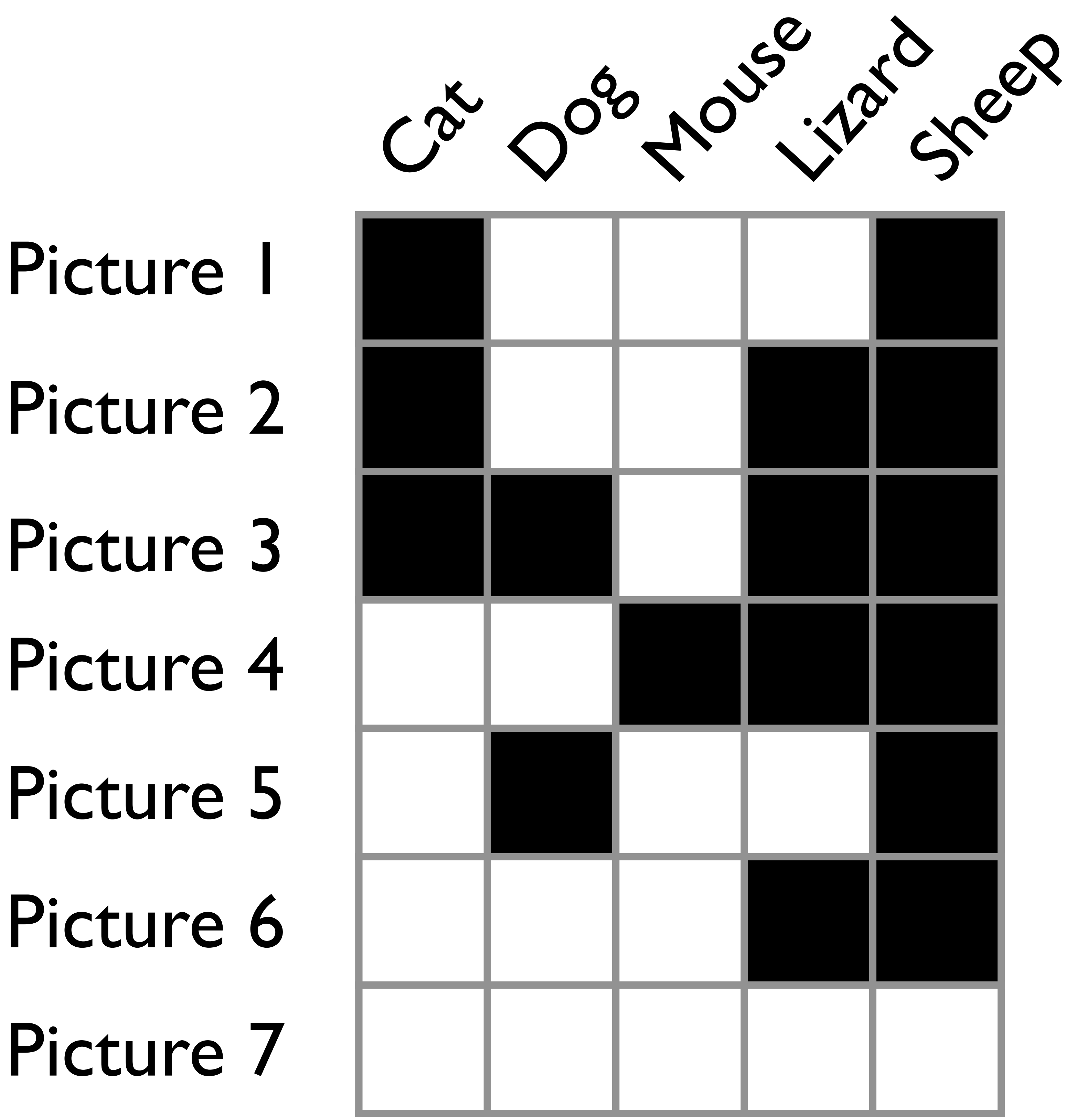
Clusters


Cat
Dog
Mouse
Lizard
Sheep
Picture 1
Picture 2
Picture 3
Picture 4
Picture 5
Picture 6
Picture 7

Features


Cat
Dog
Mouse
Lizard
Sheep
Picture 1
Picture 2
Picture 3
Picture 4
Picture 5
Picture 6
Picture 7

Features



Many other possible latent structures in data

How do we learn latent structure?

How do we learn latent structure?

K-means

How do we learn latent structure?

K-means

- Fast

How do we learn latent structure?

K-means

- Fast
- Can parallelize

How do we learn latent structure?

K-means

- Fast
- Can parallelize
- Straightforward

How do we learn latent structure?

K-means

- Fast
- Can parallelize
- Straightforward
- Only works for K
clusters

How do we learn latent structure?

K-means

- Fast
- Can parallelize
- Straightforward
- Only works for K clusters

Nonparametric Bayes

How do we learn latent structure?

K-means

- Fast
- Can parallelize
- Straightforward
- Only works for K clusters

Nonparametric Bayes

- Modular (general latent structure)

How do we learn latent structure?

K-means

- Fast
- Can parallelize
- Straightforward
- Only works for K clusters

Nonparametric Bayes

- Modular (general latent structure)
- Flexible (K can grow as data grows)

How do we learn latent structure?

K-means

- Fast
- Can parallelize
- Straightforward
- Only works for K clusters

Nonparametric Bayes

- Modular (general latent structure)
- Flexible (K can grow as data grows)
- Coherent treatment of uncertainty

How do we learn latent structure?

K-means
- Fast
- Can parallelize
- Straightforward
- Only works for K clusters

Nonparametric Bayes
- Modular (general latent structure)
- Flexible (K can grow as data grows)
- Coherent treatment of uncertainty

But...
- E.g., Silicon Valley: can have petabytes of data
- Practitioners turn to what runs

MAD-Bayes Perspectives

- Bayesian nonparametrics assists the optimization-based inference community

MAD-Bayes Perspectives

- Bayesian nonparametrics assists the optimization-based inference community
  - New, modular, flexible, nonparametric objectives & regularizers

MAD-Bayes Perspectives

- Bayesian nonparametrics assists the optimization-based inference community
  - New, modular, flexible, nonparametric objectives & regularizers
  - Alternative perspective: fast initialization

MAD-Bayes Perspectives

- Bayesian nonparametrics assists the optimization-based inference community
  - New, modular, flexible, nonparametric objectives & regularizers
  - Alternative perspective: fast initialization

Inspiration

- Consider a finite Gaussian mixture model

MAD-Bayes Perspectives

- Bayesian nonparametrics assists the optimization-based inference community
  - New, modular, flexible, nonparametric objectives & regularizers
  - Alternative perspective: fast initialization

Inspiration

- Consider a finite Gaussian mixture model
- The steps of the EM algorithm limit to the steps of the K-means algorithm as the Gaussian variance is taken to 0

MAD-Bayes

The MAD-Bayes idea

- Start with nonparametric Bayes model
- Take a similar limit to get a K-means-like objective

MAD-Bayes

The MAD-Bayes idea

- Start with nonparametric Bayes model
- Take a similar limit to get a K-means-like objective

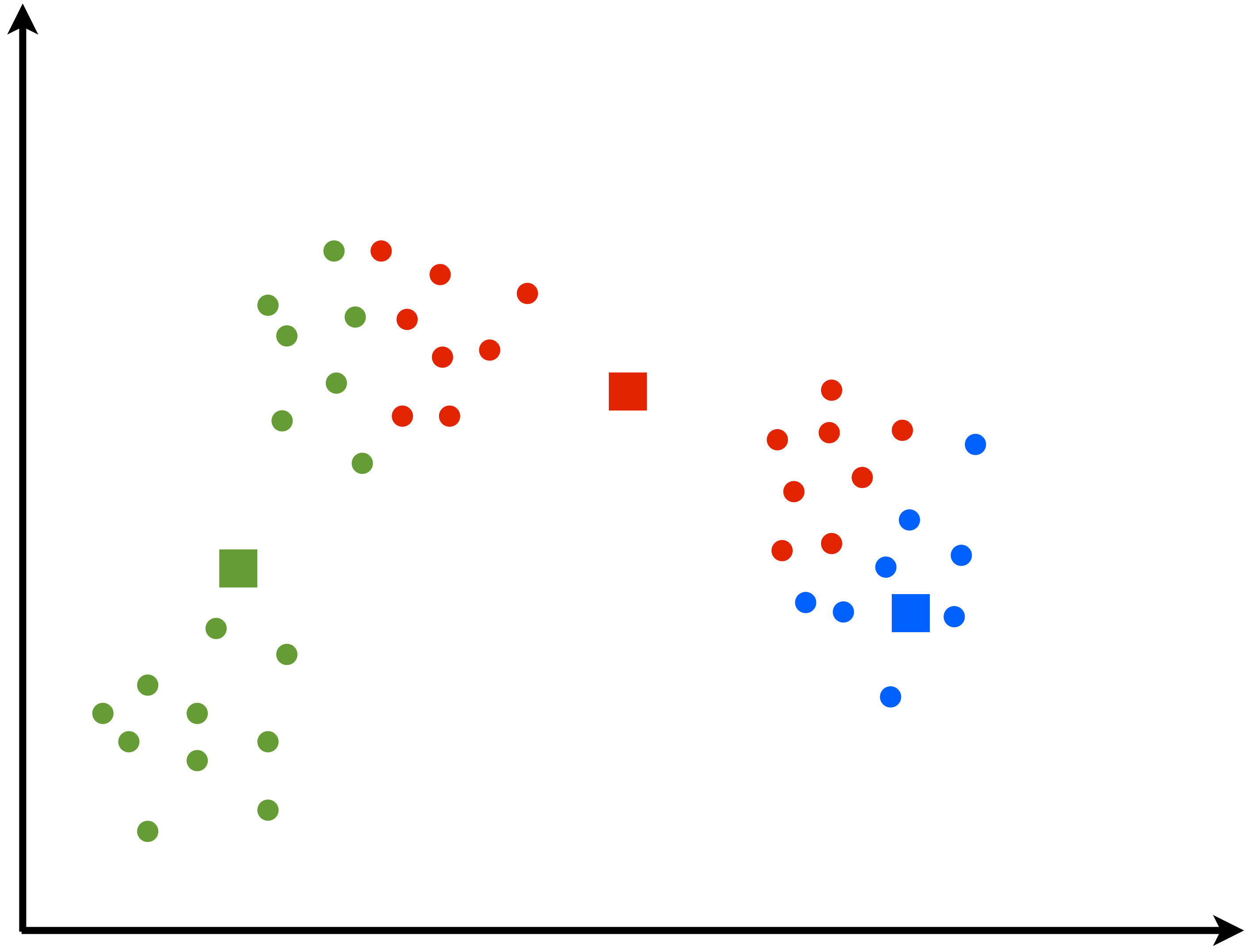
MAD-Bayes

The MAD-Bayes idea

- Start with nonparametric Bayes model
- Take a similar limit to get a K-means-like objective

K-means

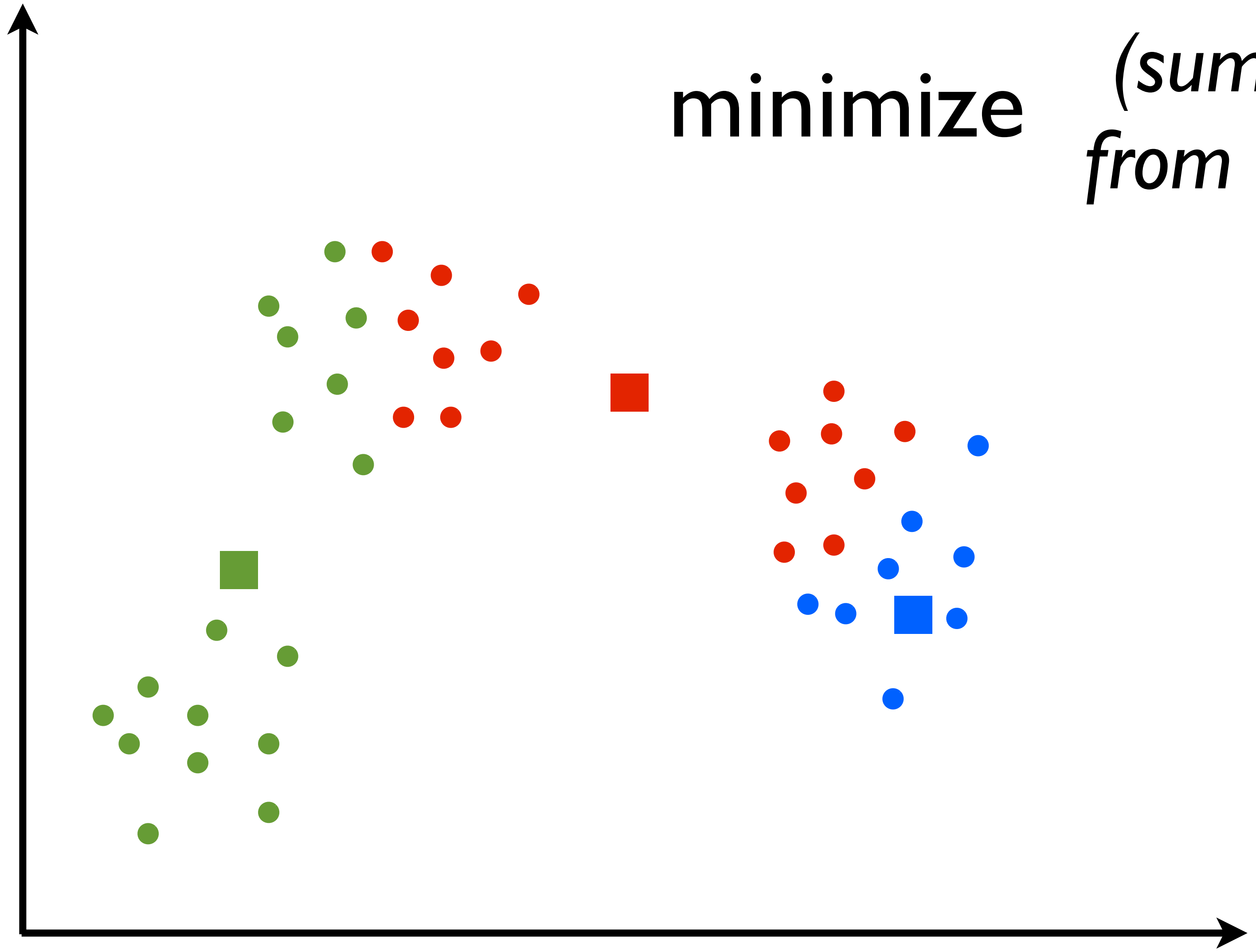
K-means clustering problem



K-means

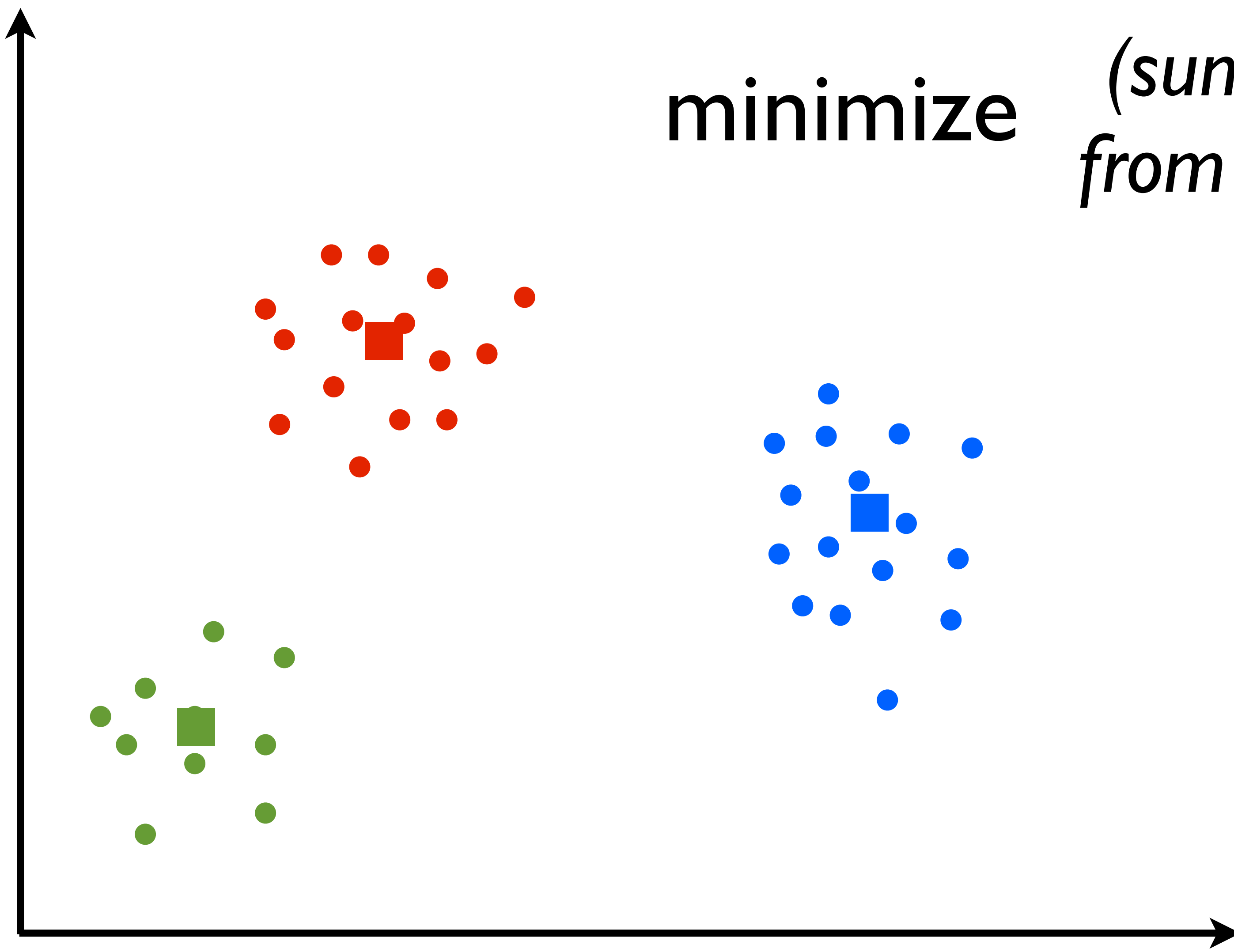
K-means clustering problem

minimize *(sum of square distances from data points to cluster centers)*



K-means

K-means clustering problem

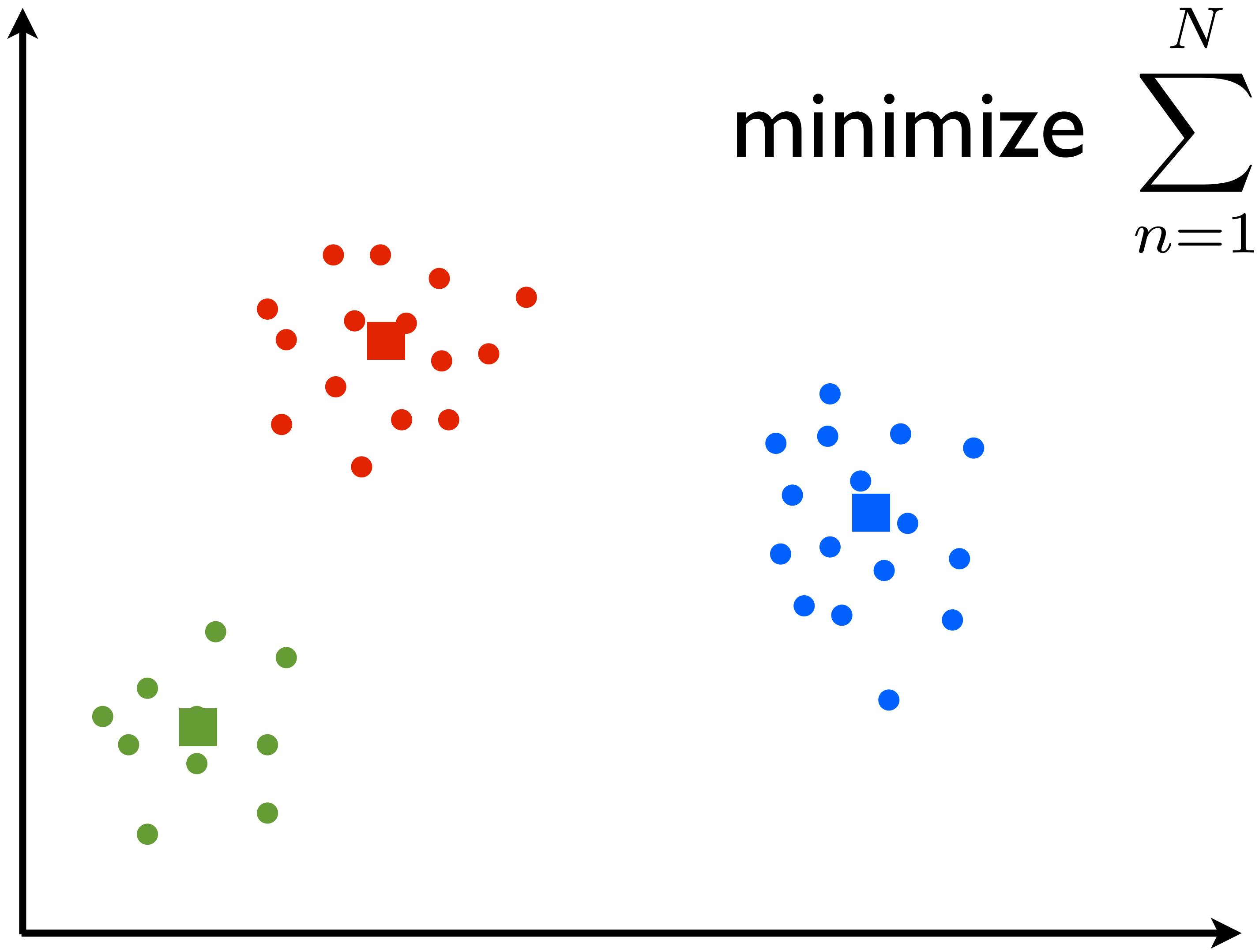


minimize *(sum of square distances from data points to cluster centers)*

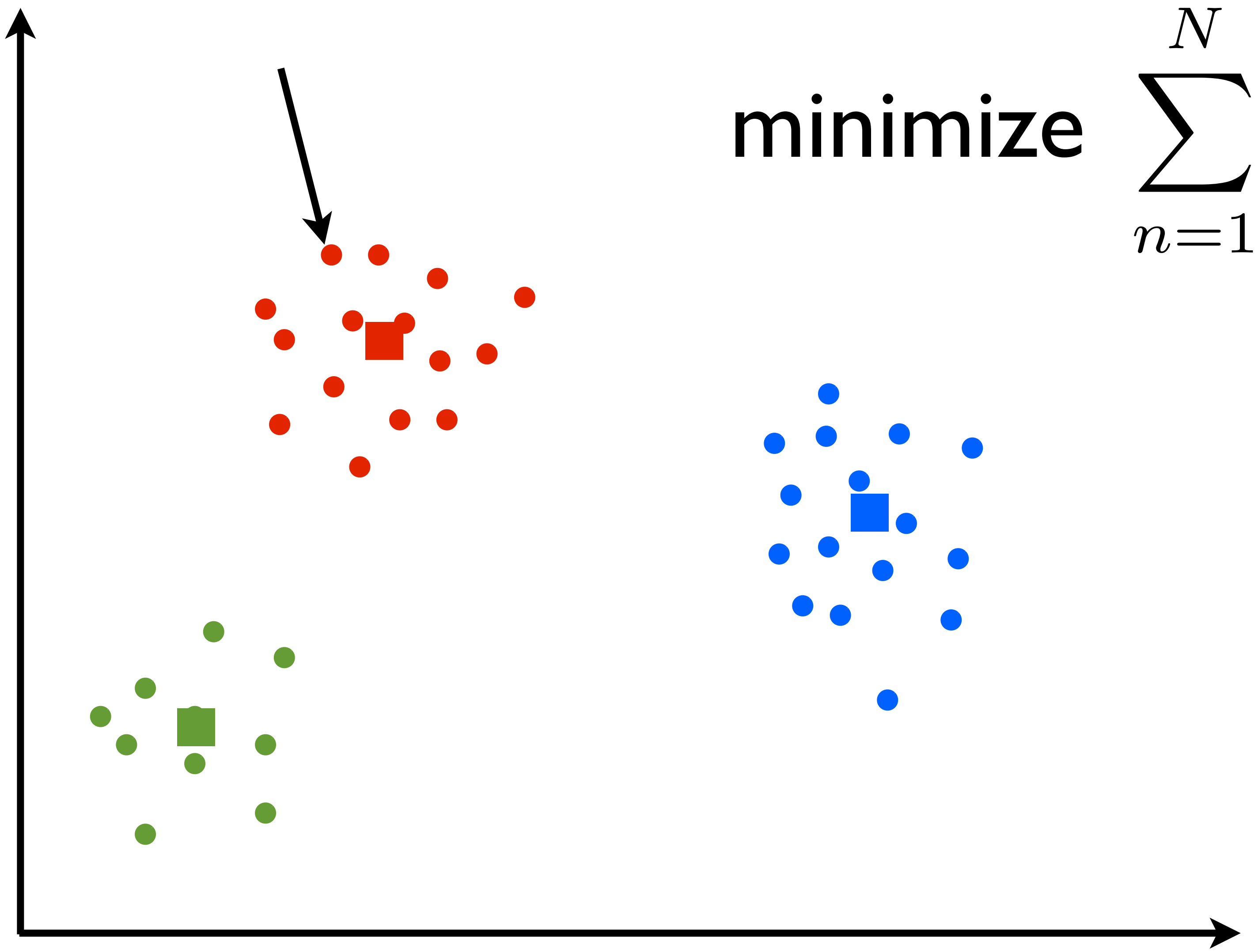
K-means

K-means clustering problem

minimize $\sum_{n=1}^{N} \|x_n - \text{center}_n\|^2$



K-means

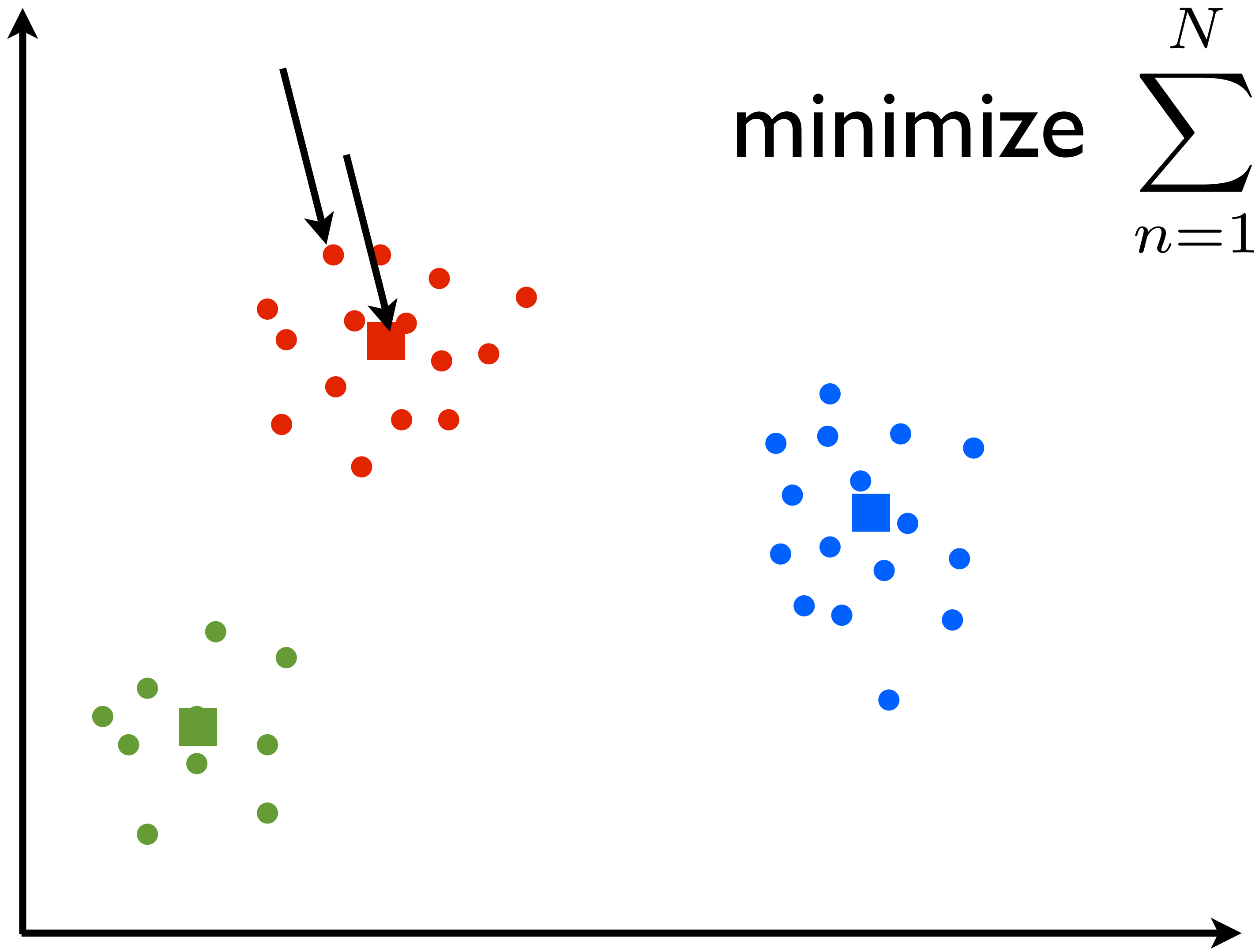


K-means clustering problem

$$\text{minimize} \sum_{n=1}^{N} \|x_n - \text{center}_n\|^2$$

K-means

K-means clustering problem

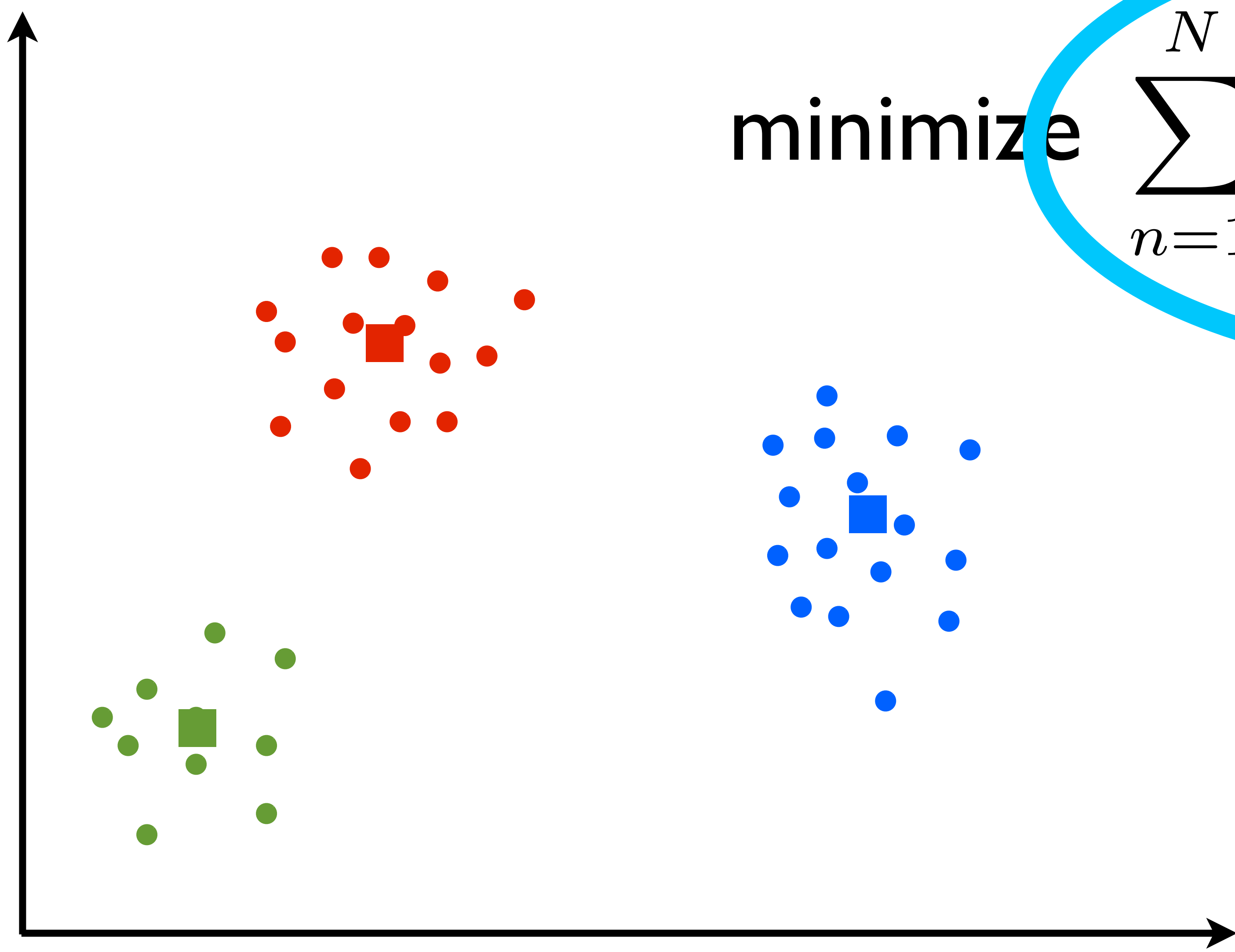


minimize $\sum_{n=1}^{N} \|x_n - \text{center}_n\|^2$

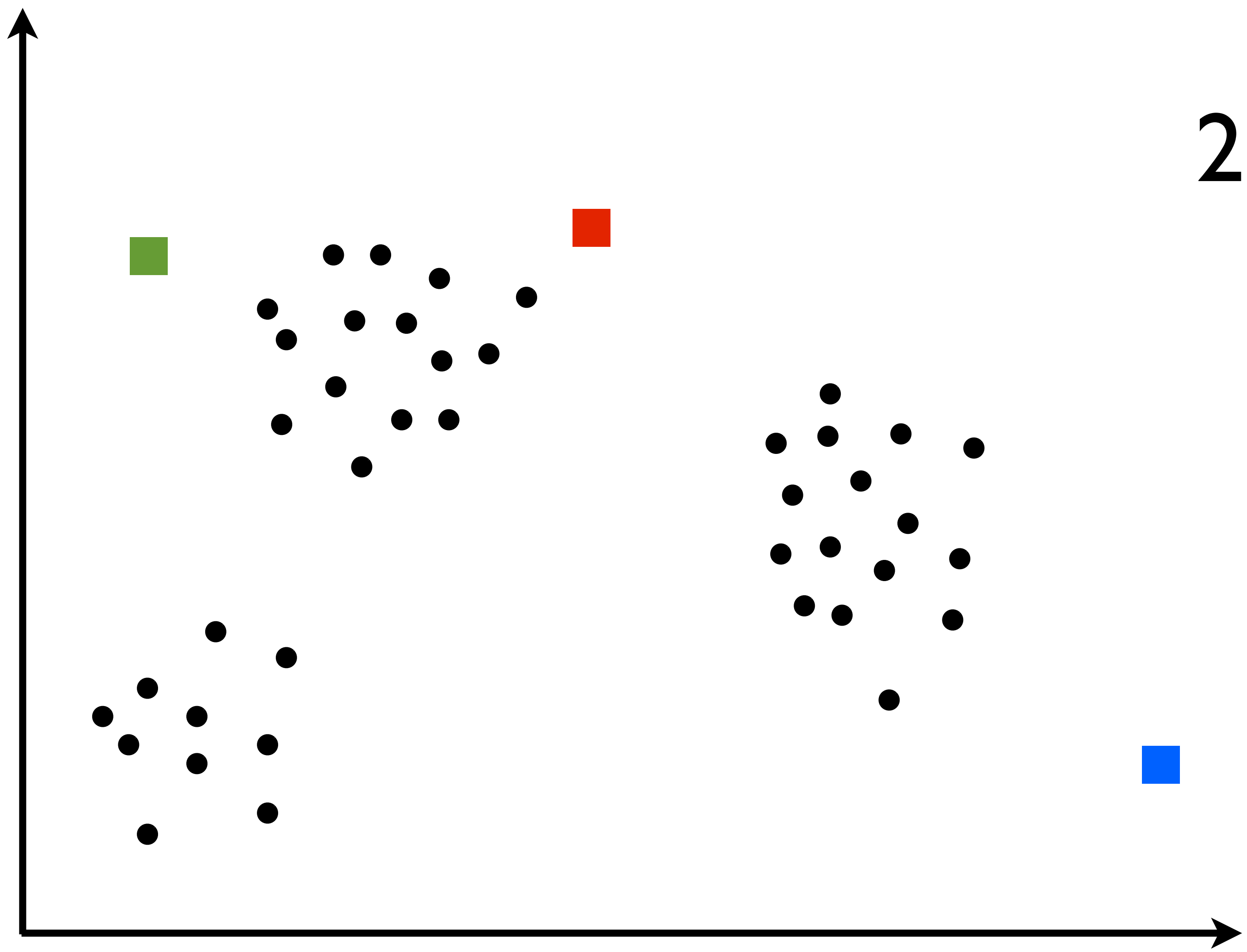
K-means

K-means objective

minimize $\sum_{n=1}^{N} \|x_n - \text{center}_n\|^2$



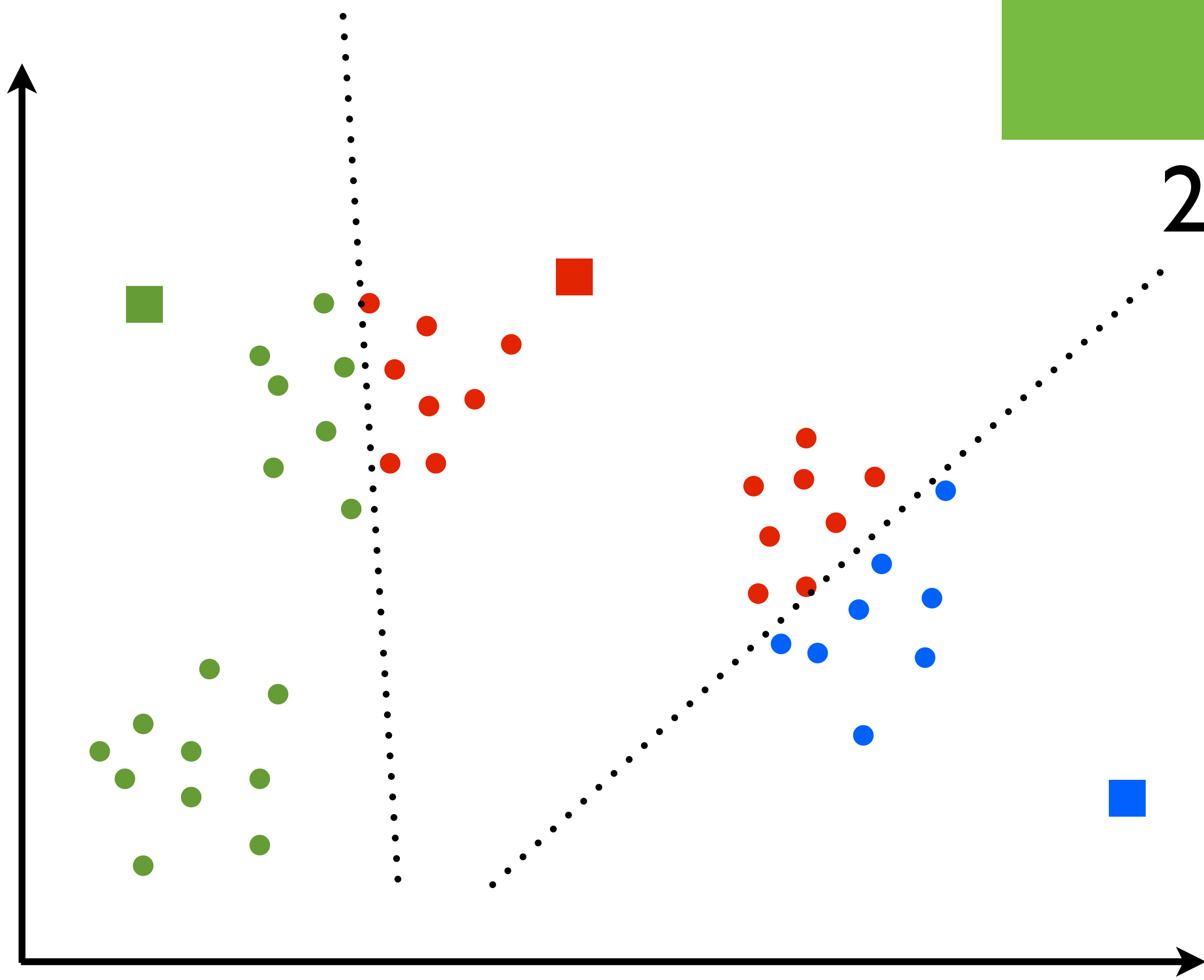
Lloyd's algorithm



Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means

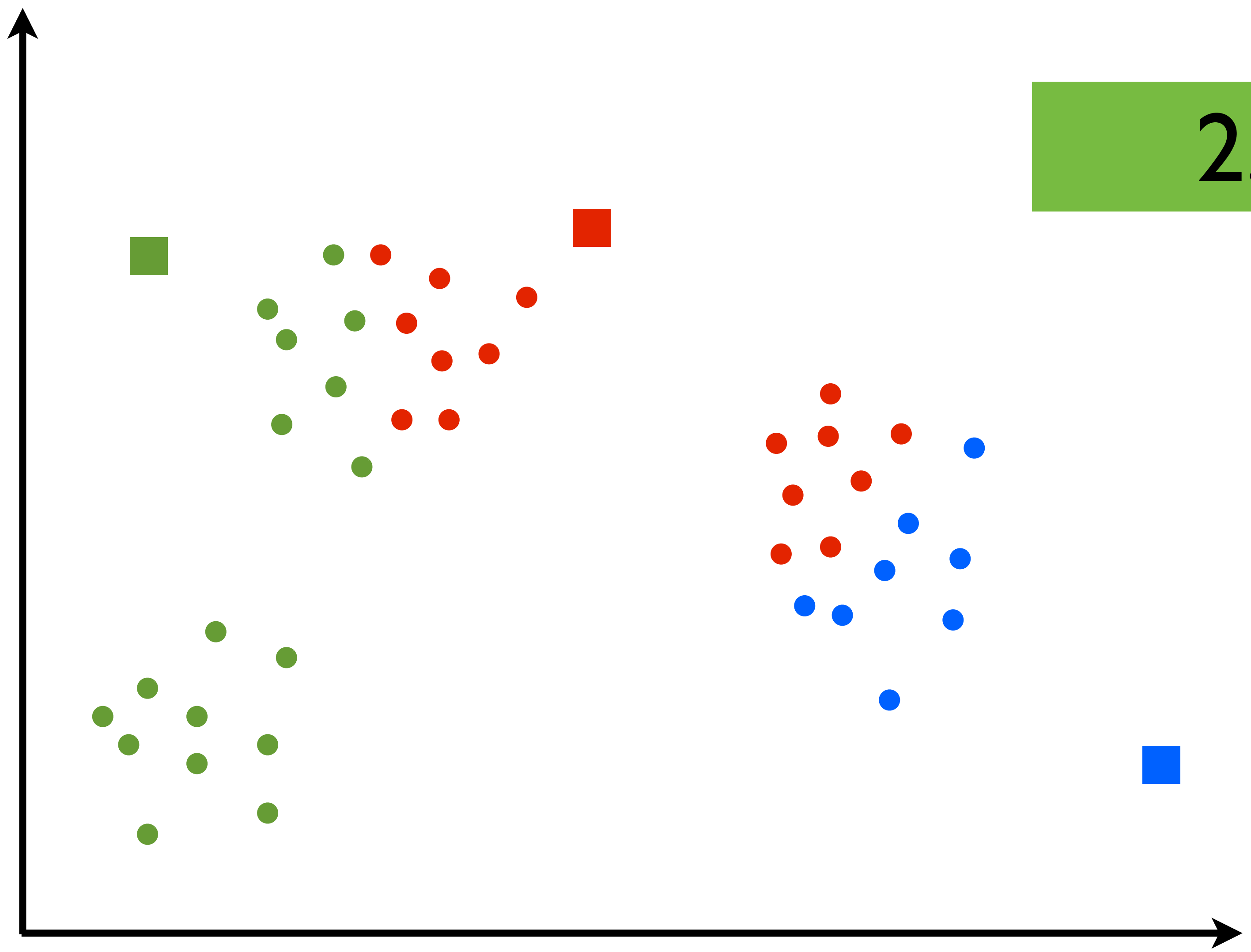
Lloyd's algorithm



Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means

Lloyd's algorithm



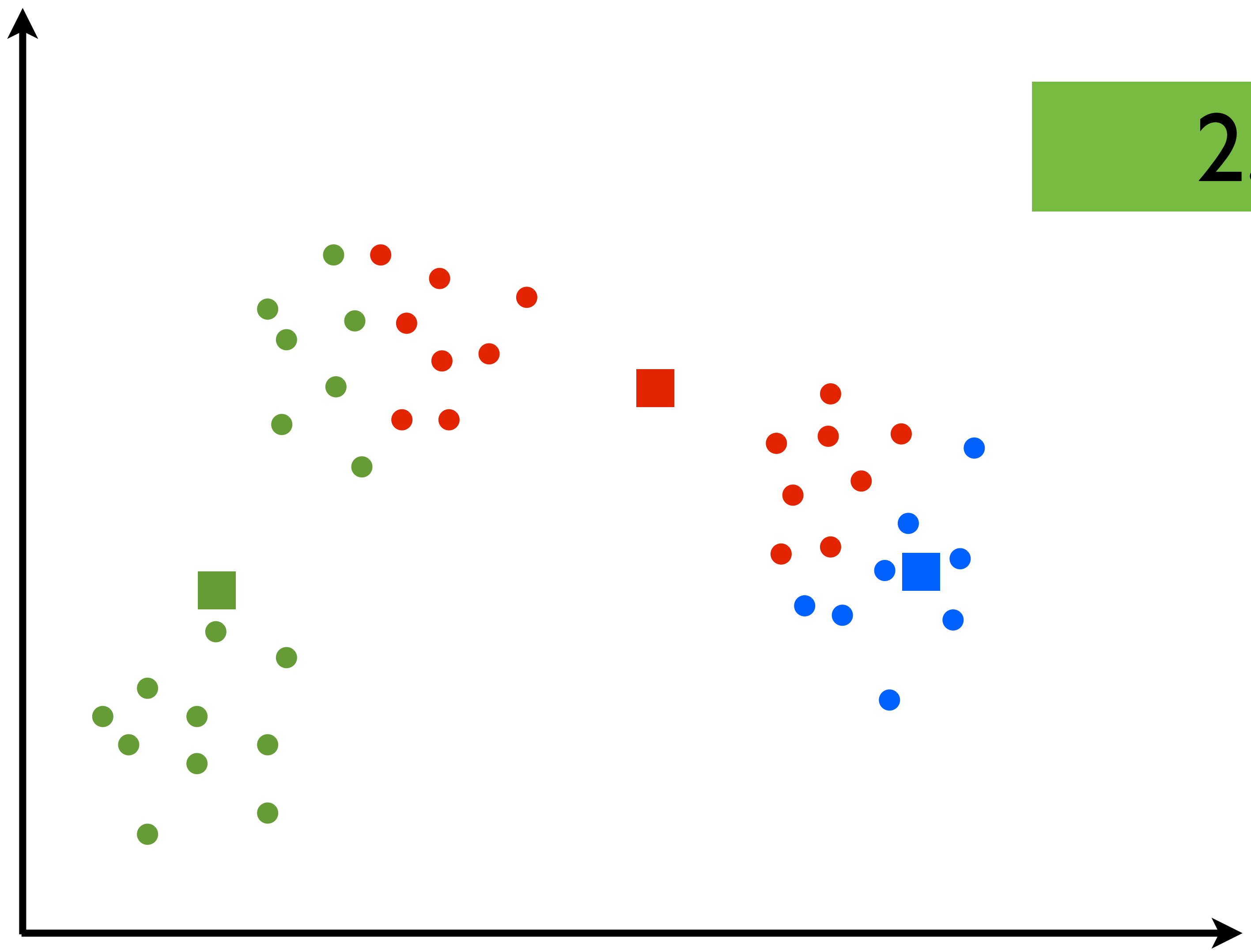
Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means

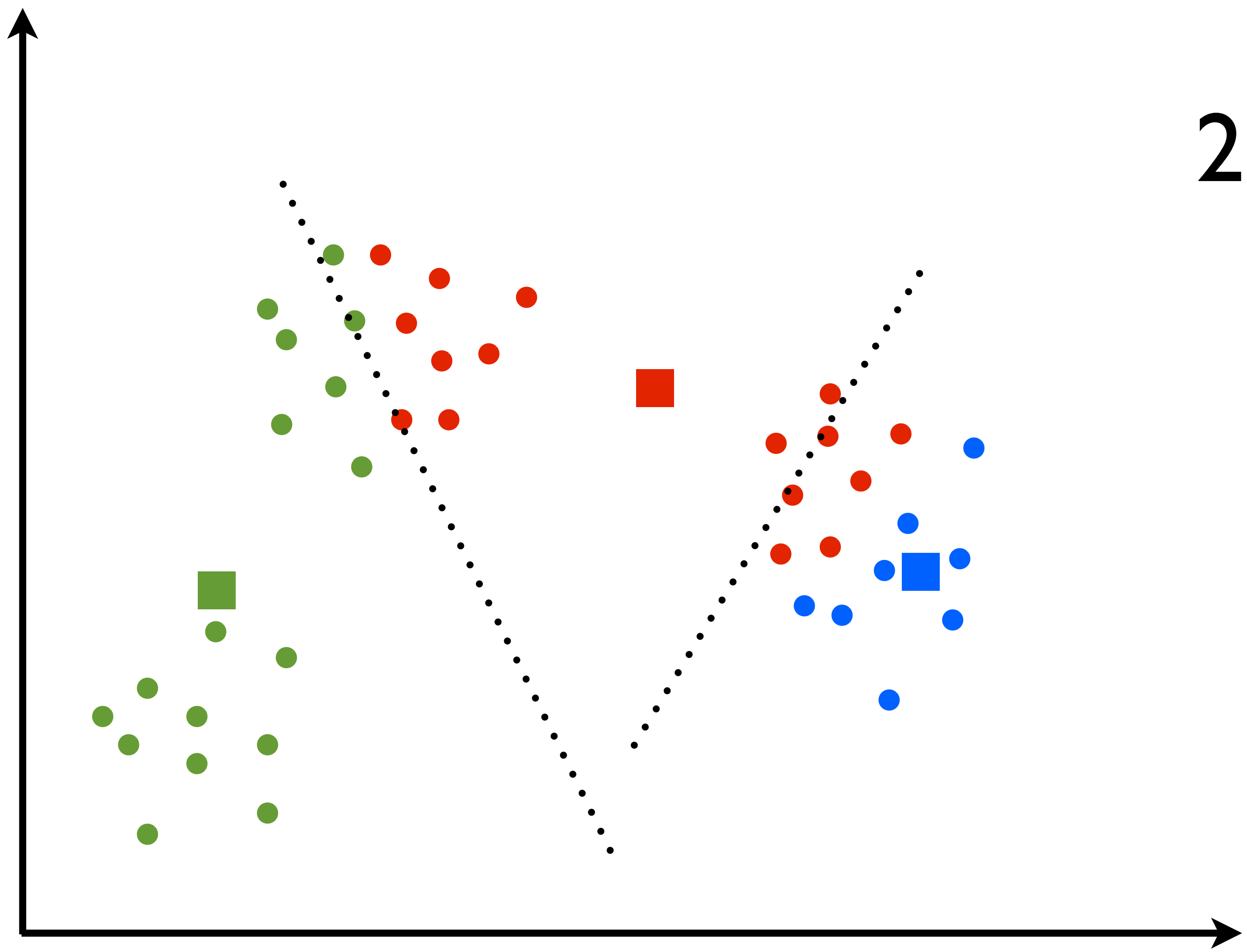
Lloyd's algorithm

Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means



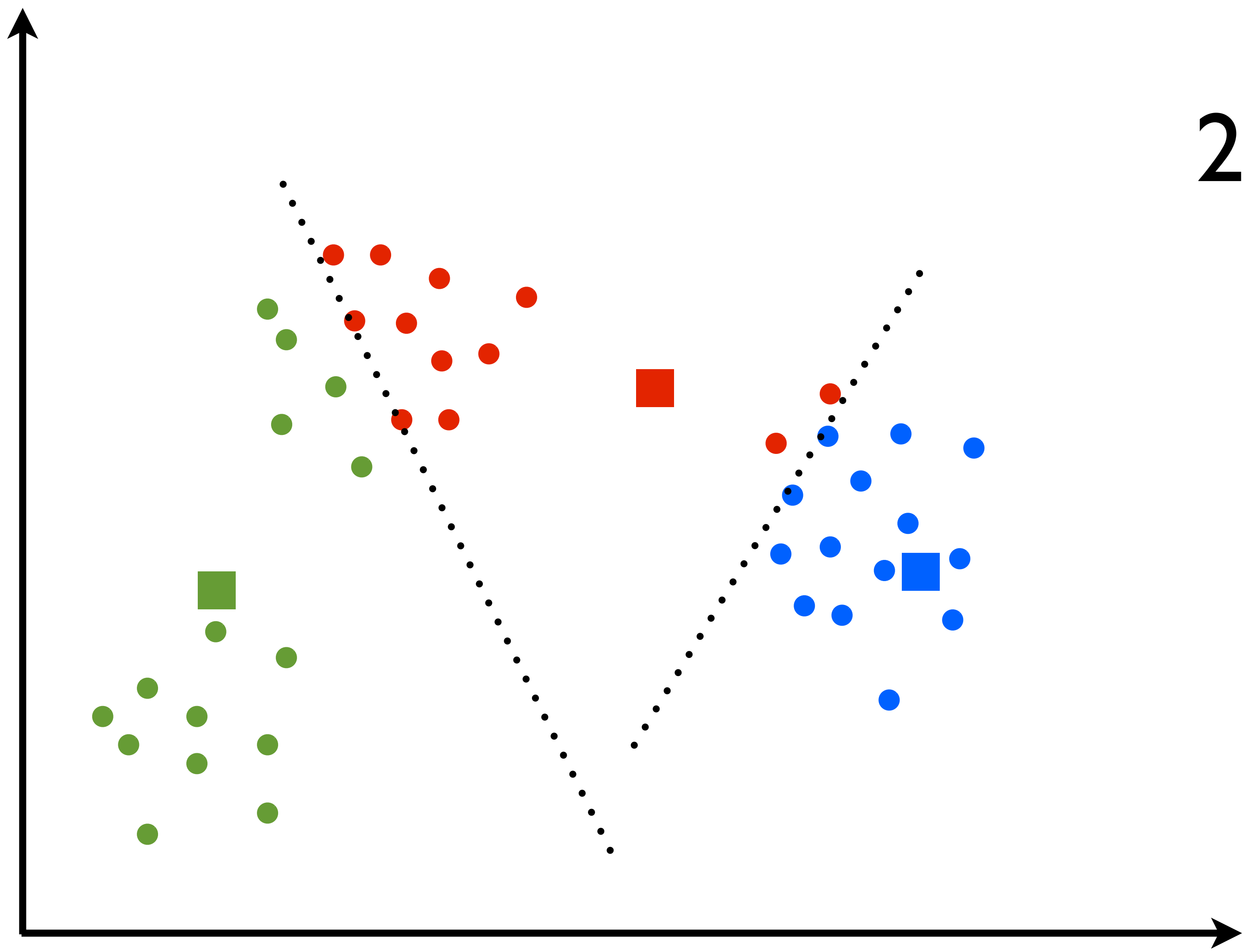
Lloyd's algorithm



Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means

Lloyd's algorithm



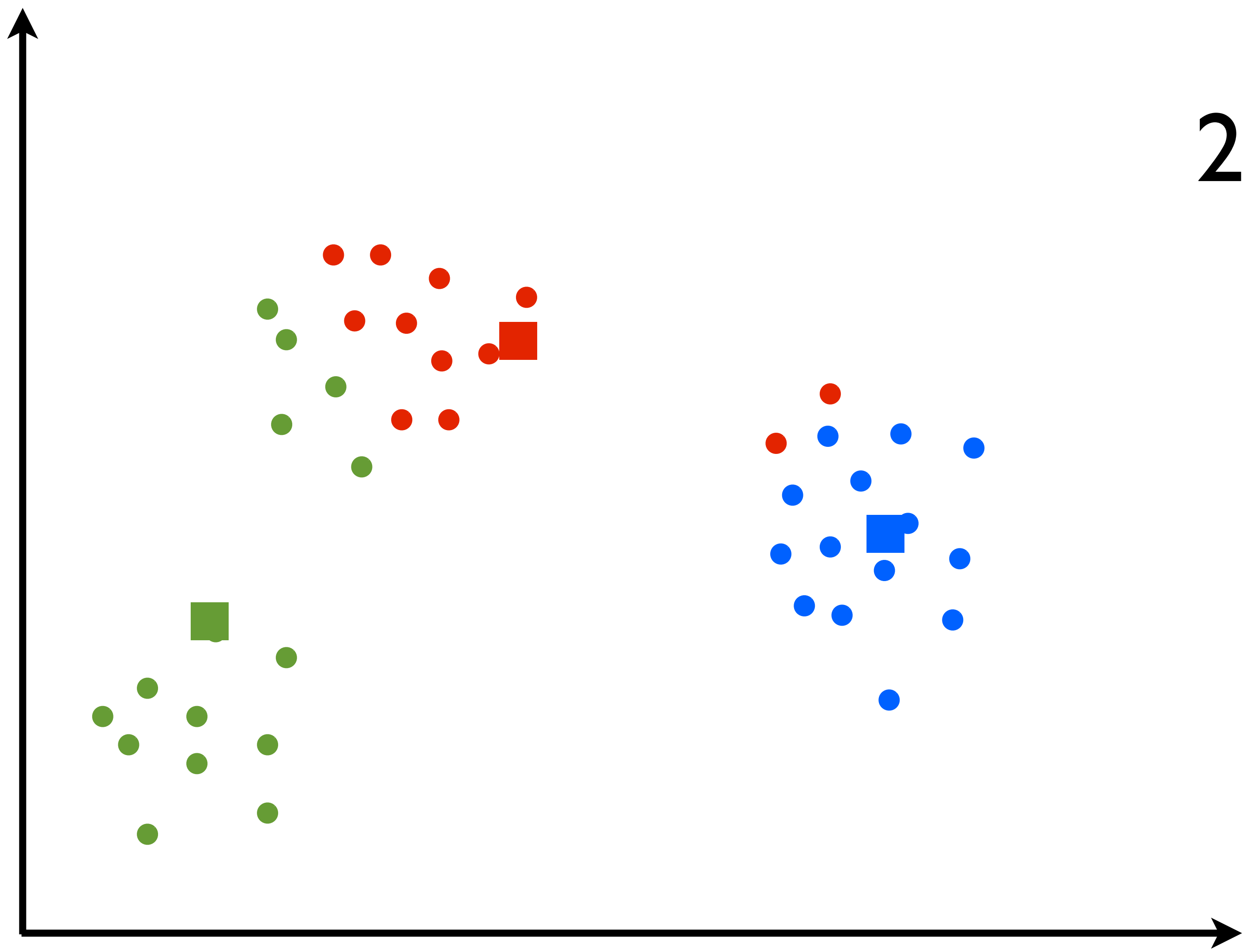
Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means

Lloyd's algorithm

Iterate until no changes:

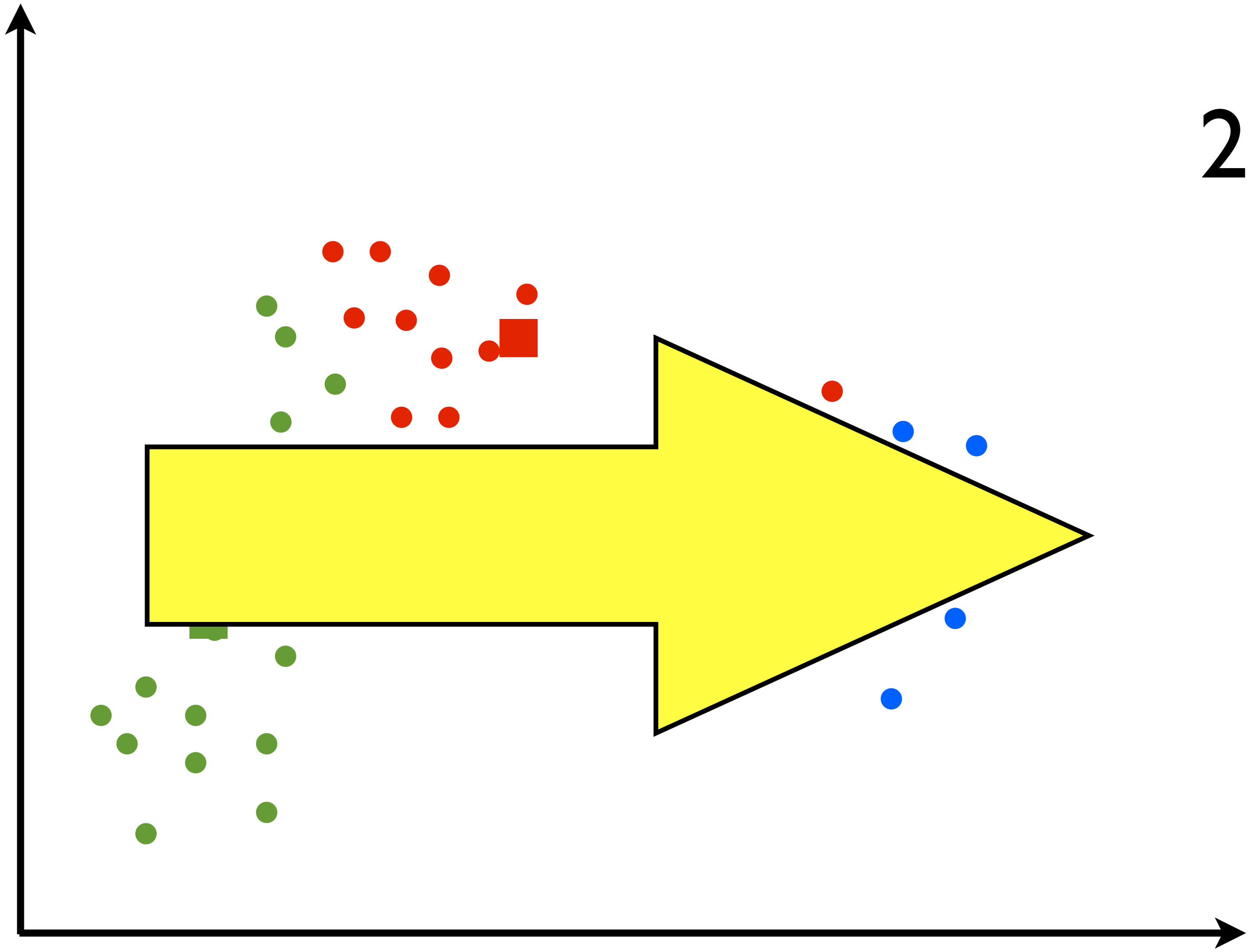
1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means



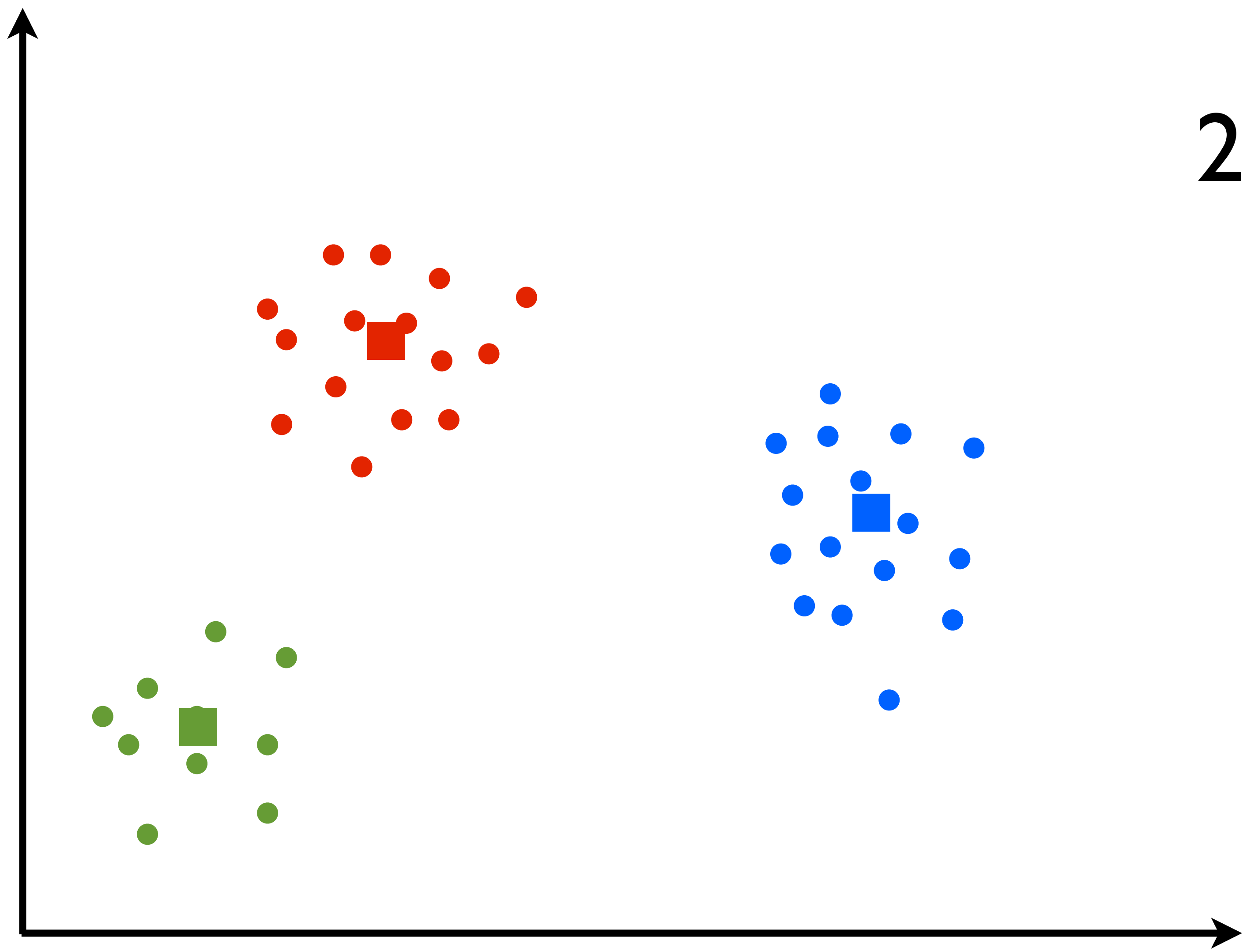
Lloyd's algorithm

Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means



Lloyd's algorithm



Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to a cluster
2. Update cluster means

MAD-Bayes

The MAD-Bayes idea

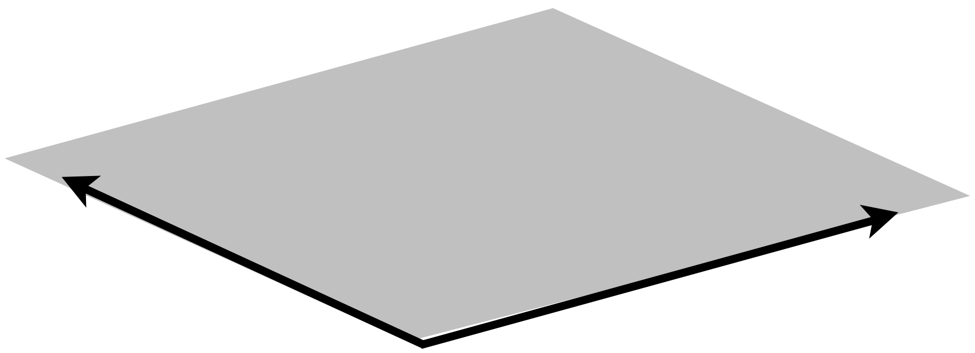
- Start with nonparametric Bayes model
- Take a similar limit to get a K-means-like objective

MAD-Bayes

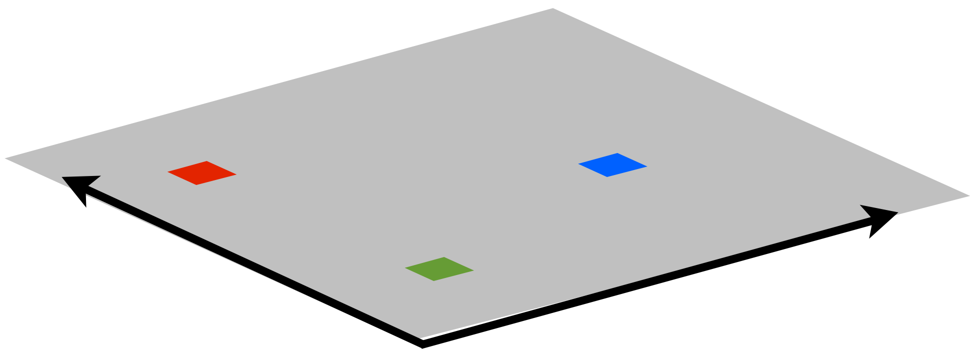
The MAD-Bayes idea

- Start with nonparametric Bayes model
- Take a similar limit to get a K-means-like objective

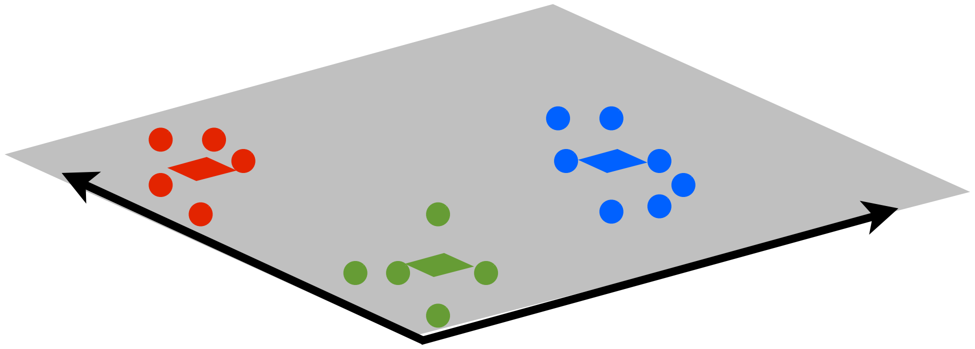
Bayesian model



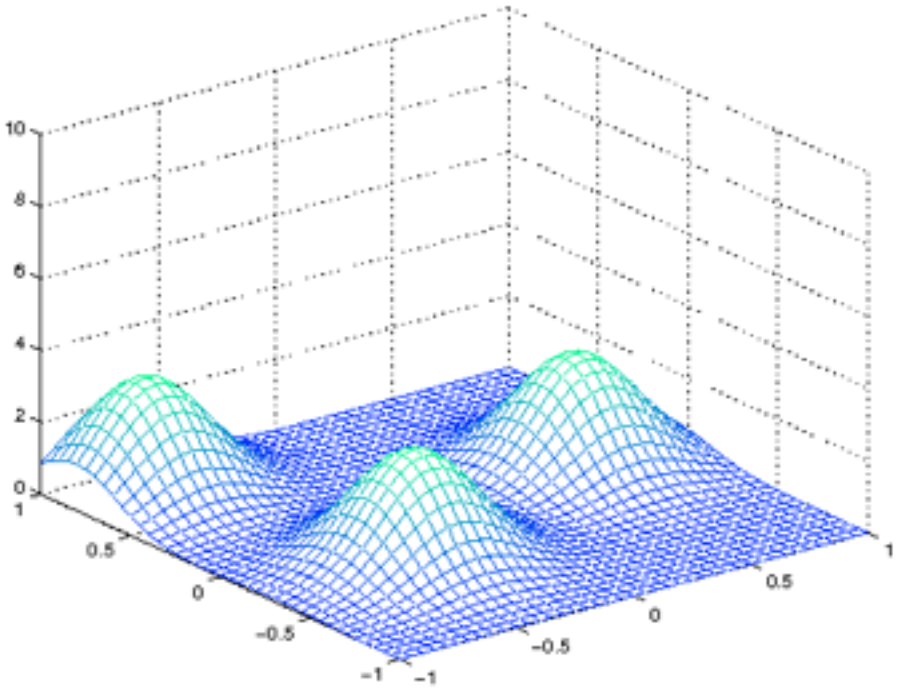
Bayesian model



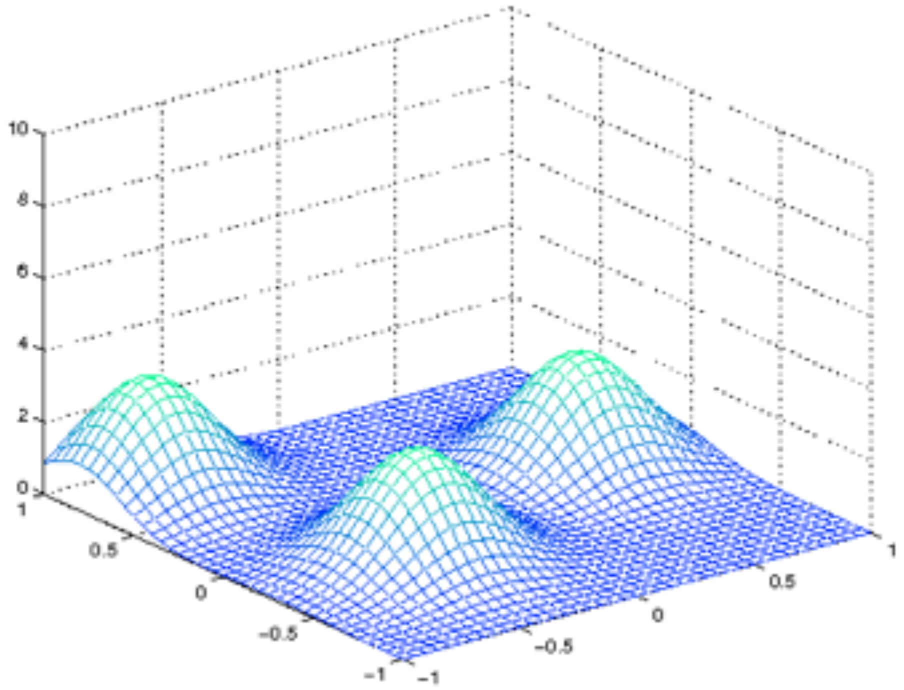
Bayesian model



Bayesian model



Bayesian model



Nonparametric

- number of parameters can grow with the number of data points

MAD-Bayes

The MAD-Bayes idea

- Start with nonparametric Bayes model
- Take a similar limit to get a K-means-like objective

MAD-Bayes

The MAD-Bayes idea

- Start with nonparametric Bayes model
- Take a similar limit to get a K-means-like objective

MAD-Bayes

MAD-Bayes

- *Maximum a Posteriori* (MAP) is an optimization problem

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

MAD-Bayes

- *Maximum a Posteriori* (MAP) is an optimization problem

$$\mathrm{argmax}_{\mathrm{parameters}} \mathbb{P}(\mathrm{parameters}|\mathrm{data})$$

- We take a limit of the objective (posterior) and get one like K-means

MAD-Bayes

- *Maximum a Posteriori* (MAP) is an optimization problem

$$\mathrm{argmax}_{\mathrm{parameters}} \mathbb{P}(\mathrm{parameters}|\mathrm{data})$$

- We take a limit of the objective (posterior) and get one like K-means
  - ◇ "Small-variance asymptotics"

MAD-Bayes

Bayesian posterior

K-means-like objectives

MAD-Bayes

Bayesian posterior		K-means-like objectives
Mixture of K Gaussians	$\Rightarrow$	K-means

MAD-Bayes

Bayesian posterior		K-means-like objectives
Mixture of K Gaussians	⇒	K-means
Dirichlet process mixture	⇒	Unbounded number of clusters

MAD-Bayes

<u>Bayesian posterior</u>	<u>K-means-like objectives</u>
Mixture of K Gaussians $\Rightarrow$	K-means
Dirichlet process mixture $\Rightarrow$	Unbounded number of clusters
Hierarchical Dirichlet process $\Rightarrow$	Multiple data sets share cluster centers

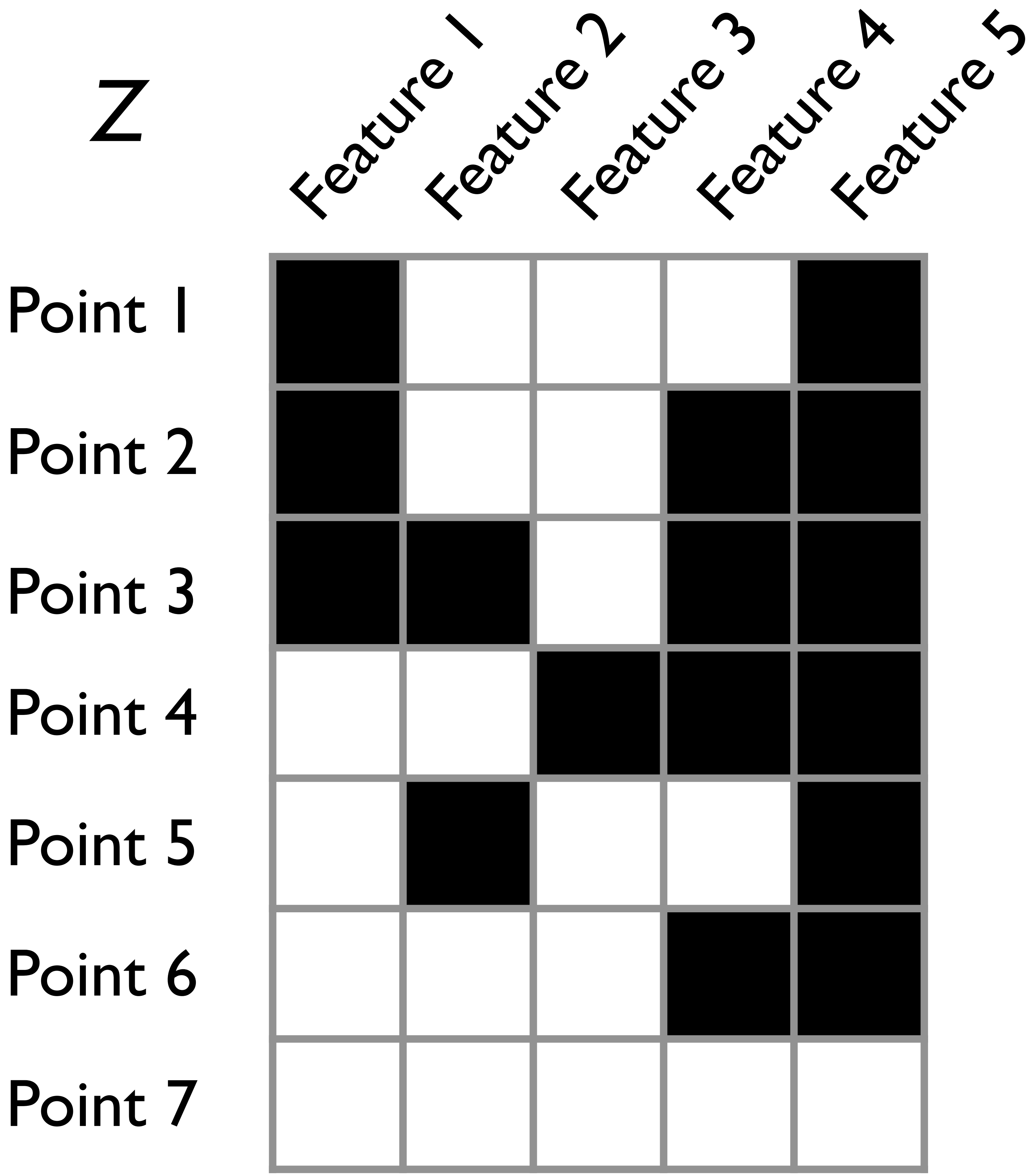
MAD-Bayes

Bayesian posterior		K-means-like objectives
Mixture of K Gaussians	→	K-means
Dirichlet process mixture	→	Unbounded number of clusters
Hierarchical Dirichlet process	→	Multiple data sets share cluster centers
⋮		⋮

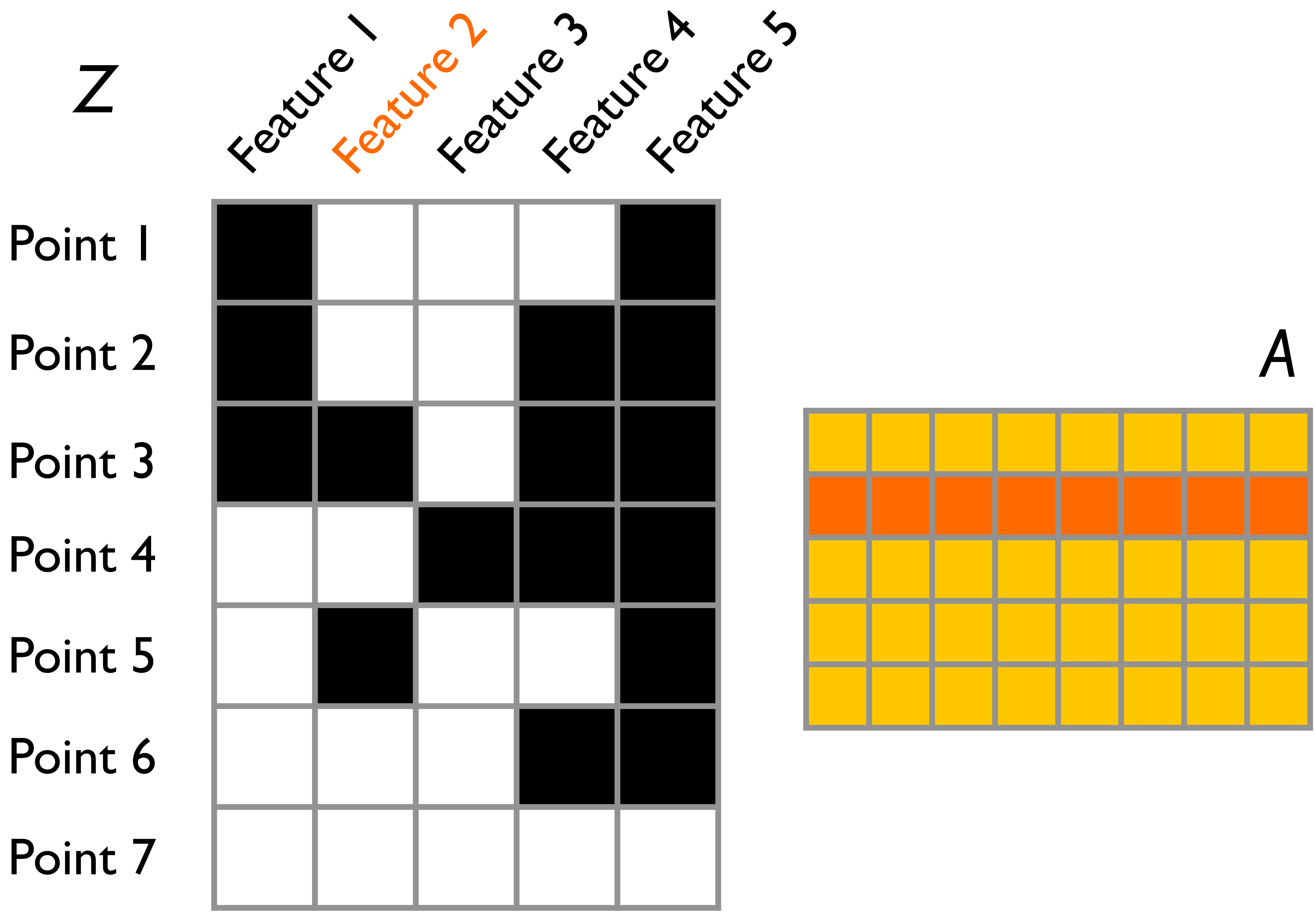
MAD-Bayes

Bayesian posterior		K-means-like objectives
Mixture of K Gaussians	→	K-means
Dirichlet process mixture	→	Unbounded number of clusters
Hierarchical Dirichlet process	→	Multiple data sets share cluster centers
⋮		⋮
Beta process	→	Features

Features

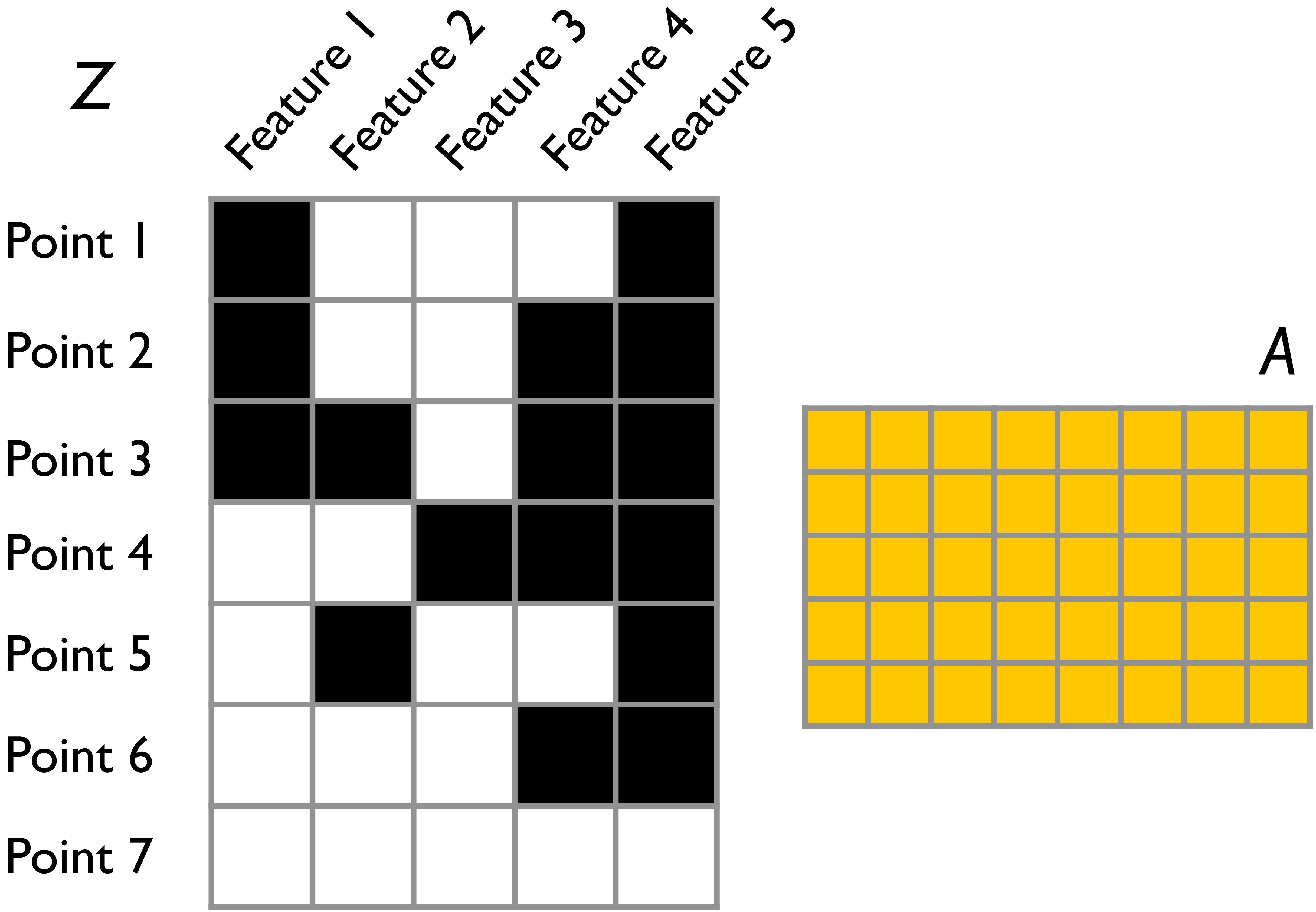

Z
Feature 1
Feature 2
Feature 3
Feature 4
Feature 5
Point 1
Point 2
Point 3
Point 4
Point 5
Point 6
Point 7

Features



Z
Feature 1
Feature 2
Feature 3
Feature 4
Feature 5
Point 1
Point 2
Point 3
Point 4
Point 5
Point 6
Point 7
A

Features


Z
Feature 1
Feature 2
Feature 3
Feature 4
Feature 5
Point 1
Point 2
Point 3
Point 4
Point 5
Point 6
Point 7
A

MAD-Bayes

Bayesian posterior

$$
\begin{aligned}
&\mathbb{P}(Z, A|X) \\
&\qquad \propto \frac{1}{(2\pi\sigma^2)^{ND/2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{tr}((X - ZA)'(X - ZA))\right\} \\
&\qquad \cdot \frac{\gamma^{K^+} \exp\left\{-\sum_{n=1}^{N} \frac{\gamma}{n}\right\}}{\prod_{h=1}^{H} \tilde{K}_h!} \prod_{k=1}^{K^+} \frac{(S_{N,k} - 1)!(N - S_{N,k})!}{N!} \\
&\qquad \cdot \frac{1}{(2\pi\rho^2)^{K^+D/2}} \exp\left\{-\frac{1}{2\rho^2}\mathbf{tr}(A'A)\right\}.
\end{aligned}
$$

MAD-Bayes

Bayesian posterior

$$\mathbb{P}(Z, A|X)$$

$$\propto \frac{1}{(2\pi\sigma^2)^{ND/2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{tr}((X-ZA)'(X-ZA))\right\}$$

$$\cdot \frac{\gamma^{K^+} \exp\left\{-\sum_{n=1}^{N} \frac{\gamma}{n}\right\}}{\prod_{h=1}^{H} \tilde{K}_h!} \prod_{k=1}^{K^+} \frac{(S_{N,k}-1)!(N-S_{N,k})!}{N!}$$

$$\cdot \frac{1}{(2\pi\rho^2)^{K^+D/2}} \exp\left\{-\frac{1}{2\rho^2}\mathbf{tr}(A'A)\right\}.$$

MAD-Bayes

Bayesian posterior

$$
\begin{aligned}
&\mathbb{P}(Z, A | X) \\
&\quad \propto \frac{1}{(2\pi\sigma^2)^{ND/2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{tr}((X - ZA)'(X - ZA))\right\} \\
&\quad \cdot \frac{\gamma^{K^+} \exp\left\{-\sum_{n=1}^{N} \frac{\gamma}{n}\right\}}{\prod_{h=1}^{H} \tilde{K}_h!} \prod_{k=1}^{K^+} \frac{(S_{N,k} - 1)!(N - S_{N,k})!}{N!} \\
&\quad \cdot \frac{1}{(2\pi\rho^2)^{K^+D/2}} \exp\left\{-\frac{1}{2\rho^2}\mathbf{tr}(A'A)\right\}.
\end{aligned}
$$

MAD-Bayes

Bayesian posterior

$$\mathbb{P}(Z, A | X)$$

$$\propto \frac{1}{(2\pi\sigma^2)^{ND/2}} \exp\left\{ -\frac{1}{2\sigma^2} \mathbf{tr}((X - ZA)'(X - ZA)) \right\}$$

$$\cdot \frac{\gamma^{K^+} \exp\left\{ -\sum_{n=1}^{N} \frac{\gamma}{n} \right\}}{\prod_{h=1}^{H} \tilde{K}_h!} \prod_{k=1}^{K^+} \frac{(S_{N,k} - 1)!(N - S_{N,k})!}{N!}$$

$$\cdot \frac{1}{(2\pi\rho^2)^{K^+ D/2}} \exp\left\{ -\frac{1}{2\rho^2} \mathbf{tr}(A'A) \right\}.$$

MAD-Bayes

Bayesian posterior

$$\mathbb{P}(Z, A | X)$$

$$\propto \frac{1}{(2\pi\sigma^2)^{ND/2}} \exp\left\{ -\frac{1}{2\sigma^2} \mathbf{tr}((X - ZA)'(X - ZA)) \right\}$$

$$\cdot \frac{\gamma^{K^+} \exp\left\{ -\sum_{n=1}^{N} \frac{\gamma}{n} \right\}}{\prod_{h=1}^{H} \tilde{K}_h!} \prod_{k=1}^{K^+} \frac{(S_{N,k} - 1)!(N - S_{N,k})!}{N!}$$

$$\cdot \frac{1}{(2\pi\rho^2)^{K^+ D/2}} \exp\left\{ -\frac{1}{2\rho^2} \mathbf{tr}(A'A) \right\}.$$

MAD-Bayes

Bayesian posterior

$$
\begin{aligned}
&\mathbb{P}(Z, A|X) \\
&\quad \propto \frac{1}{(2\pi\sigma^2)^{ND/2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{tr}((X - ZA)'(X - ZA))\right\} \\
&\quad \cdot \frac{\gamma^{K^+} \exp\left\{-\sum_{n=1}^{N} \frac{\gamma}{n}\right\}}{\prod_{h=1}^{H} \tilde{K}_h!} \prod_{k=1}^{K^+} \frac{(S_{N,k} - 1)!(N - S_{N,k})!}{N!} \\
&\quad \cdot \frac{1}{(2\pi\rho^2)^{K^+ D/2}} \exp\left\{-\frac{1}{2\rho^2}\mathbf{tr}(A'A)\right\}.
\end{aligned}
$$

MAD-Bayes

Bayesian posterior

$$\mathbb{P}(Z, A|X)$$

$$\propto \frac{1}{(2\pi\sigma^2)^{ND/2}} \exp\left\{-\frac{1}{2\sigma^2}\mathbf{tr}((X - ZA)'(X - ZA))\right\}$$

$$\cdot \frac{\gamma^{K^+} \exp\left\{-\sum_{n=1}^{N} \frac{\gamma}{n}\right\}}{\prod_{h=1}^{H} \tilde{K}_h!} \prod_{k=1}^{K^+} \frac{(S_{N,k}-1)!(N-S_{N,k})!}{N!}$$

$$\cdot \frac{1}{(2\pi\rho^2)^{K^+D/2}} \exp\left\{-\frac{1}{2\rho^2}\mathbf{tr}(A'A)\right\}.$$

MAD-Bayes

BP-means objective

$$\mathrm{argmin}_{K^+,Z,A} \mathbf{tr}[(X - ZA)'(X - ZA)] + K^+\lambda^2.$$

MAD-Bayes

BP-means objective

$$\operatorname{argmin}_{K^+, Z, A} \mathbf{tr}[(X - ZA)'(X - ZA)] + K^+\lambda^2.$$

MAD-Bayes

BP-means objective

$$\mathrm{argmin}_{K^+,Z,A} \mathbf{tr}[(X - ZA)'(X - ZA)] + K^+\lambda^2.$$

MAD-Bayes

BP-means objective

$$\text{argmin}_{K^+,Z,A}\mathbf{tr}[(X-ZA)'(X-ZA)]+K^+\lambda^2.$$

BP-means algorithm

Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to features
   - Create a new feature if it lowers the objective
2. Update feature means $A \leftarrow (Z'Z)^{-1}Z'X$

MAD-Bayes

BP-means objective

$$\mathrm{argmin}_{K^+, Z, A} \mathbf{tr}[(X - ZA)'(X - ZA)] + K^+ \lambda^2.$$

BP-means algorithm

Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to features
   - Create a new feature if it lowers the objective
2. Update feature means $A \leftarrow (Z'Z)^{-1} Z'X$

MAD-Bayes

BP-means objective

$$\text{argmin}_{K^+, Z, A} \mathbf{tr}[(X - ZA)'(X - ZA)] + K^+\lambda^2.$$

BP-means algorithm

Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to features
   - Create a new feature if it lowers the objective
2. Update feature means $A \leftarrow (Z'Z)^{-1}Z'X$

MAD-Bayes

BP-means objective

$$\text{argmin}_{K^+,Z,A}\mathbf{tr}[(X-ZA)'(X-ZA)]+K^+\lambda^2.$$

BP-means algorithm

Iterate until no changes:

1. For n = 1, ..., N
   - Assign point n to features
   - Create a new feature if it lowers the objective
2. Update feature means $A \leftarrow (Z'Z)^{-1}Z'X$

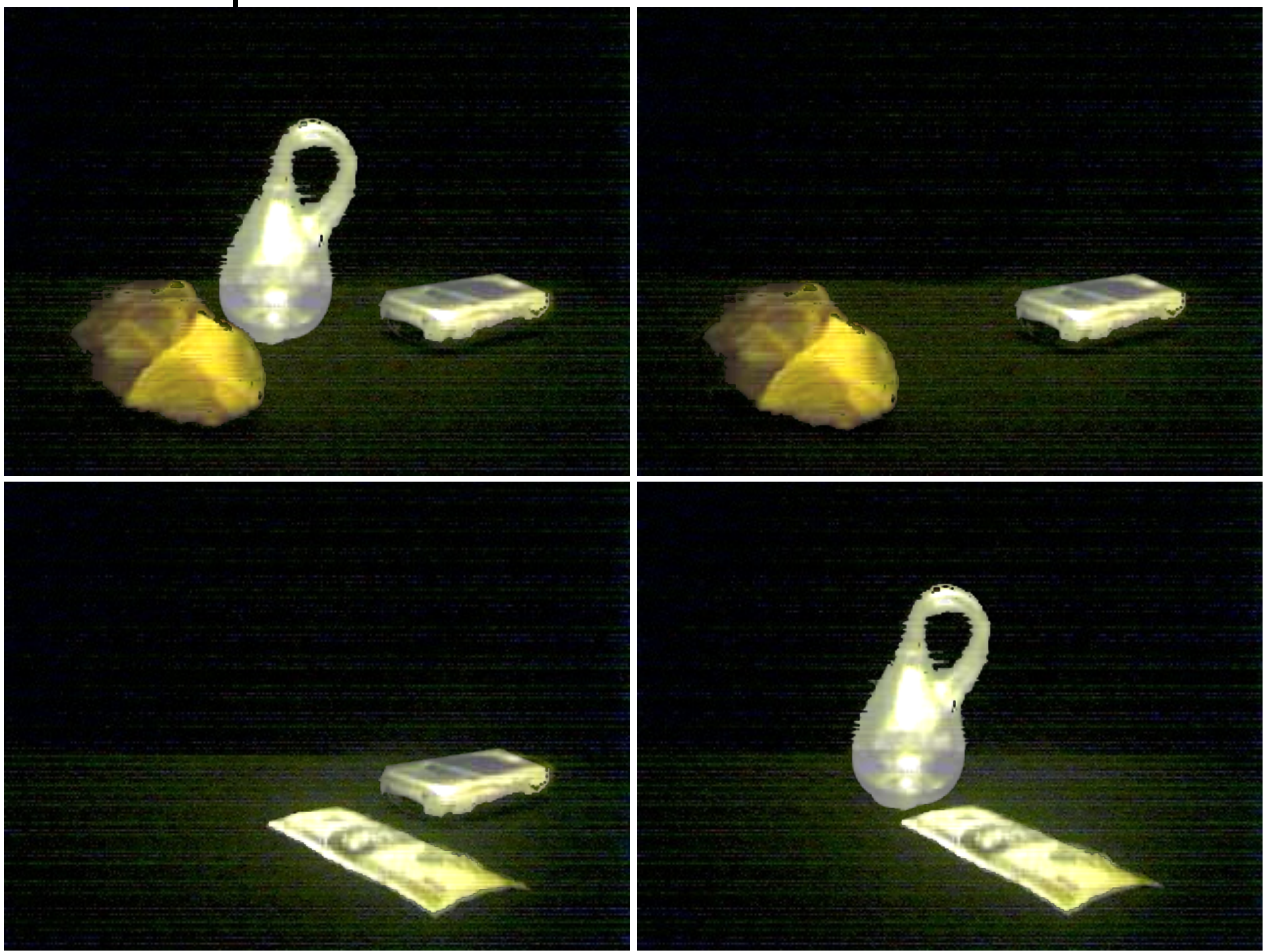
MAD-Bayes

Griffiths & Ghahramani (2006) computer vision problem “tabletop data”

[Griffiths, Ghahramani 2006]

MAD-Bayes

Griffiths & Ghahramani (2006) computer vision problem “tabletop data”

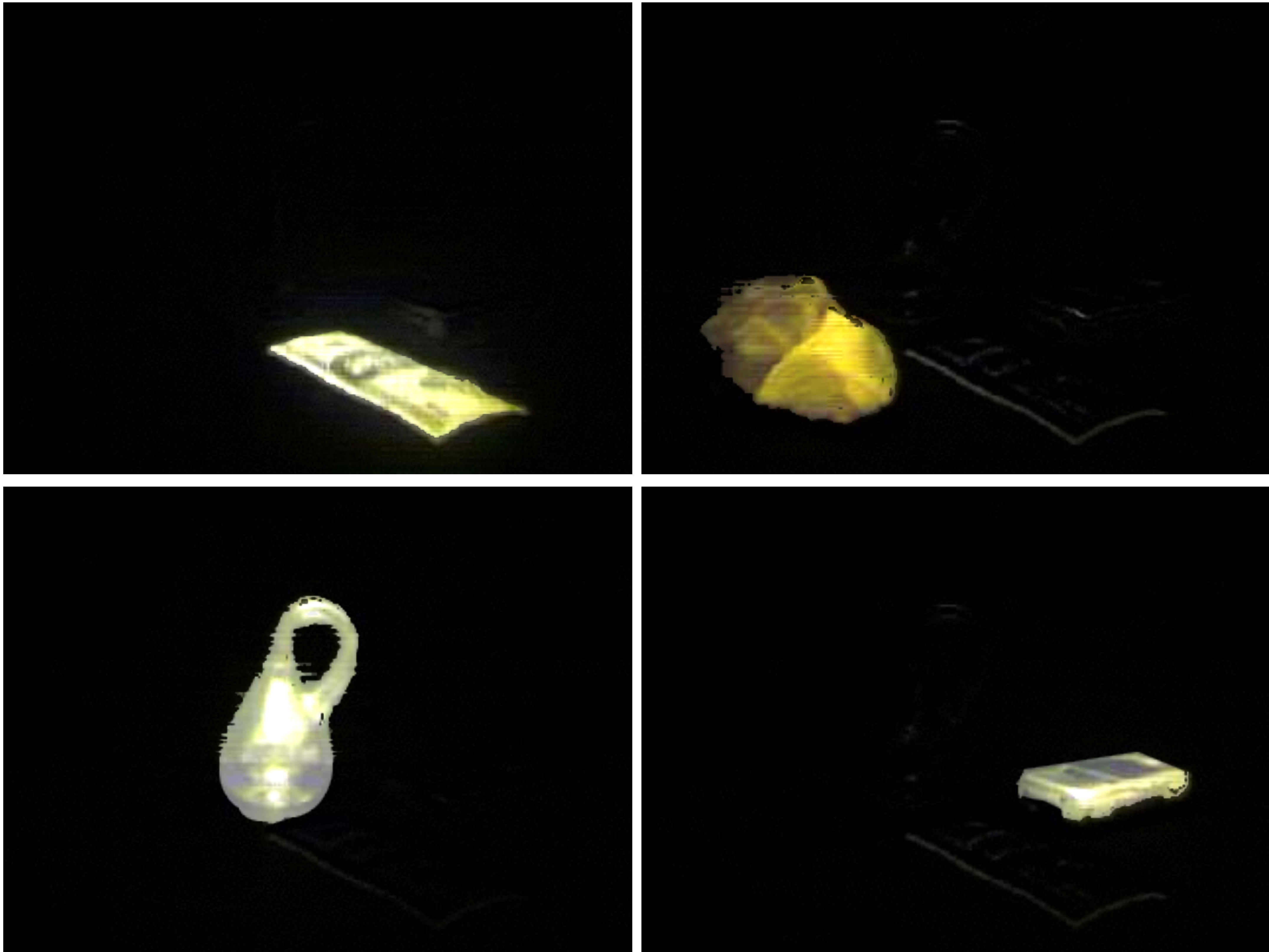


MAD-Bayes

BP-means features: table and four objects

MAD-Bayes

BP-means features: table and four objects



MAD-Bayes

Griffiths & Ghahramani (2006) computer vision problem "tabletop data"

Bayesian posterior
Gibbs sampler

BP-means algorithm

$8.5 * 10^3$ sec

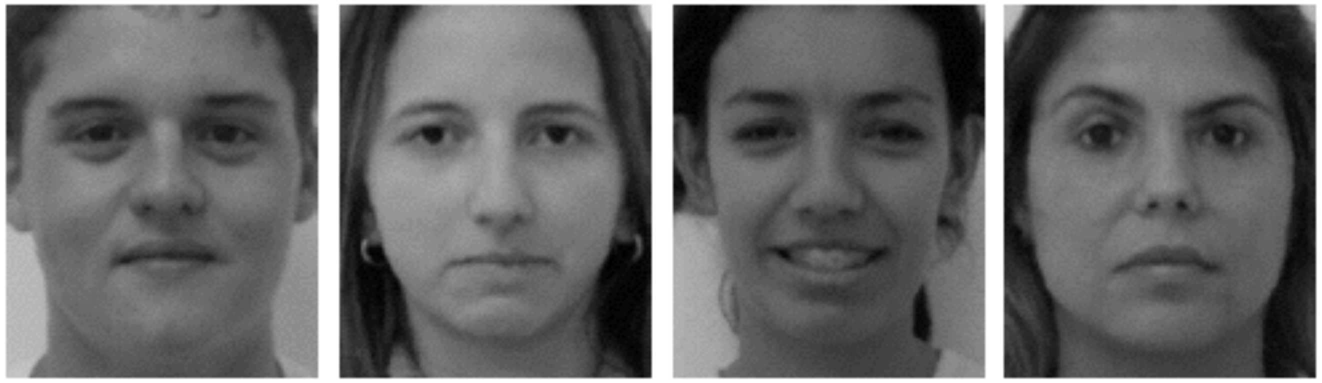
0.36 sec

Still faster by order of magnitude if restart 1000 times

Face data

Pre-aligned faces

Samples

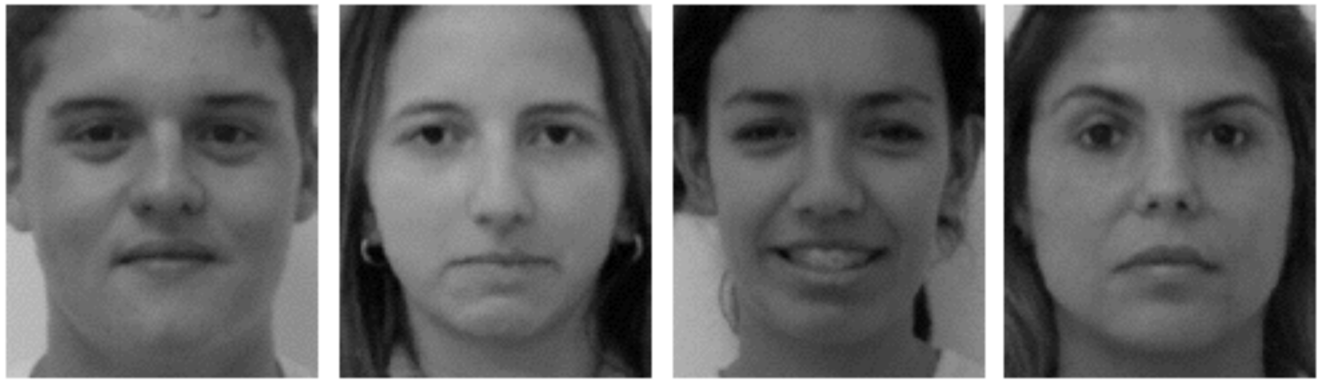


[Thomaz, Giraldi 2010]

Face data

Pre-aligned faces

Samples



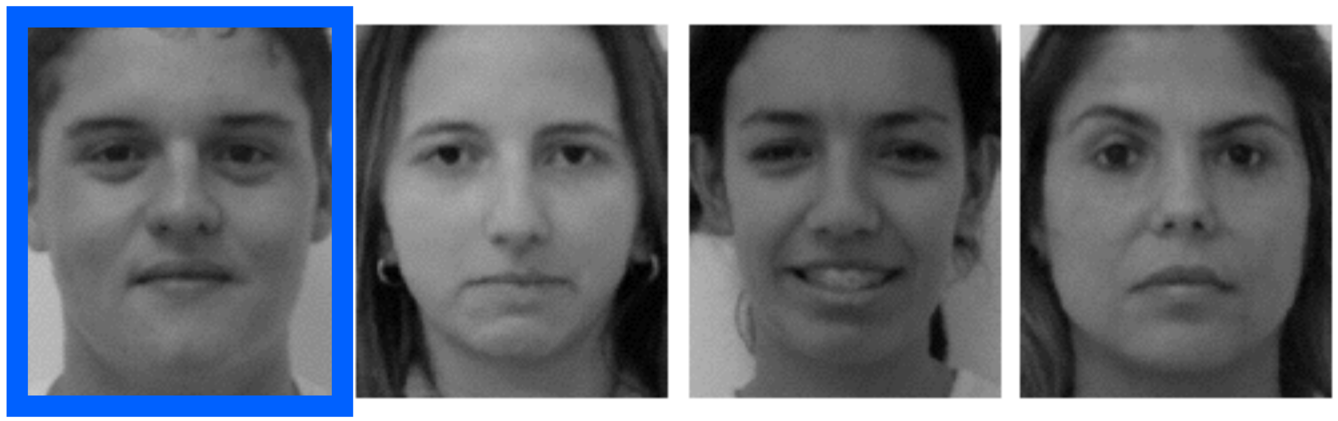
3 <u>**features**</u>

(BP-means)

Face data

Pre-aligned faces

Samples

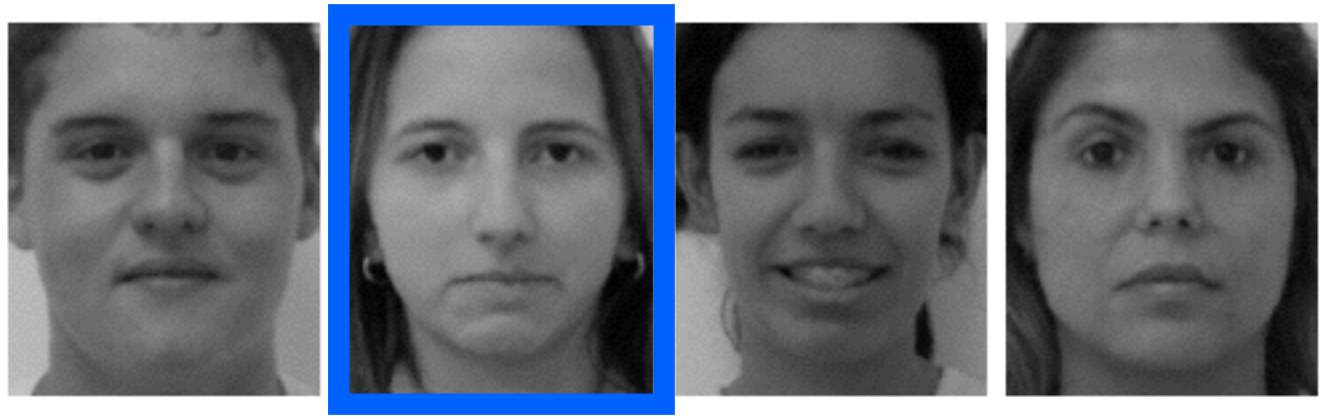


3 **<u>features</u>**
(BP-means)

Face data

Pre-aligned faces

Samples



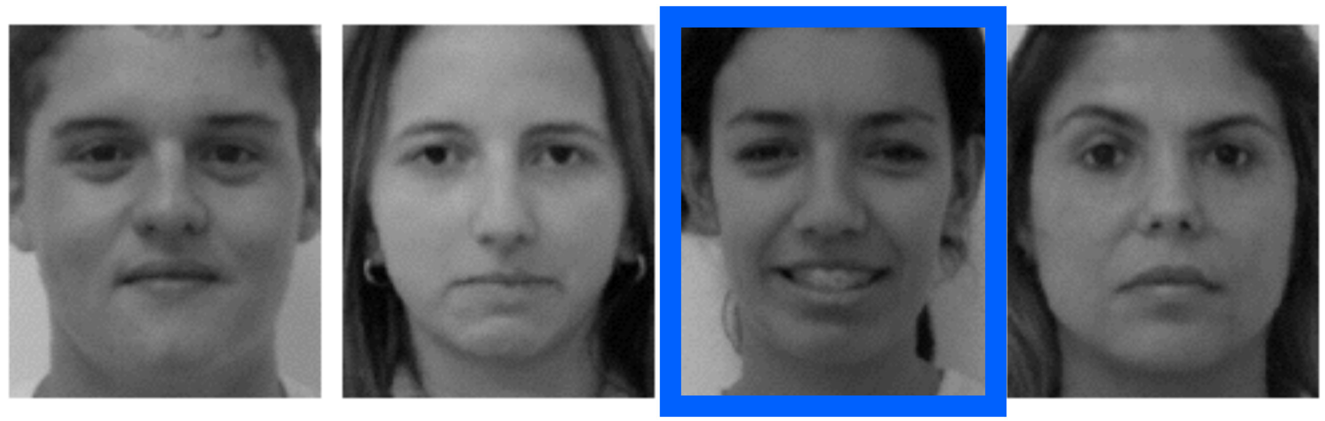
3 **features**

(BP-means)

Face data

Pre-aligned faces

Samples



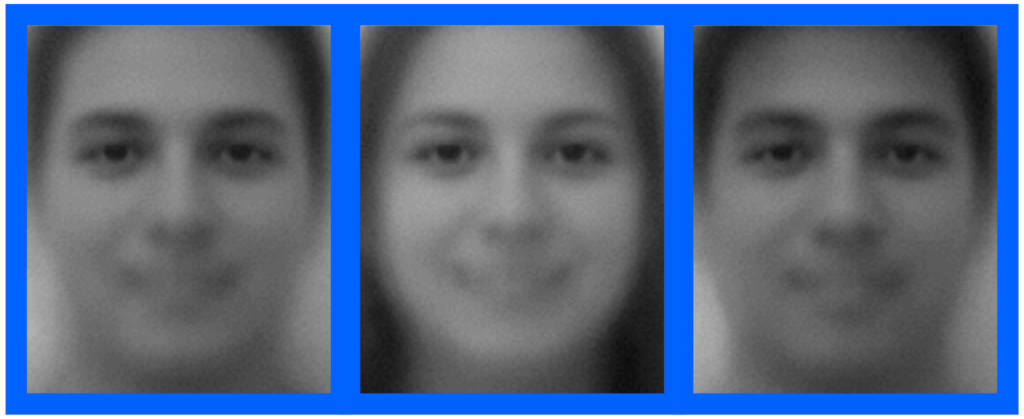
3 **features**
(BP-means)

Face data

Pre-aligned faces

Samples

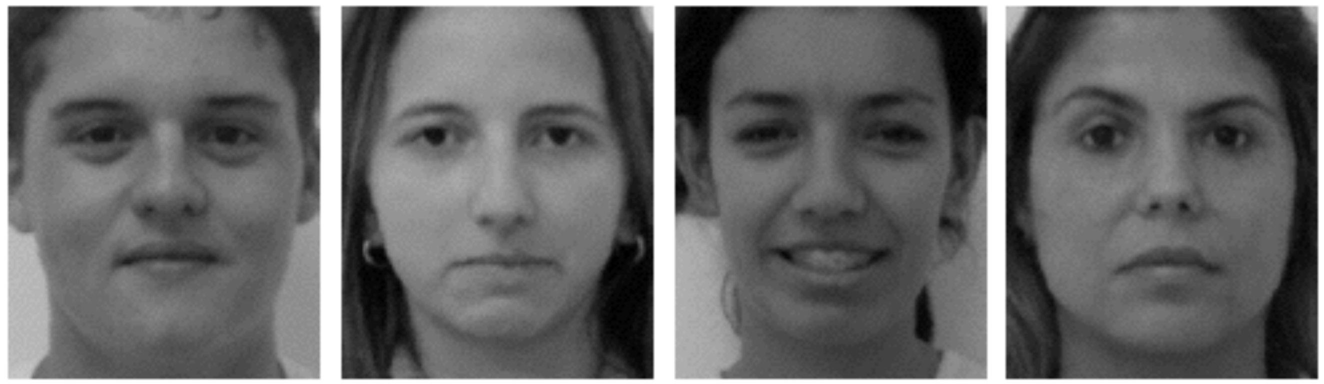
3 <u>**features**</u>
(BP-means)



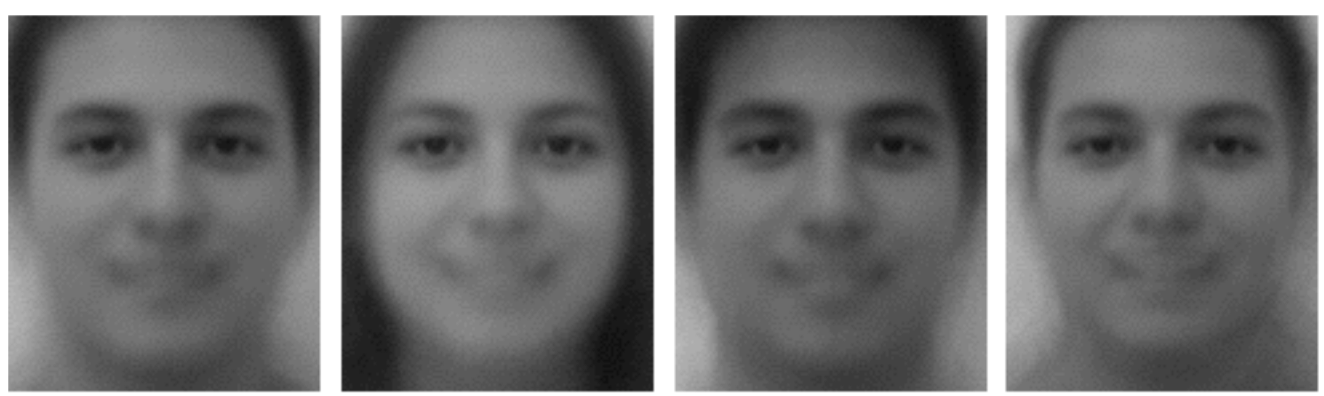
Face data

Pre-aligned faces

Samples



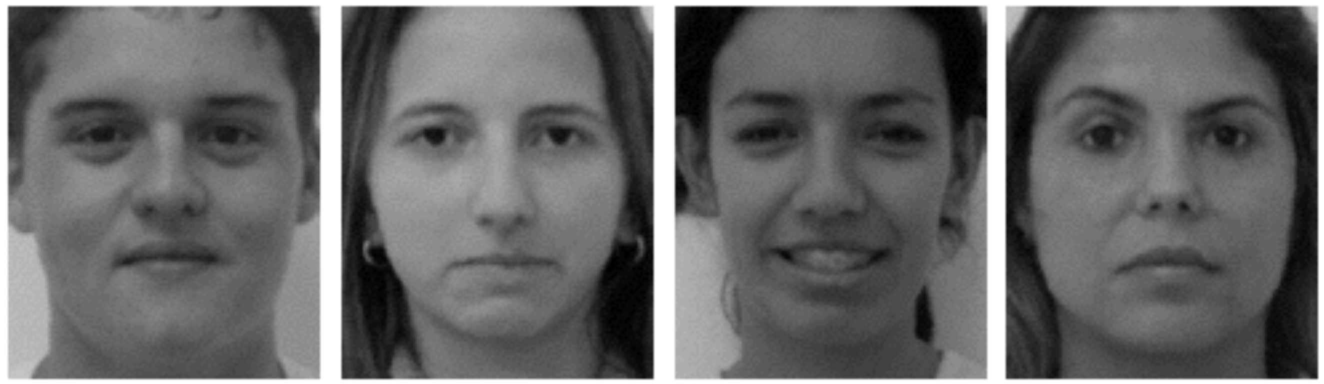
4 clusters
(K-means, K=4)



Face data

Pre-aligned faces

Samples



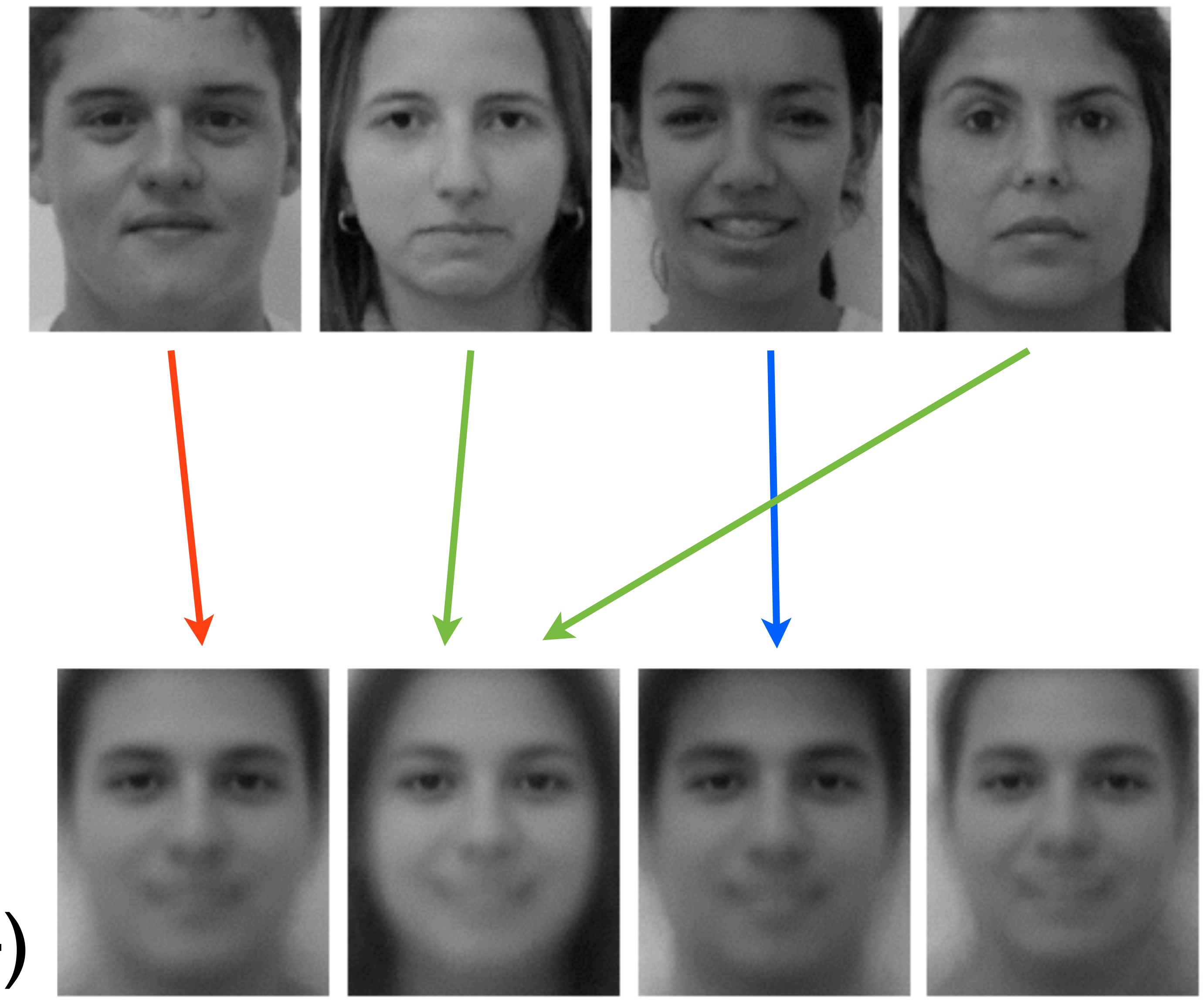
4 clusters
(K-means, K=4)

Face data

Pre-aligned faces

Samples

4 clusters
(K-means, K=4)



MAD-Bayes

Parallelism and optimistic concurrency control

	DP-means alg.	BP-means alg.
# data points	**134M**	**8M**
time per iteration	**5.5 min**	**4.3 min**

MAD-Bayes

Bayesian posterior		K-means-like objectives
Mixture of K Gaussians	→	K-means
Dirichlet process mixture	→	Unbounded number of clusters
Hierarchical Dirichlet process	→	Multiple data sets share cluster centers
⋮		⋮
Beta process	→	Features

MAD-Bayes conclusions

MAD-Bayes conclusions

- We provide new optimization objectives and regularizers

MAD-Bayes conclusions

- We provide new optimization objectives and regularizers
    - In fact, general means of obtaining more

MAD-Bayes conclusions

- We provide new optimization objectives and regularizers
  - In fact, general means of obtaining more
  - Straightforward, fast algorithms

References

T. Broderick, B. Kulis, and M. I. Jordan. MAD-Bayes: MAP-based asymptotic derivations from Bayes. In *International Conference on Machine Learning*, 2013.

X. Pan, J. E. Gonzales, S. Jegelka, T. Broderick, and M. I. Jordan. Optimistic concurrency control for distributed unsupervised learning. In *Neural Information Processing Systems*, 2013.

T. Broderick, N. Boyd, A. Wibisono, A. C. Wilson, and M. I. Jordan. Streaming variational Bayes. In *Neural Information Processing Systems*, 2013.

R. Giordano and T. Broderick. Linear response methods for accurate covariance estimates from mean field variational Bayes. In *Neural Information Processing Systems*, 2015.

Further References

T. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *Neural Information Processing Systems*, 2006.

N. L. Hjort. Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics*, 18(3):1259–1294, 1990.

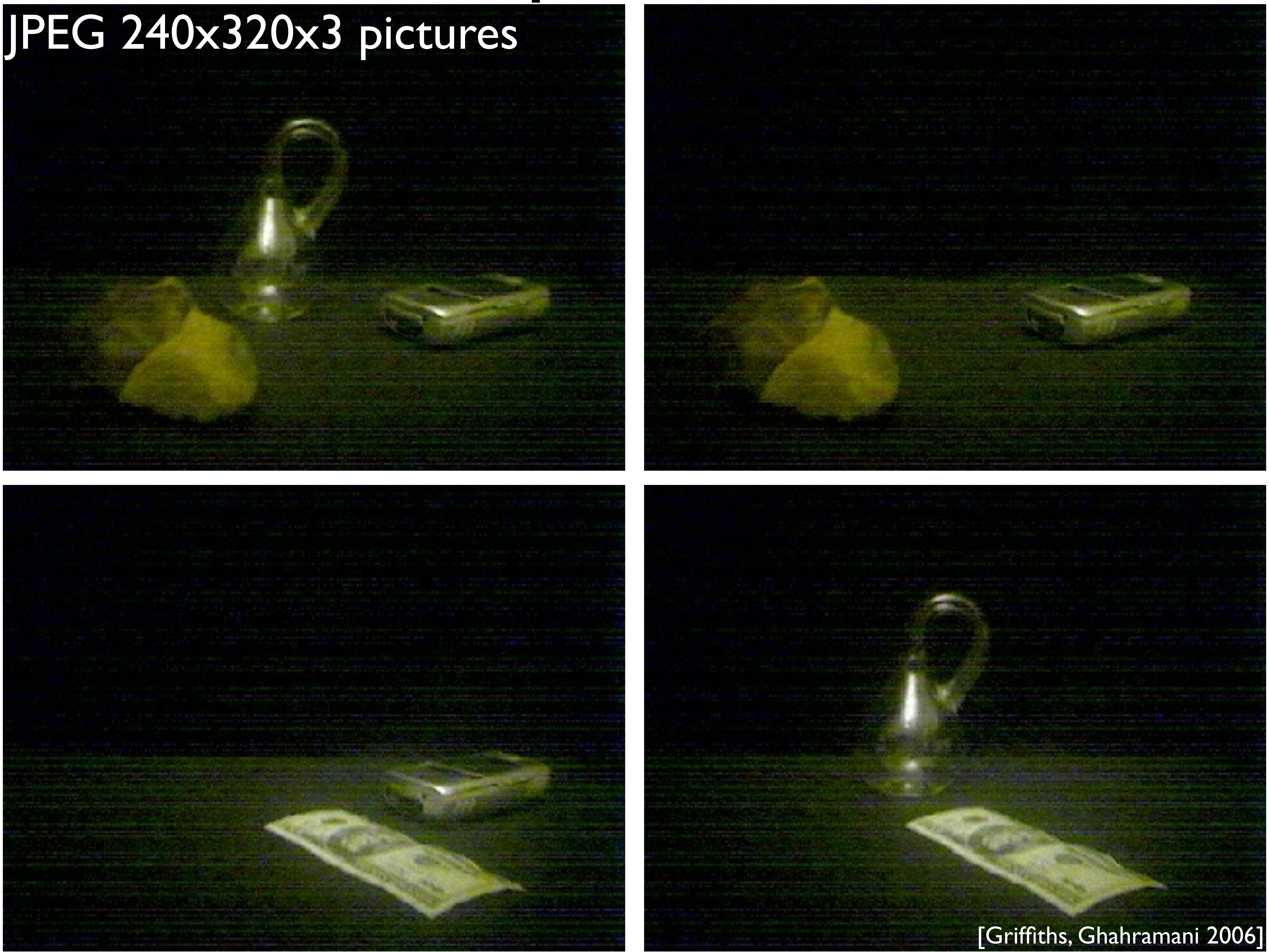
J. F. C. Kingman. The representation of partition structures. *Journal of the London Mathematical Society*, 2(2):374, 1978.

B. Kulis and M. I. Jordan. Revisiting k-means: New algorithms via Bayesian nonparametrics. In *International Conference on Machine Learning*, 2012.

J. Pitman. Exchangeable and partially exchangeable random partitions. *Probability Theory and Related Fields*, 102(2):145–158, 1995.

R. Thibaux and M. I. Jordan. Hierarchical beta processes and the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, 2007.

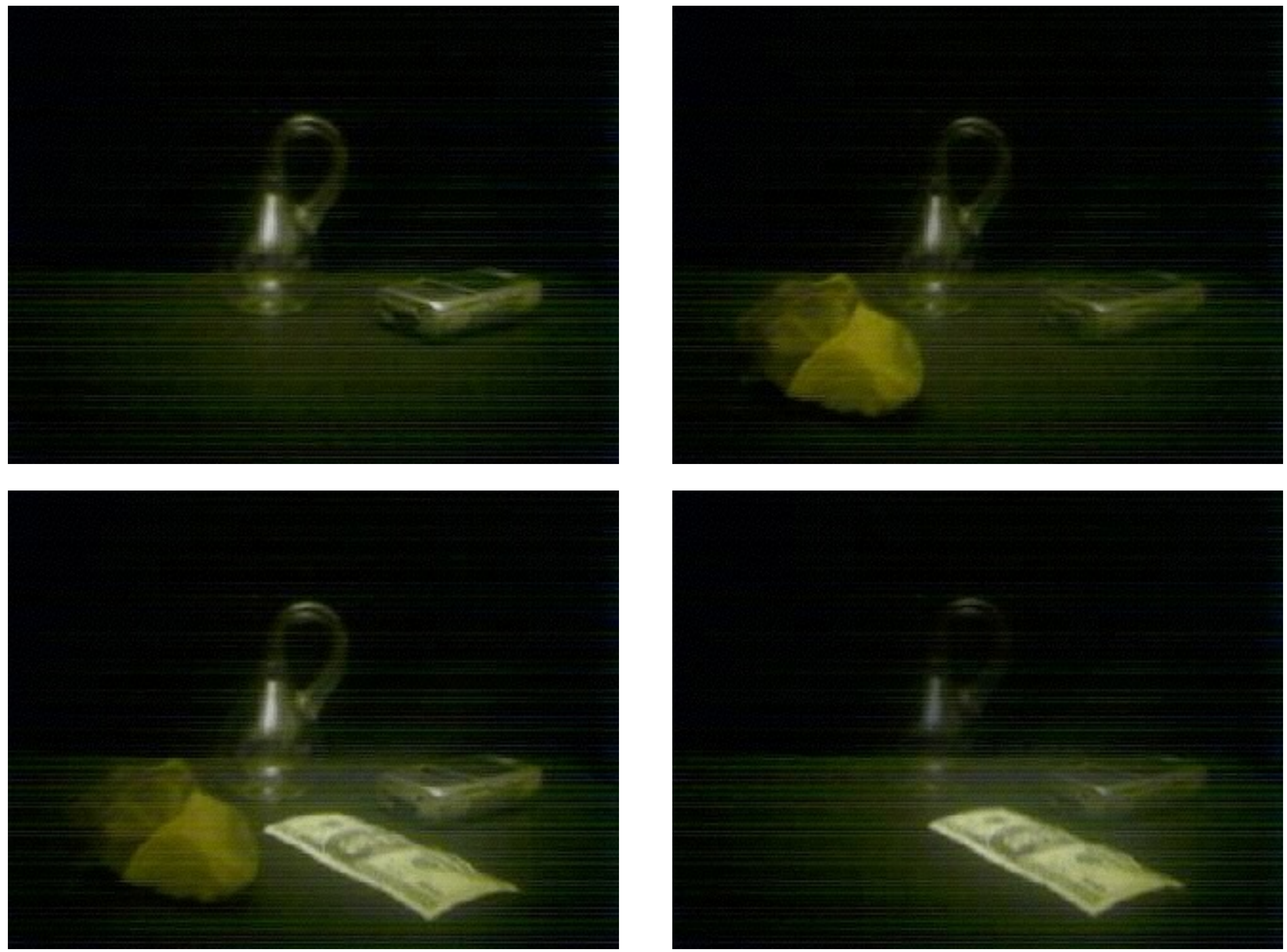
BP-means: Tabletop data



[Griffiths, Ghahramani 2006]

BP-means: Tabletop data results

K-means (K=4) cluster centers:



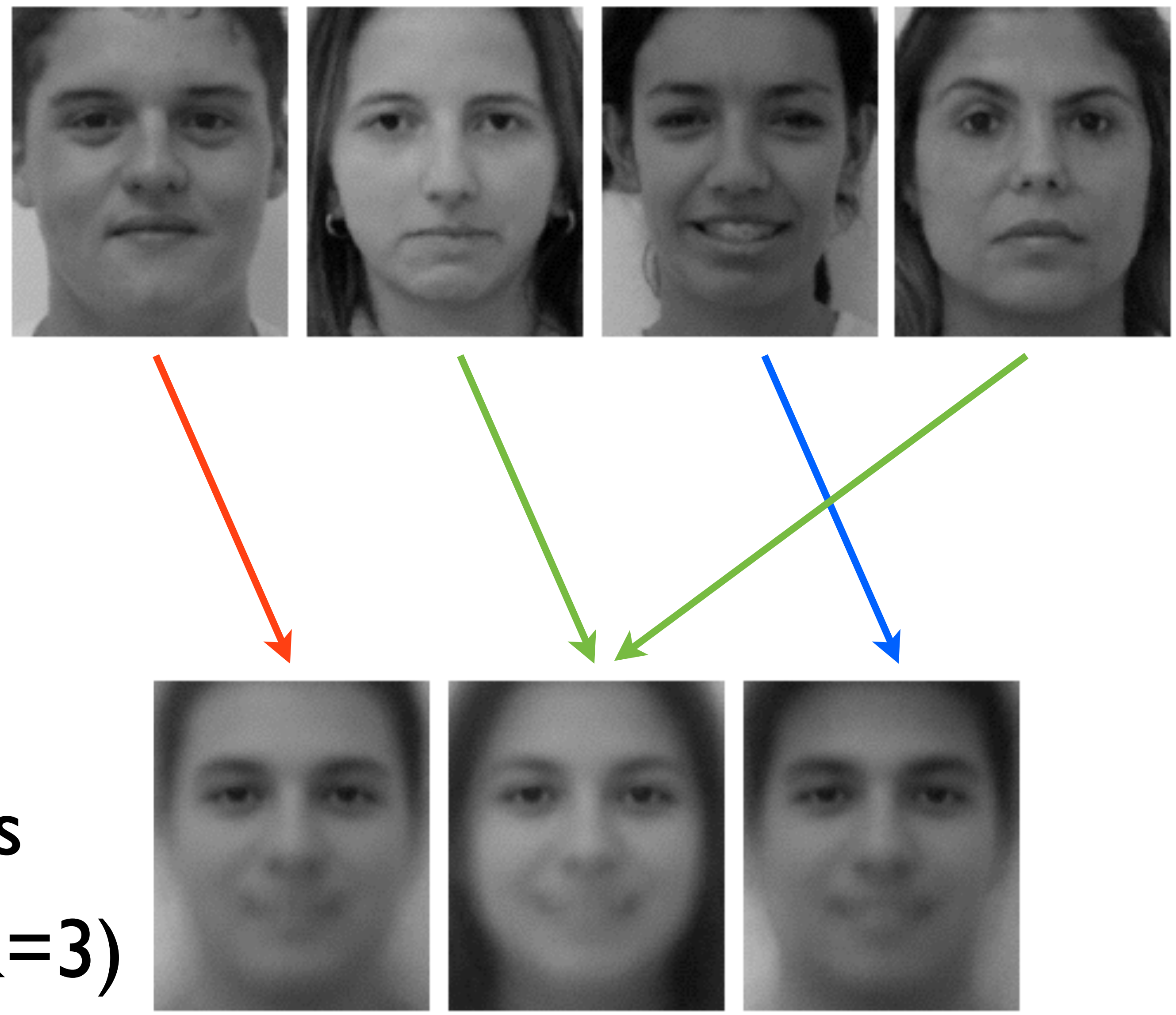
BP-means: Tabletop data results

BP-means features: table and four objects

Face data

Pre-aligned faces

Samples



3 clusters

(K-means, K=3)

Face data

Pre-aligned faces

Samples

4 clusters

(K-means, K=4)