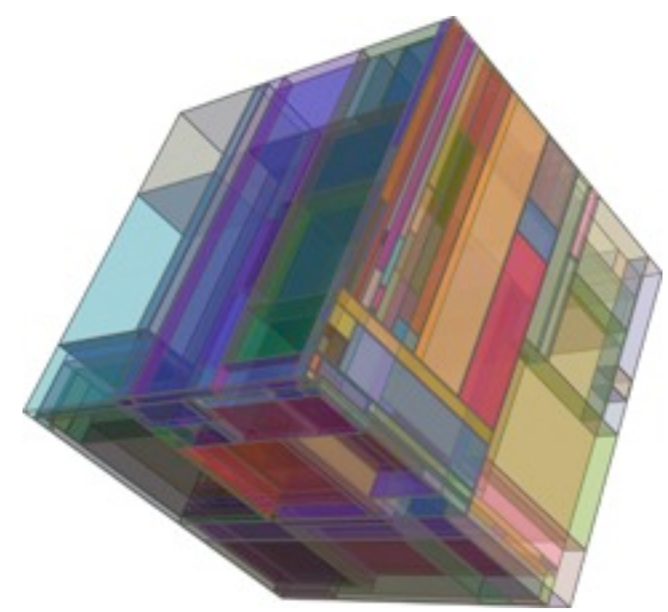
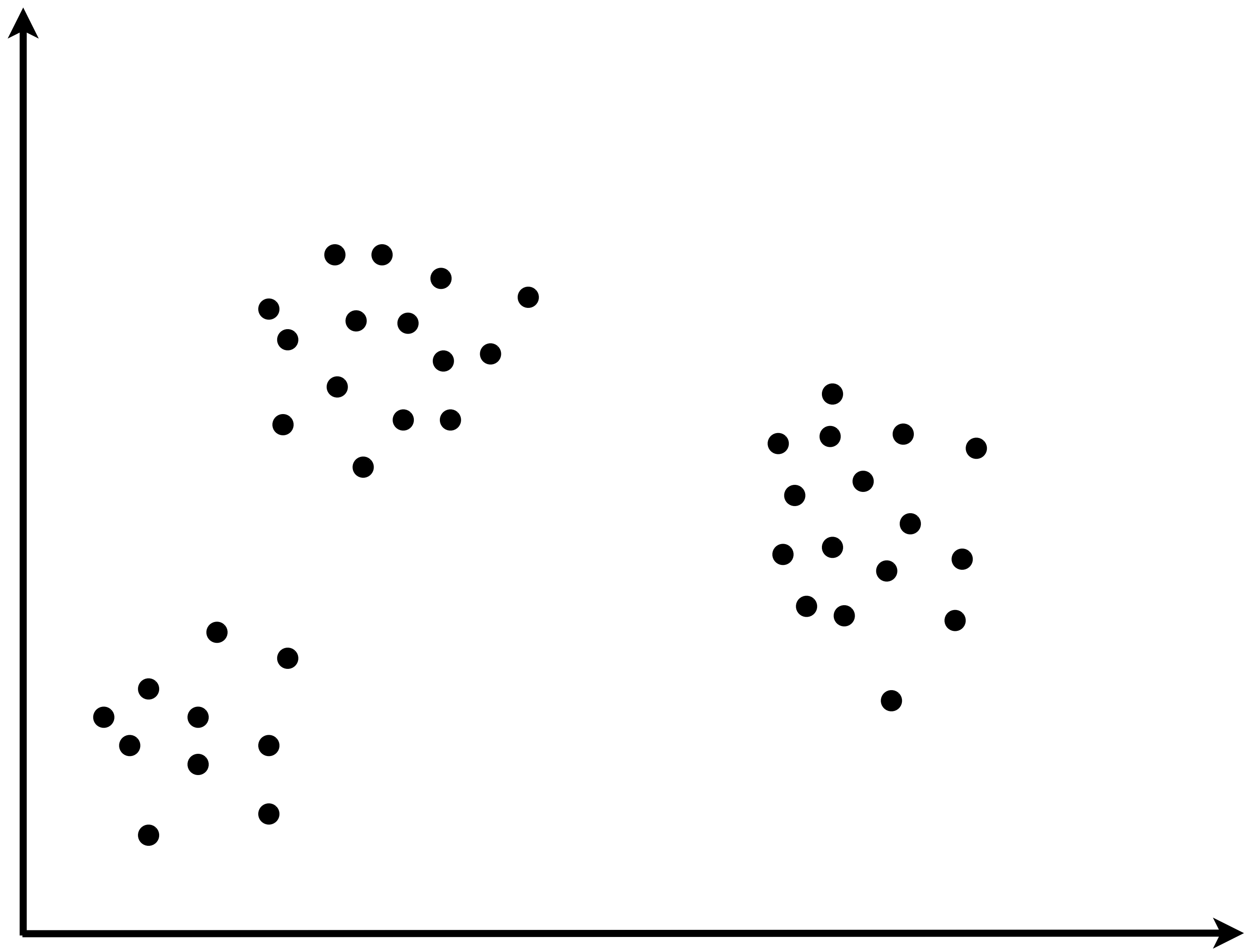
Clusters and features from combinatorial stochastic processes

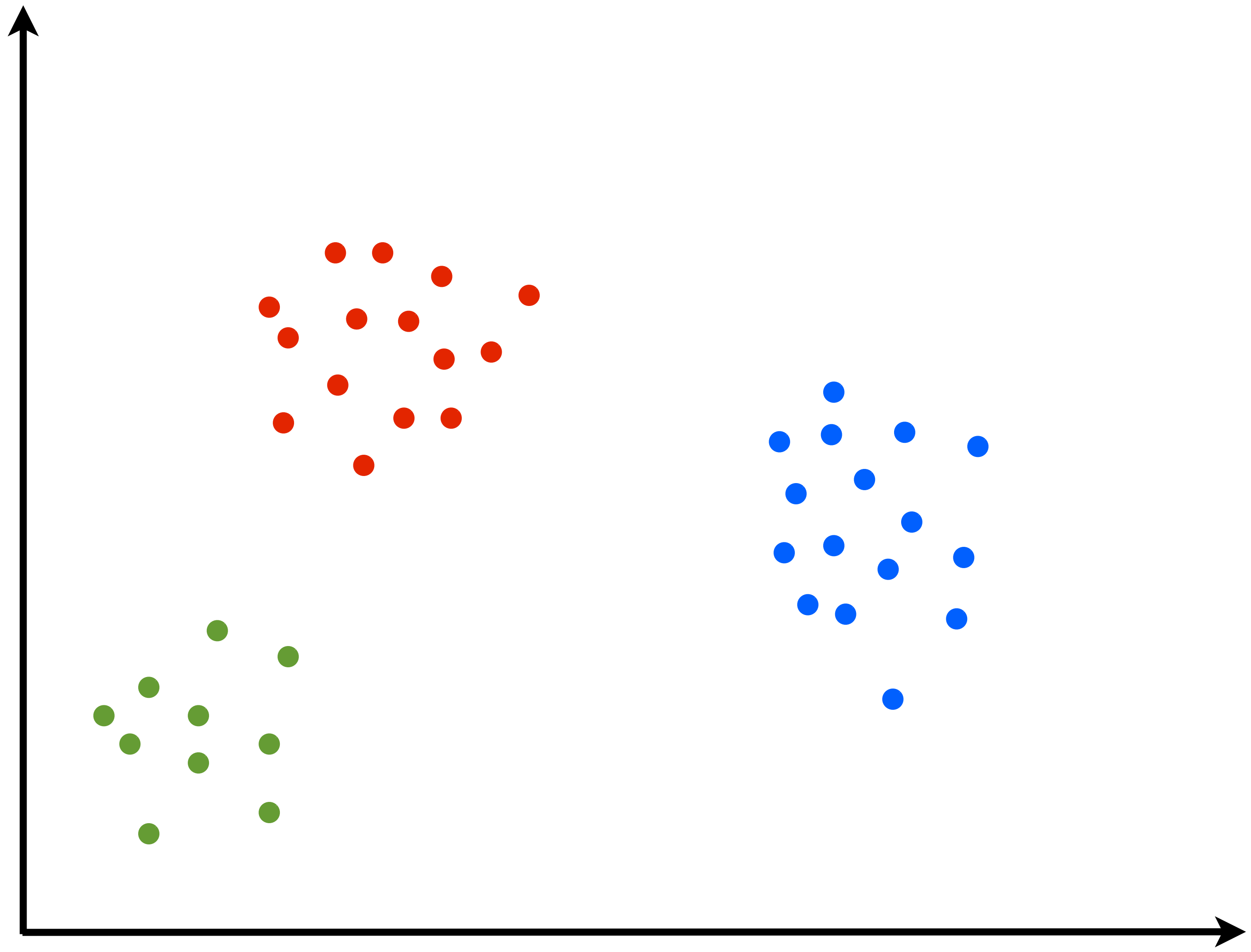
Tamara Broderick, Michael I. Jordan, Jim Pitman
UC Berkeley



Clustering/Partition

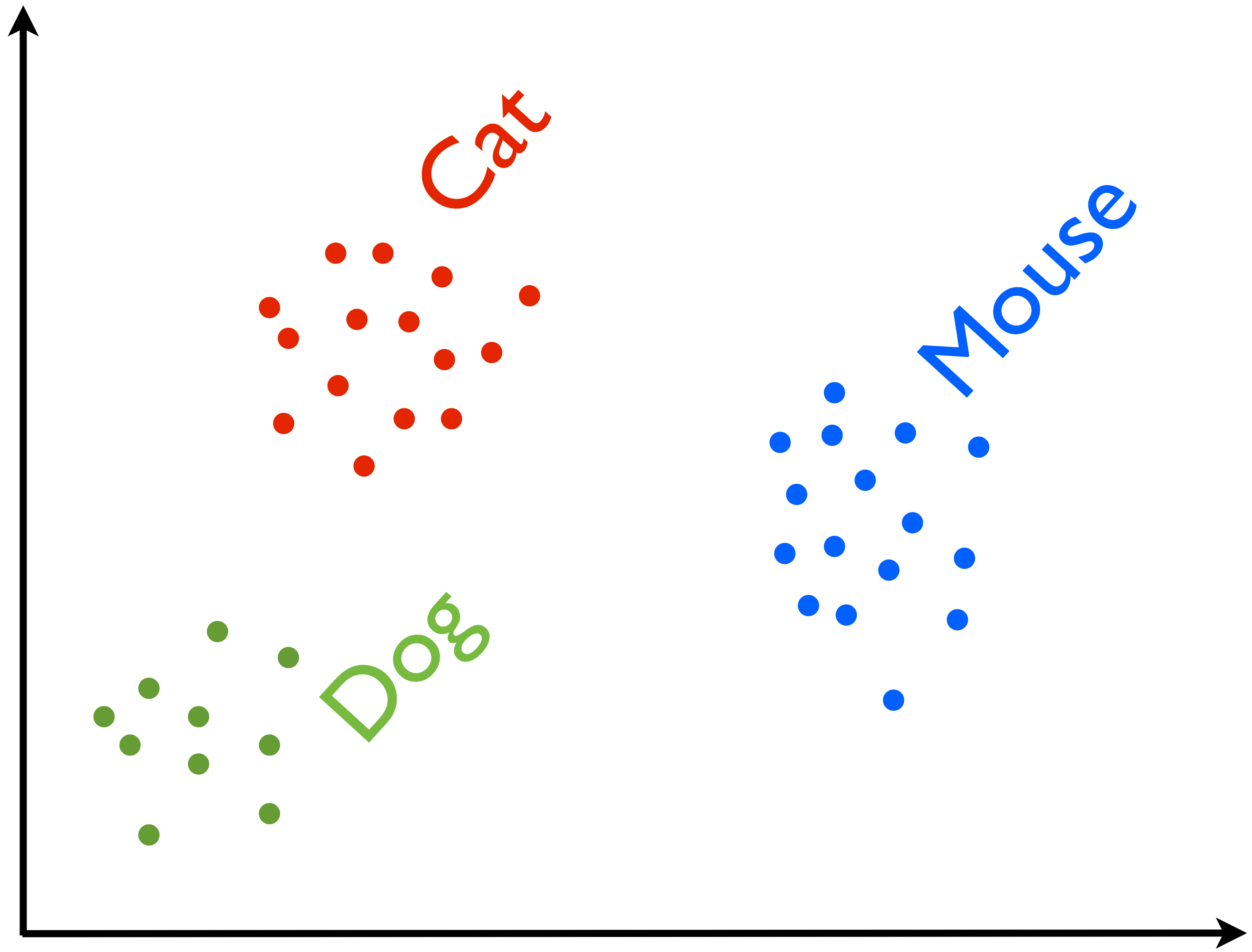


Clustering/Partition



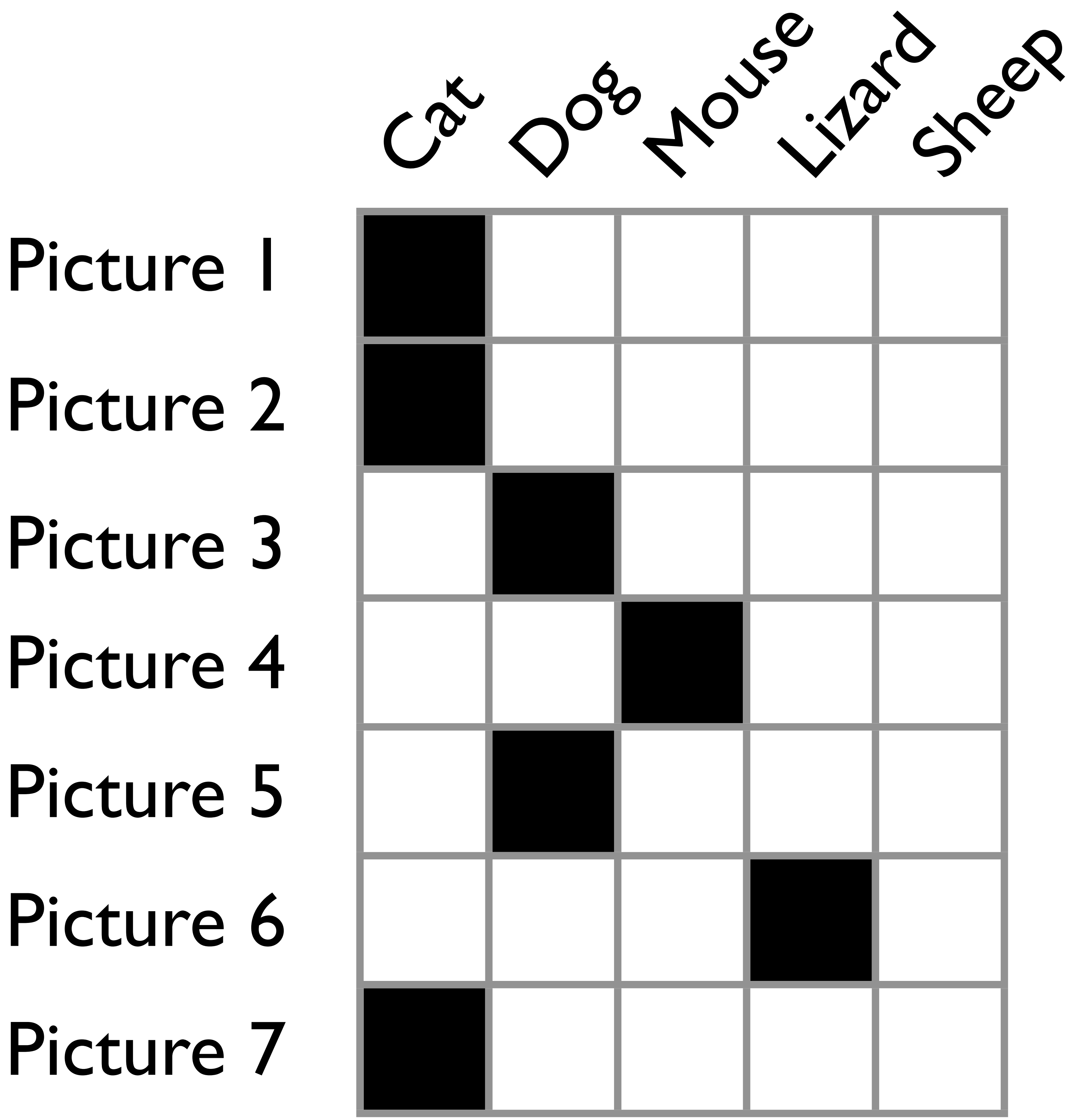
"clusters",
"classes",
"blocks (of a partition)"

Clustering/Partition

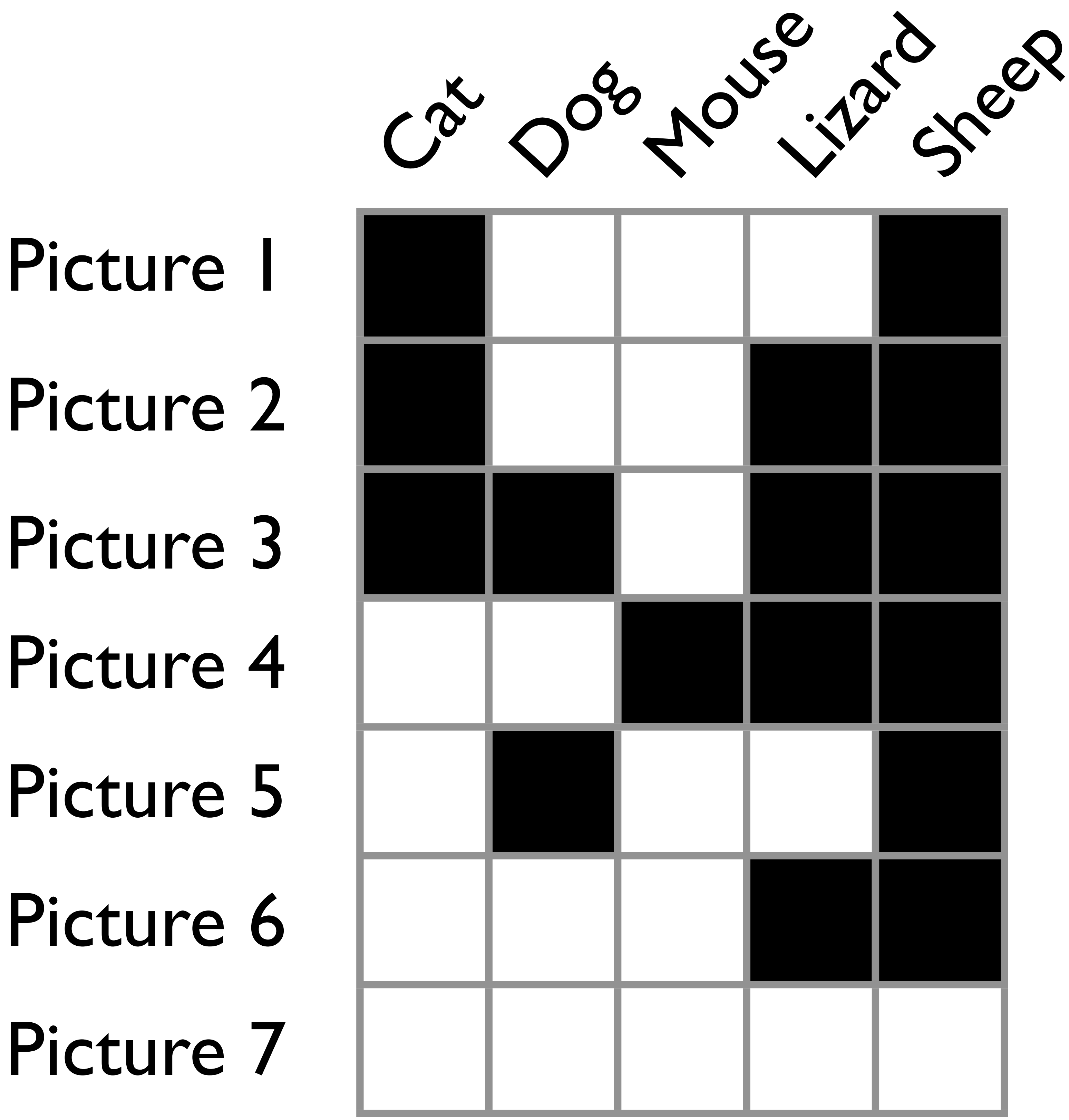


"clusters",
"classes",
"blocks (of a partition)"

Clustering/Partition



Latent feature allocation



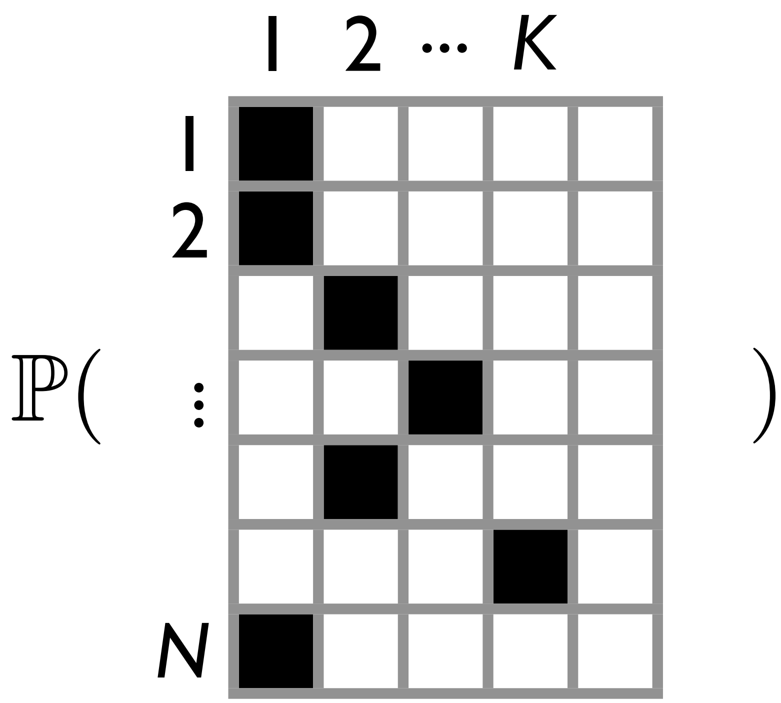
“features”,
“topics”

- Exchangeable
- Finite # of features per data point

Characterizations

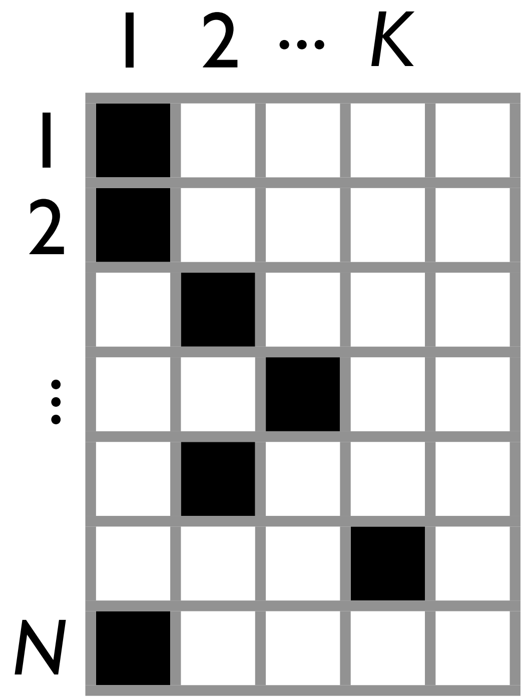
- Exchangeable cluster distributions are characterized
- What about exchangeable feature distributions?

Exchangeable probability functions



Exchangeable probability functions

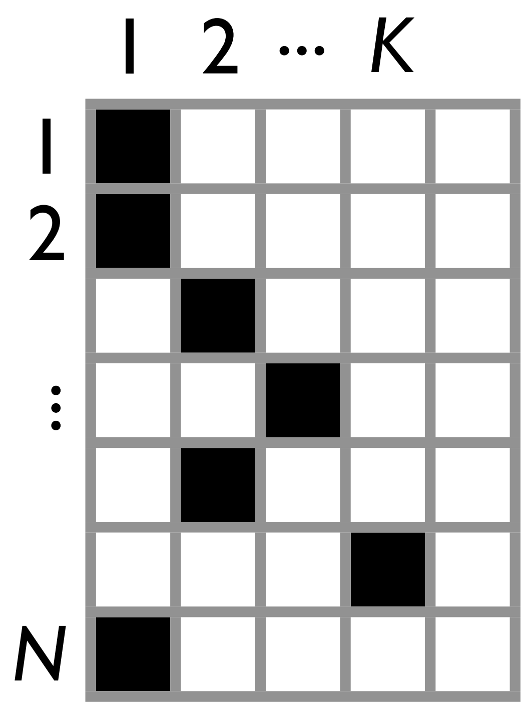
$\mathbb{P}($



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

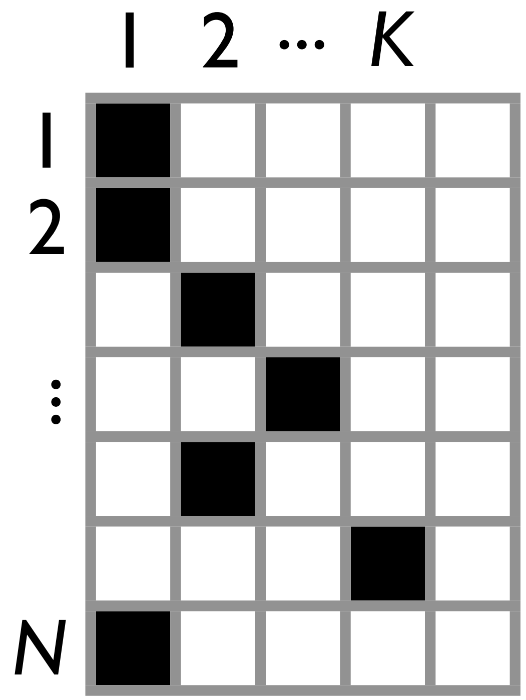
Exchangeable probability functions

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]

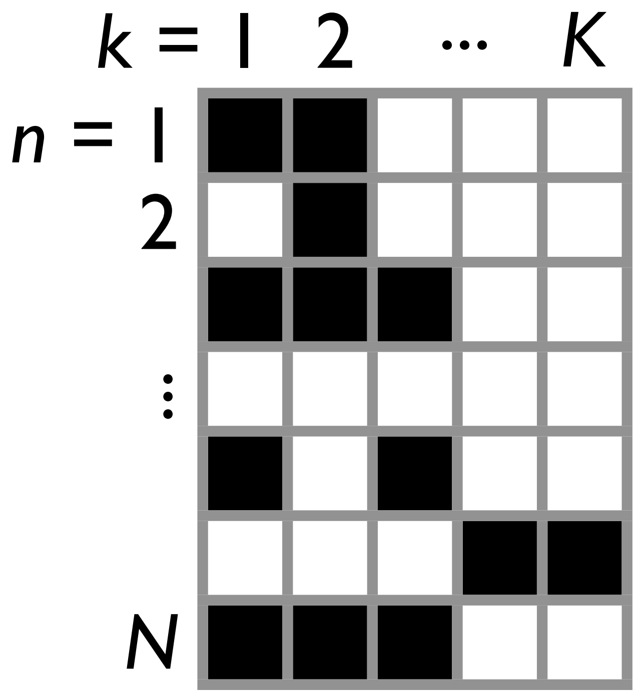
Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process

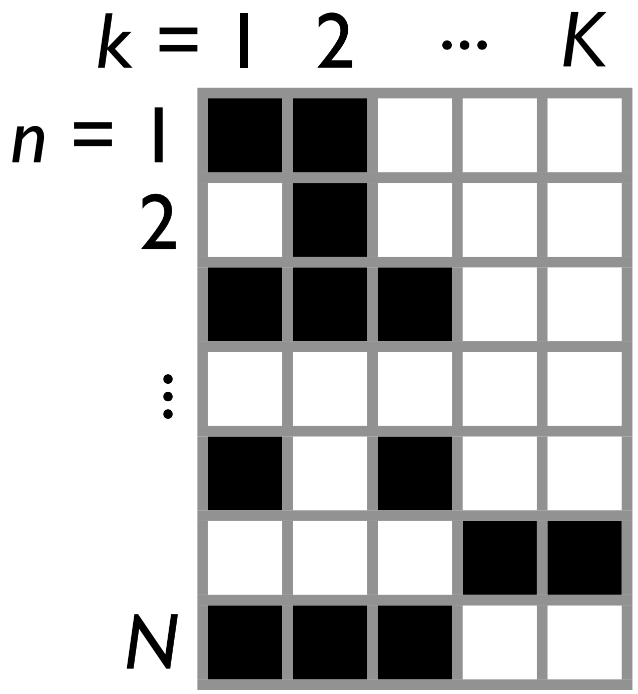
[Griffiths, Ghahramani 2005]

Example: Indian buffet process



[Griffiths, Ghahramani 2005]

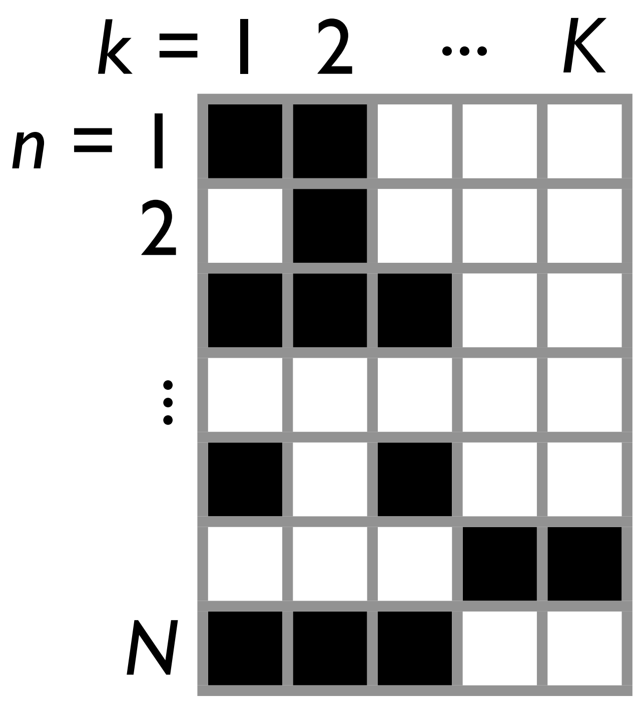
Example: Indian buffet process



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

[Griffiths, Ghahramani 2005]

Example: Indian buffet process

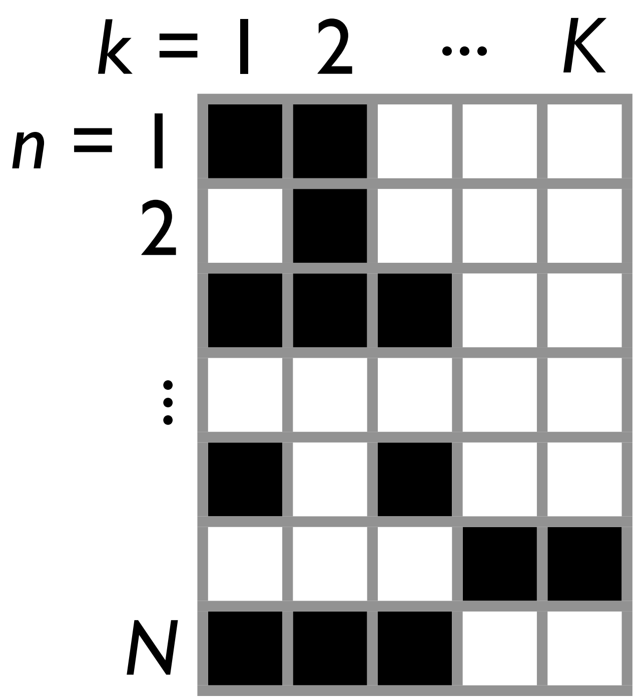


For n = 1, 2, ..., N

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 2005]

Example: Indian buffet process

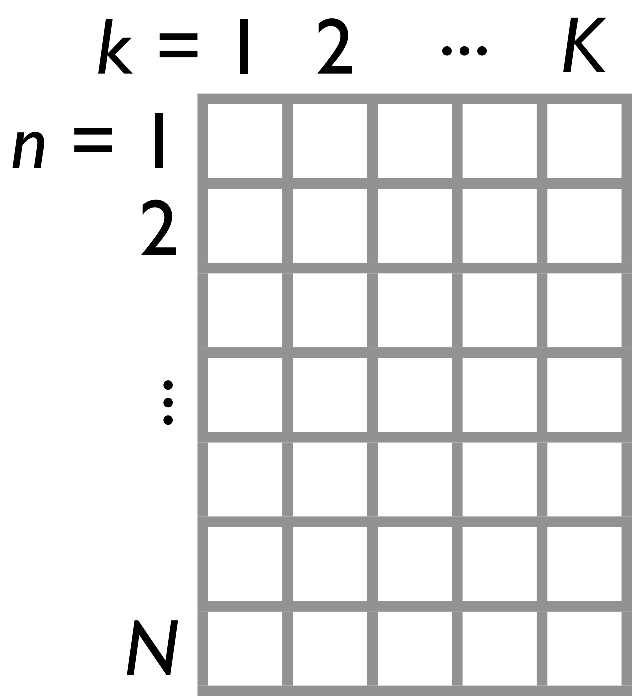


For n = 1, 2, ..., N

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 2005]

Example: Indian buffet process

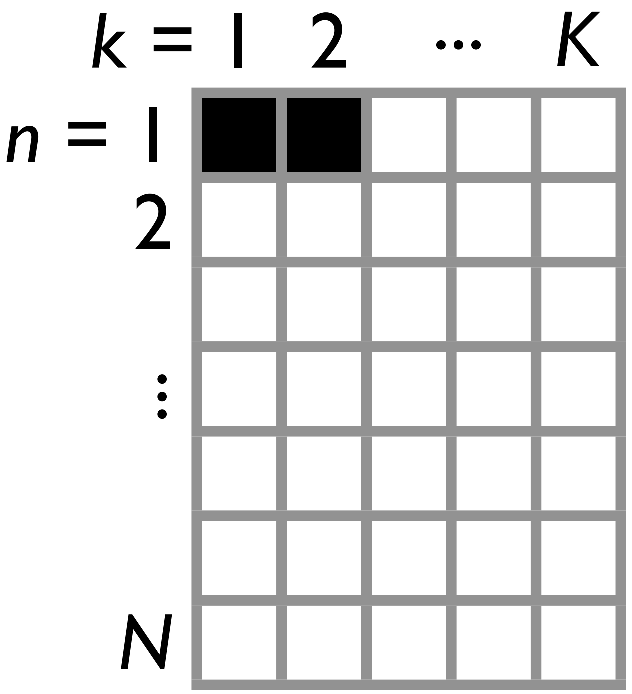


For n = 1, 2, ..., N

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 2005]

Example: Indian buffet process

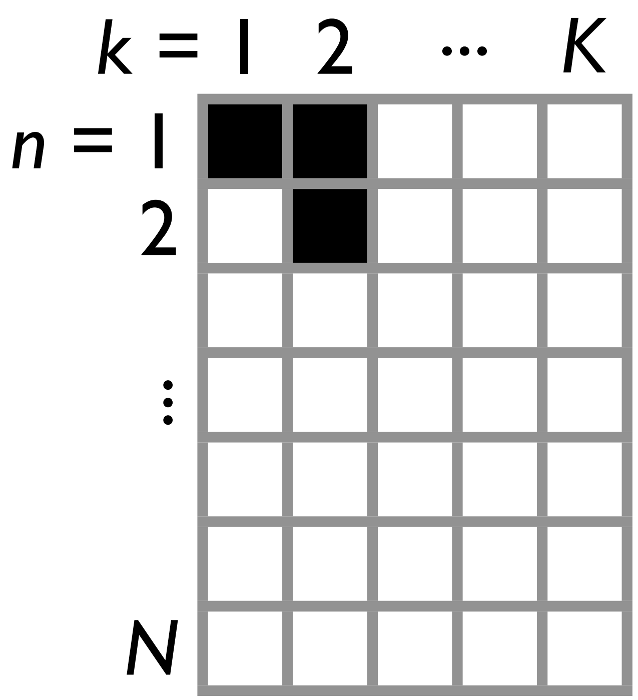


For n = 1, 2, ..., N

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 2005]

Example: Indian buffet process

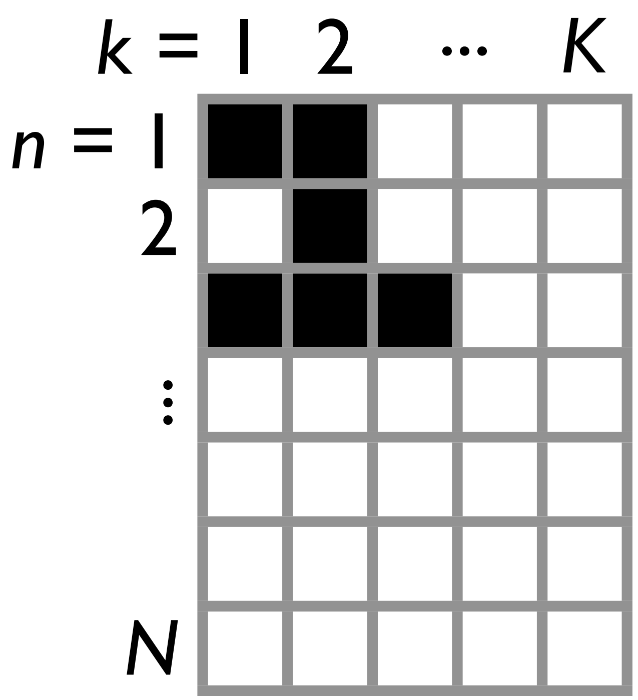


For n = 1, 2, ..., N

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 2005]

Example: Indian buffet process

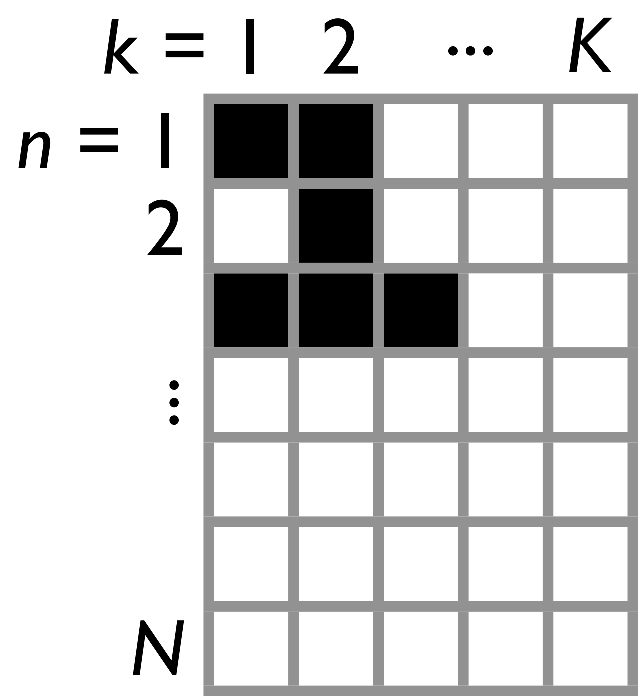


For n = 1, 2, ..., N

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 2005]

Example: Indian buffet process

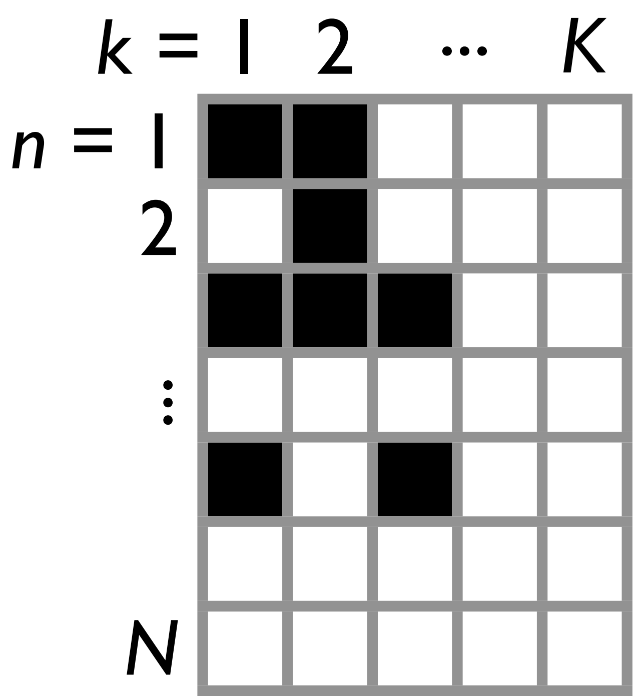


For n = 1, 2, ..., N

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

[Griffiths, Ghahramani 2005]

Example: Indian buffet process



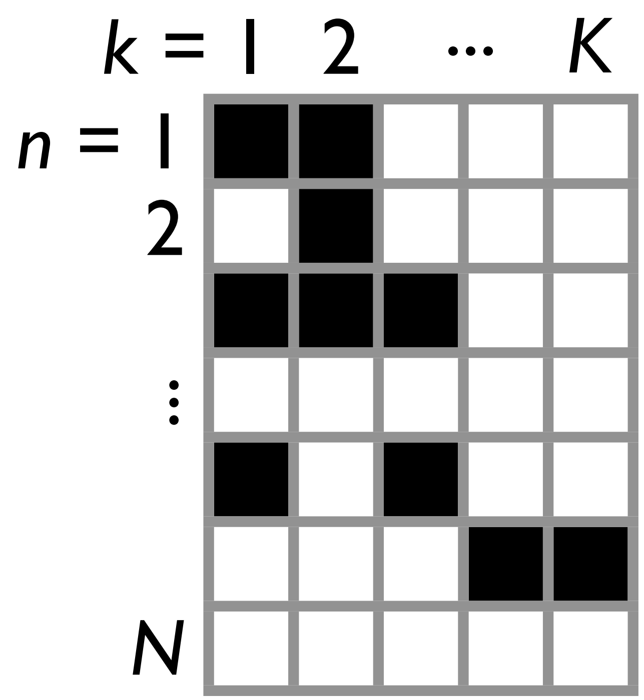
For n = 1, 2, ..., N

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 2005]

Example: Indian buffet process



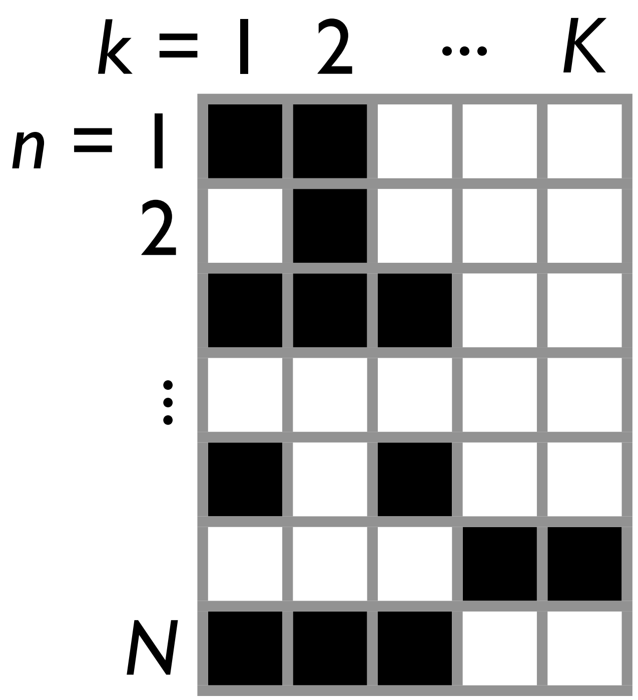
For n = 1, 2, ..., N

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 2005]

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 2005]

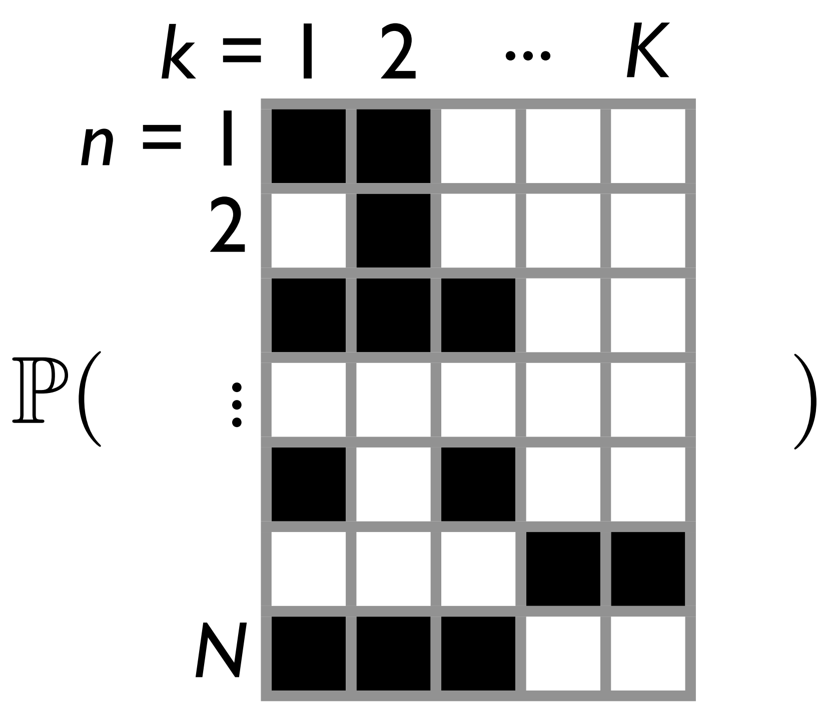
Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

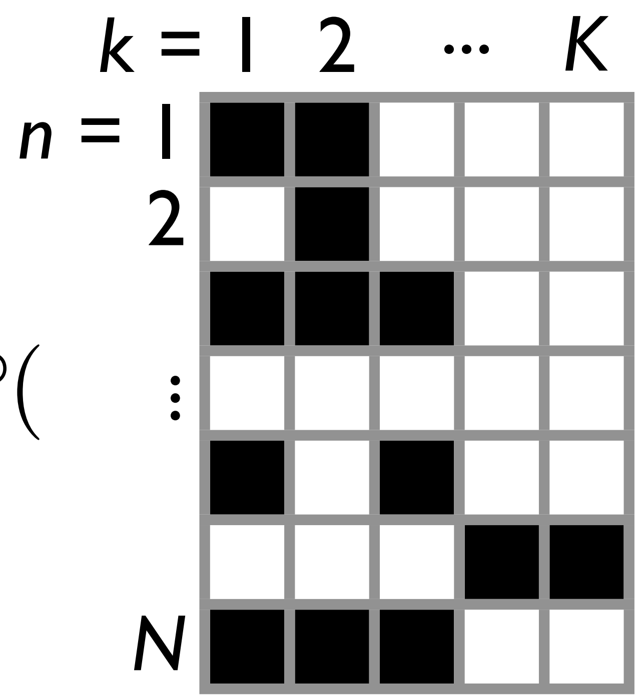


Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($



$)$

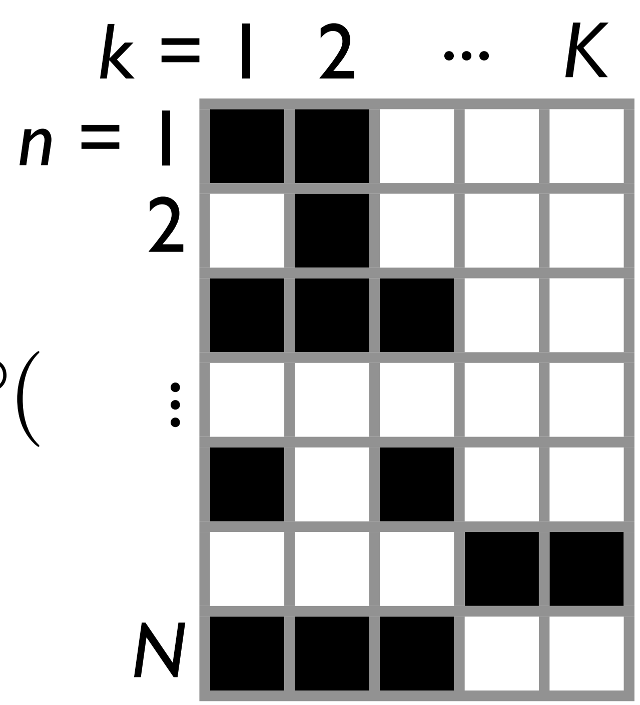
$$= \frac{1}{K_N!} (\theta\gamma)^{K_N} \exp\left(-\theta\gamma \sum_{n=1}^{N} (\theta + n - 1)^{-1} \right) \prod_{k=1}^{K_N} \frac{\Gamma(S_{N,k}) \Gamma(N - S_{N,k} + \theta)}{\Gamma(N + \theta)}$$

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($



$)$

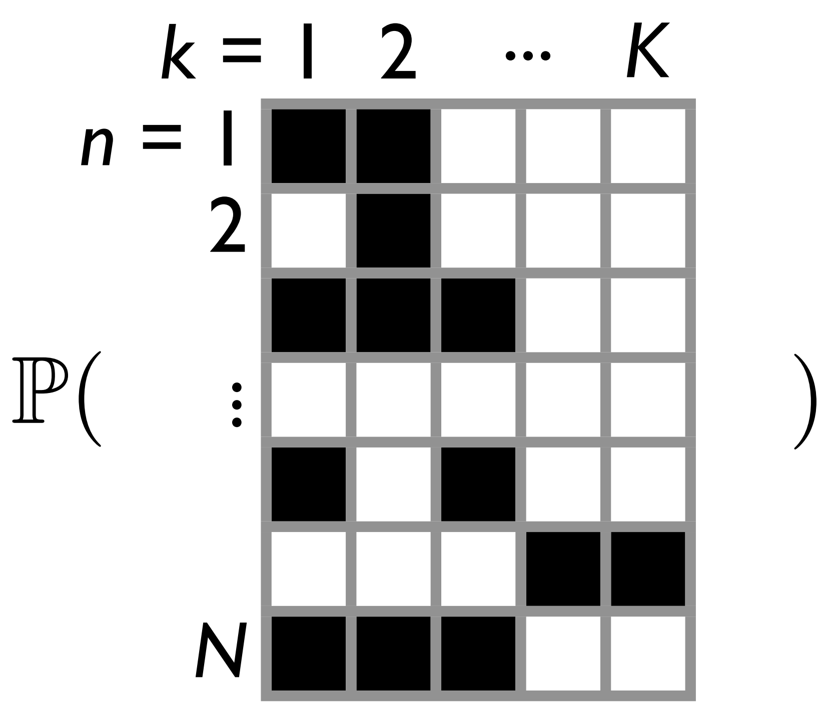
Size of *k*th feature

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

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)



Size of *k*th feature

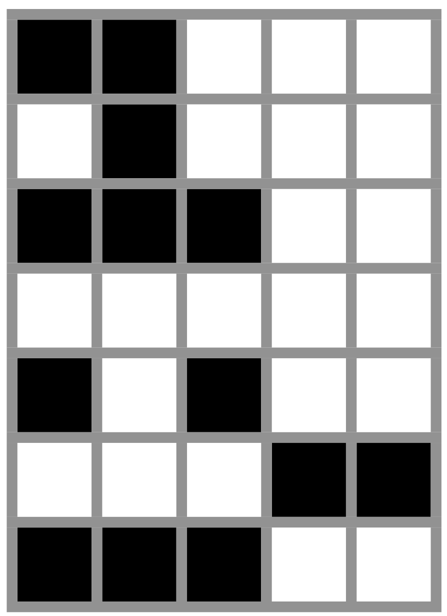
Number of features

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

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($  $)$

Number of data points

Size of *k*th feature

Number of features

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

Exchangeable probability functions

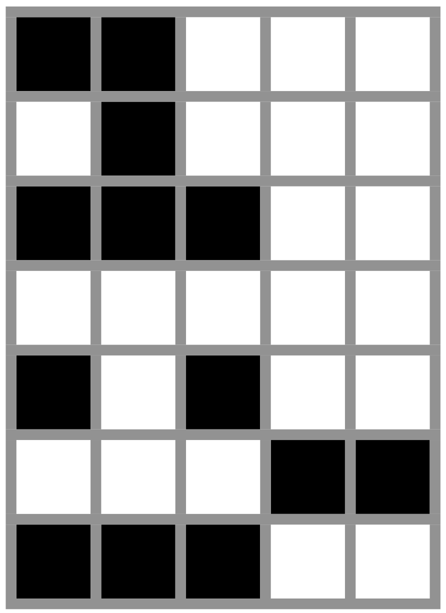
"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

k = 1 2 ... K

n = 1

2

$\mathbb{P}($ ⋮  $)$

N

Number of data points

Size of kth feature

Number of features

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

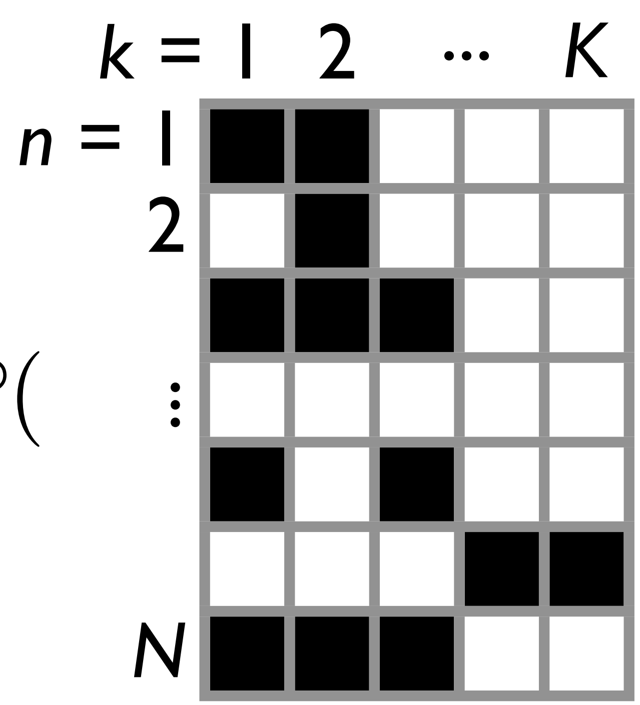
$$= p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})$$

[Broderick, Jordan, Pitman 2012]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Example: Indian buffet process (IBP)

$\mathbb{P}($


$)$

Number of data points

Size of *k*th feature

Number of features

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

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

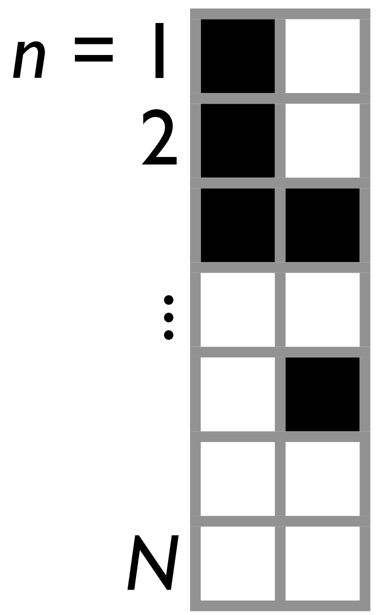
"EFPF"

[Broderick, Jordan, Pitman 2012]

Exchangeable probability functions

"Exchangeable feature probability function" (EFPF)?

Counterexample

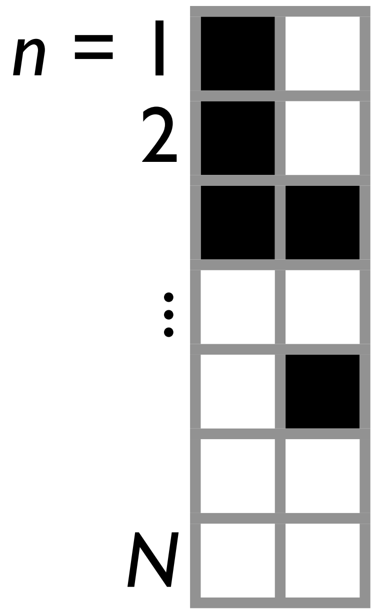


[Broderick, Jordan, Pitman 2012]

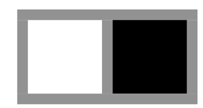
Exchangeable probability functions

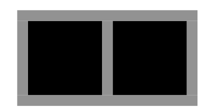
"Exchangeable feature probability function" (EFPF)?

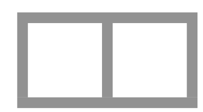
Counterexample



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

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

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

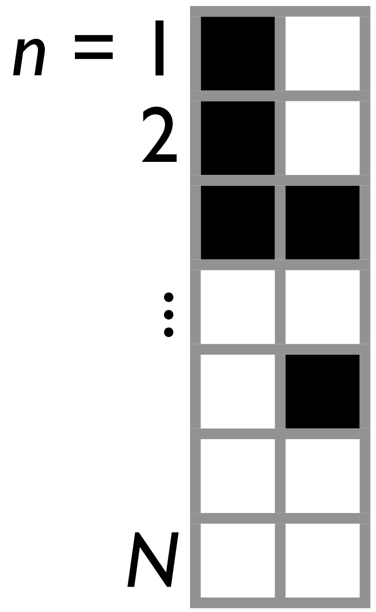
$\mathbb{P}(\text{row} =$  $) = p_4$

[Broderick, Jordan, Pitman 2012]

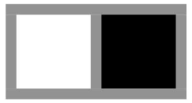
Exchangeable probability functions

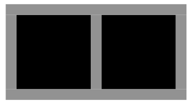
"Exchangeable feature probability function" (EFPF)?

Counterexample

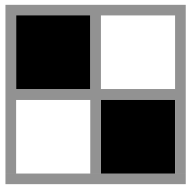


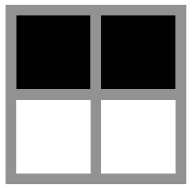
$\mathbb{P}(\text{row} =$  $) = p_1$

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

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

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



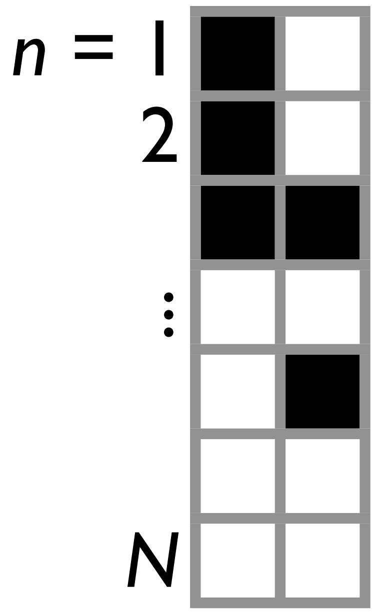


[Broderick, Jordan, Pitman 2012]

Exchangeable probability functions

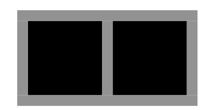
"Exchangeable feature probability function" (EFPF)?

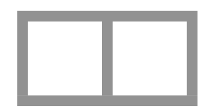
Counterexample

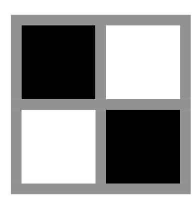
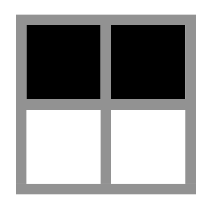


$\mathbb{P}(\text{row} =$  $) = p_1$

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

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

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

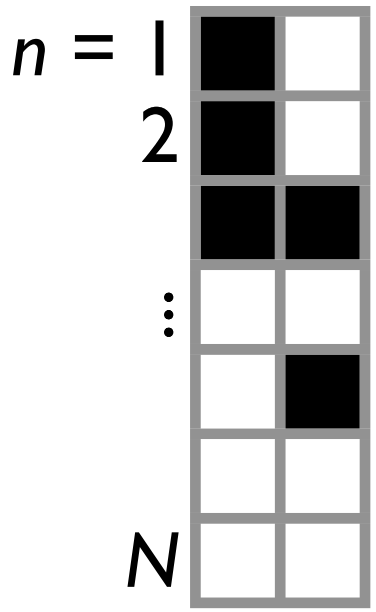
$\mathbb{P}($  $)$ $\mathbb{P}($  $)$

[Broderick, Jordan, Pitman 2012]

Exchangeable probability functions

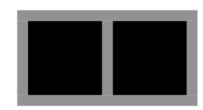
"Exchangeable feature probability function" (EFPF)?

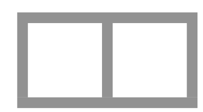
Counterexample

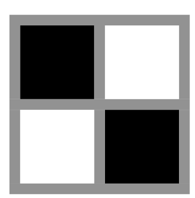
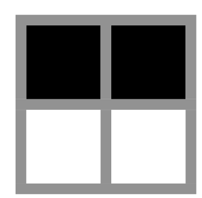


$\mathbb{P}(\text{row} =$  $) = p_1$

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

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

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

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

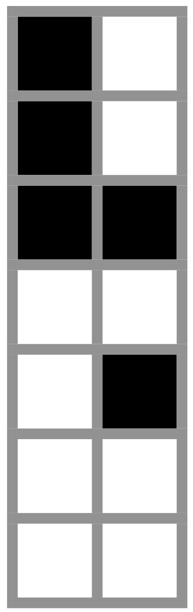
$p_1 p_2 \qquad p_3 p_4$

[Broderick, Jordan, Pitman 2012]

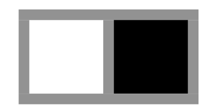
Exchangeable probability functions

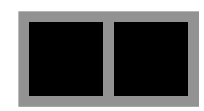
"Exchangeable feature probability function" (EFPF)?

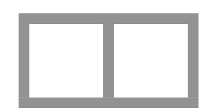
Counterexample

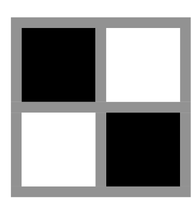
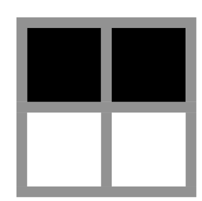


$\mathbb{P}(\text{row} =$  $) = p_1$

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

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

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

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

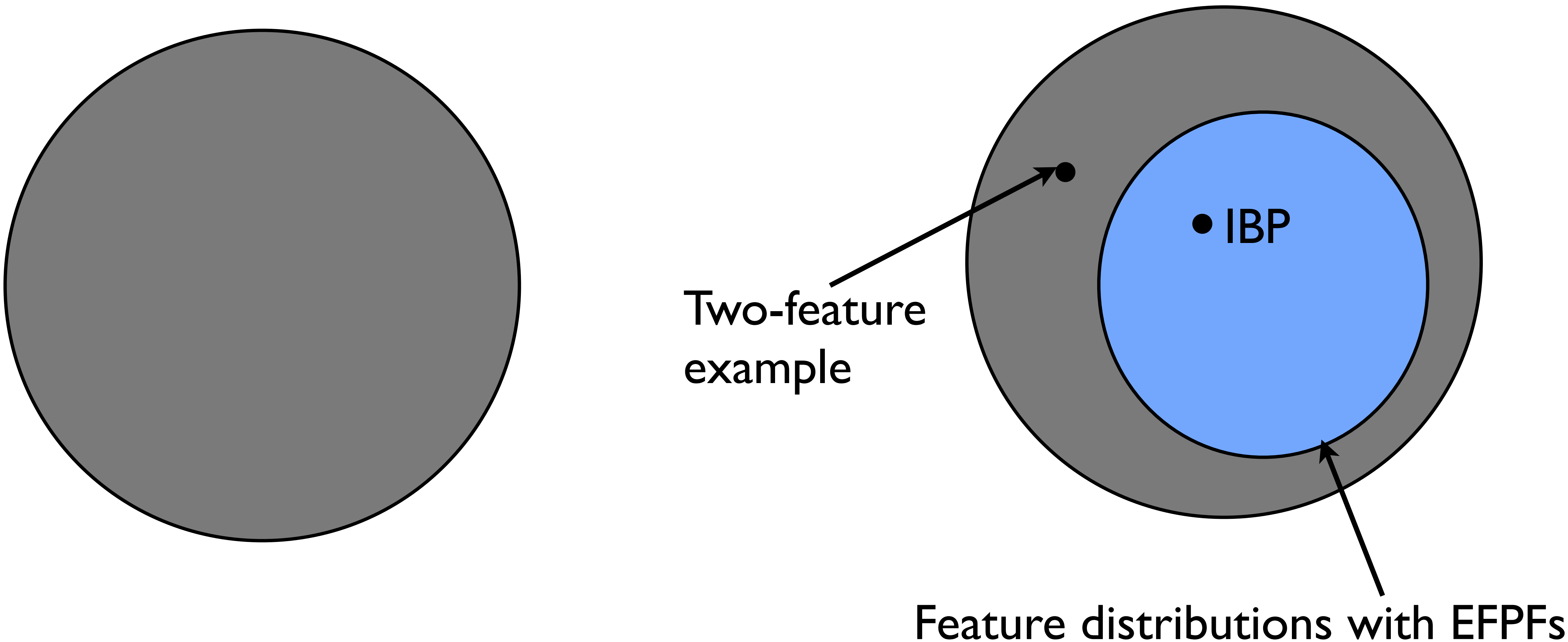
$$p_1 p_2 \neq p_3 p_4$$

[Broderick, Jordan, Pitman 2012]

Exchangeable probability functions

Exchangeable cluster distributions
= Cluster distributions with EPPFs

Exchangeable feature distributions



[Broderick, Jordan, Pitman 2012]

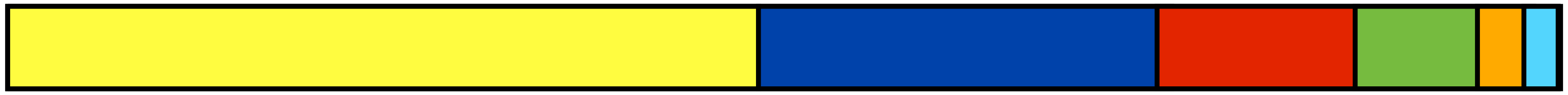
Paintboxes

Exchangeable partition: Kingman paintbox

[Kingman 1978]

Paintboxes

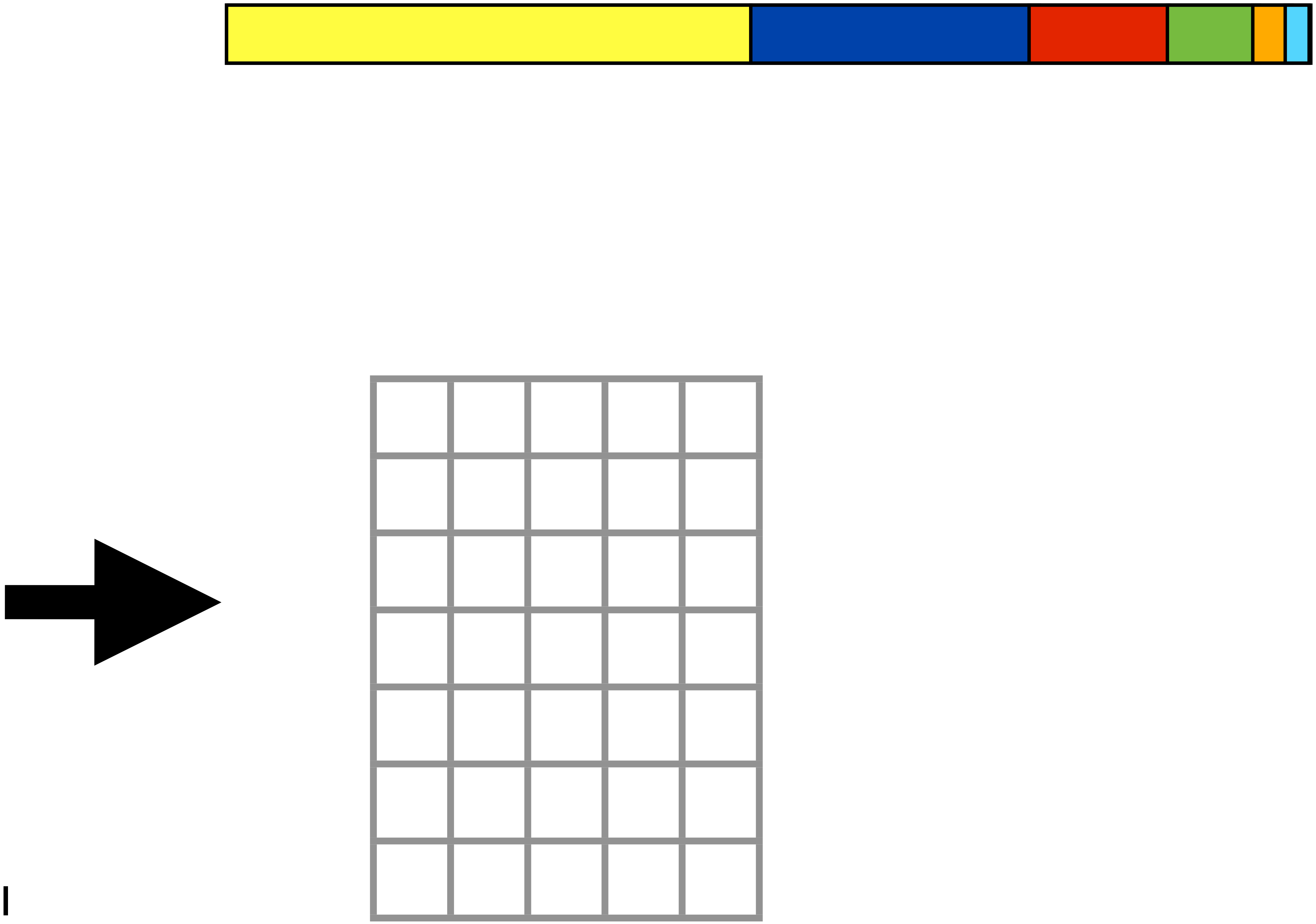
Exchangeable partition: Kingman paintbox



[Kingman 1978]

Paintboxes

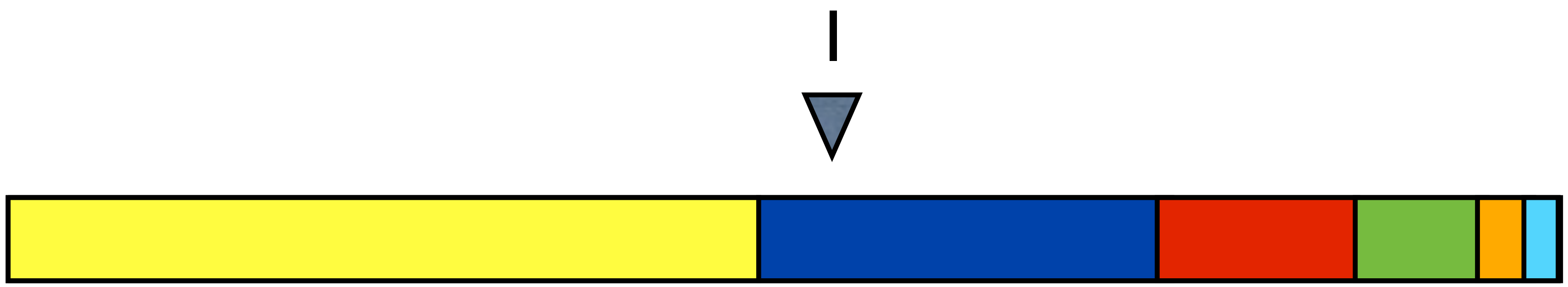
Exchangeable partition: Kingman paintbox

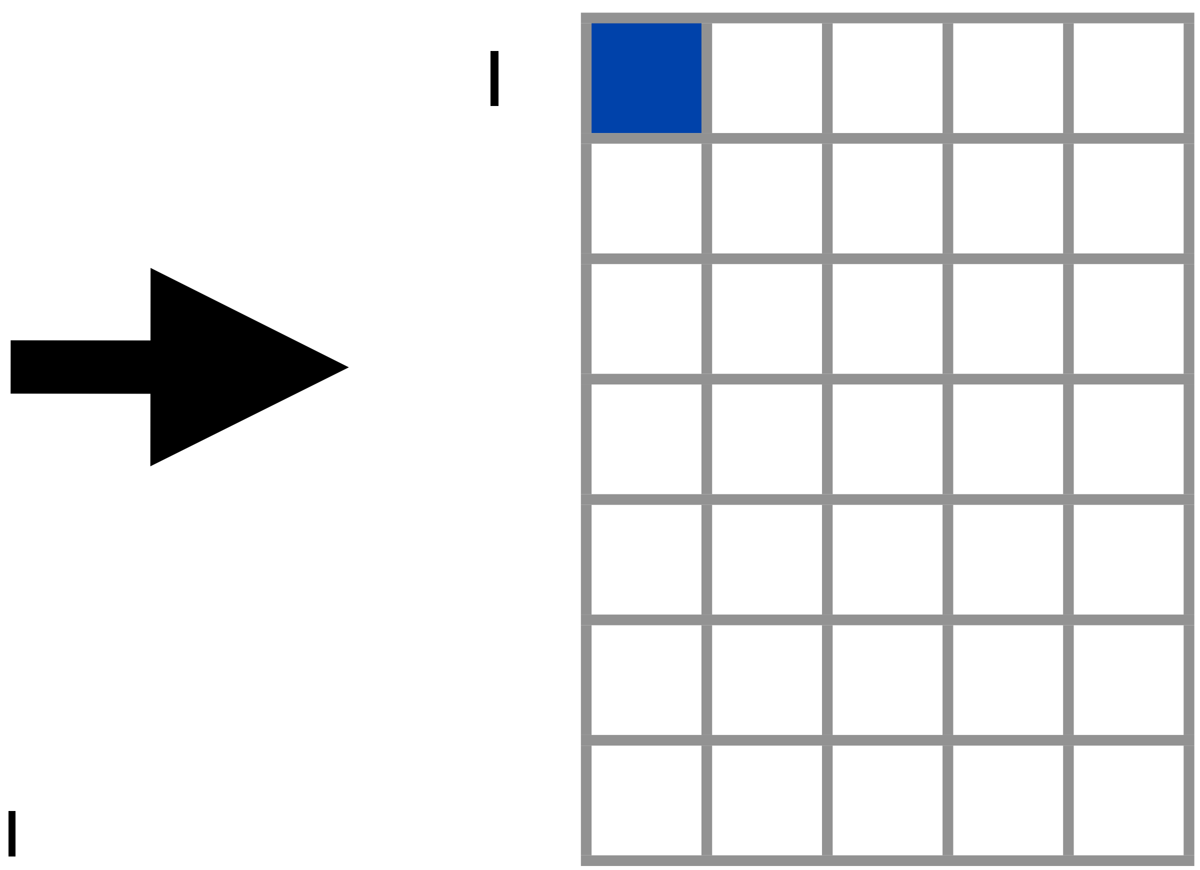


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

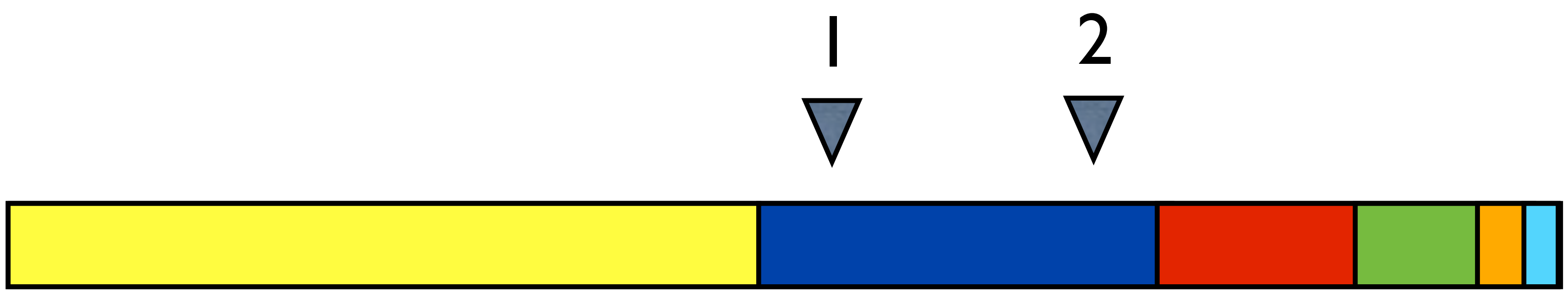


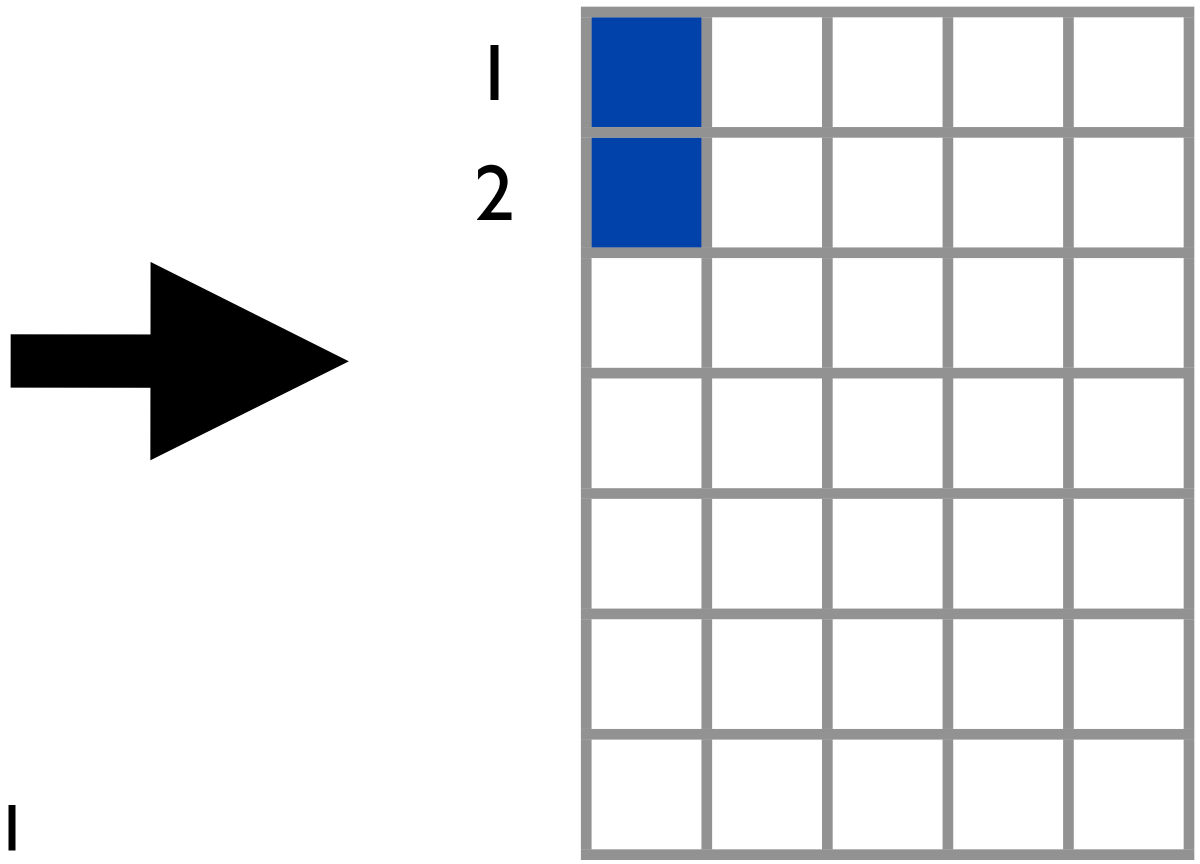


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

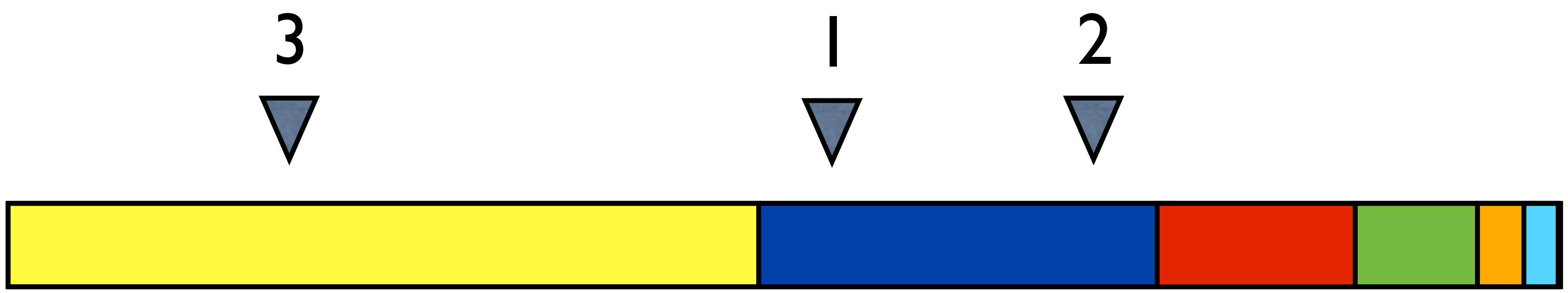


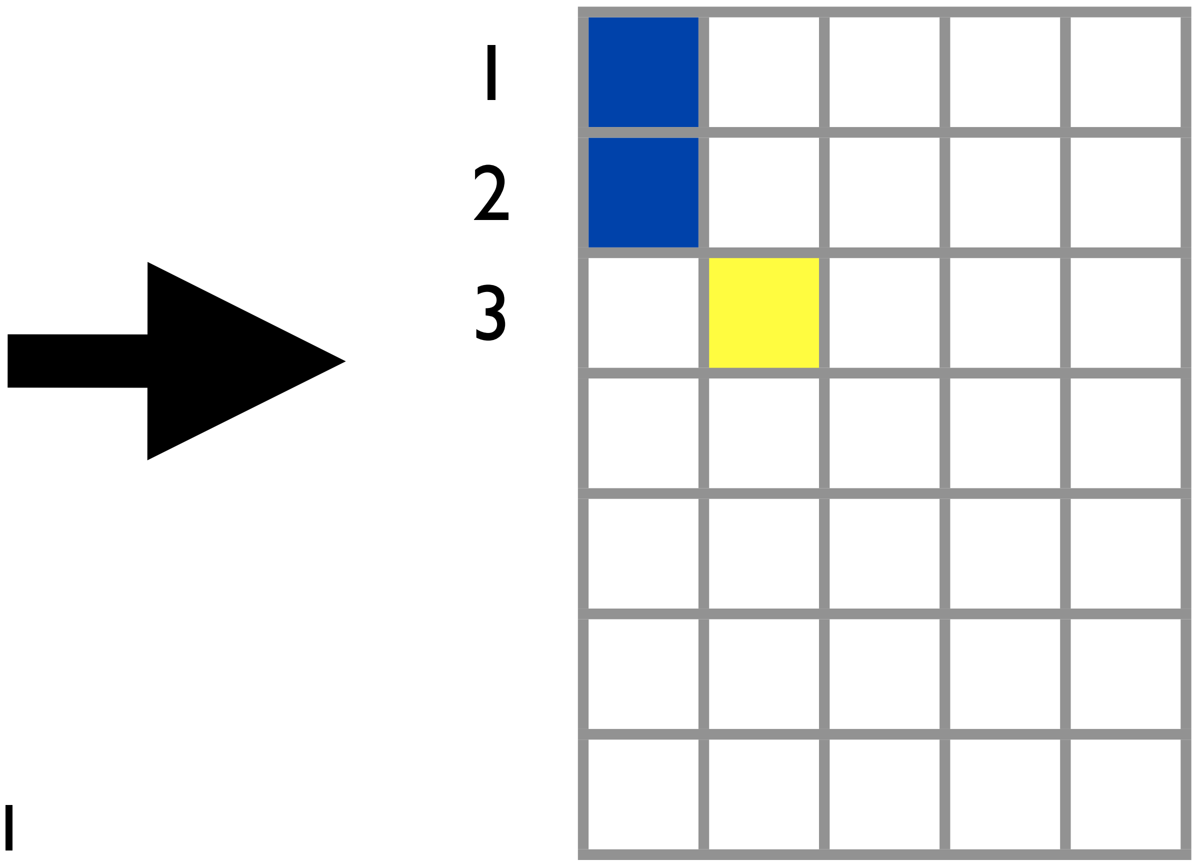


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

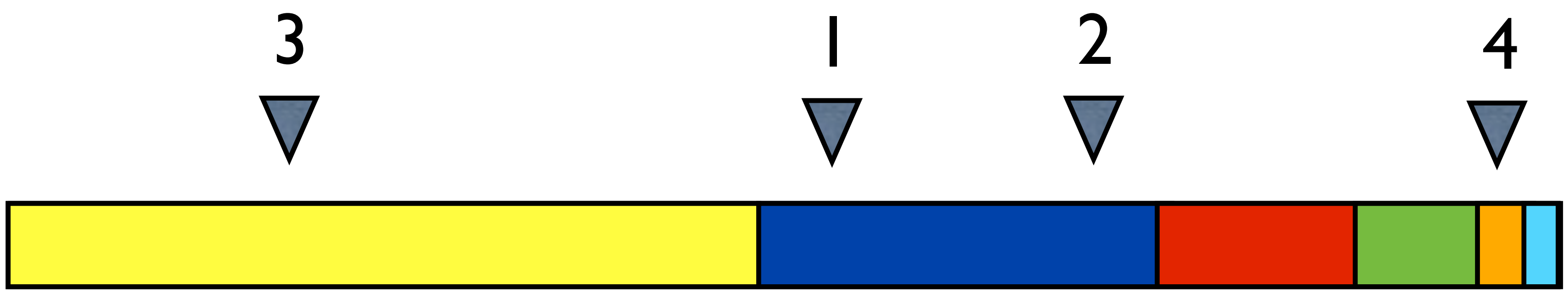


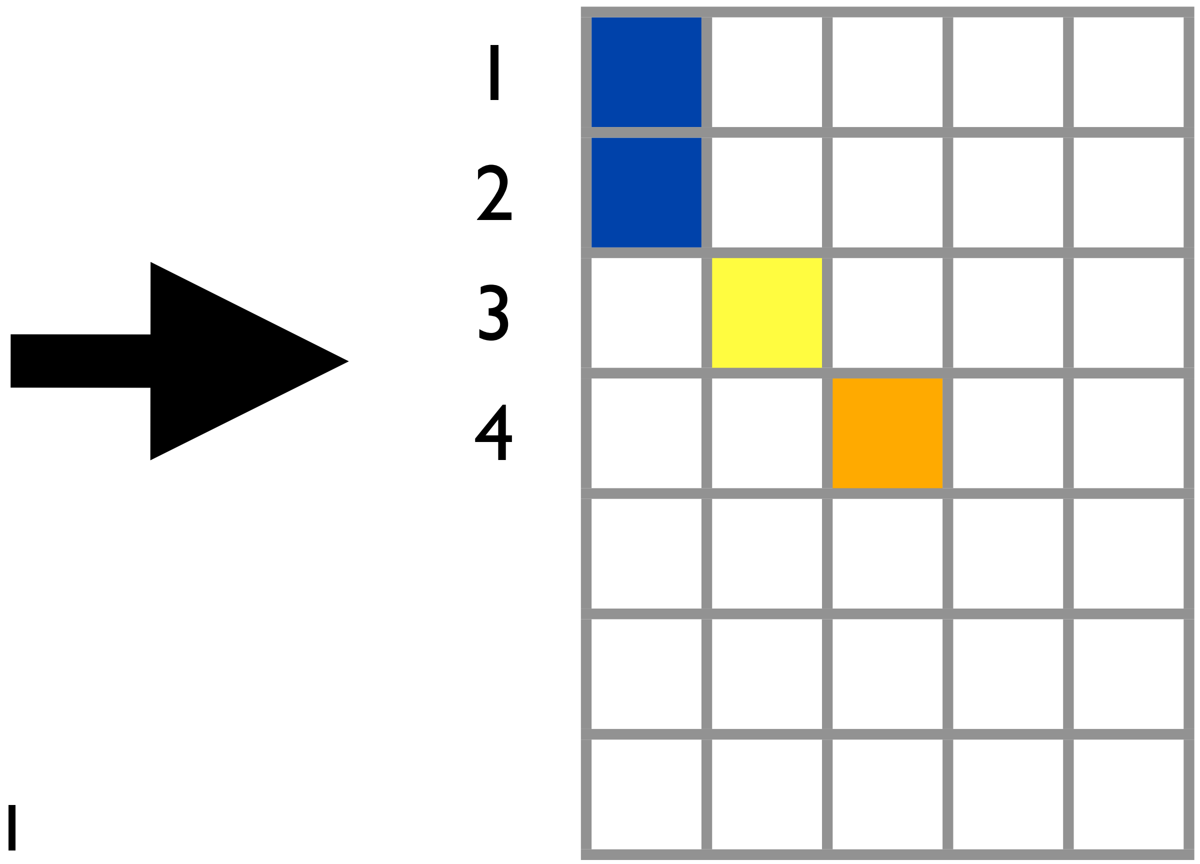


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

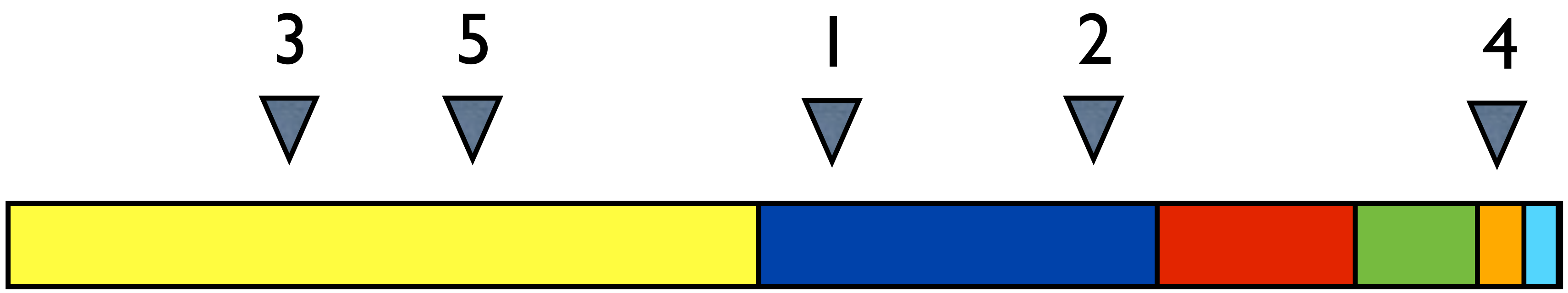


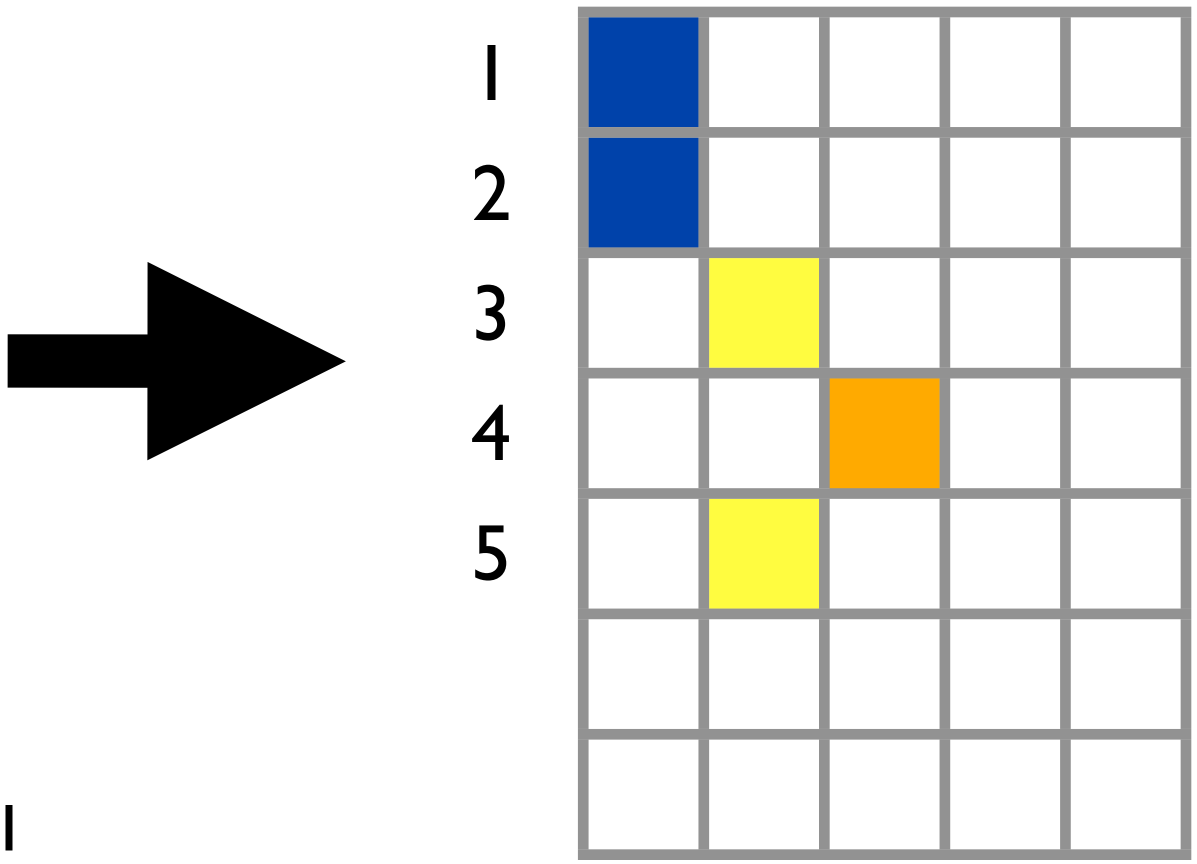


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

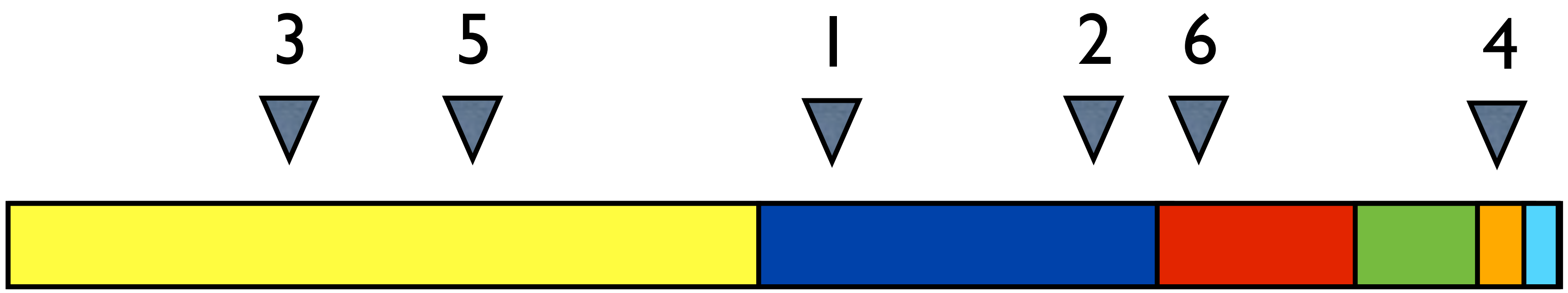


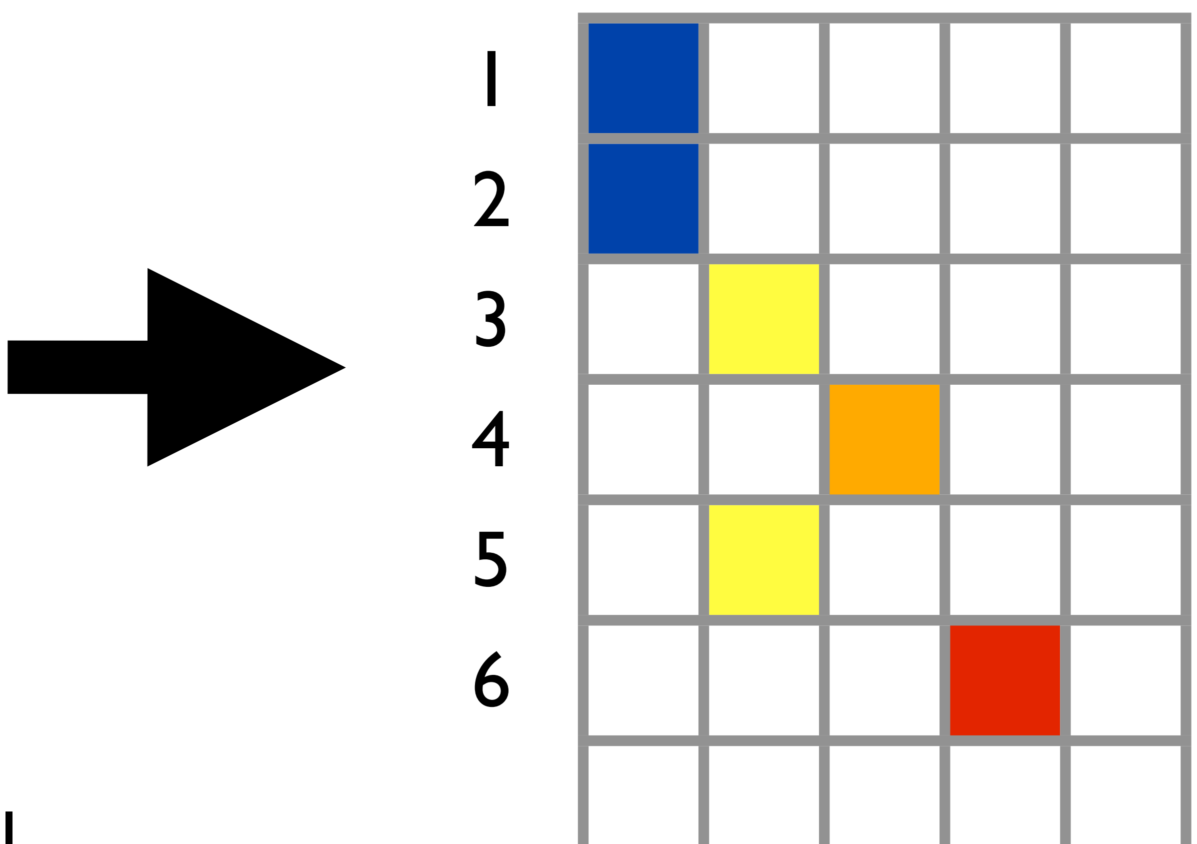


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

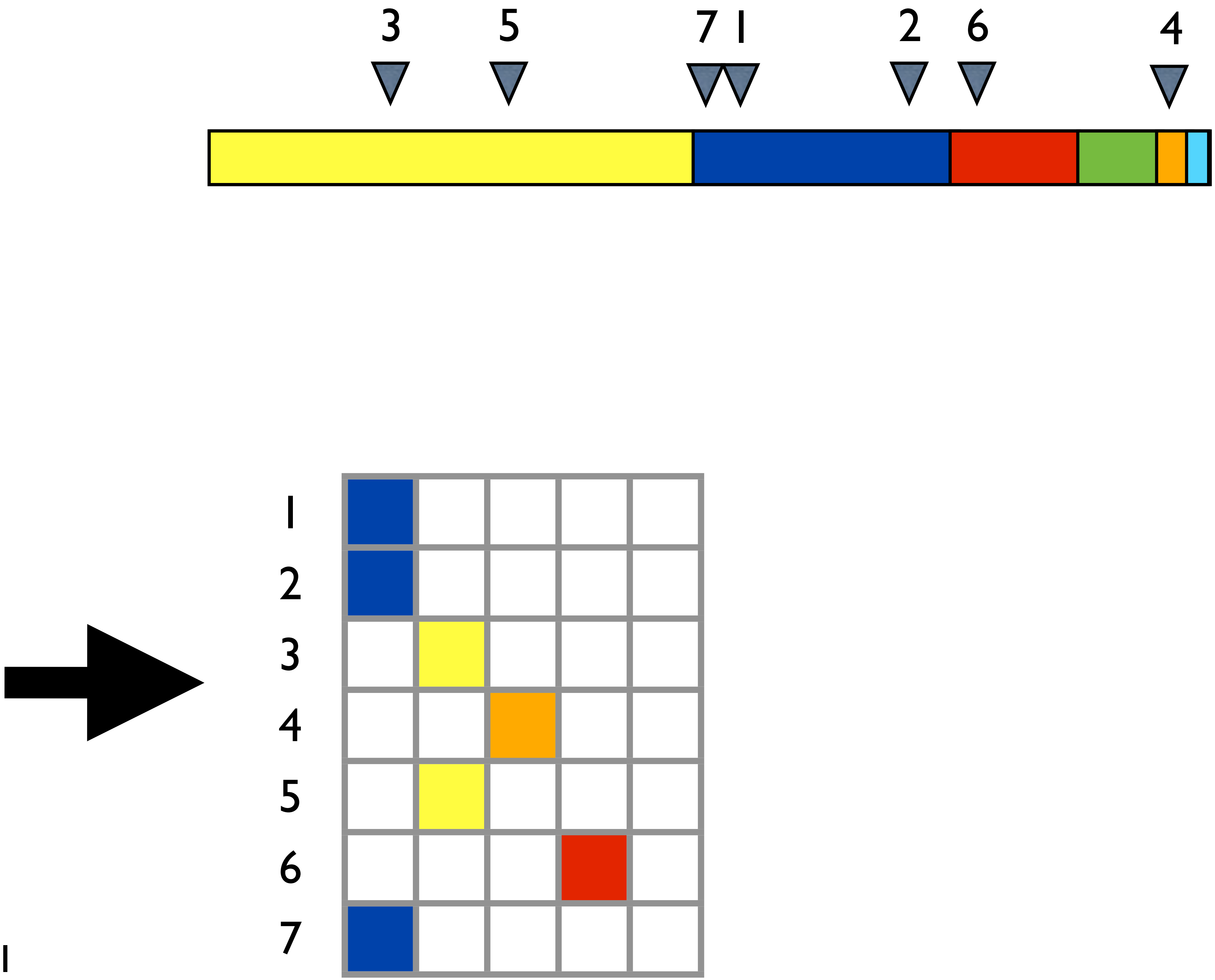




[Kingman 1978]

Paintboxes

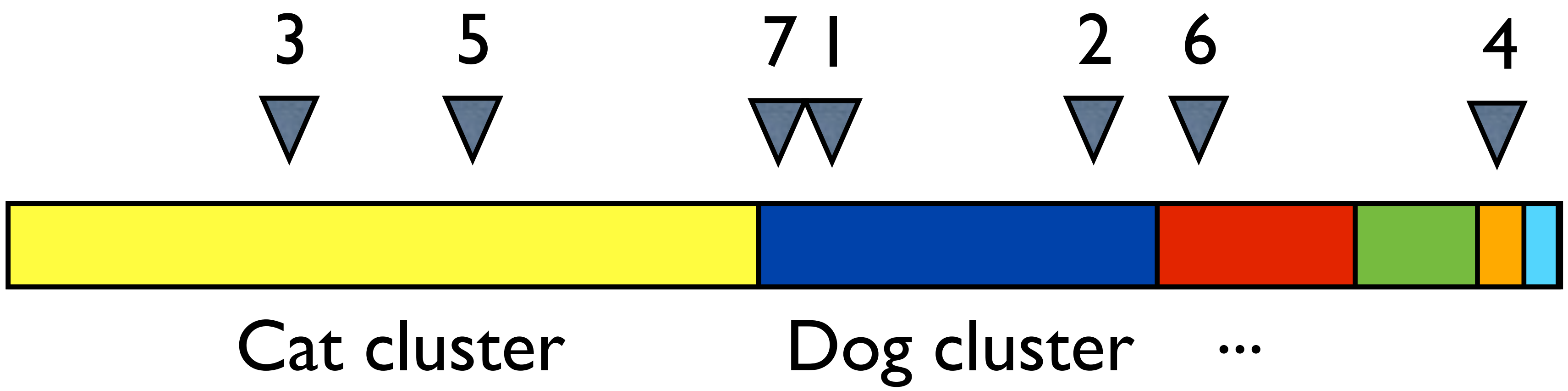
Exchangeable partition: Kingman paintbox

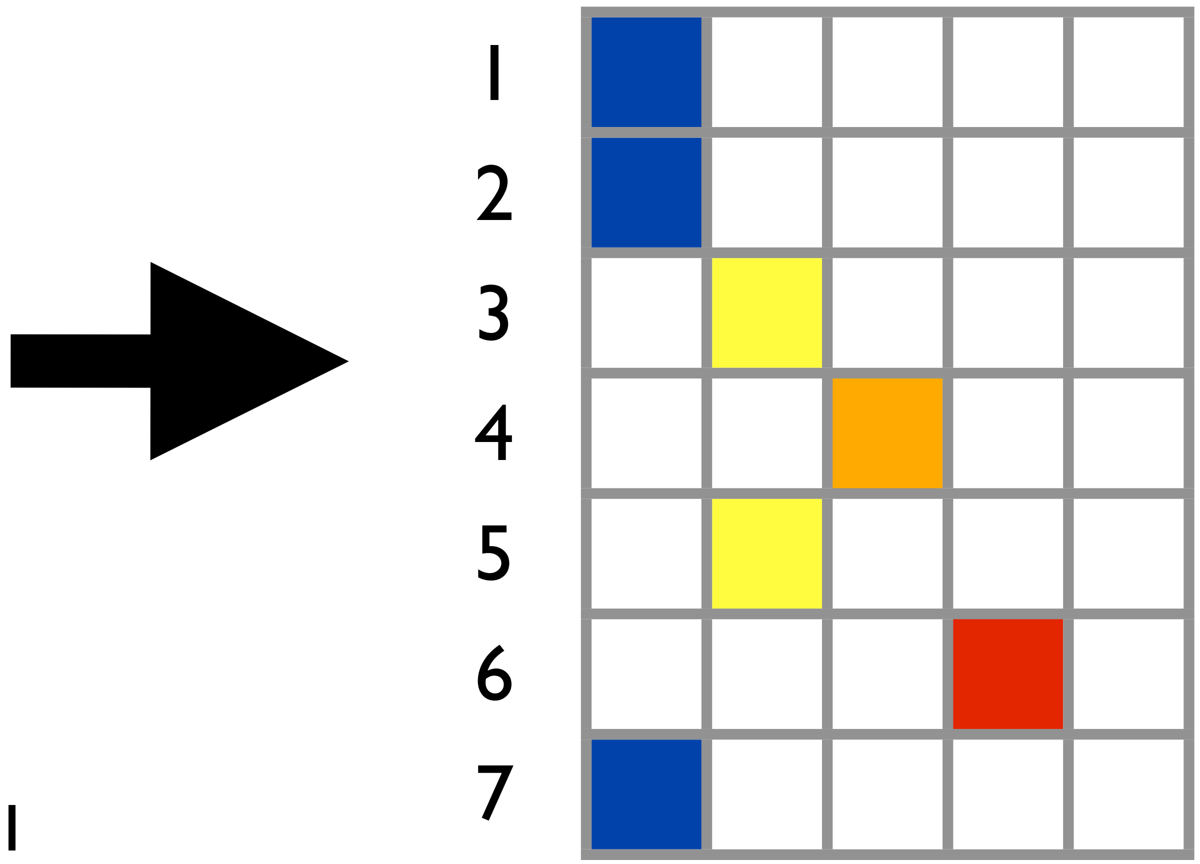


[Kingman 1978]

Paintboxes

Exchangeable partition: Kingman paintbox

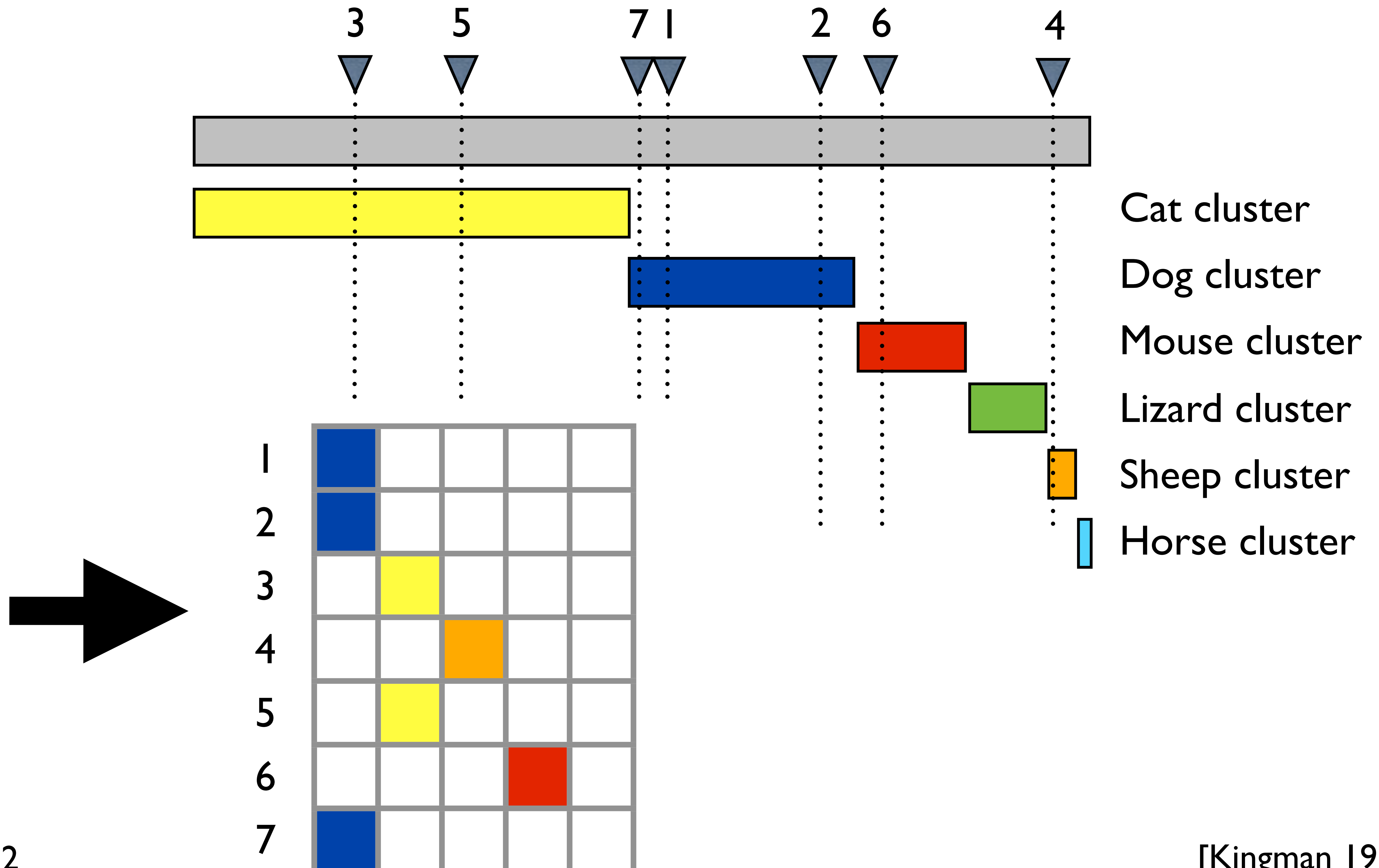




[Kingman 1978]

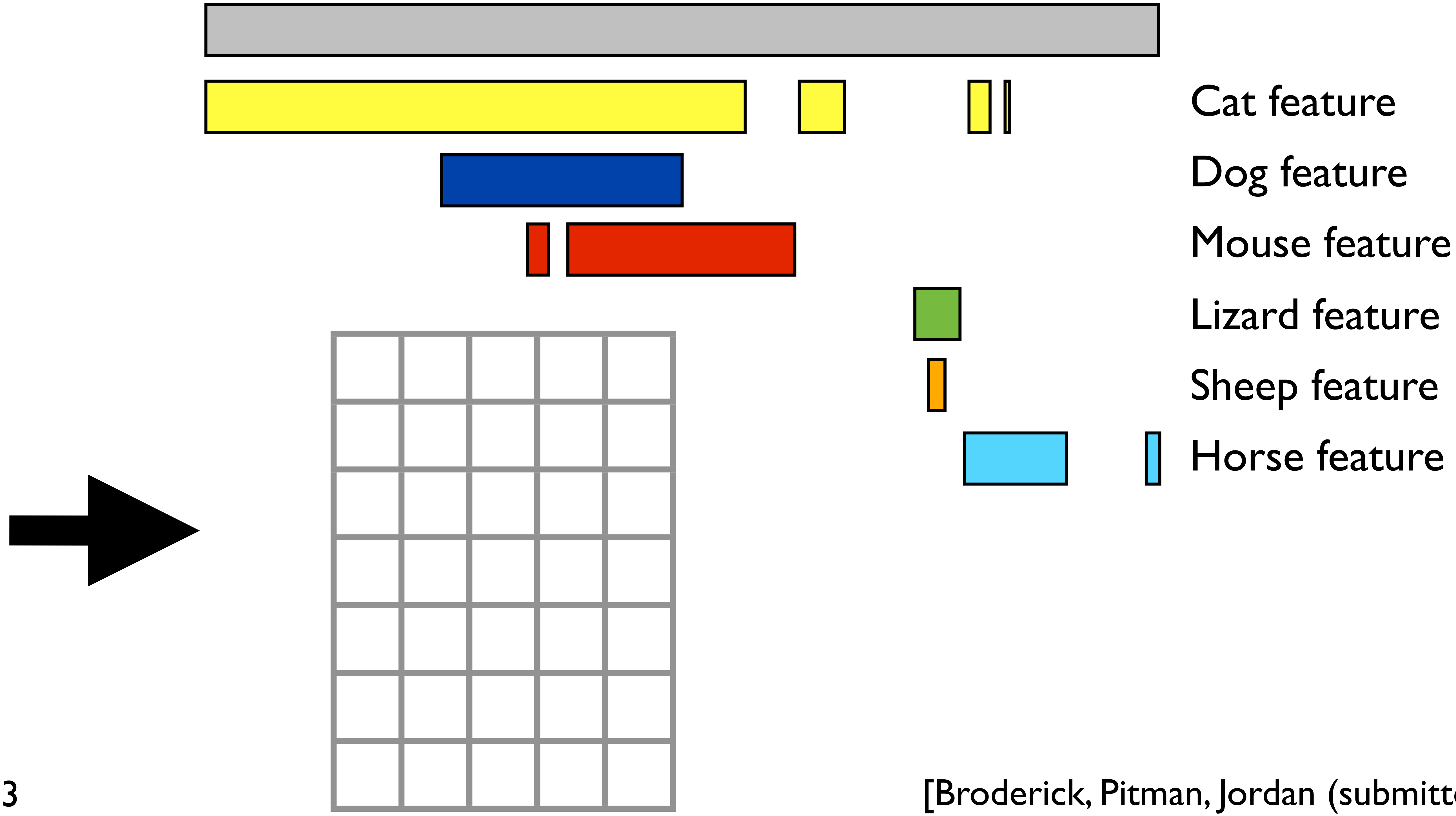
Paintboxes

Exchangeable partition: Kingman paintbox



[Kingman 1978]

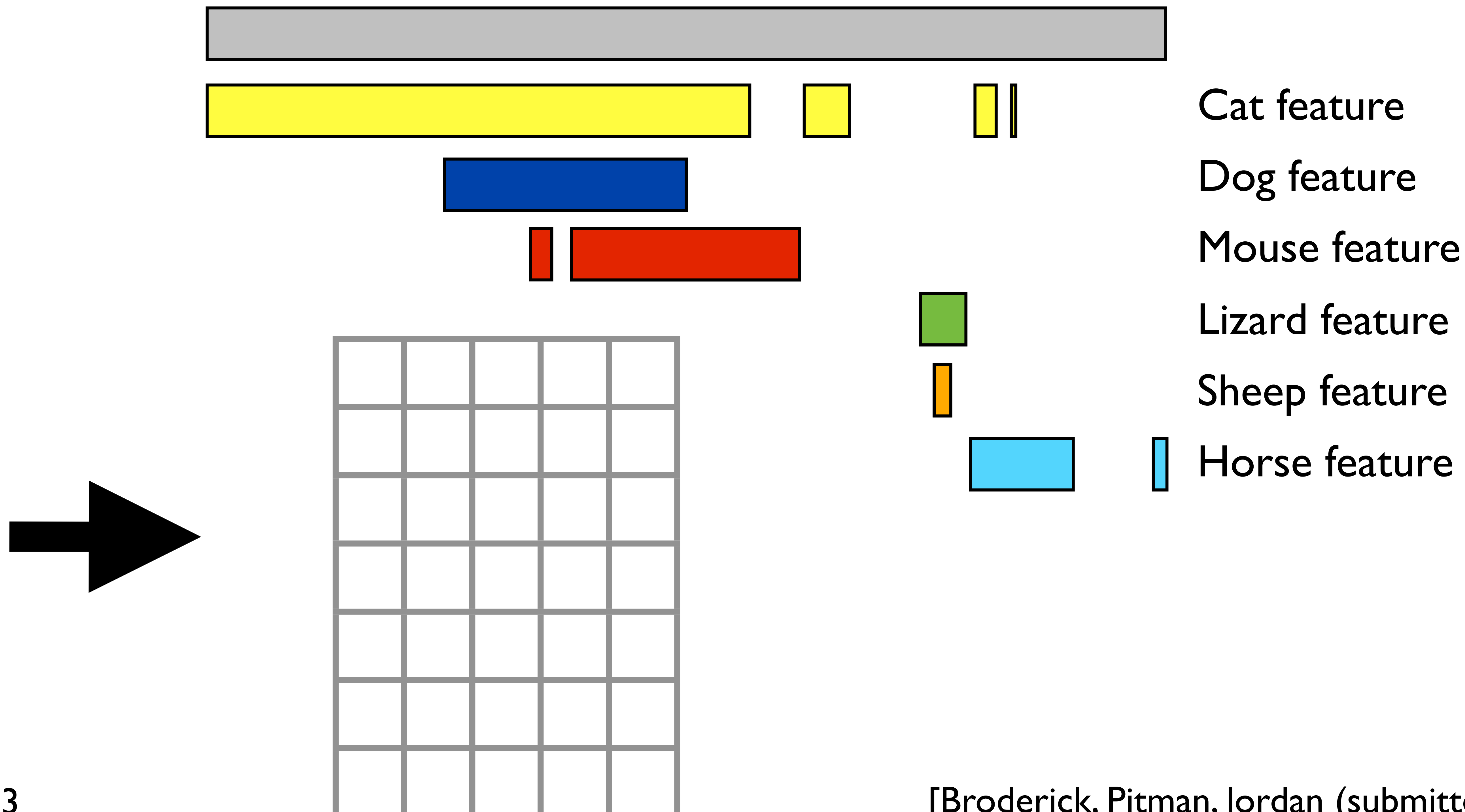
Paintboxes



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

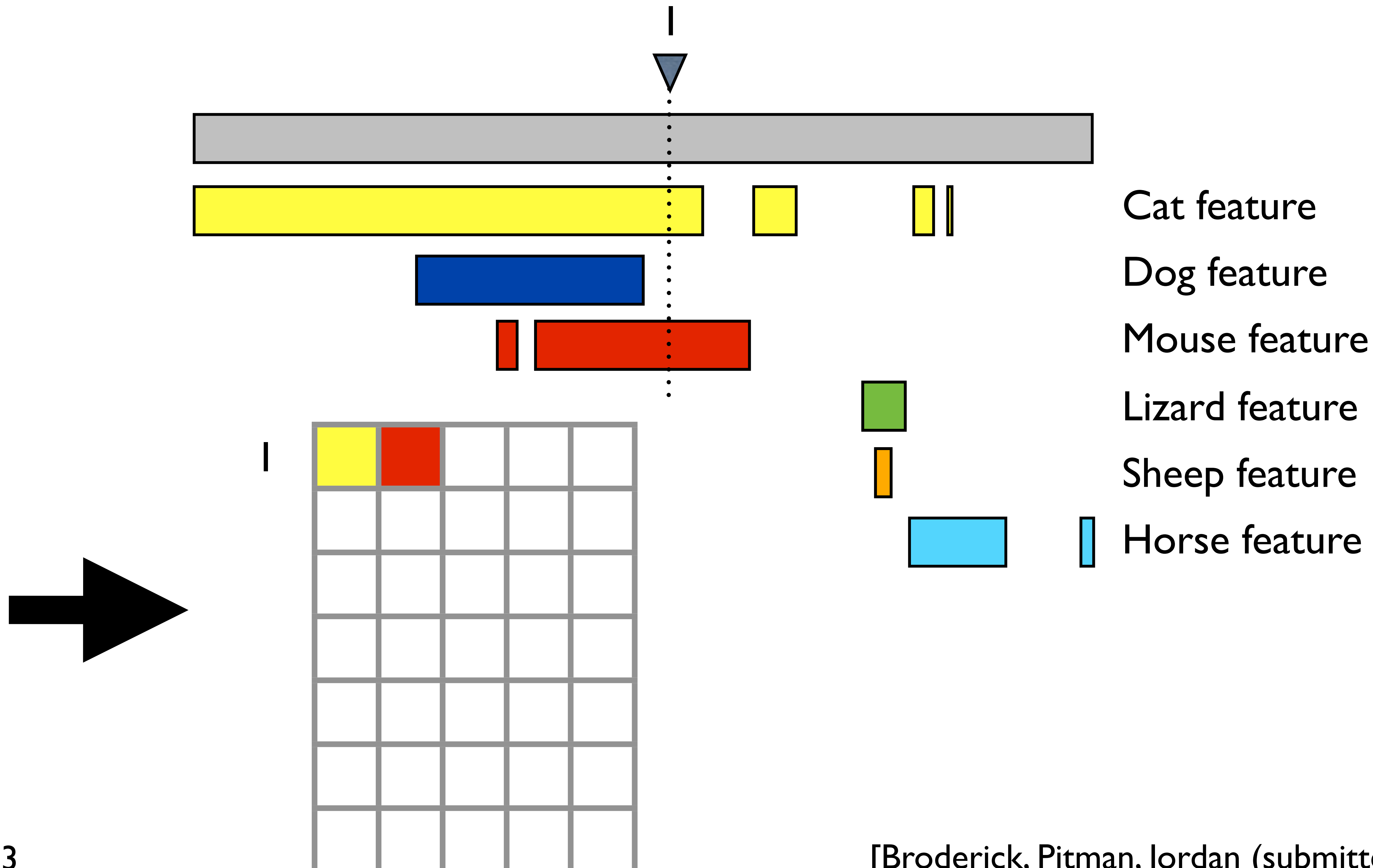
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

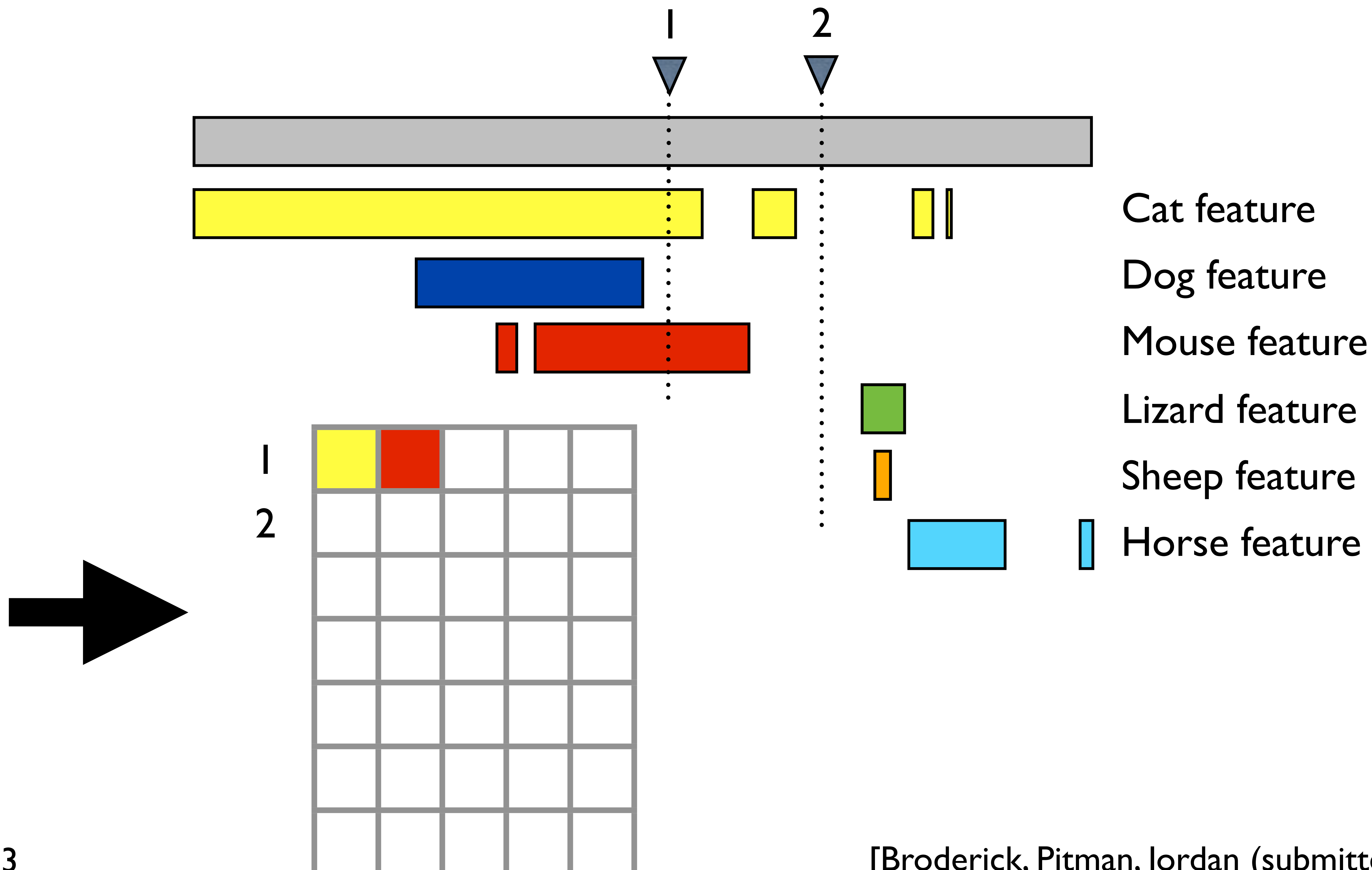
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

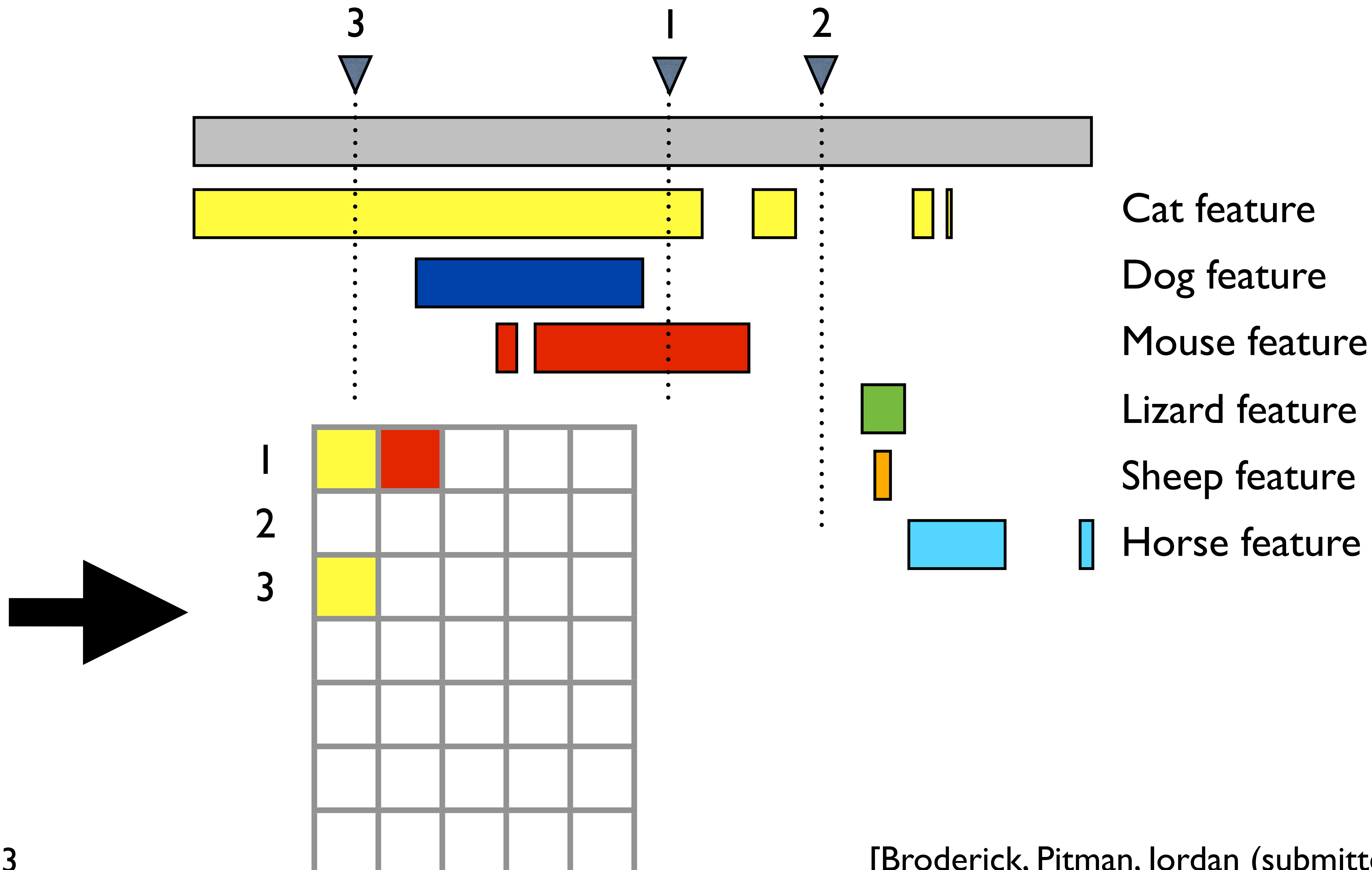
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

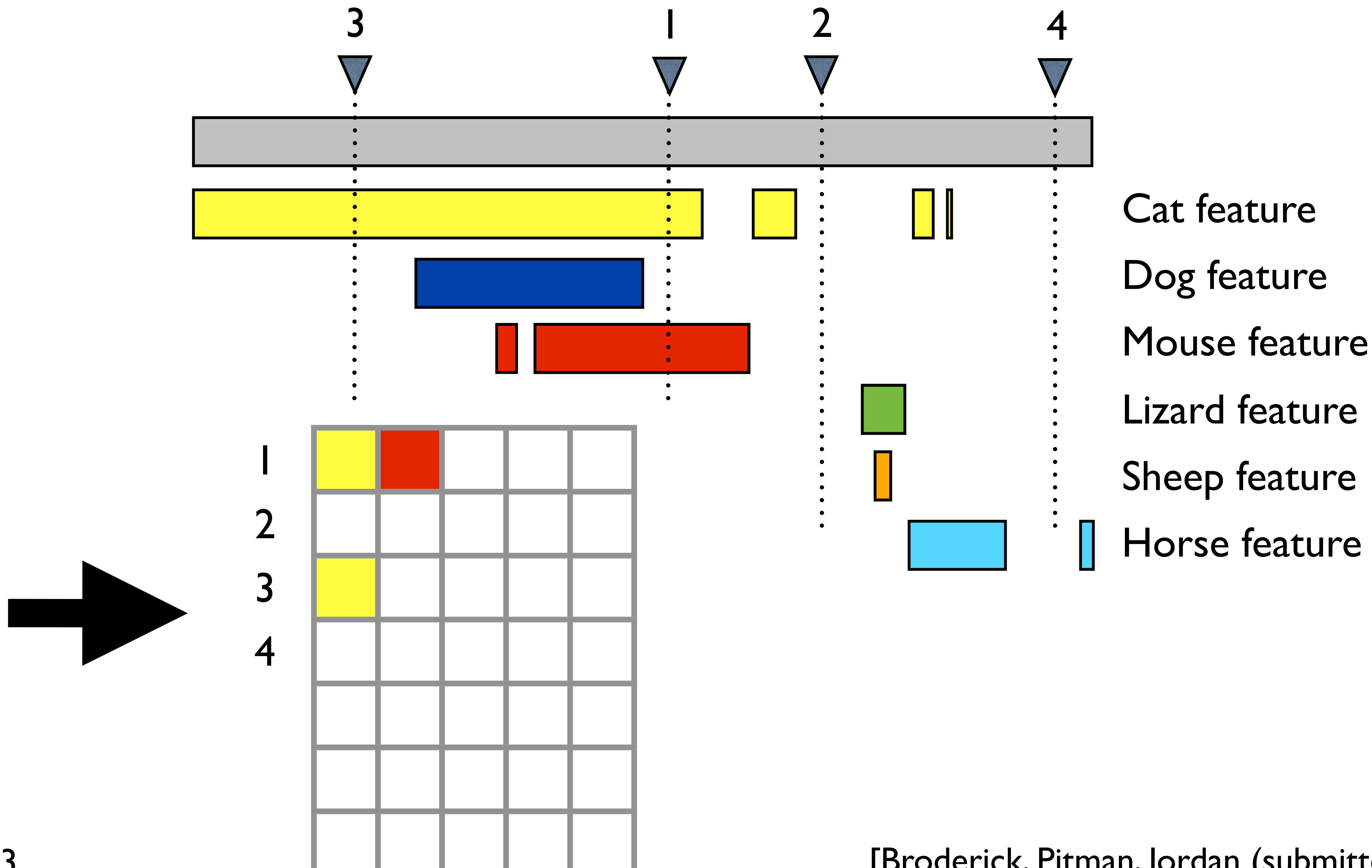
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

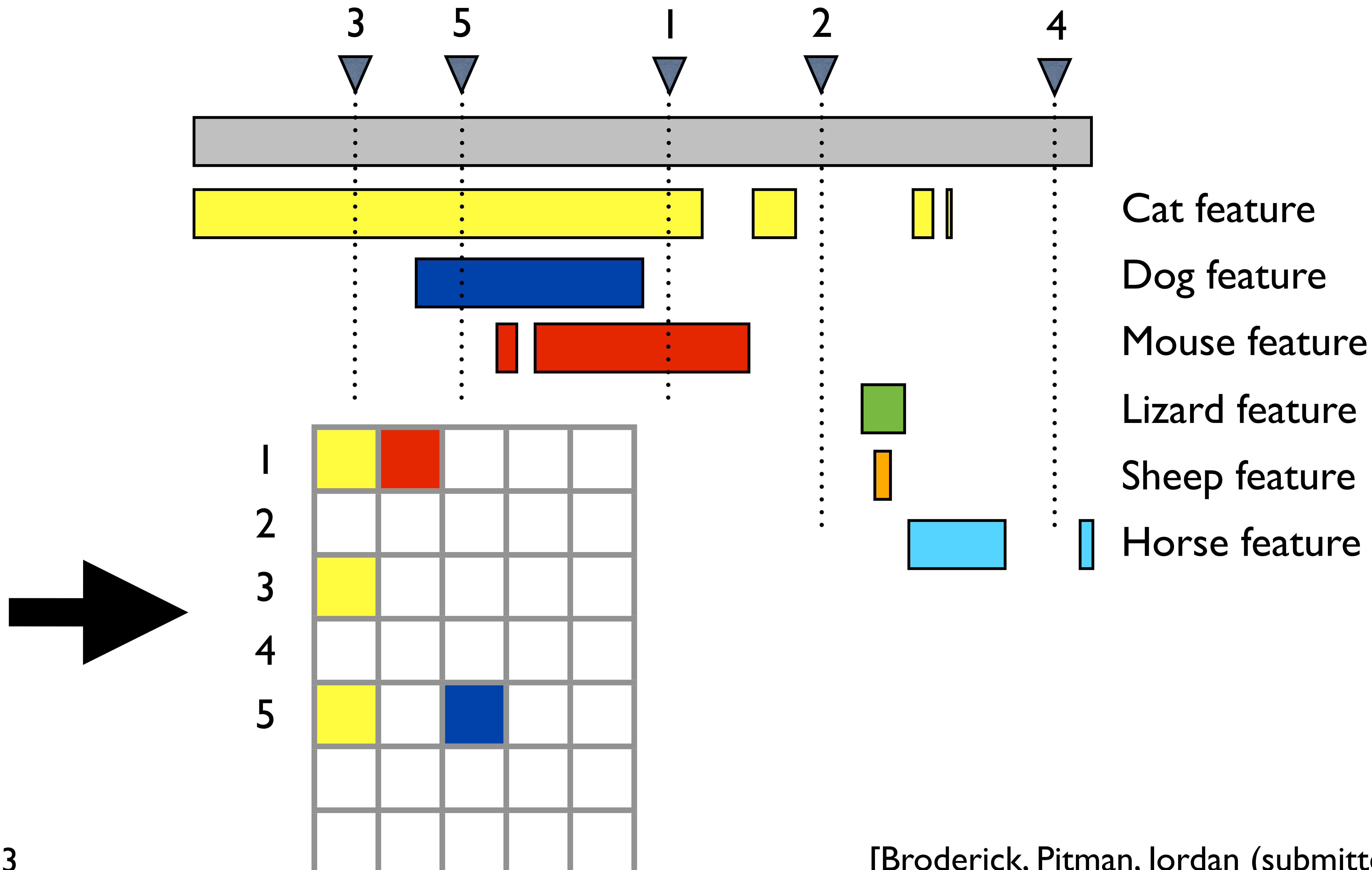
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

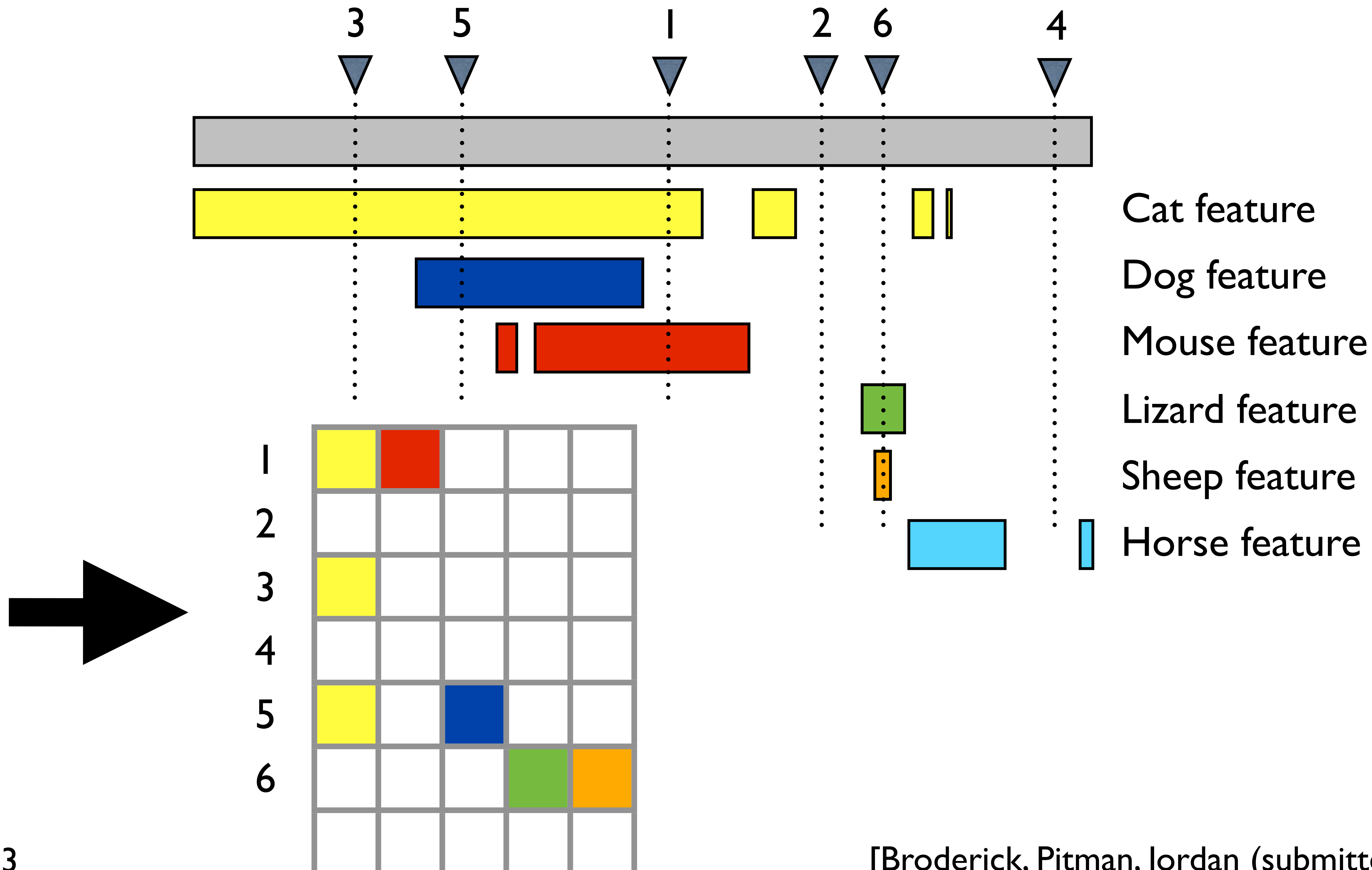
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

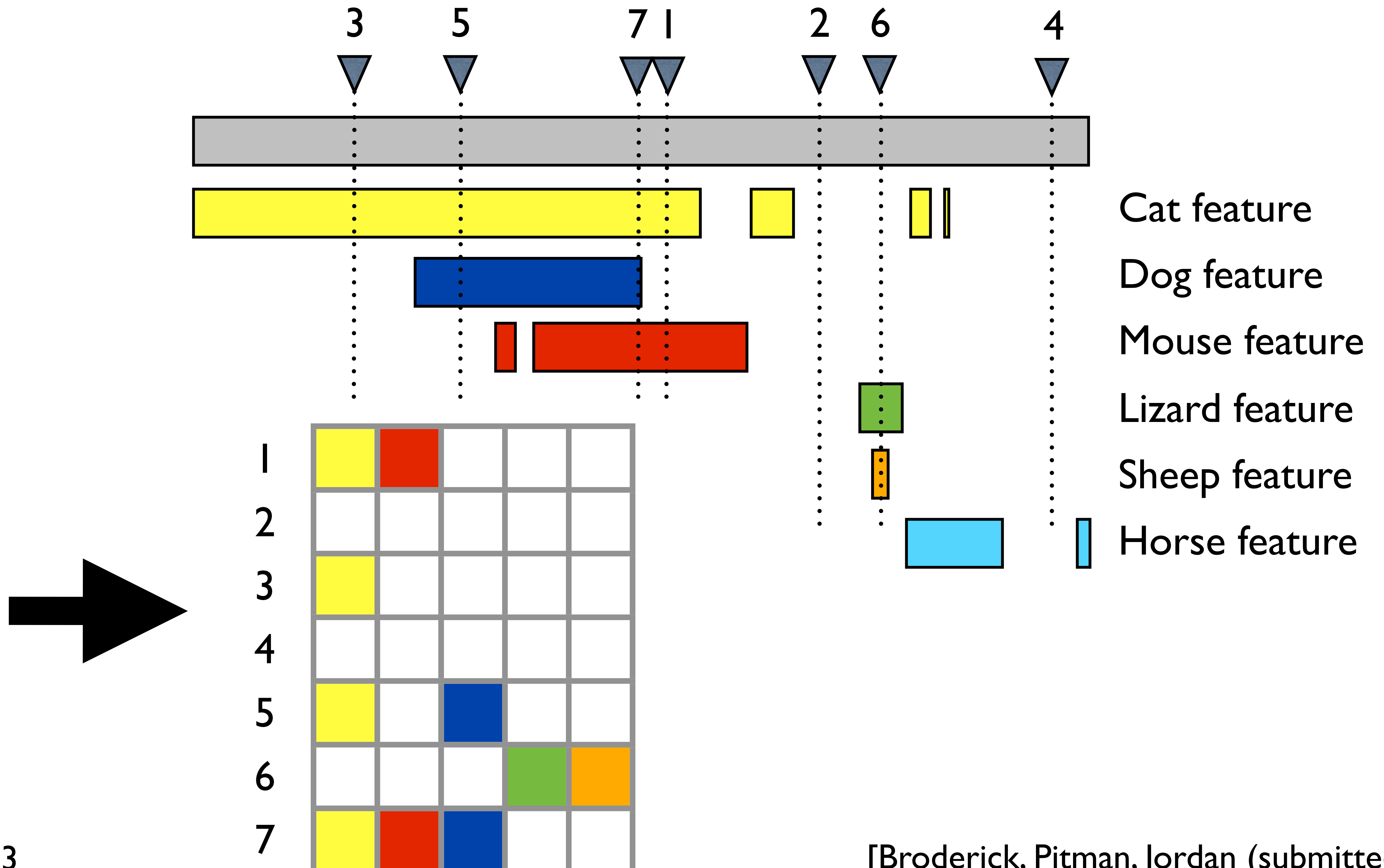
Exchangeable feature allocation: feature paintbox



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

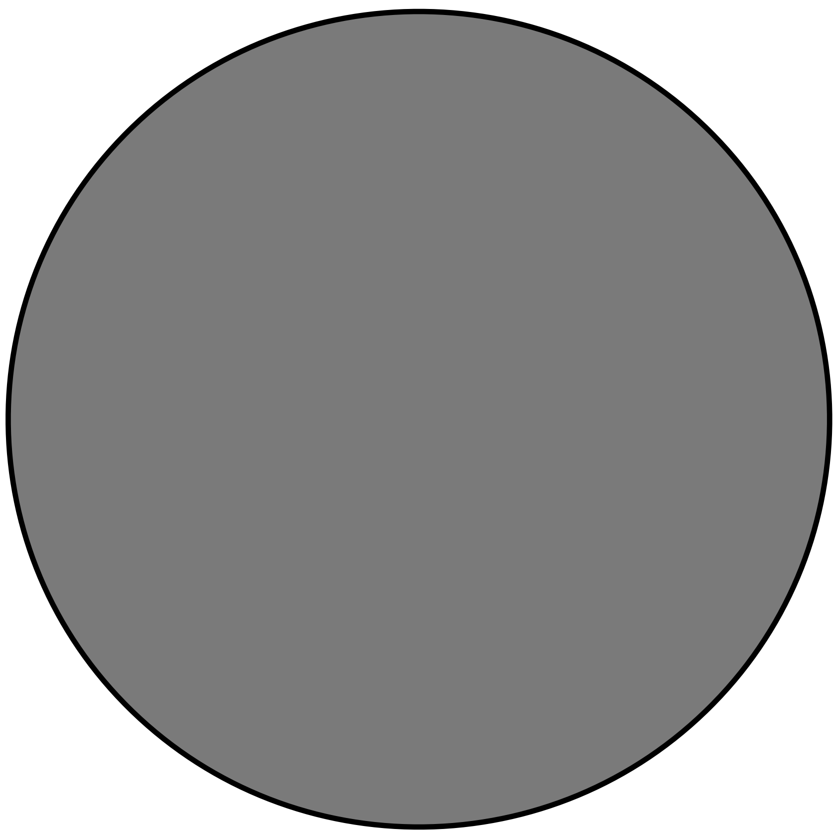
Exchangeable feature allocation: feature paintbox



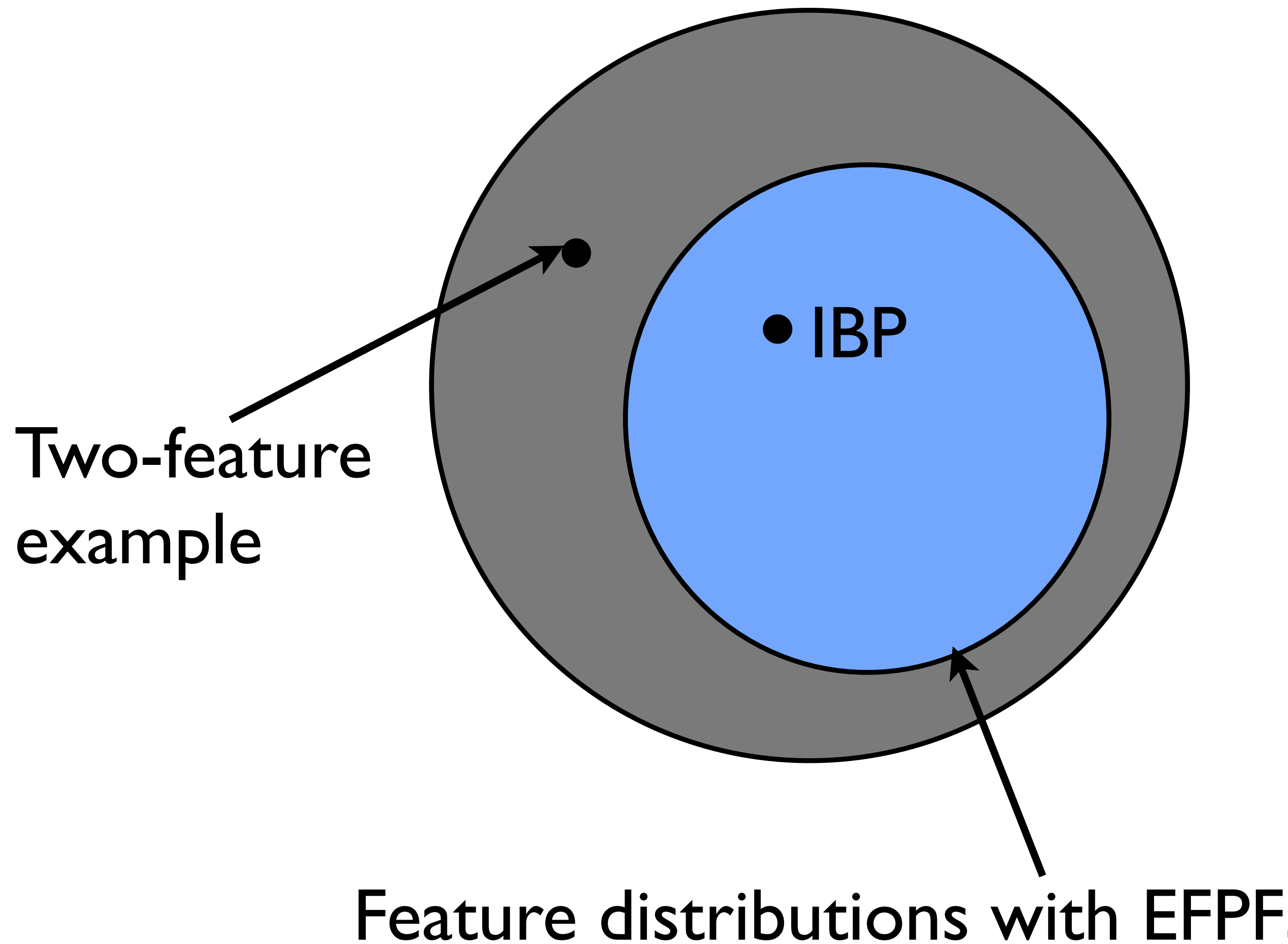
[Broderick, Pitman, Jordan (submitted)]

Paintboxes

Exchangeable cluster distributions
= Cluster distributions with EPPFs



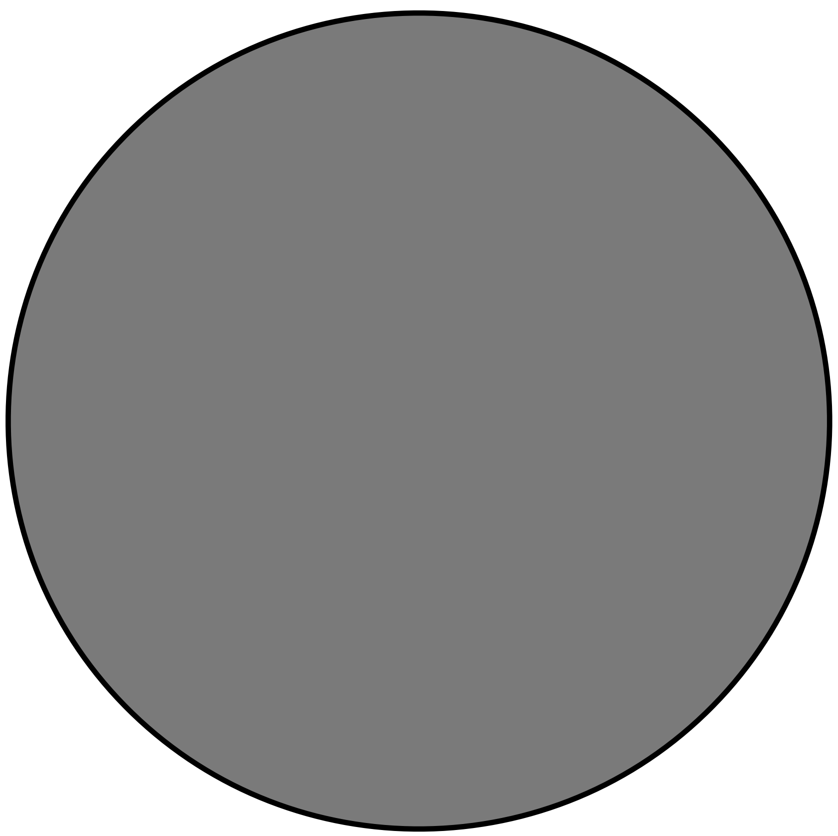
Exchangeable feature distributions



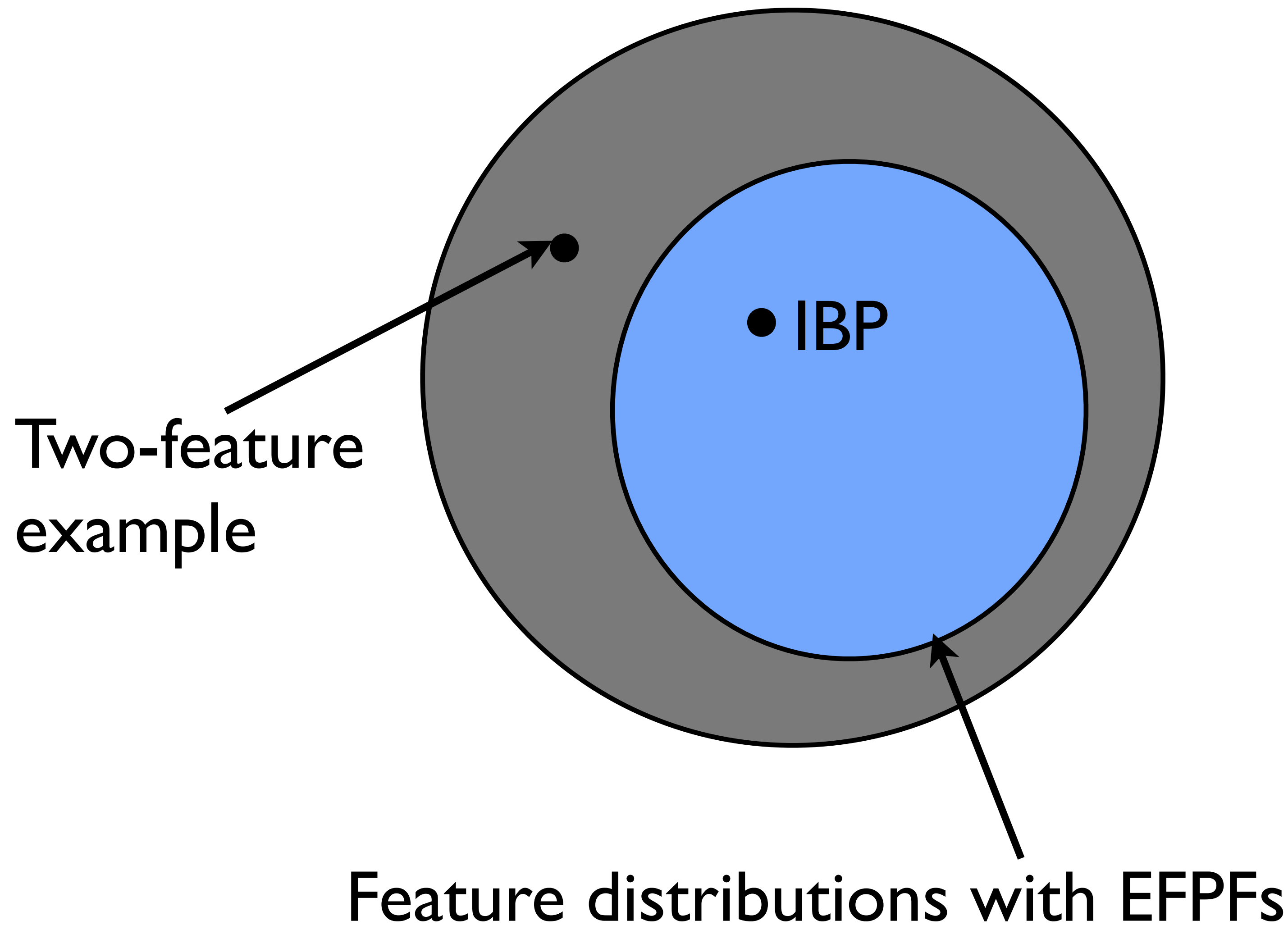
[Broderick, Jordan, Pitman 2012]

Paintboxes

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



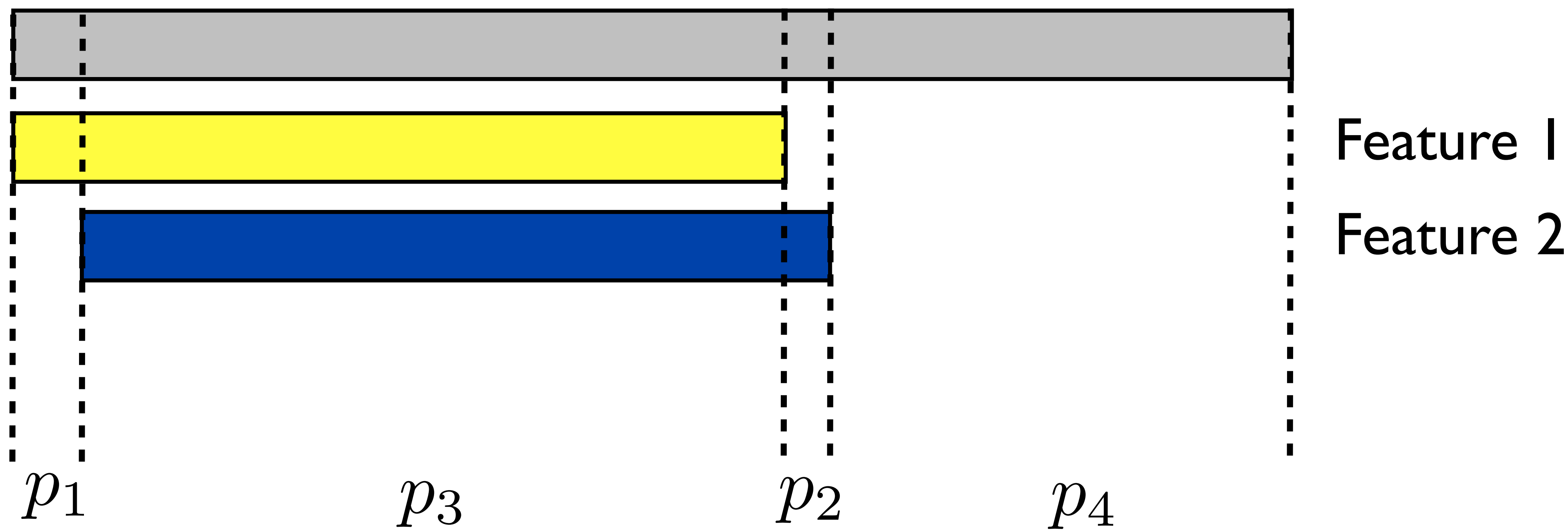
Exchangeable feature distributions
= Feature paintbox allocations

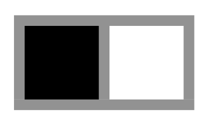


[Broderick, Pitman, Jordan (submitted)]

Paintboxes

Two feature example



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

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

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

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

Paintboxes

Indian buffet process: beta feature frequencies

[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

For m = 1, 2, ...
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

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

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

   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, ...

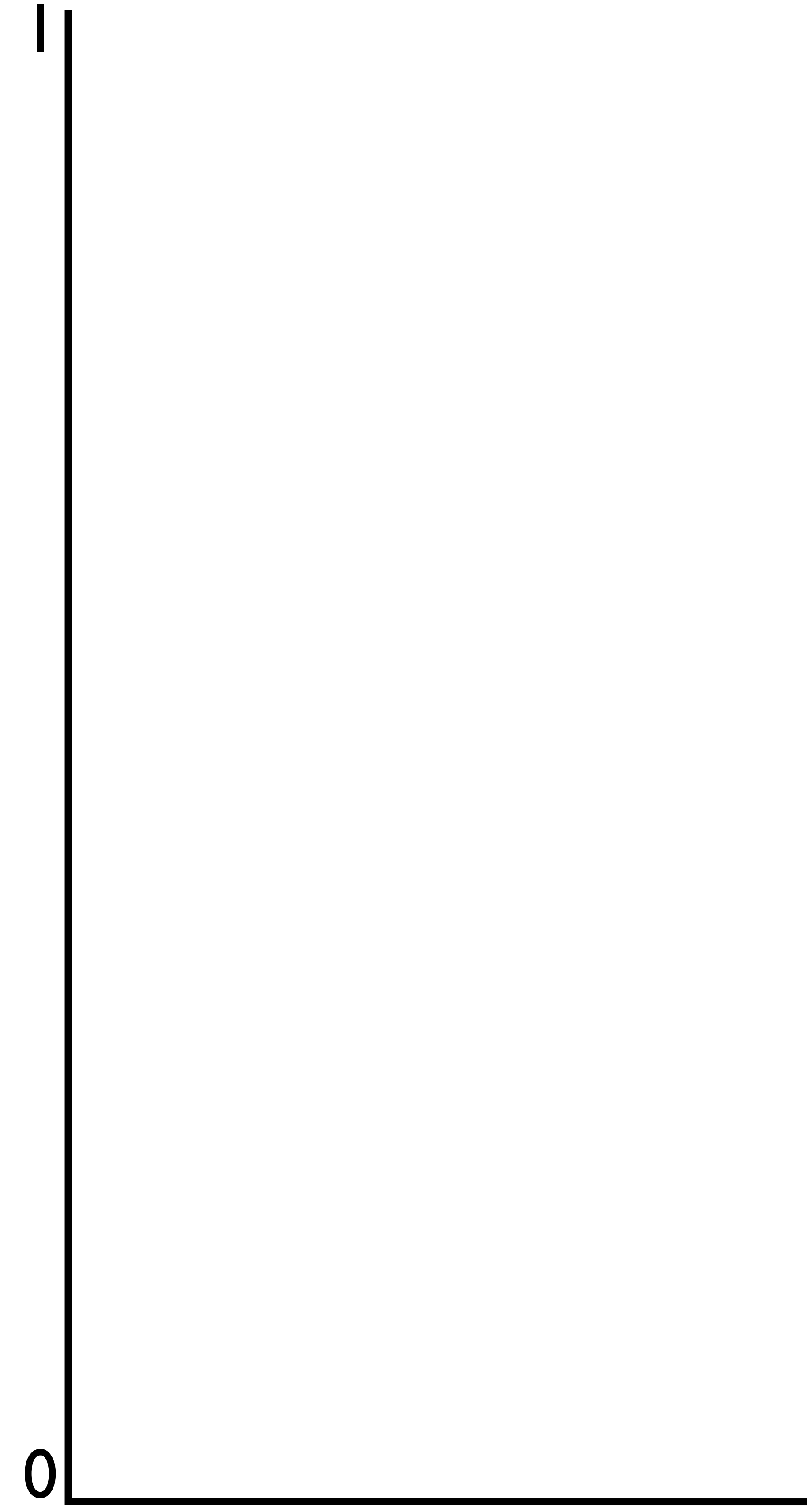
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

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

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

   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, ...

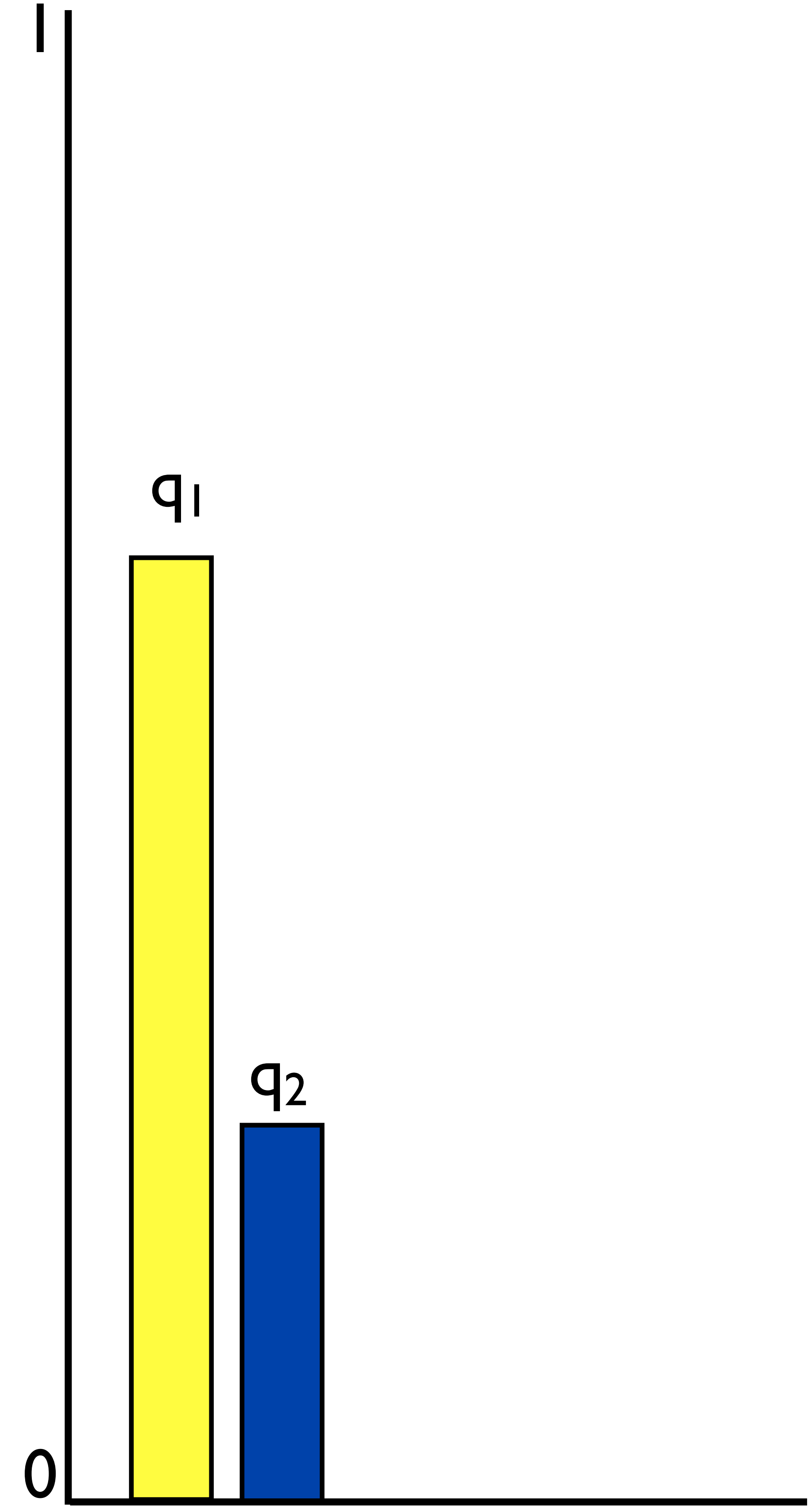
1. Draw $K_m^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

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

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

   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, ...

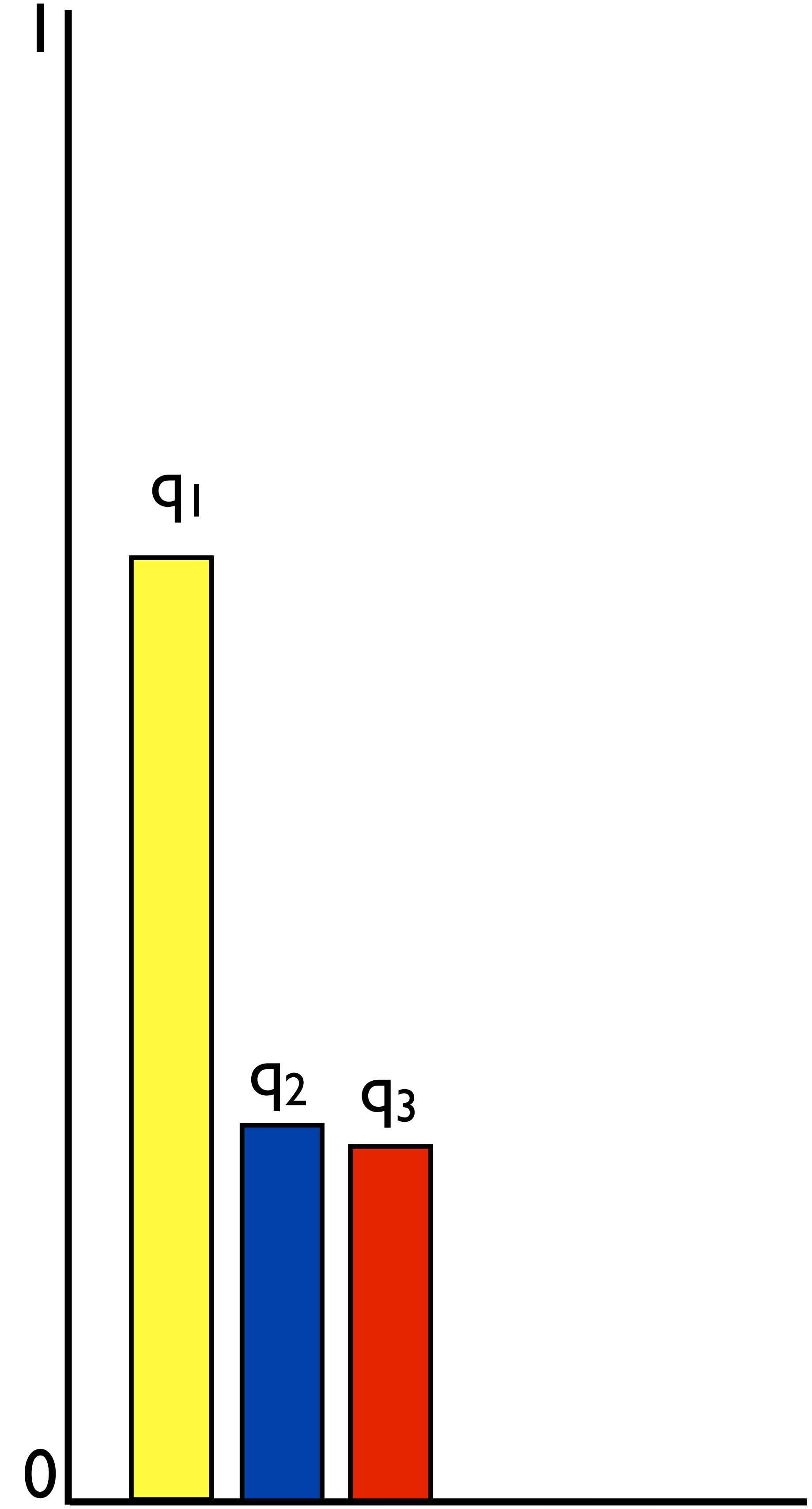
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + m - 1}\right)$

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

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

   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, ...

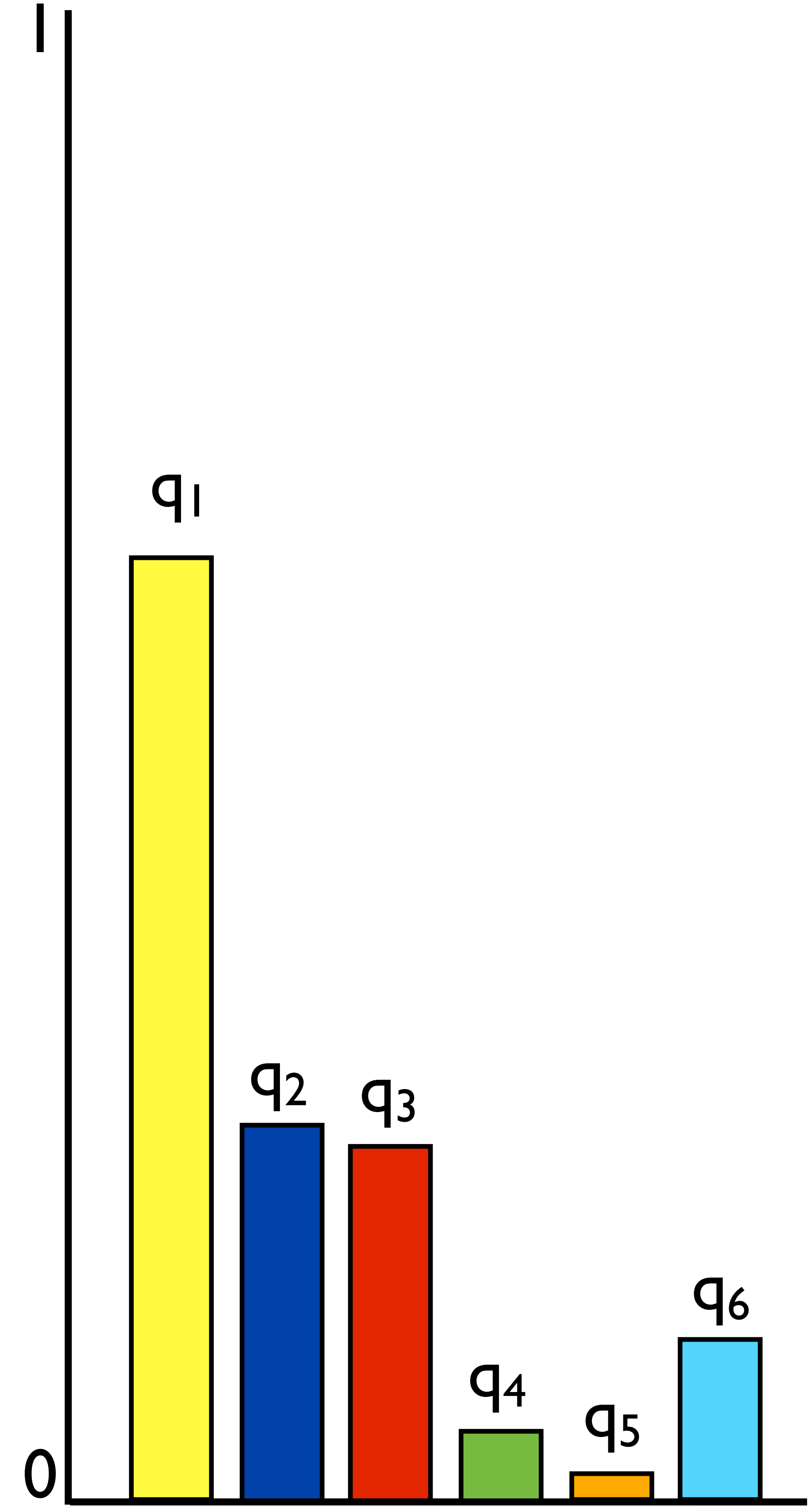
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

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

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

   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, ...

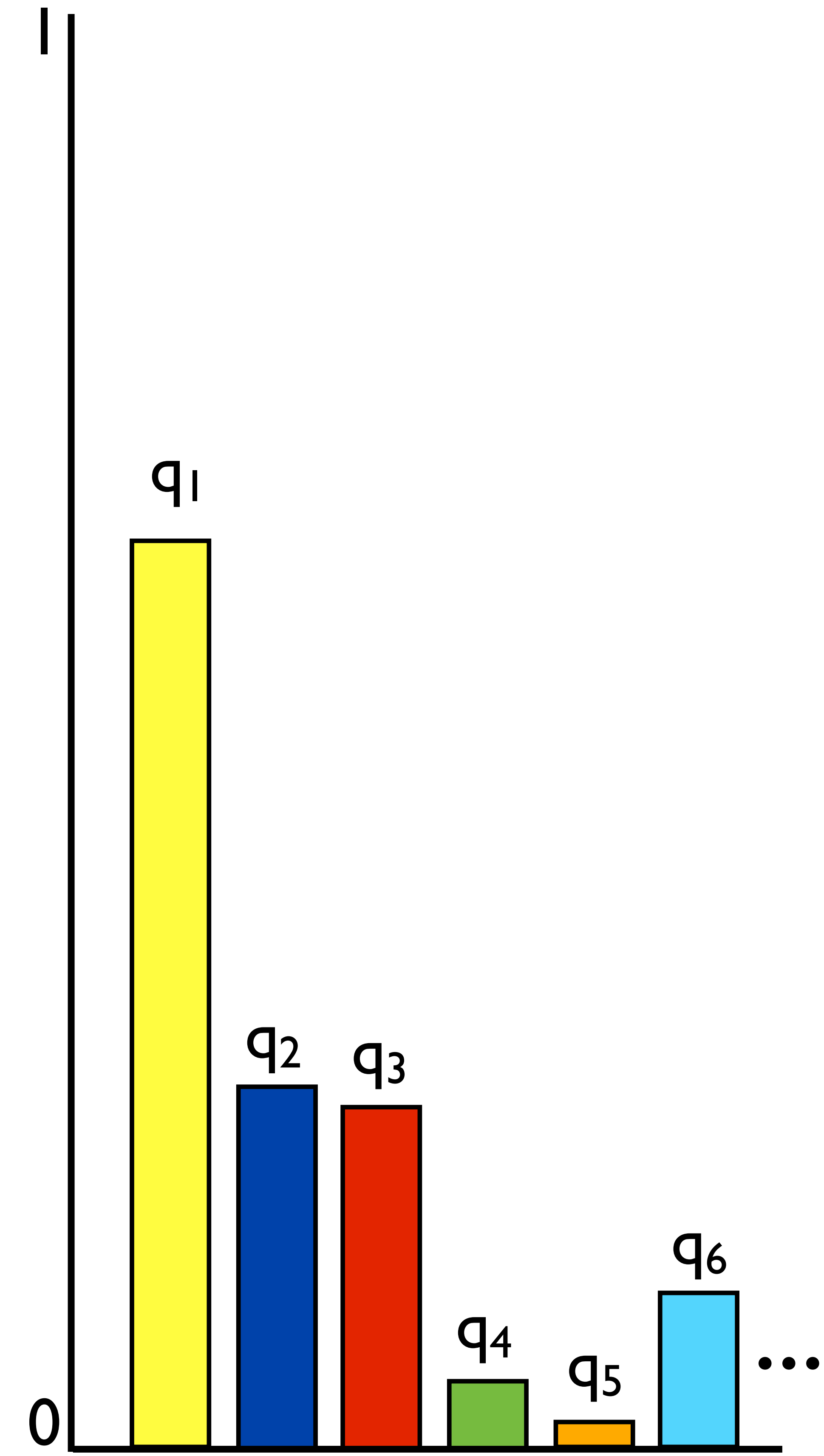
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \dfrac{\theta}{\theta + m - 1}\right)$

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

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

   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, ...

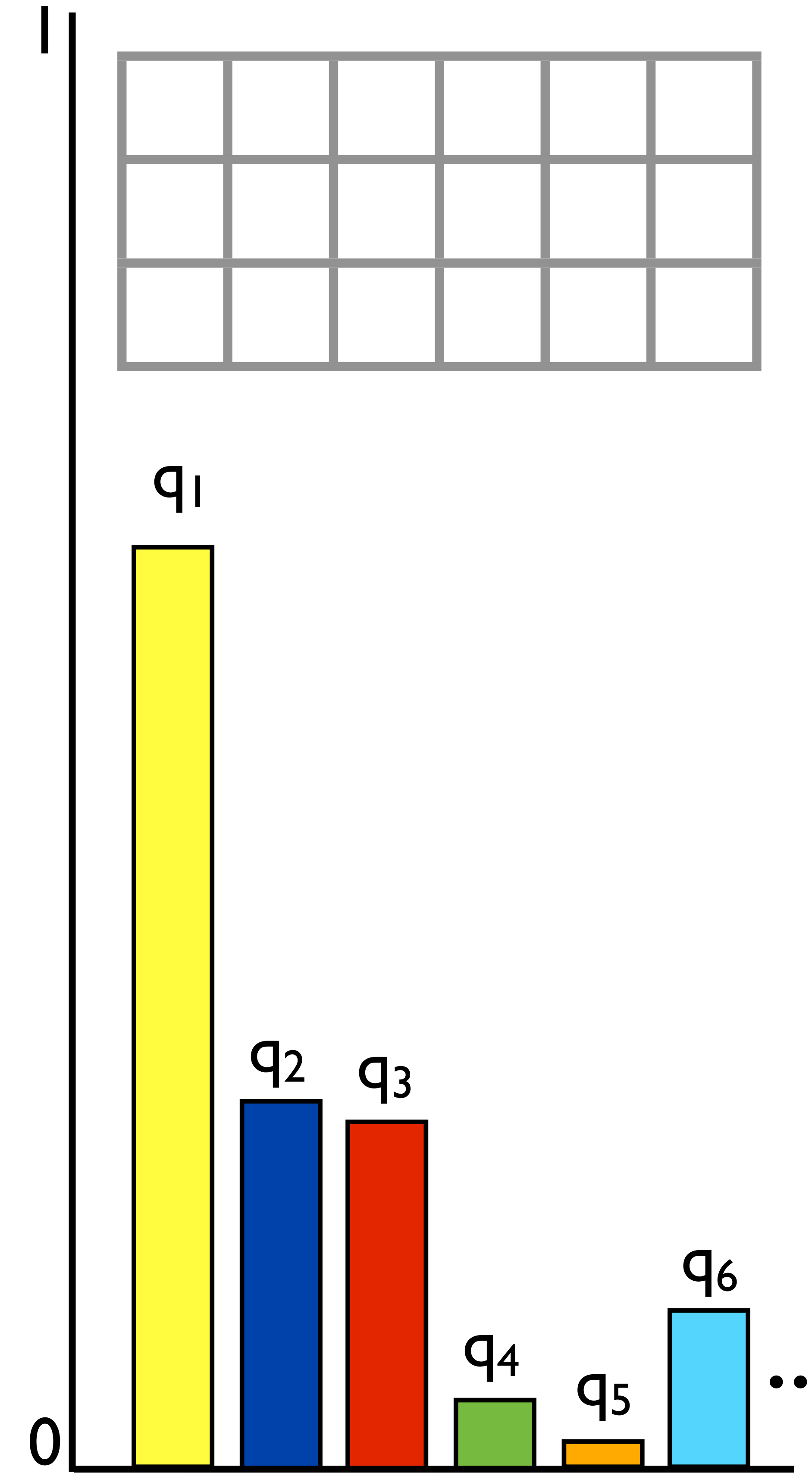
1. Draw $K_m^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

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

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

   Draw a frequency of size

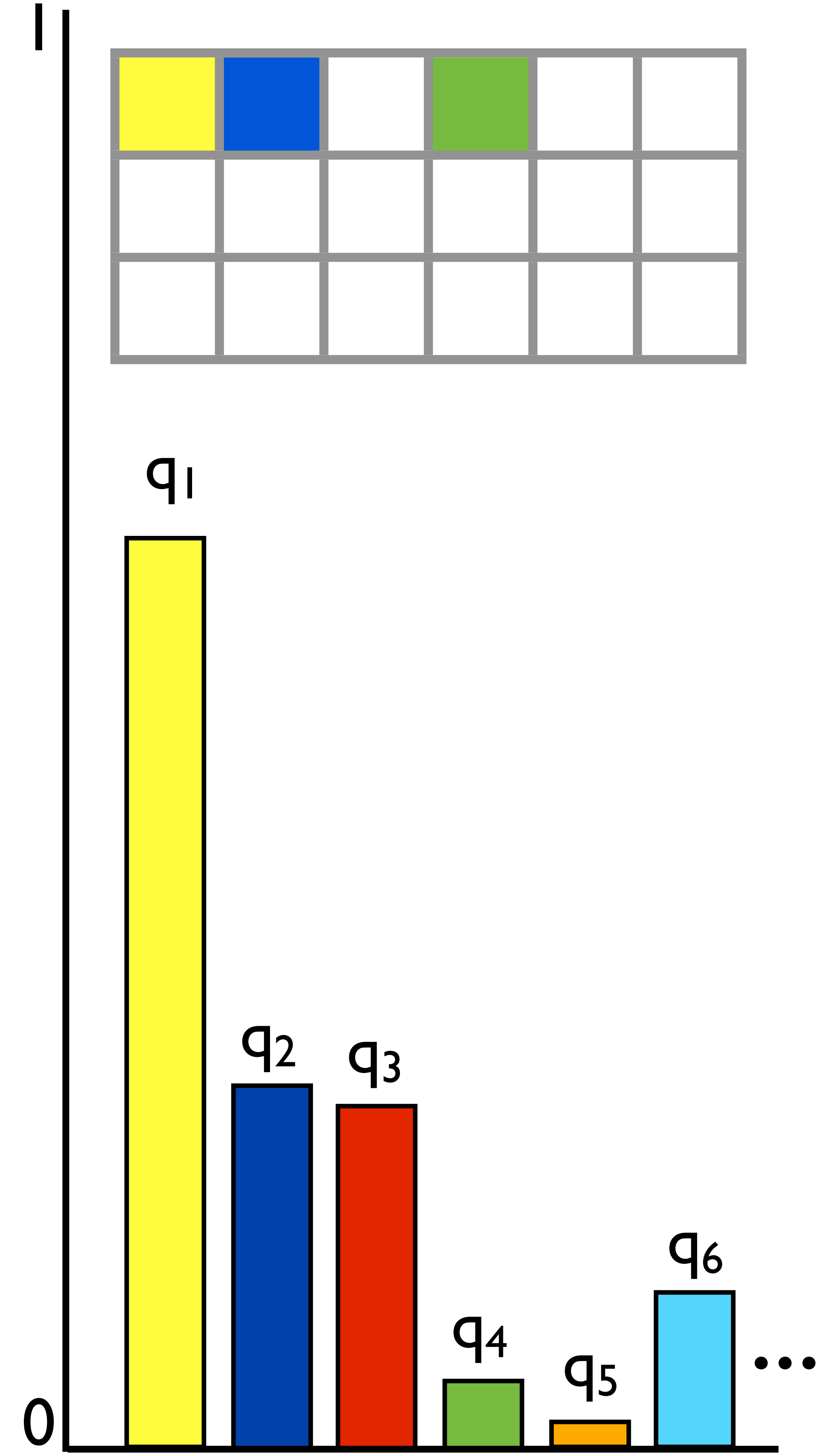
$$q_k \sim \text{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies



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

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

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

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

   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, ...

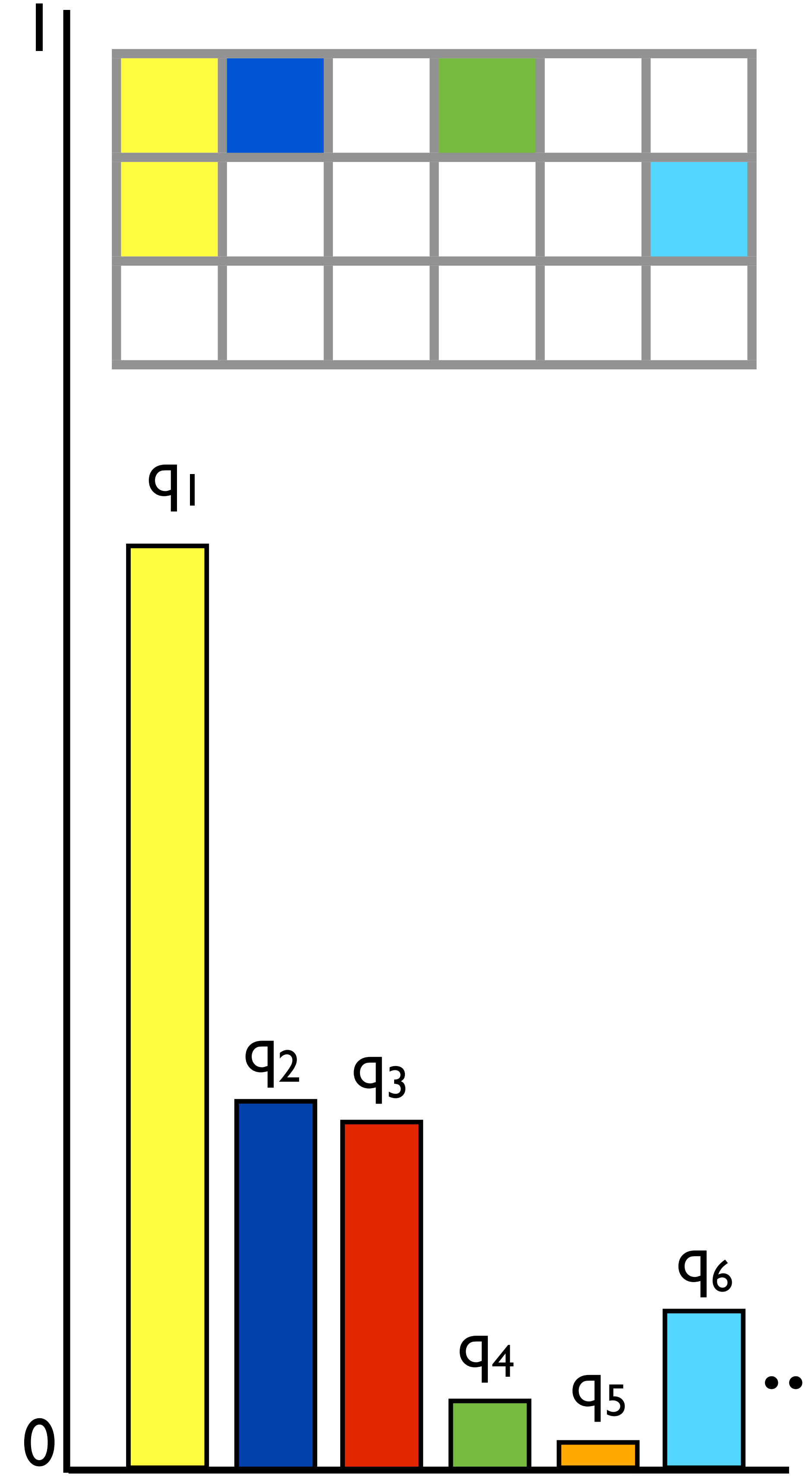
1. Draw $K_m^+ = \text{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

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

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

   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, ...

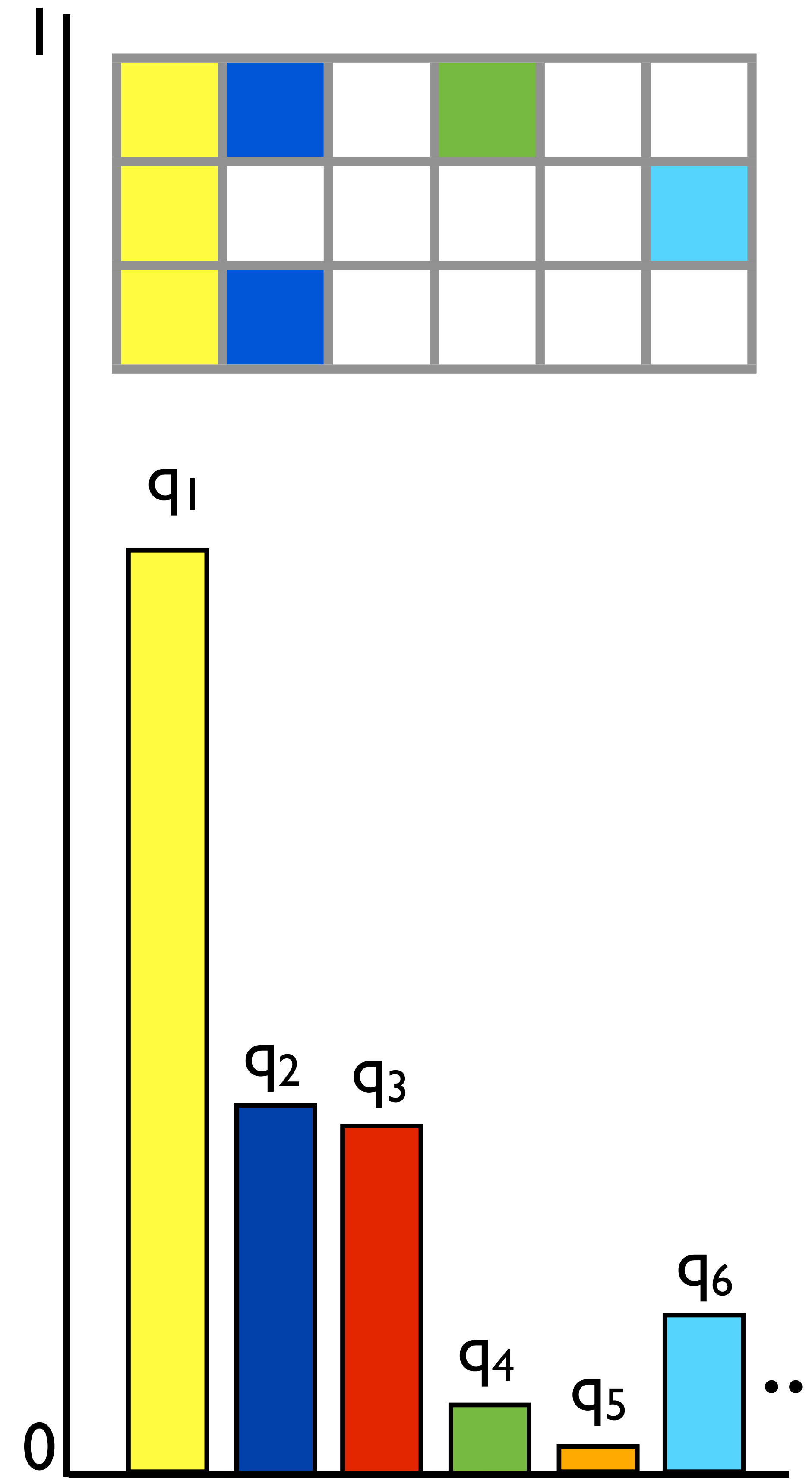
1. Draw $K_m^+ = \mathrm{Poisson}\left(\gamma \frac{\theta}{\theta + m - 1}\right)$

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

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

   Draw a frequency of size

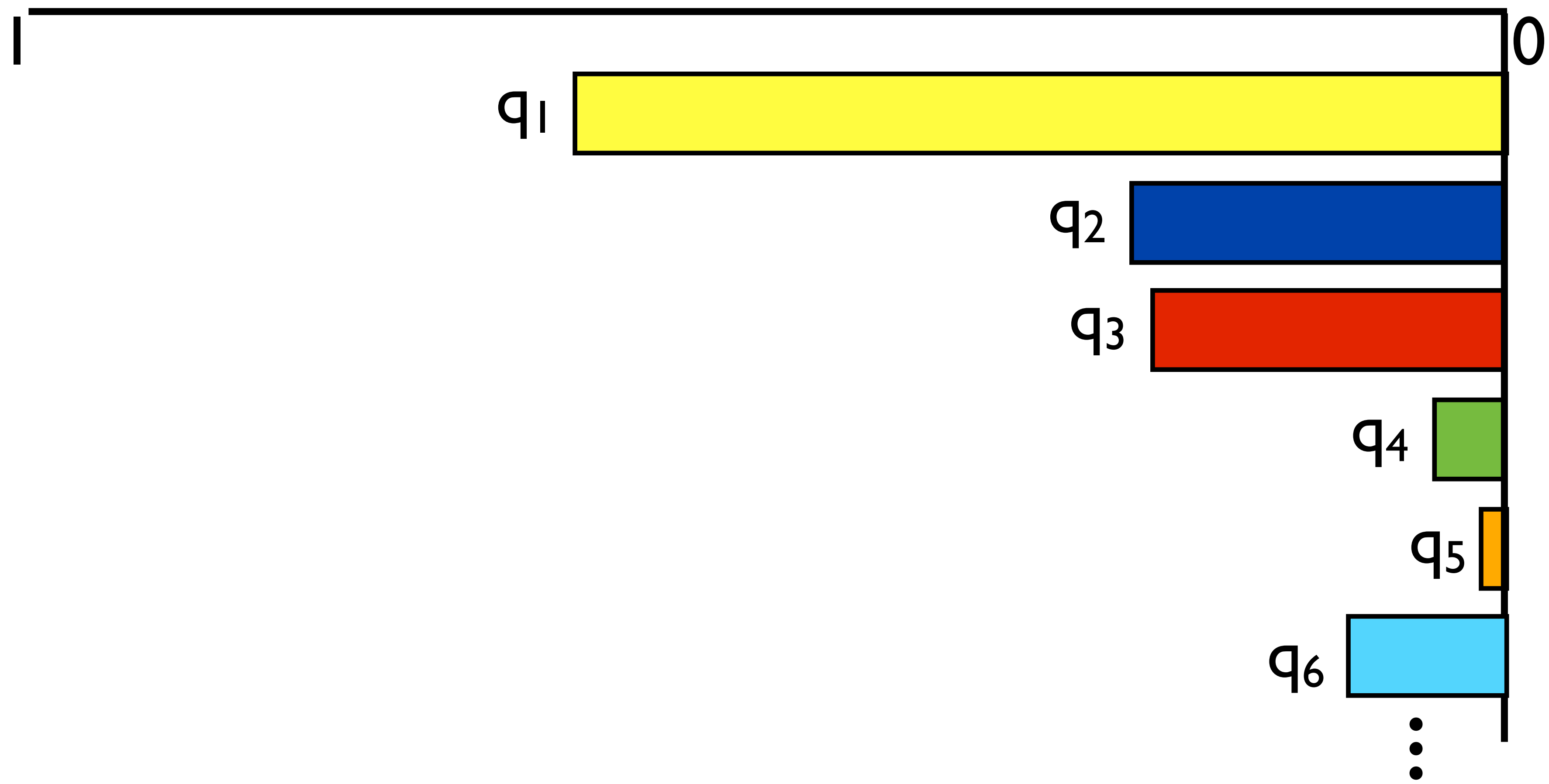
   $$q_k \sim \mathrm{Beta}(1, \theta + m - 1)$$



[Thibaux, Jordan 2007]

Paintboxes

Indian buffet process: beta feature frequencies

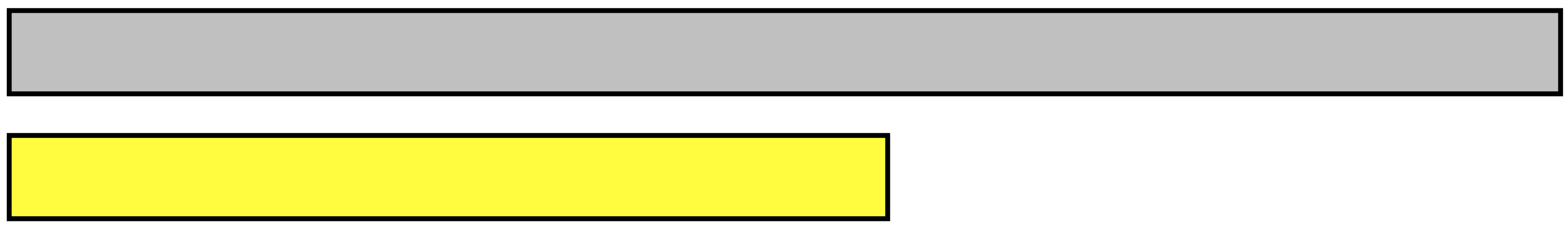


Paintboxes

Indian buffet process: beta feature frequencies

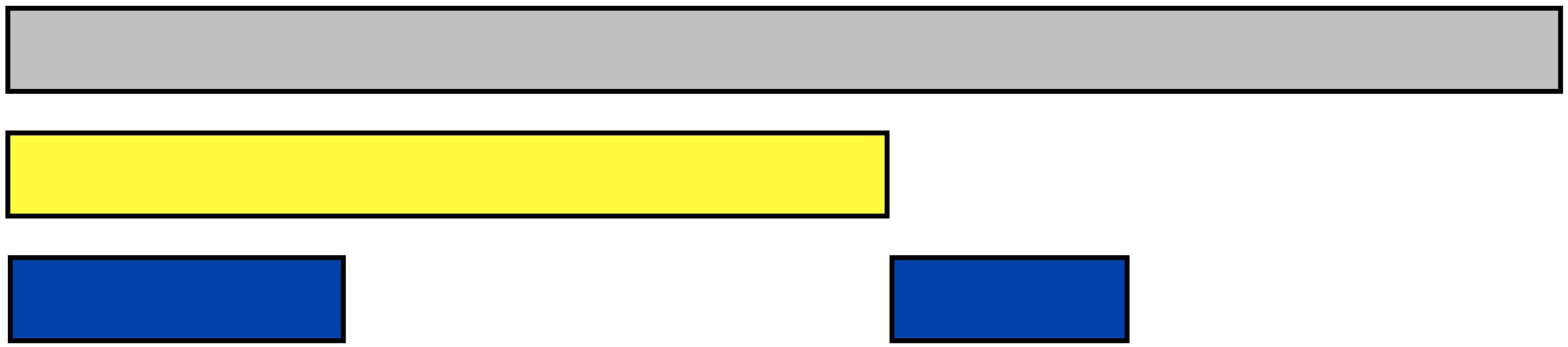
Paintboxes

Indian buffet process: beta feature frequencies



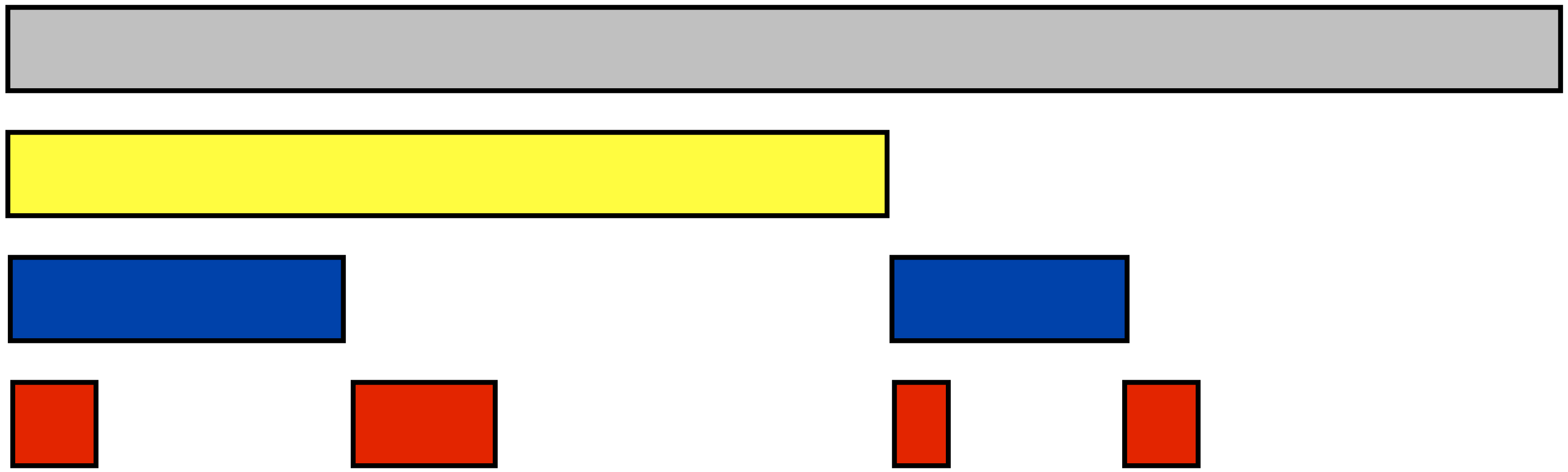
Paintboxes

Indian buffet process: beta feature frequencies



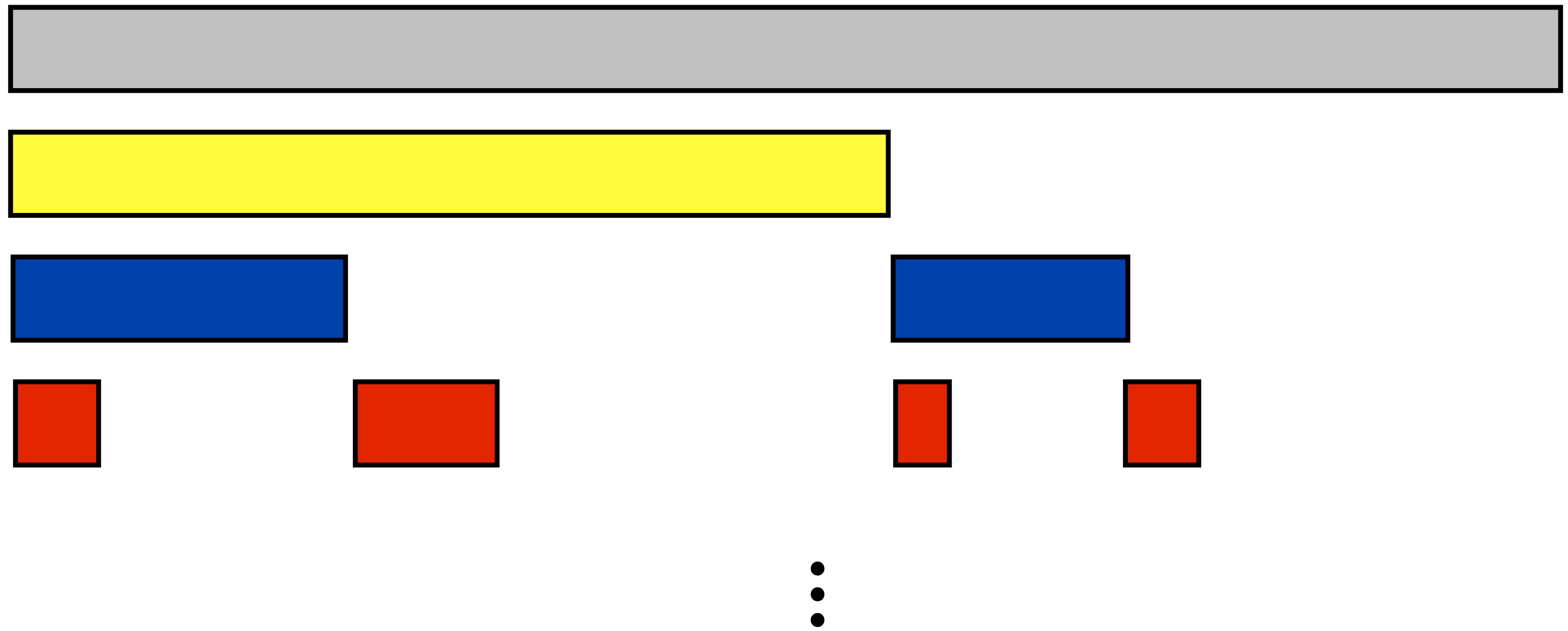
Paintboxes

Indian buffet process: beta feature frequencies

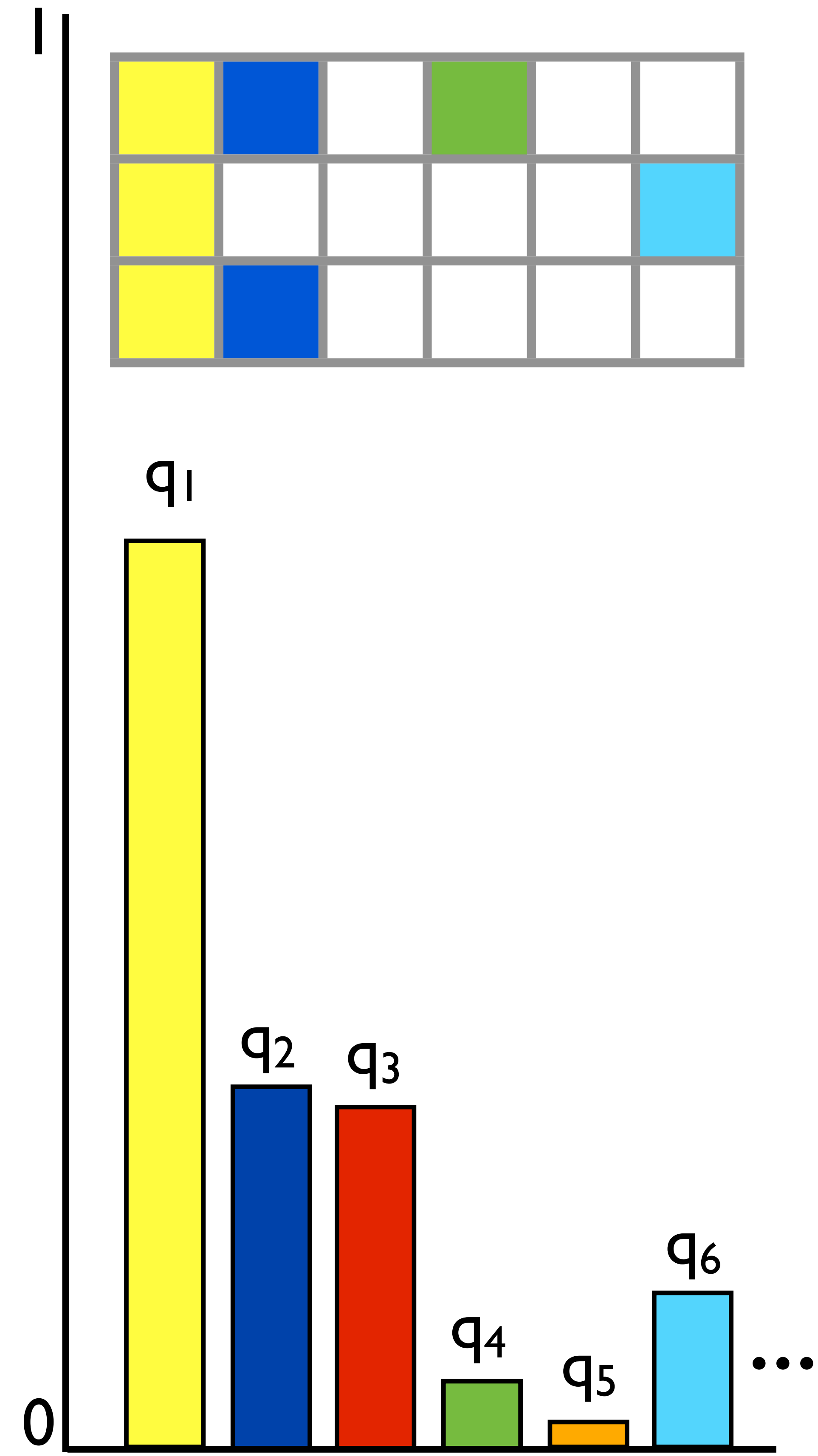


Paintboxes

Indian buffet process: beta feature frequencies

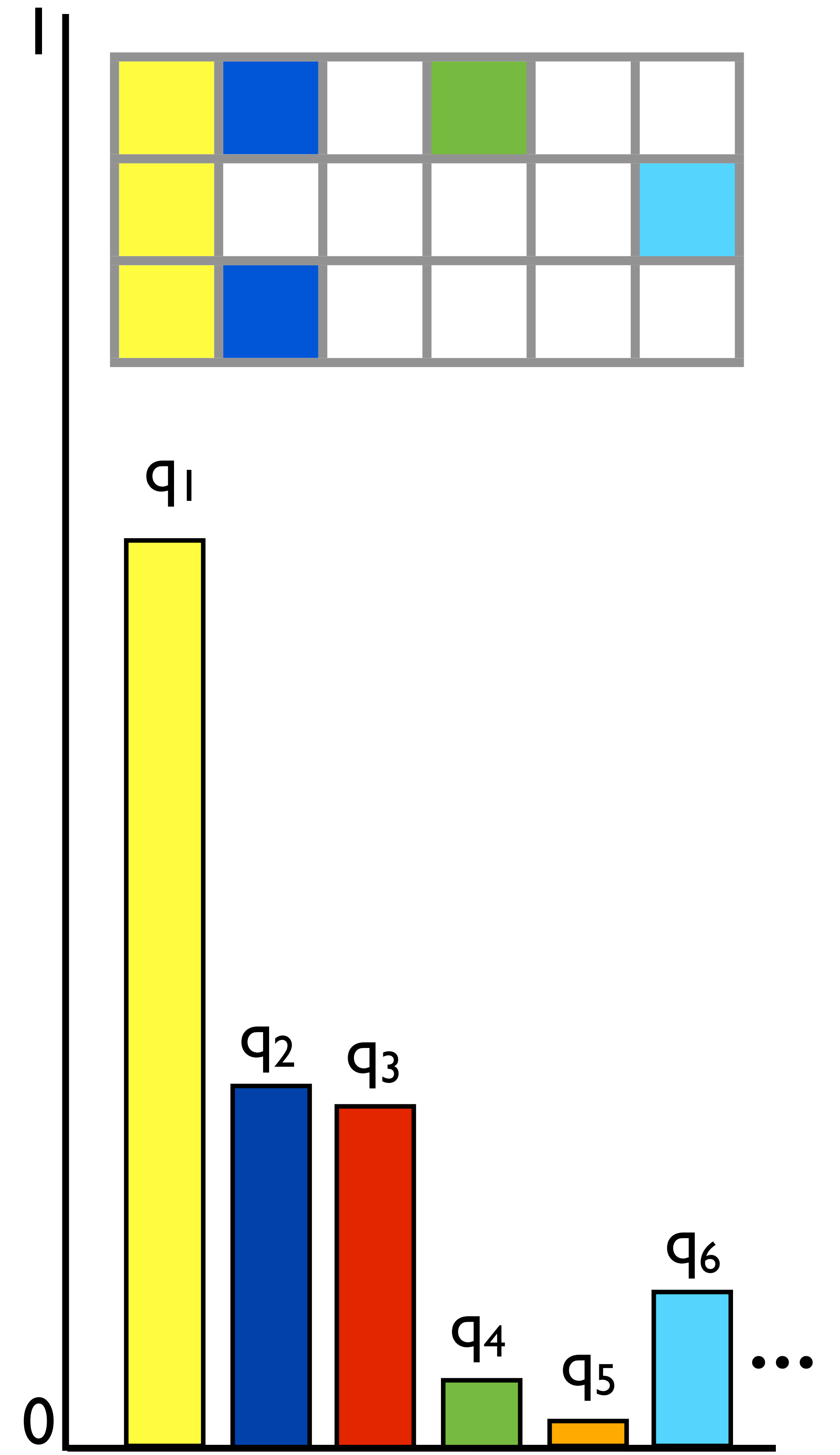


Paintboxes



Paintboxes

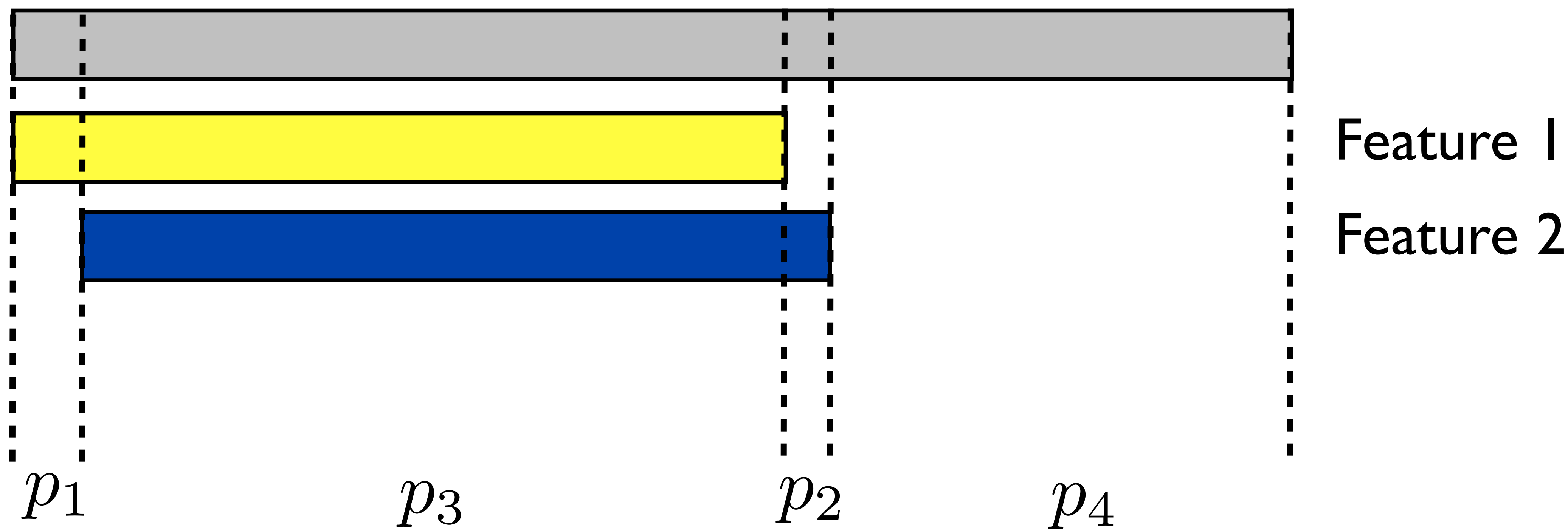
“Frequency models”



[Broderick, Pitman, Jordan (submitted)]

Paintboxes

Two feature example



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

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

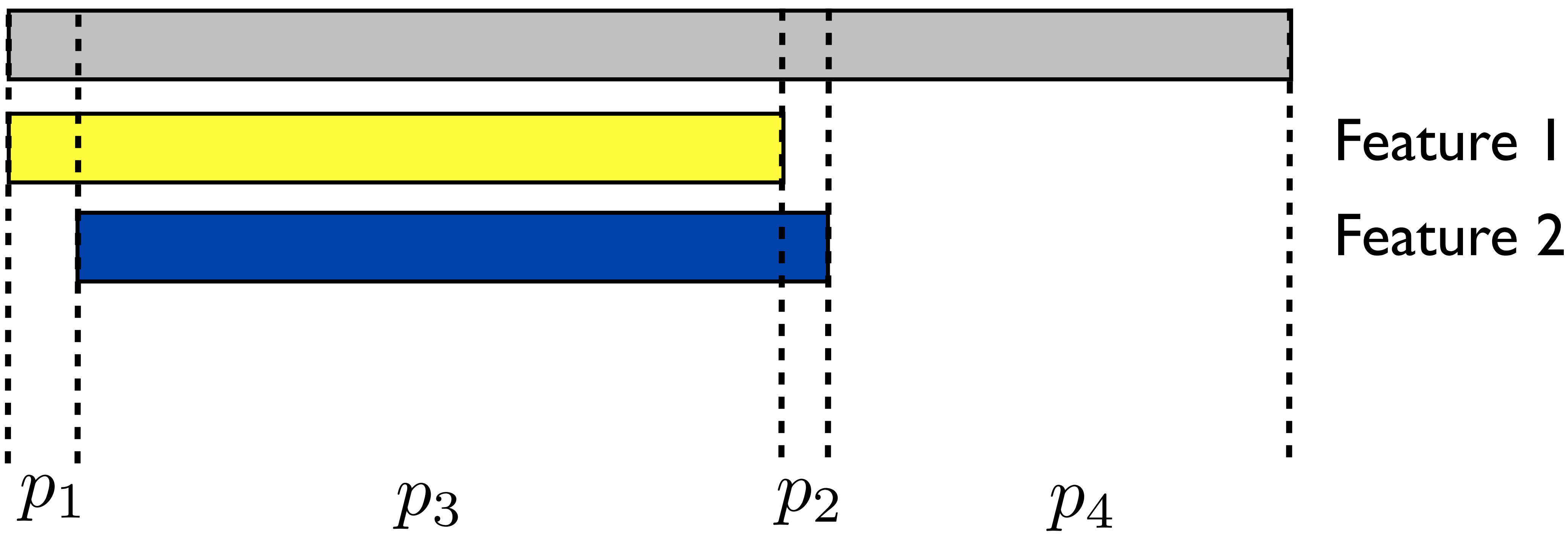
$$\mathbb{P}(\text{row} = \blacksquare\blacksquare) = p_3$$

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

Paintboxes

Two feature example

Not a frequency model



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

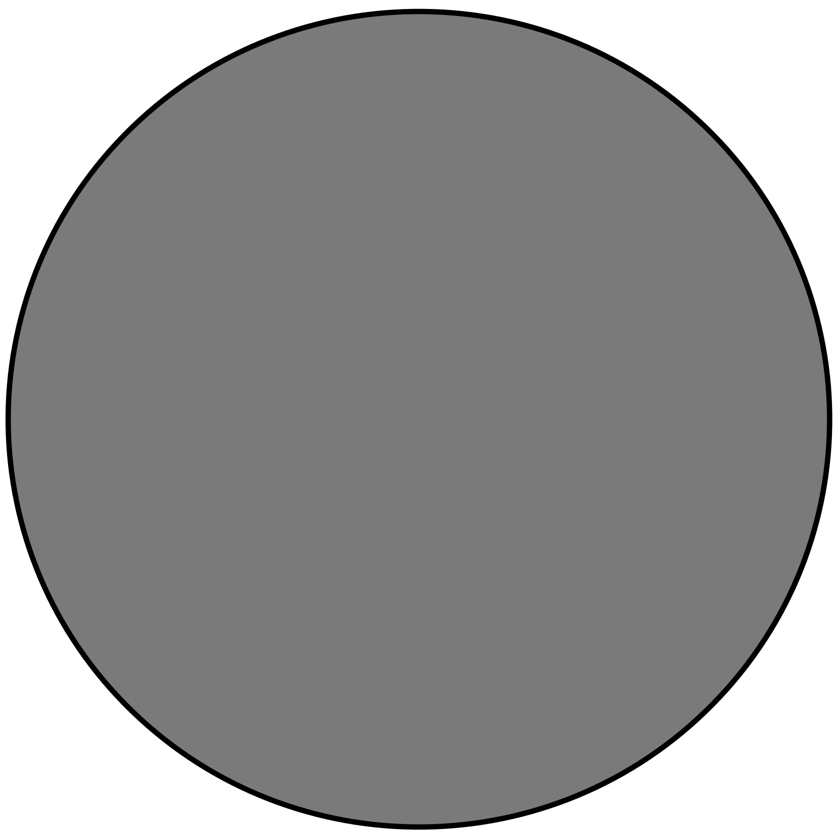
$$\mathbb{P}(\text{row} = \square\blacksquare) = p_2$$

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

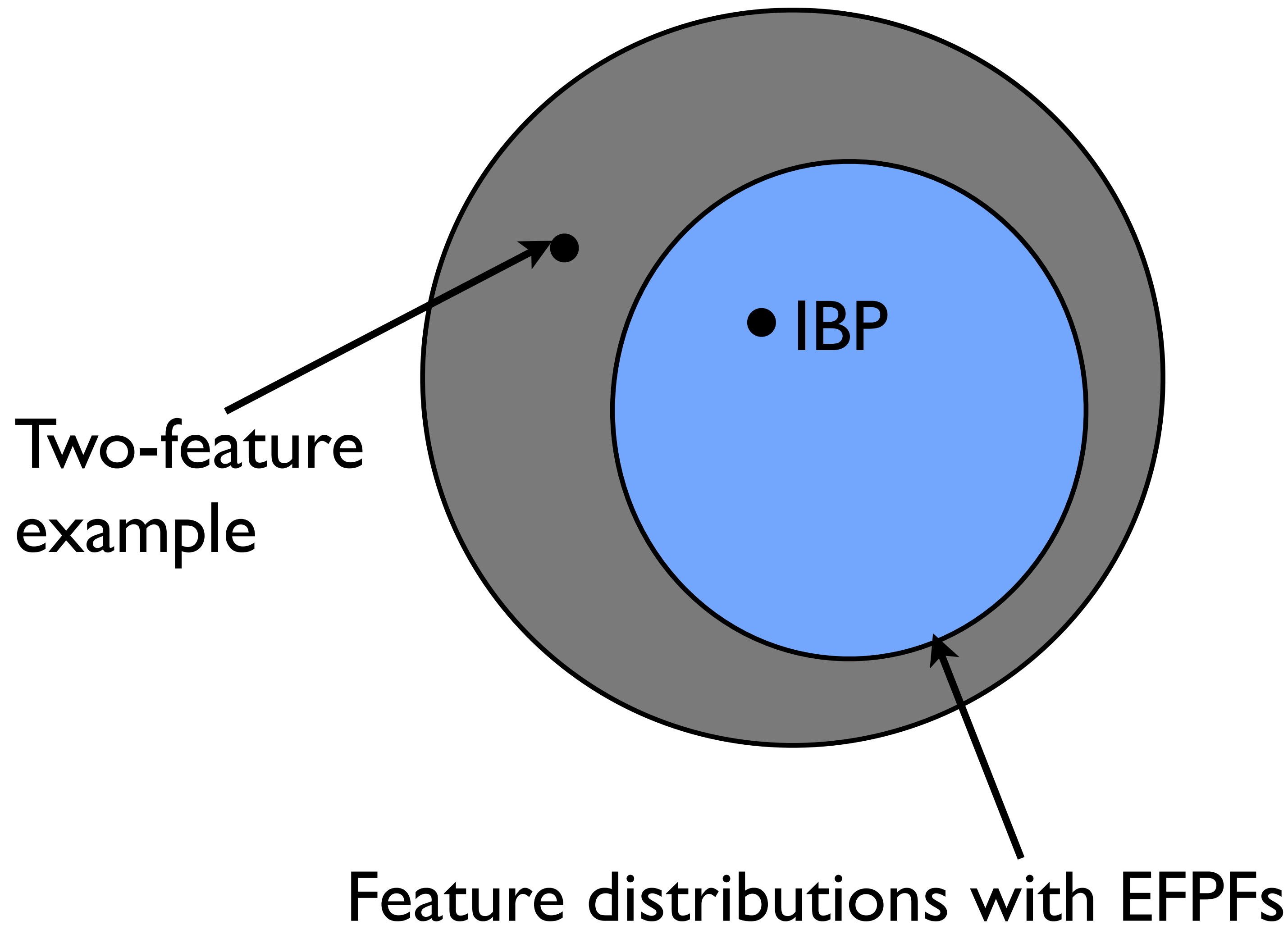
$$\mathbb{P}(\text{row} = \square\square) = p_4$$

Paintboxes

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



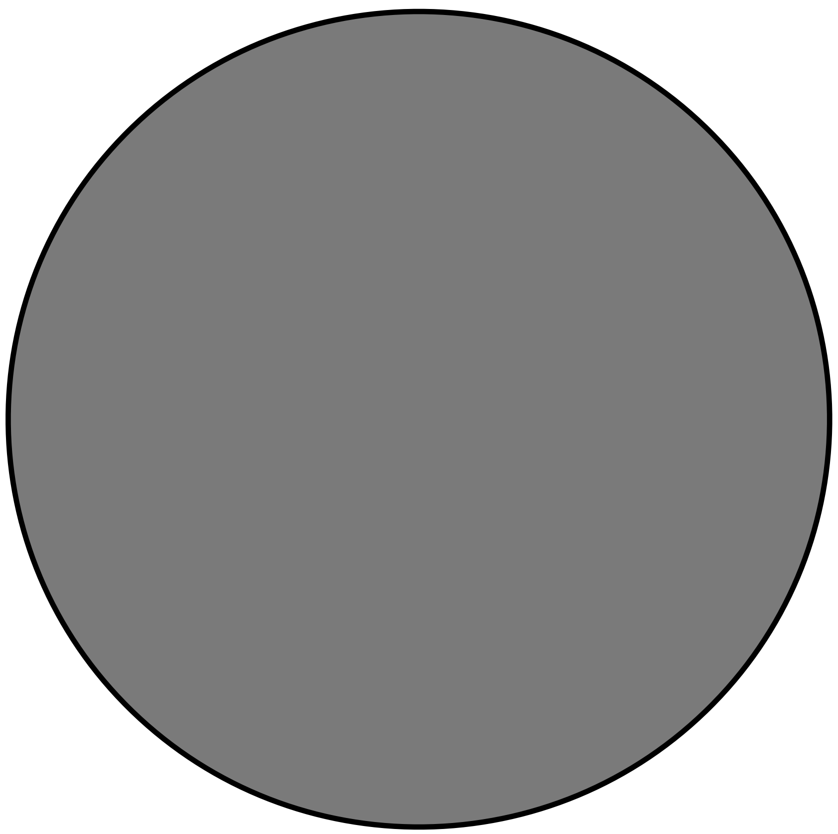
Exchangeable feature distributions
= Feature paintbox allocations



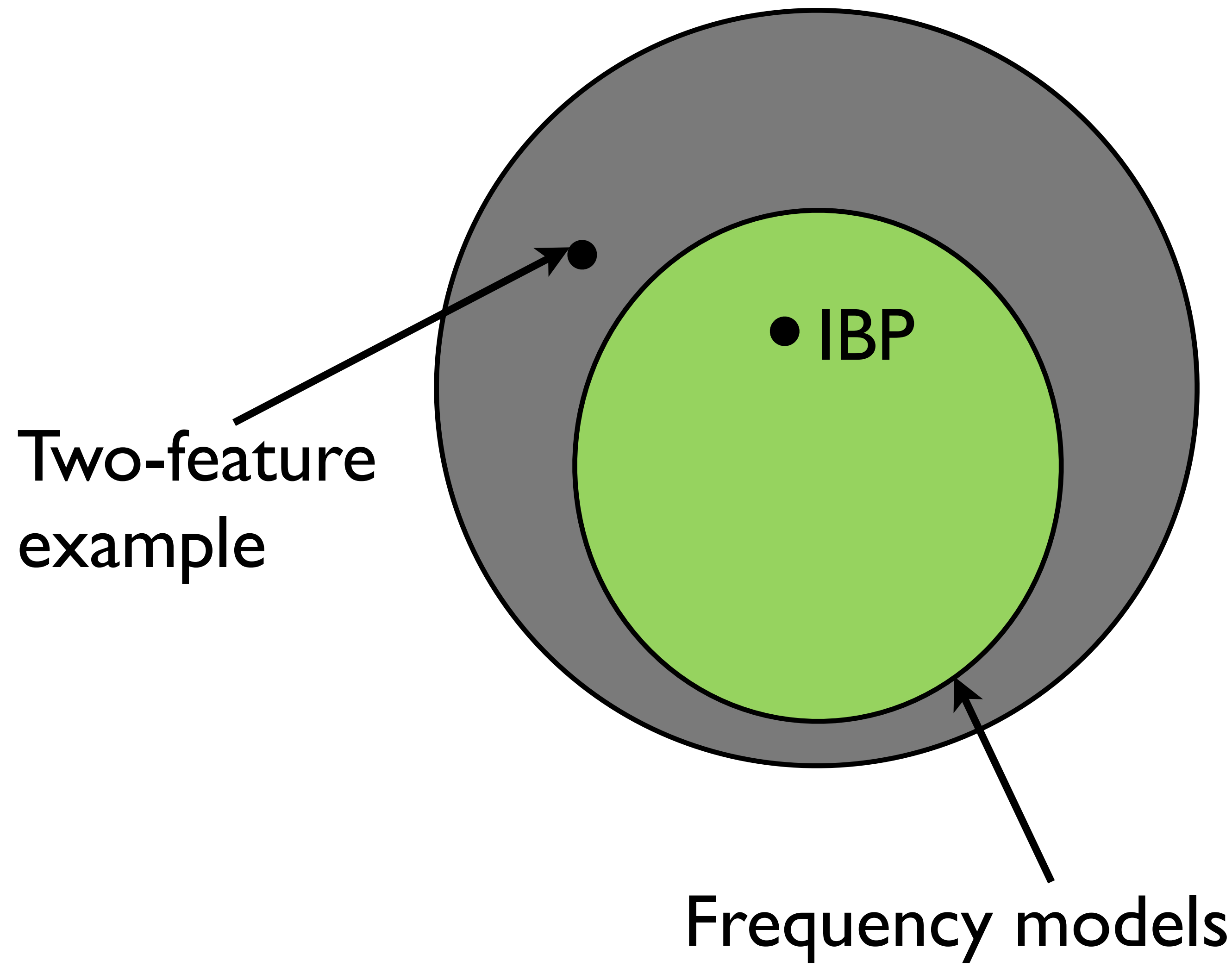
[Broderick, Pitman, Jordan (submitted)]

Paintboxes

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

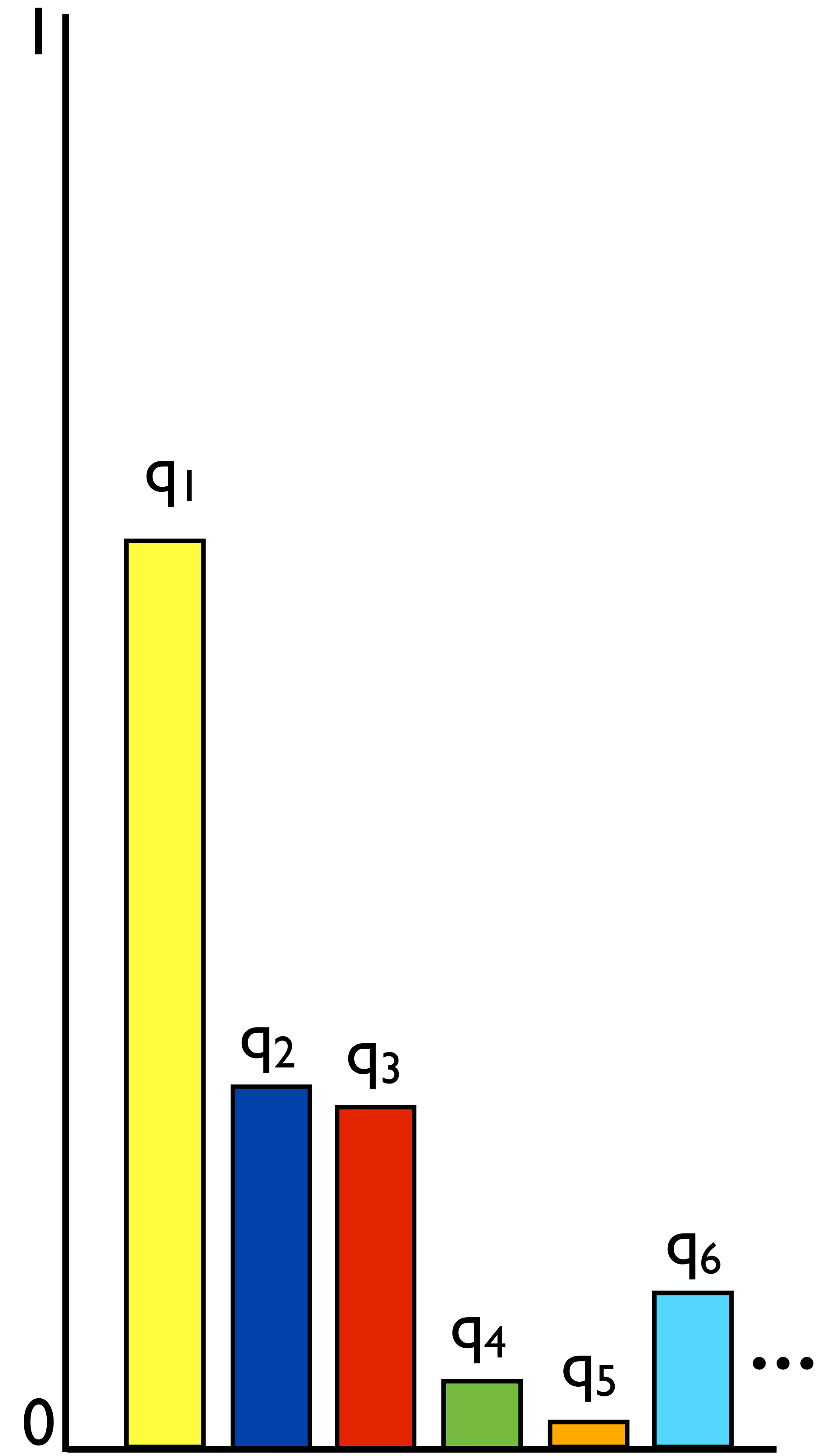


Exchangeable feature distributions
= Feature paintbox allocations

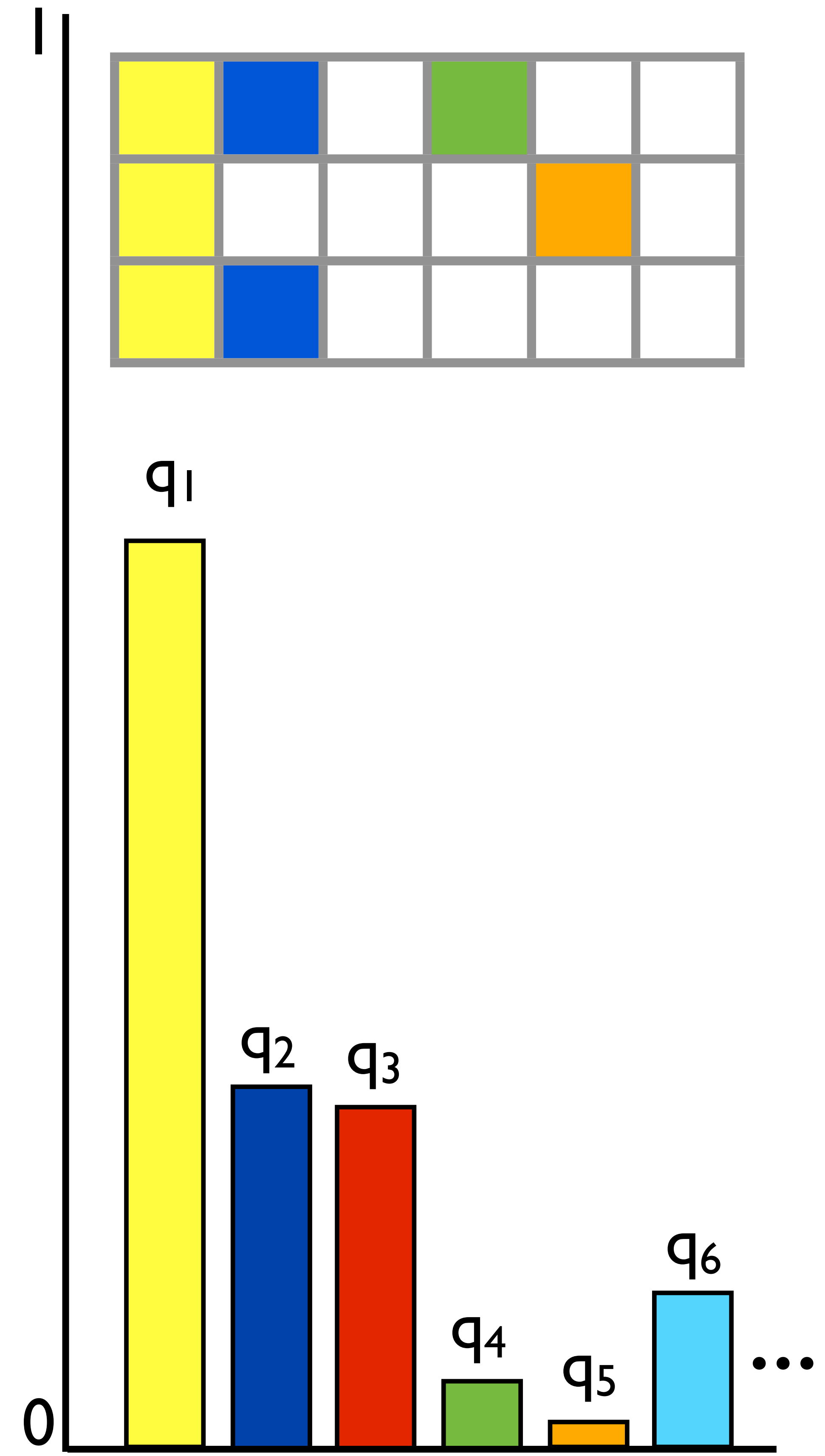


[Broderick, Pitman, Jordan (submitted)]

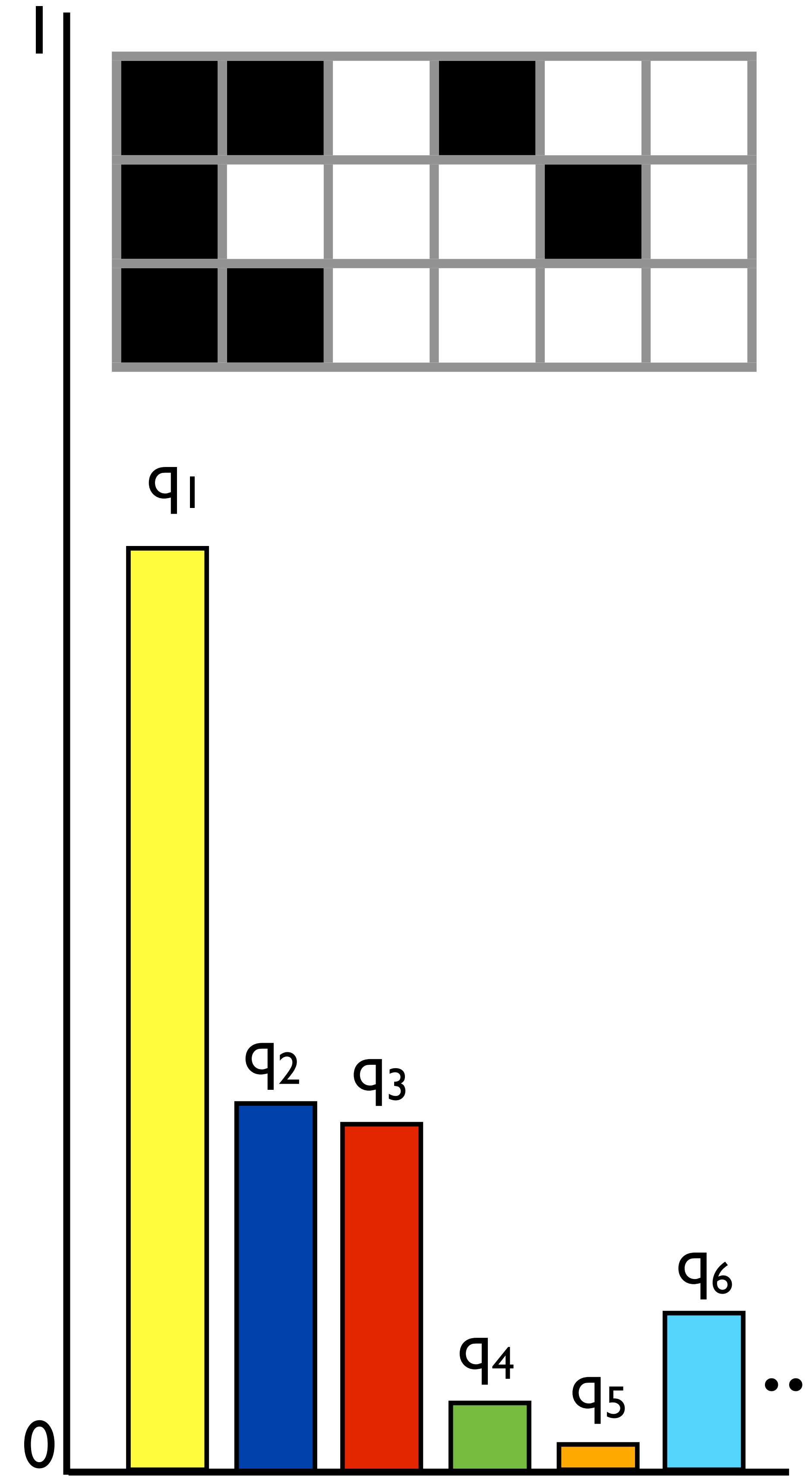
Frequency models: EFPFs?



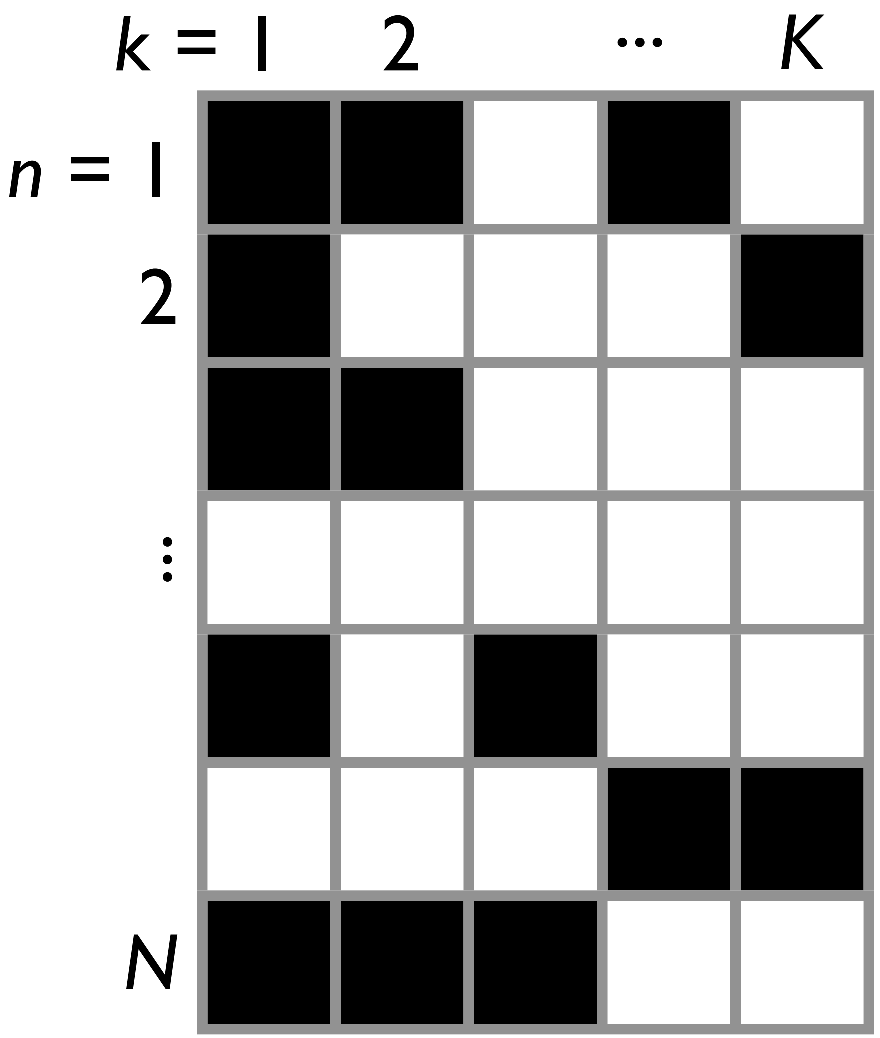
Frequency models: EFPFs?

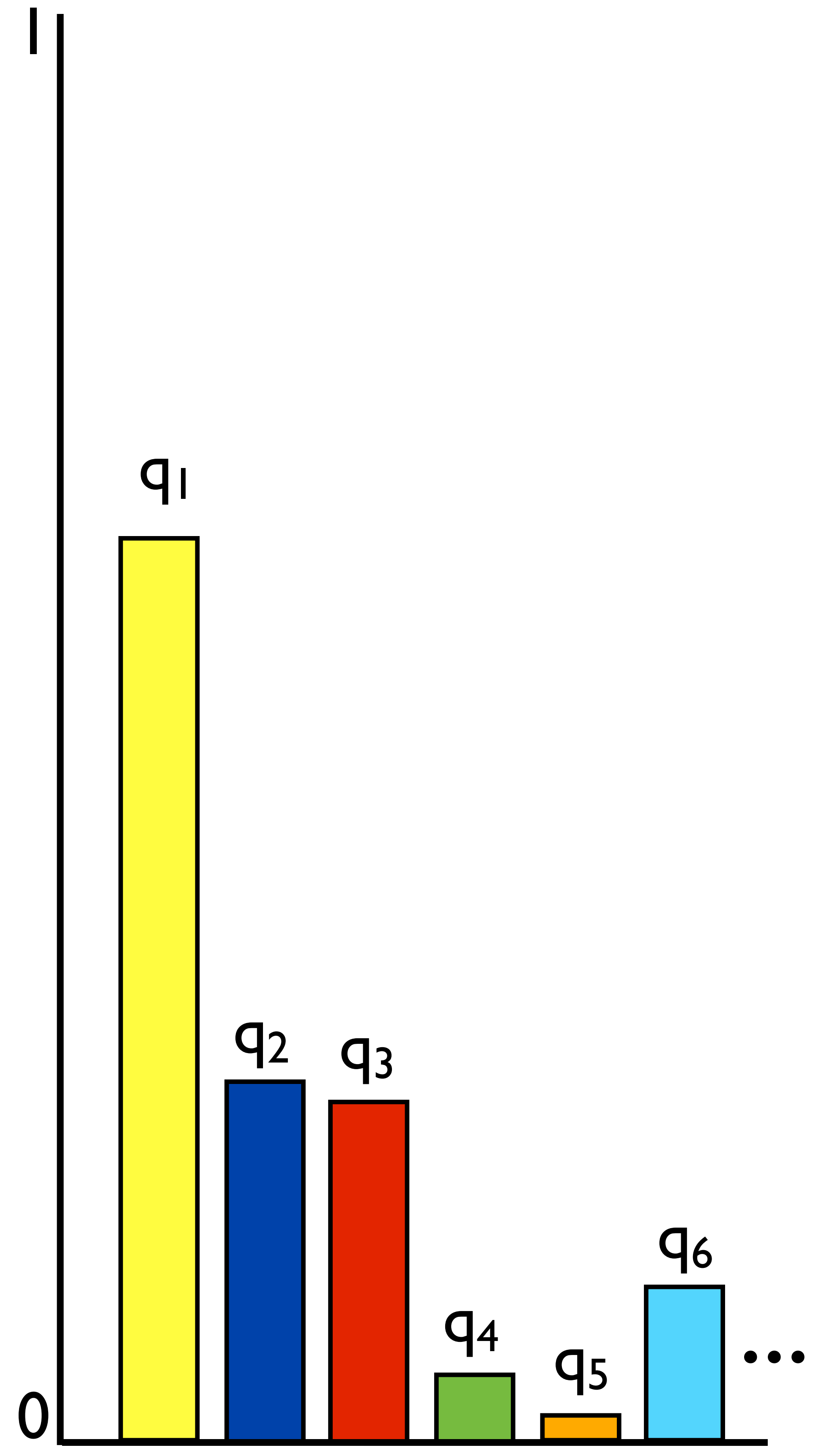


Frequency models: EFPFs?

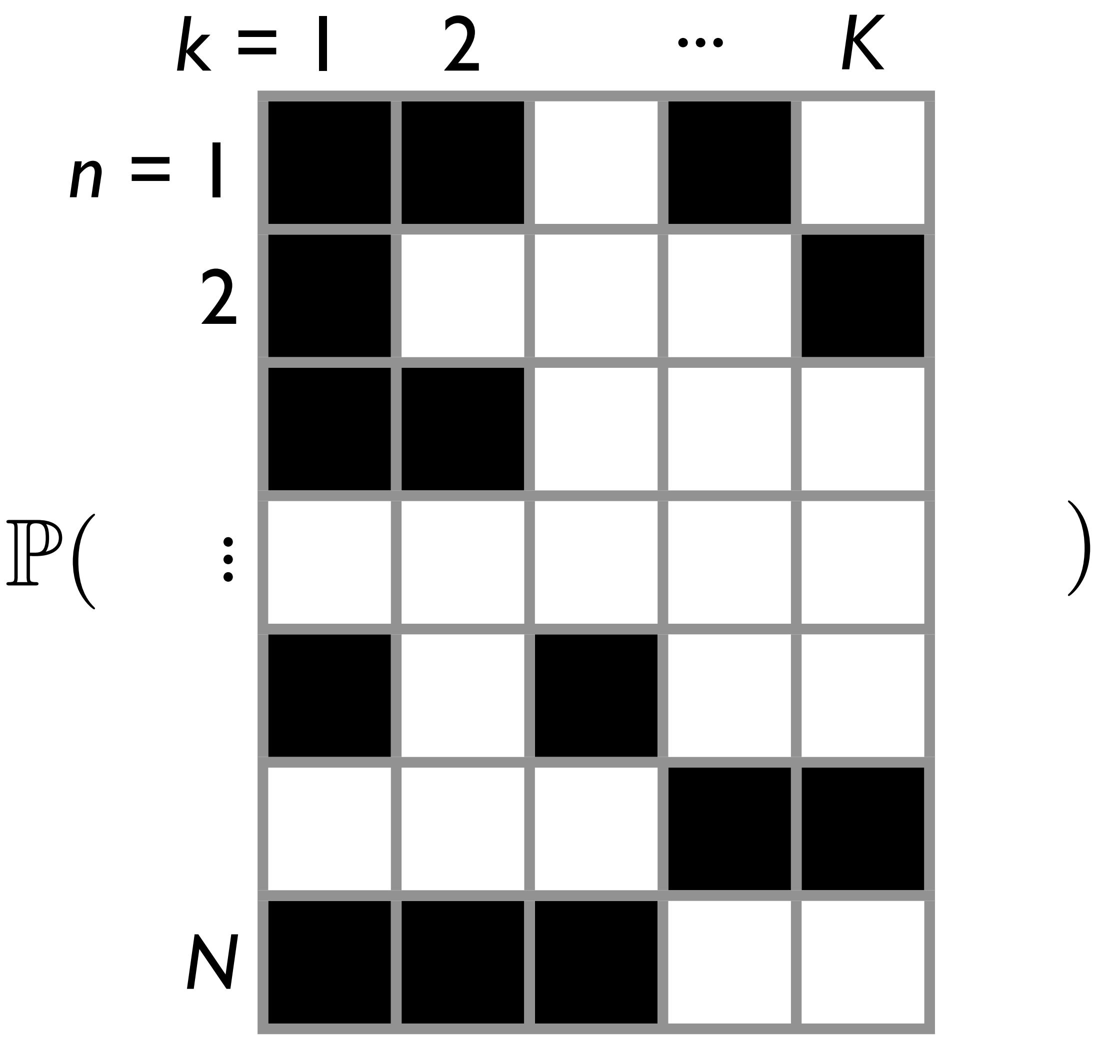


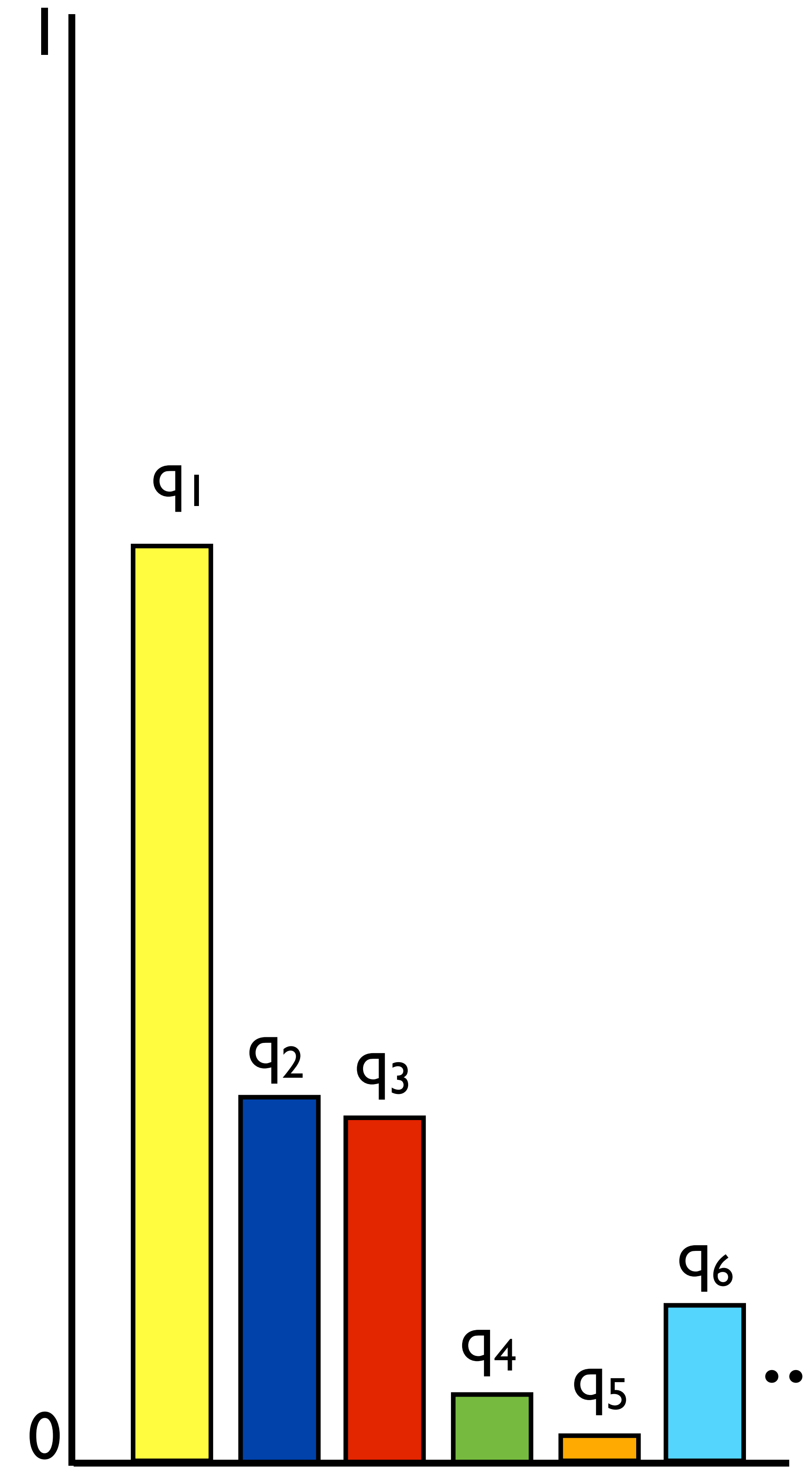
Frequency models: EFPFs?



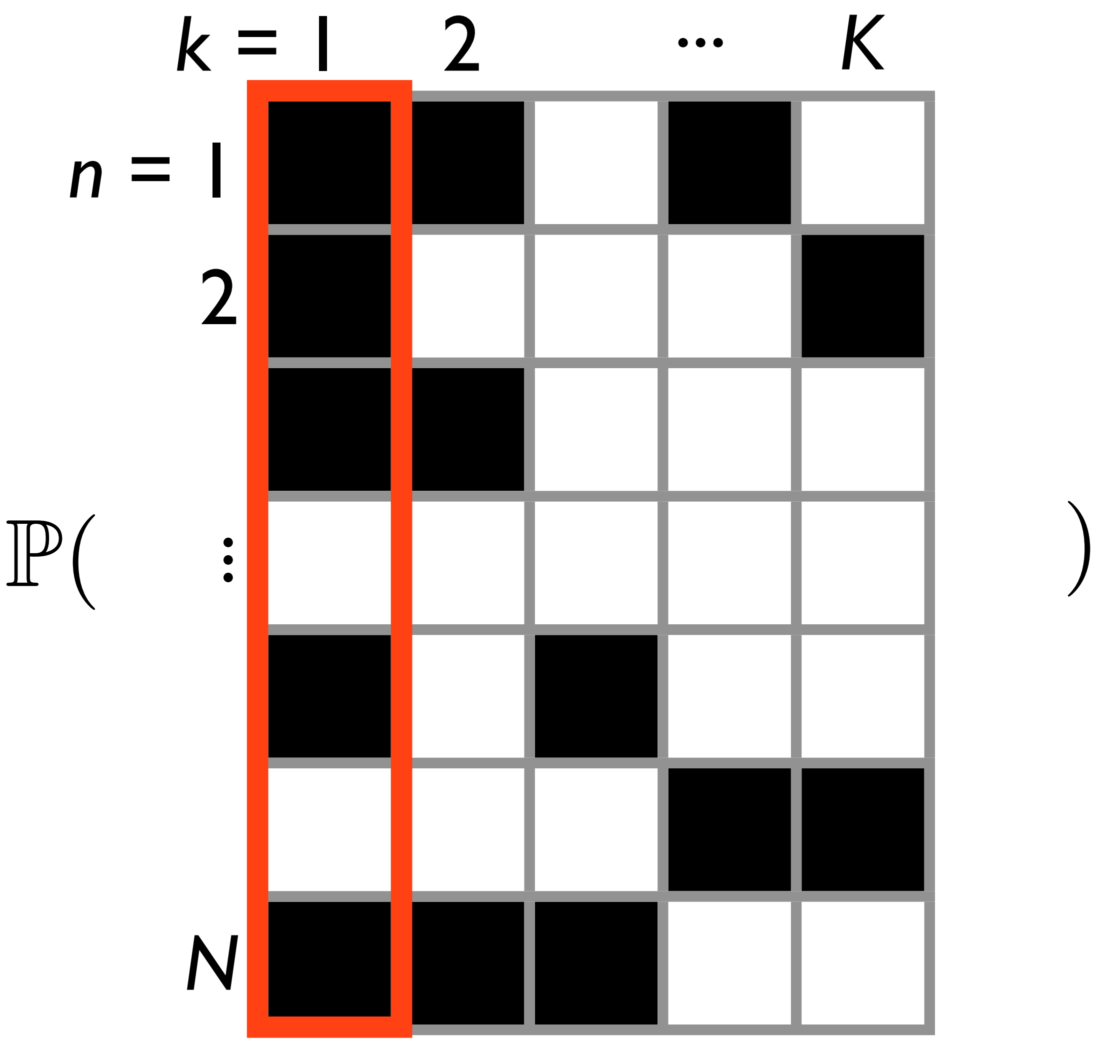


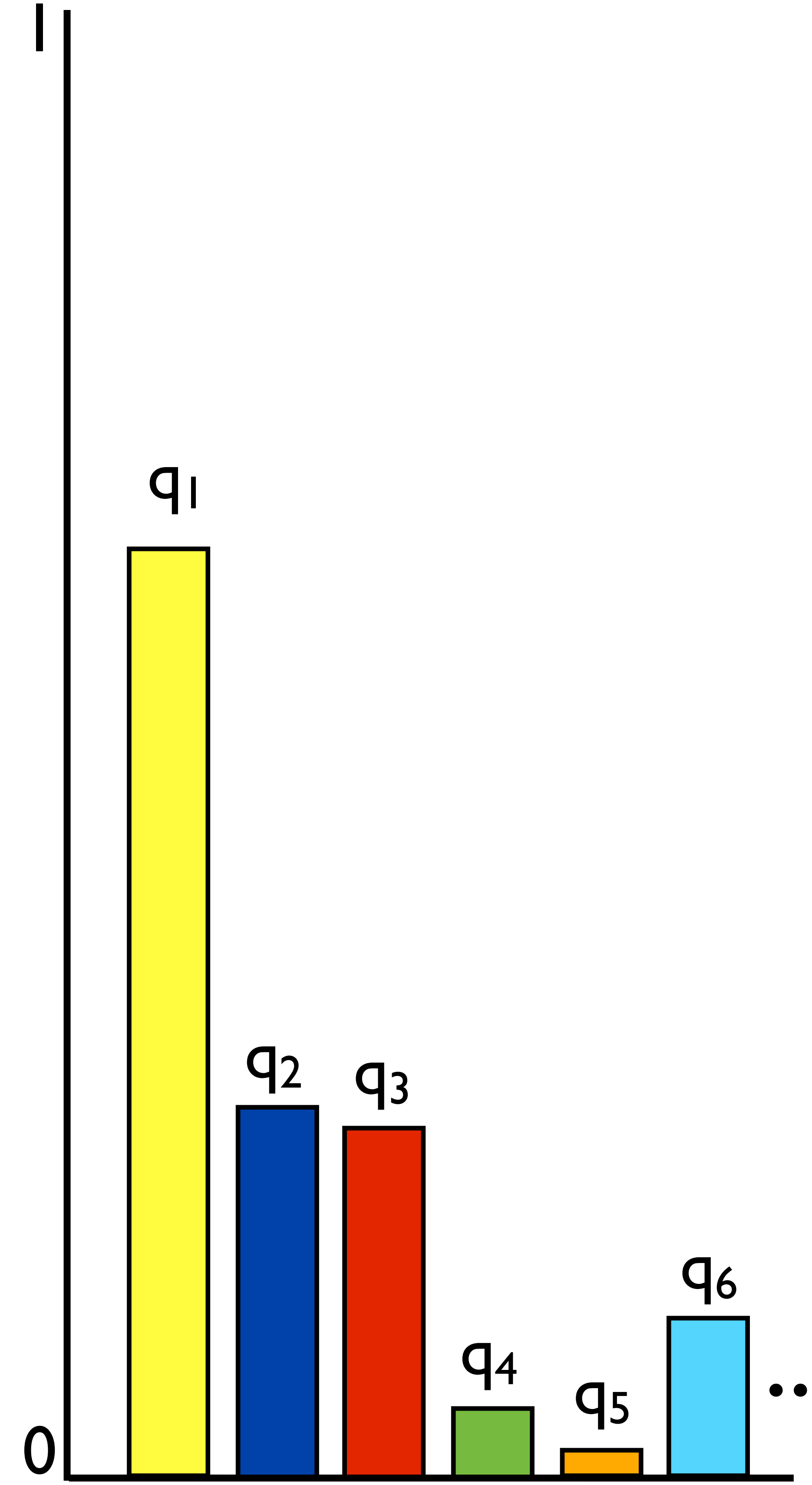
Frequency models: EFPFs?



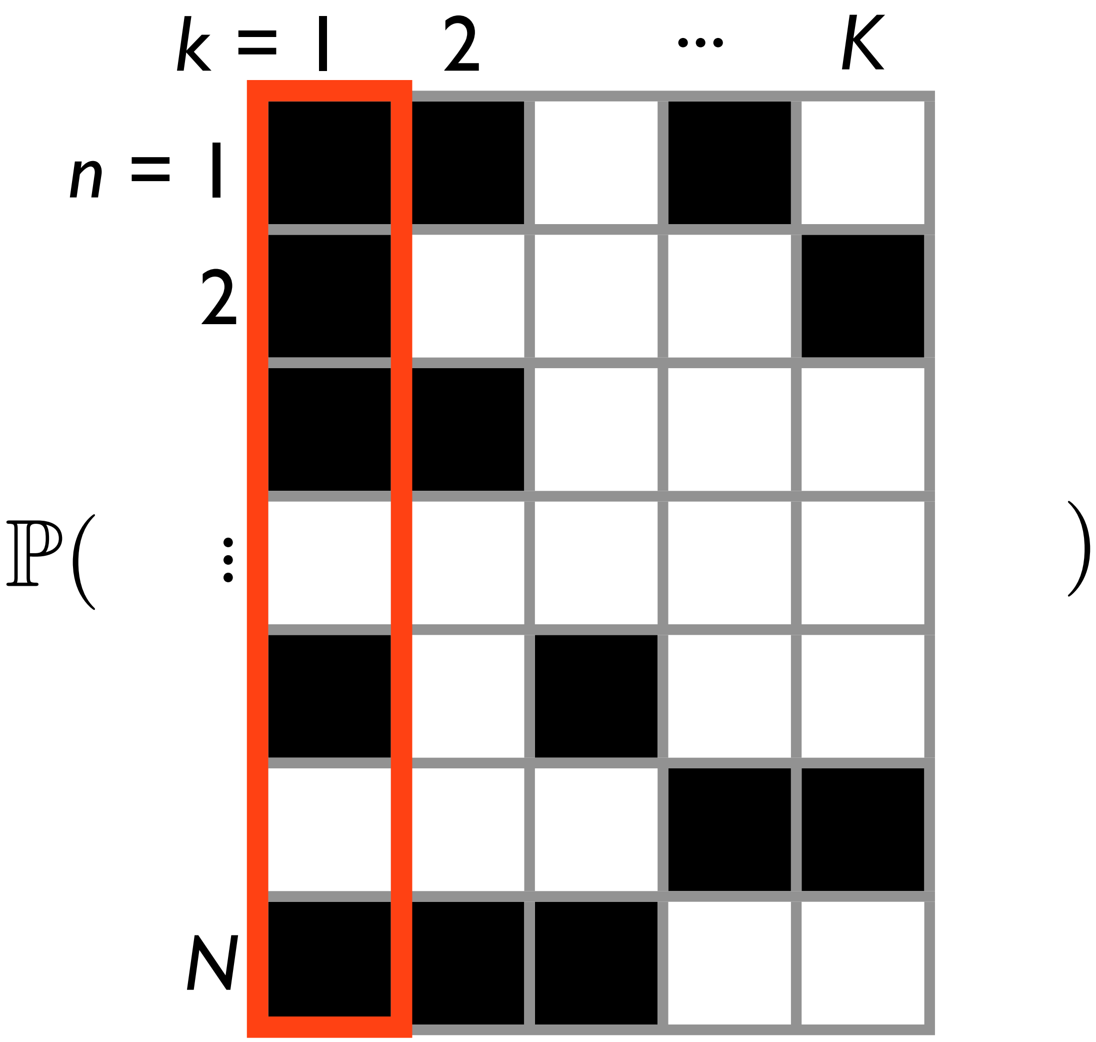


Frequency models: EFPFs?

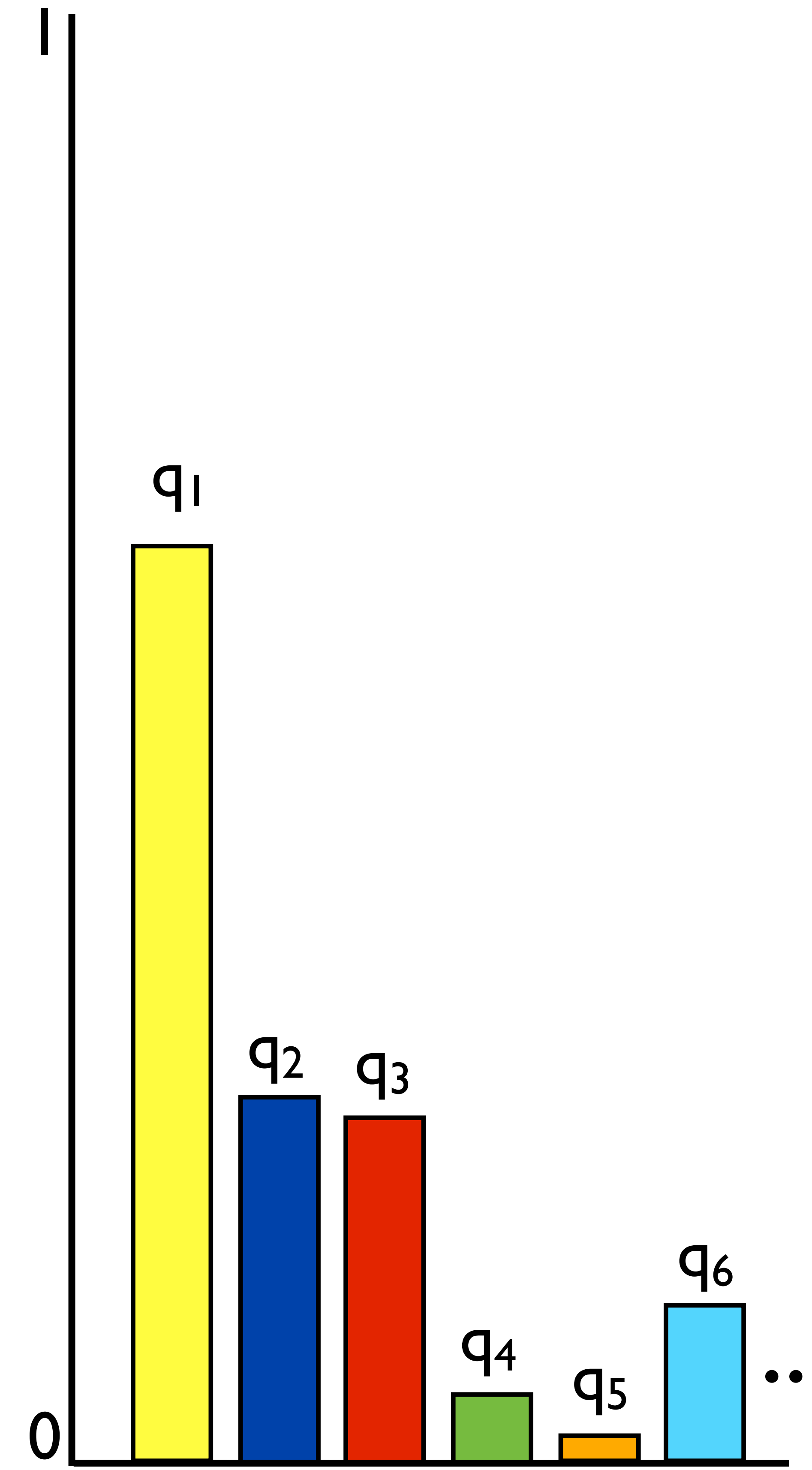




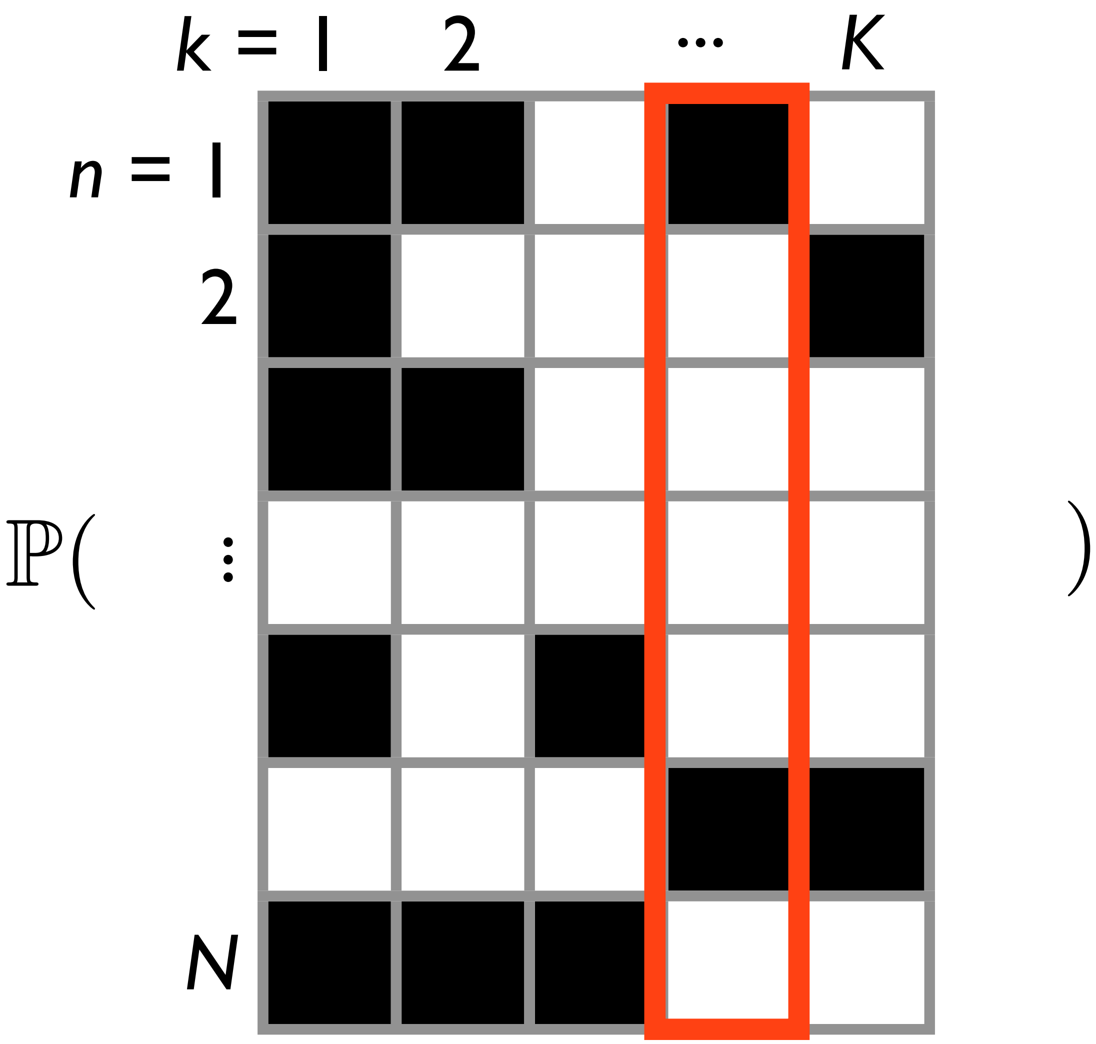
Frequency models: EFPFs?



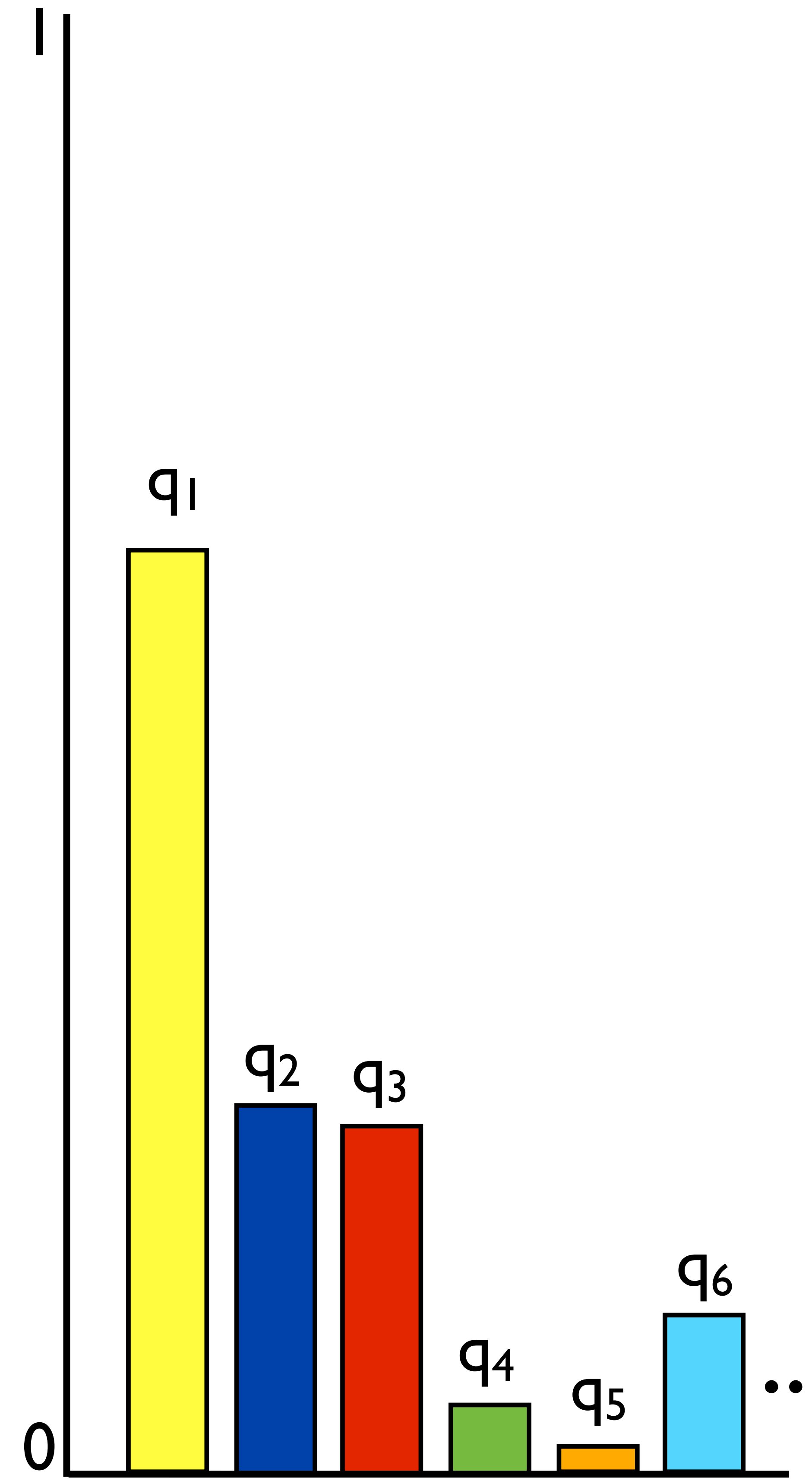
$$q_1^{S_{N,1}}(1-q_1)^{N-S_{N,1}}$$



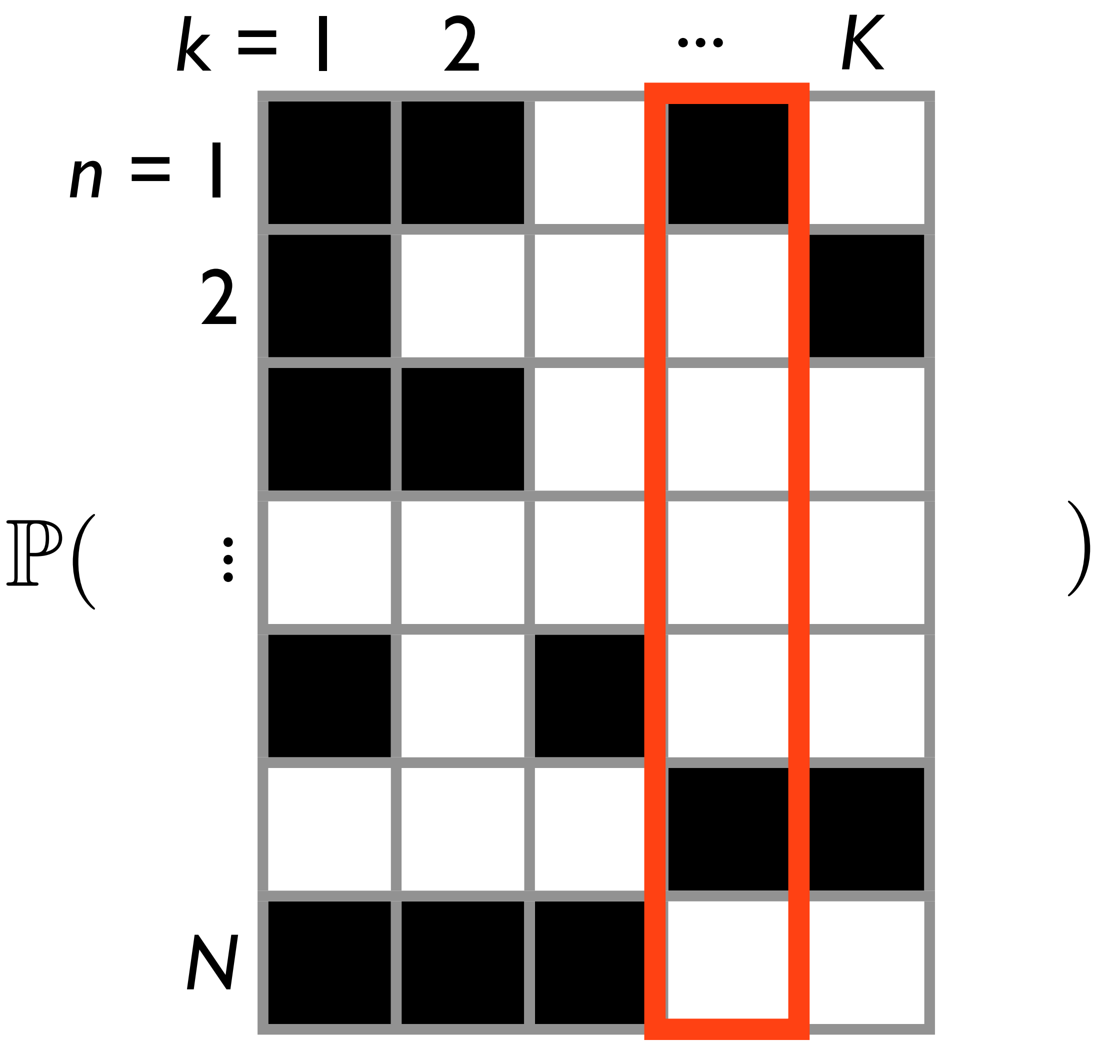
Frequency models: EFPFs?



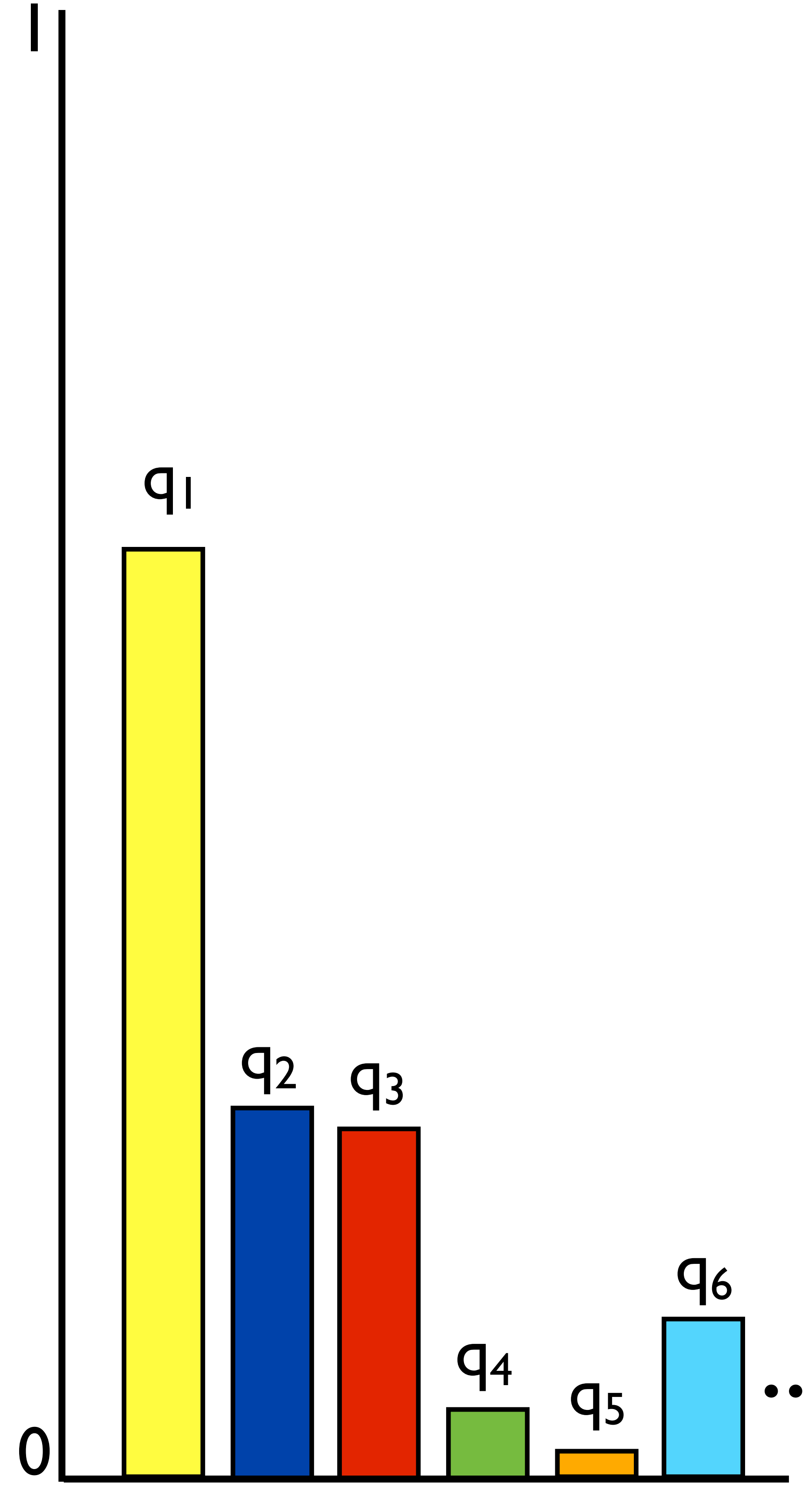
$$q_k^{S_{N,k}}(1-q_k)^{N-S_{N,k}}$$



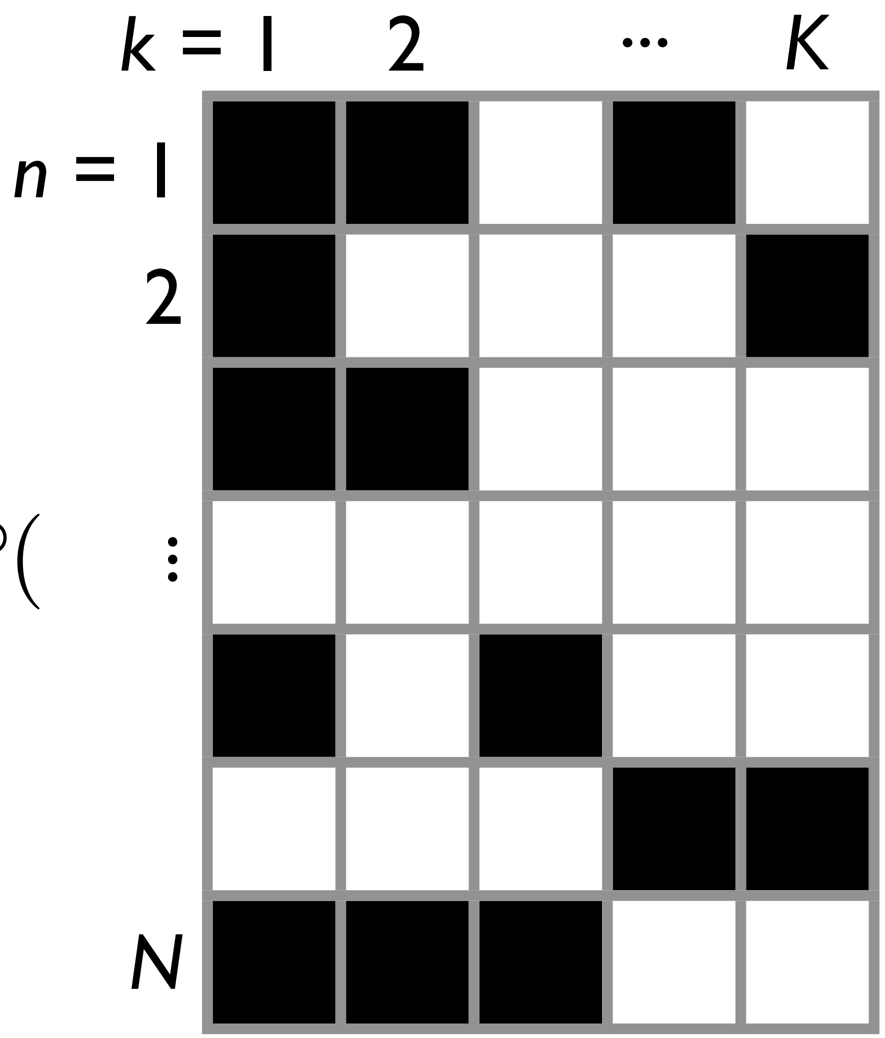
Frequency models: EFPFs?



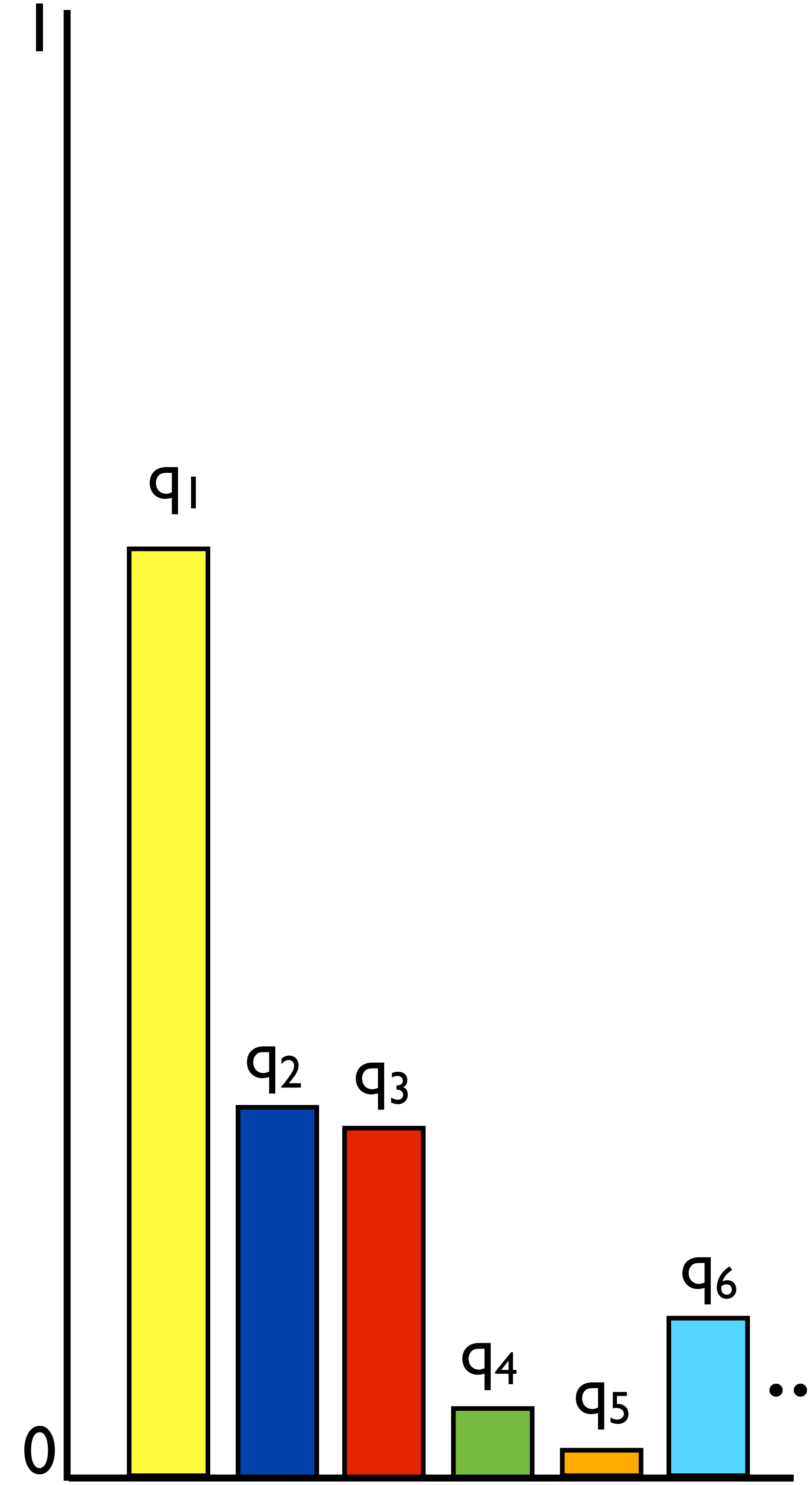
$$q_{i_k}^{S_{N,k}}(1-q_{i_k})^{N-S_{N,k}}$$



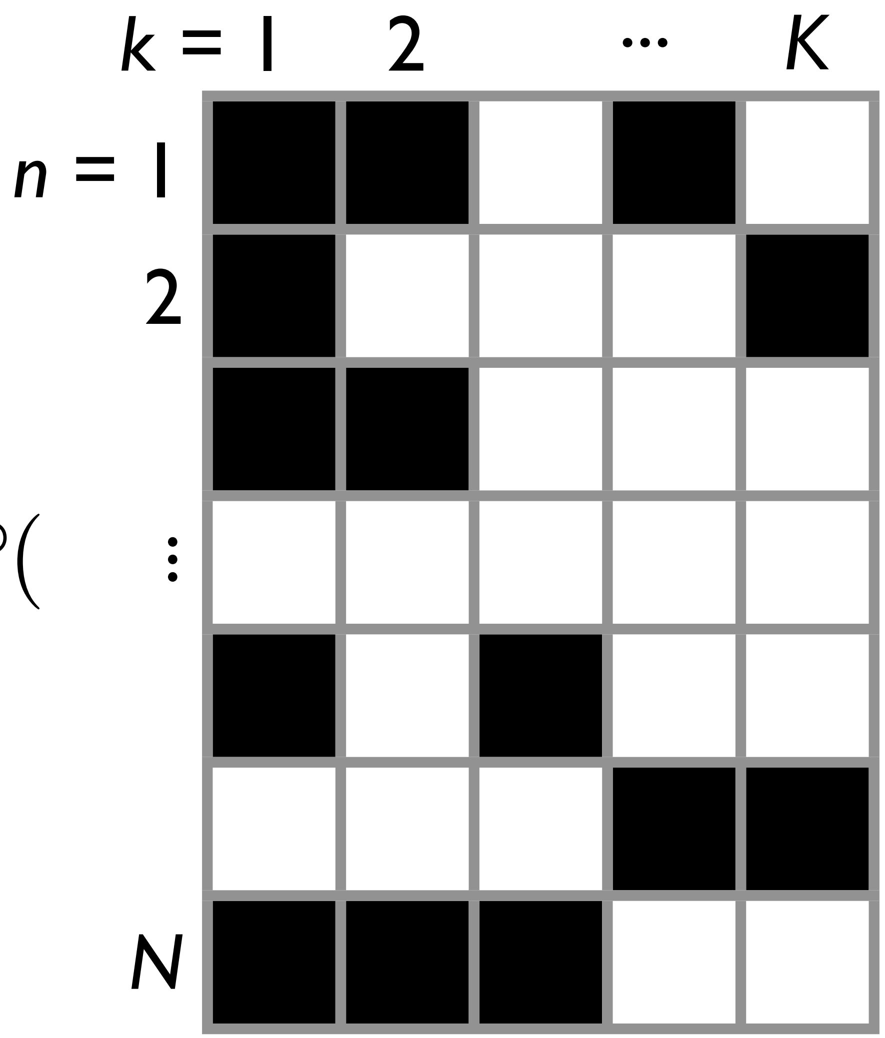
Frequency models: EFPFs?

$\mathbb{P}($  $)$

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



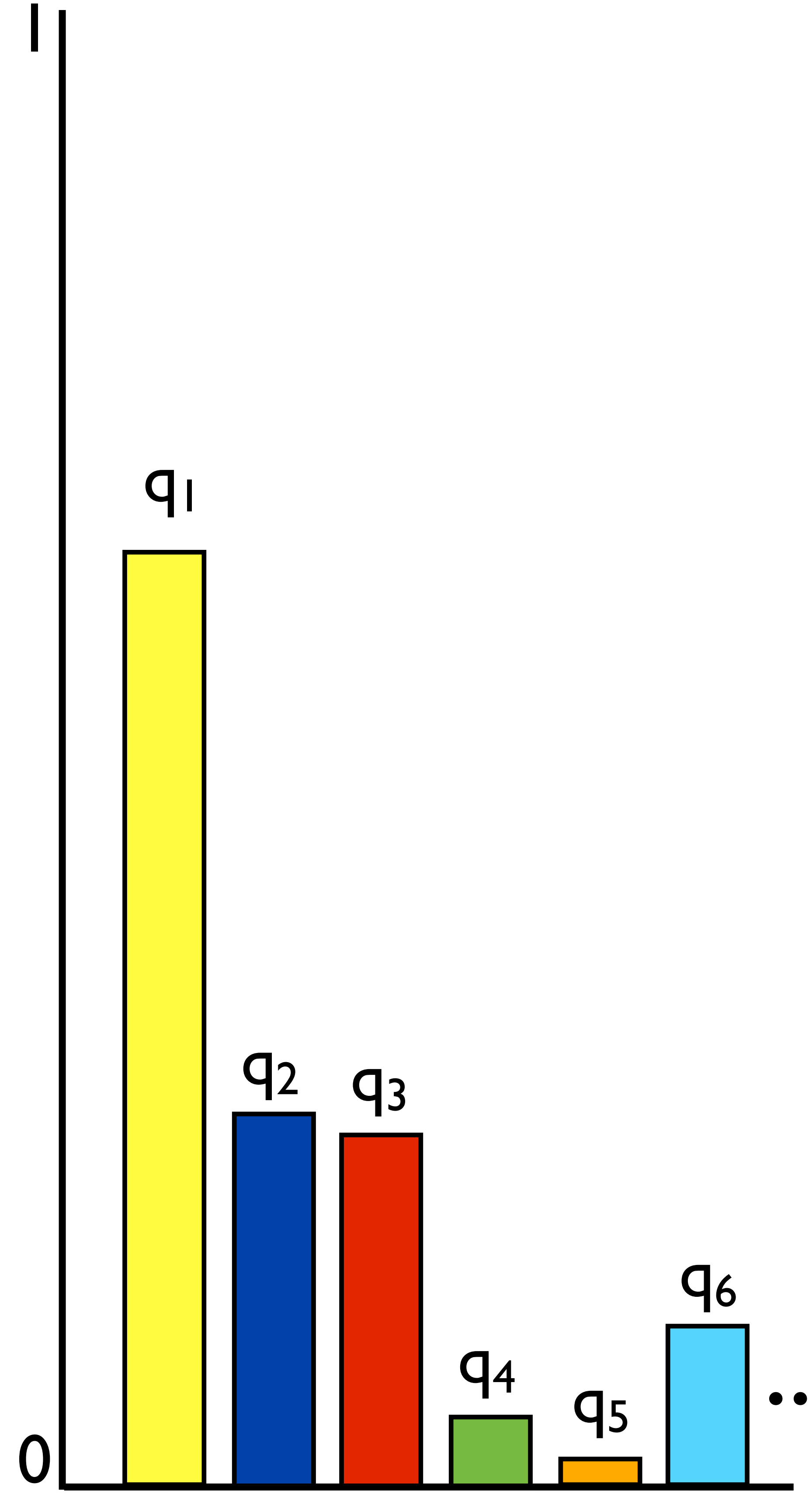
Frequency models: EFPFs?

$\mathbb{P}($


$)$

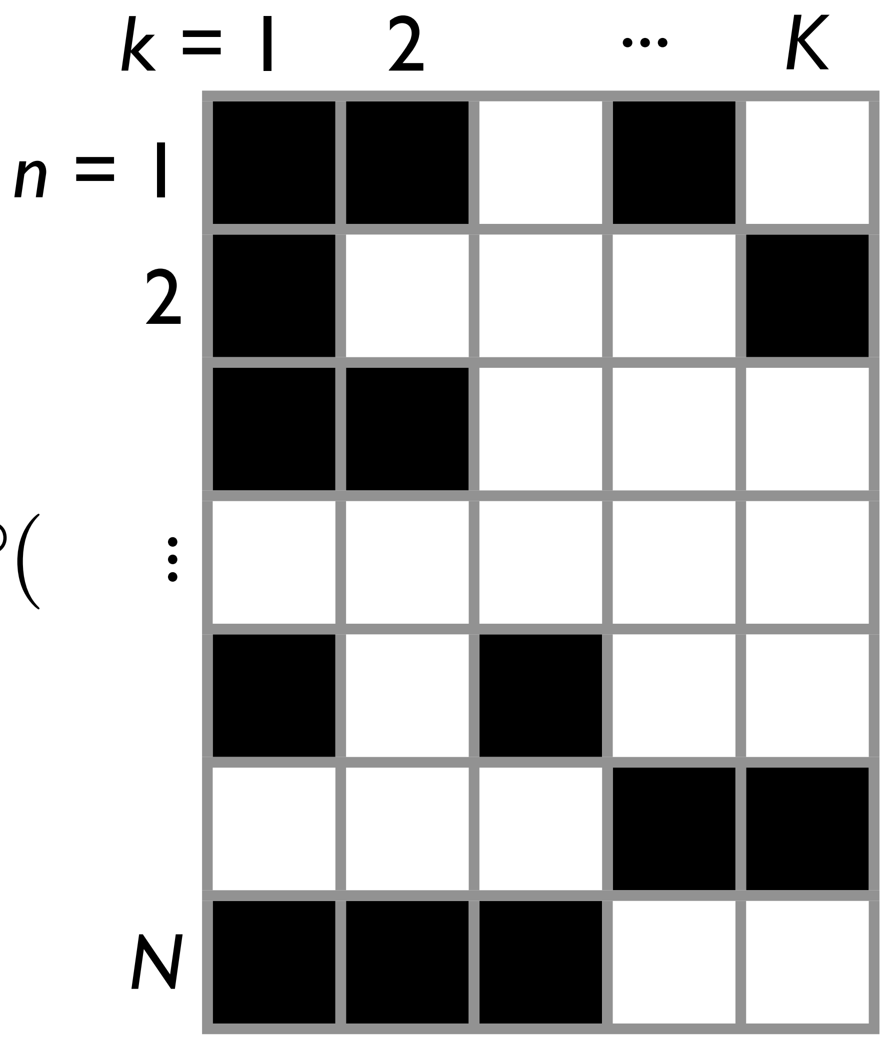
$$\prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}}$$

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



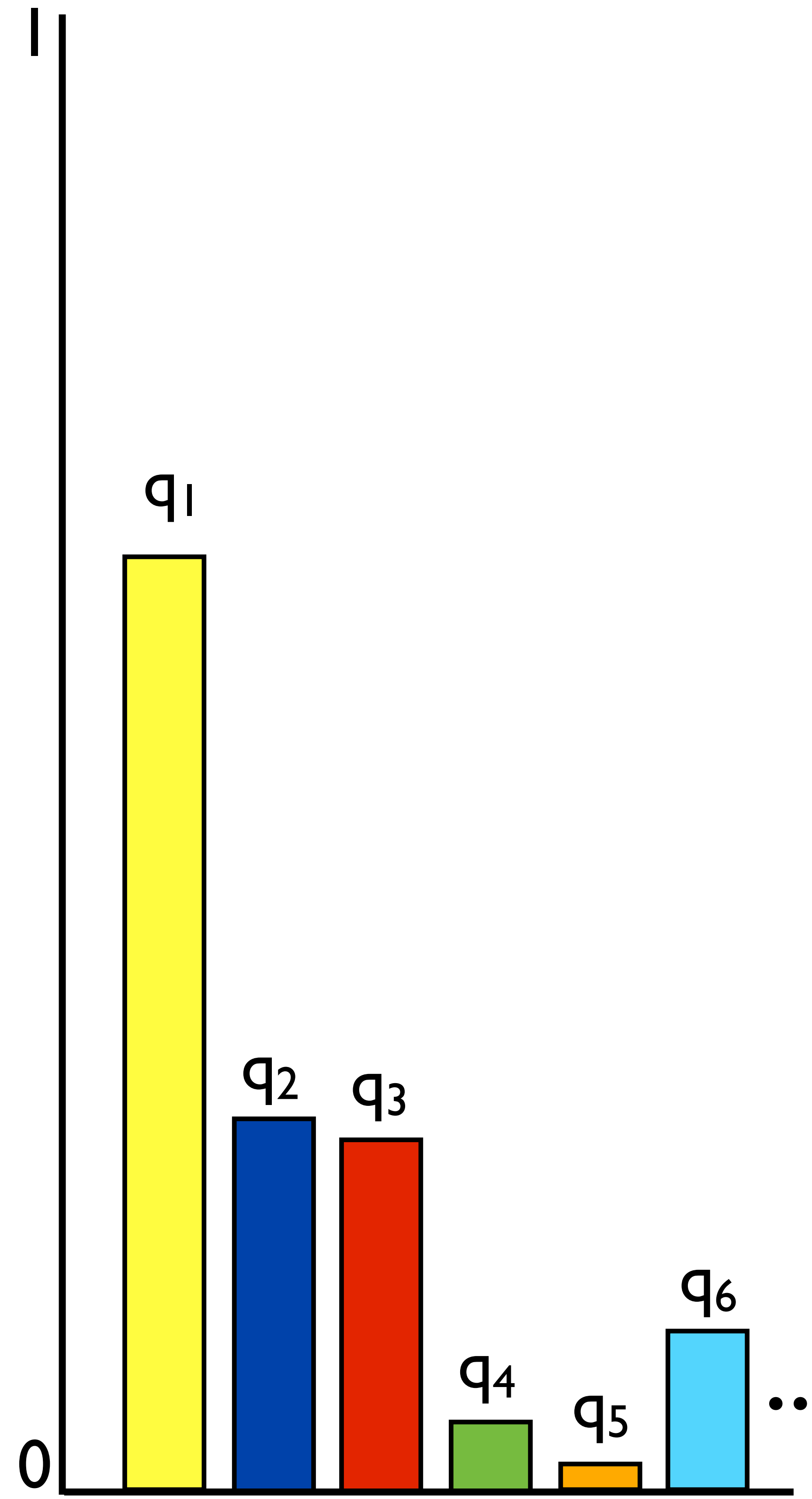
Frequency models: EFPFs?

$\mathbb{P}($

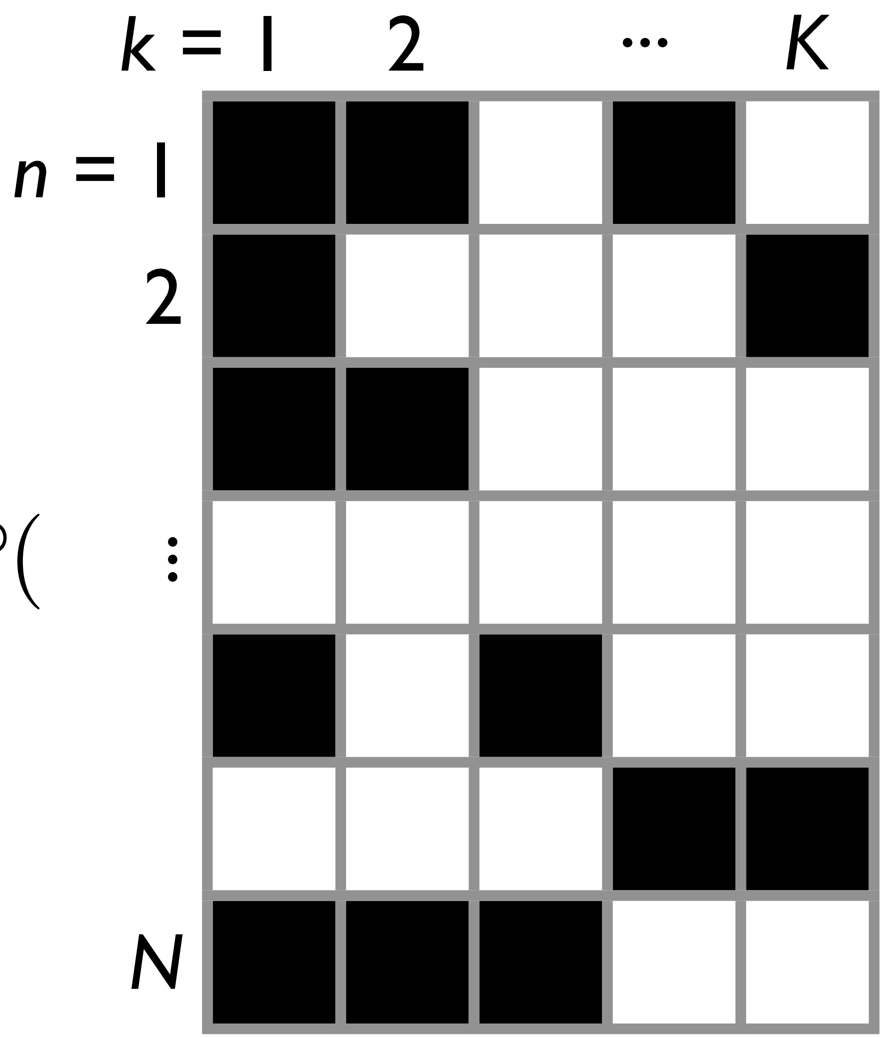


$)$

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



Frequency models: EFPFs?

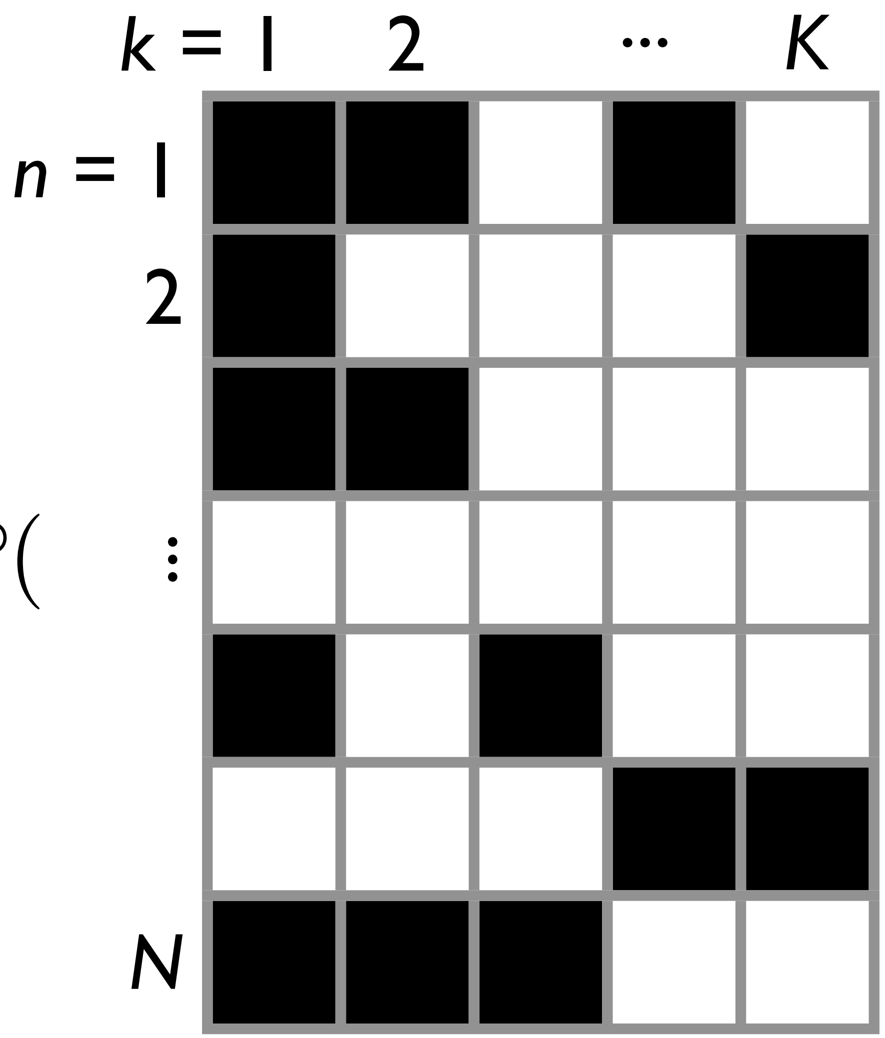
$\mathbb{P}($  $)$

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

[Broderick, Pitman, Jordan (submitted)]

Frequency models: EFPFs?

$\mathbb{P}($



$)$

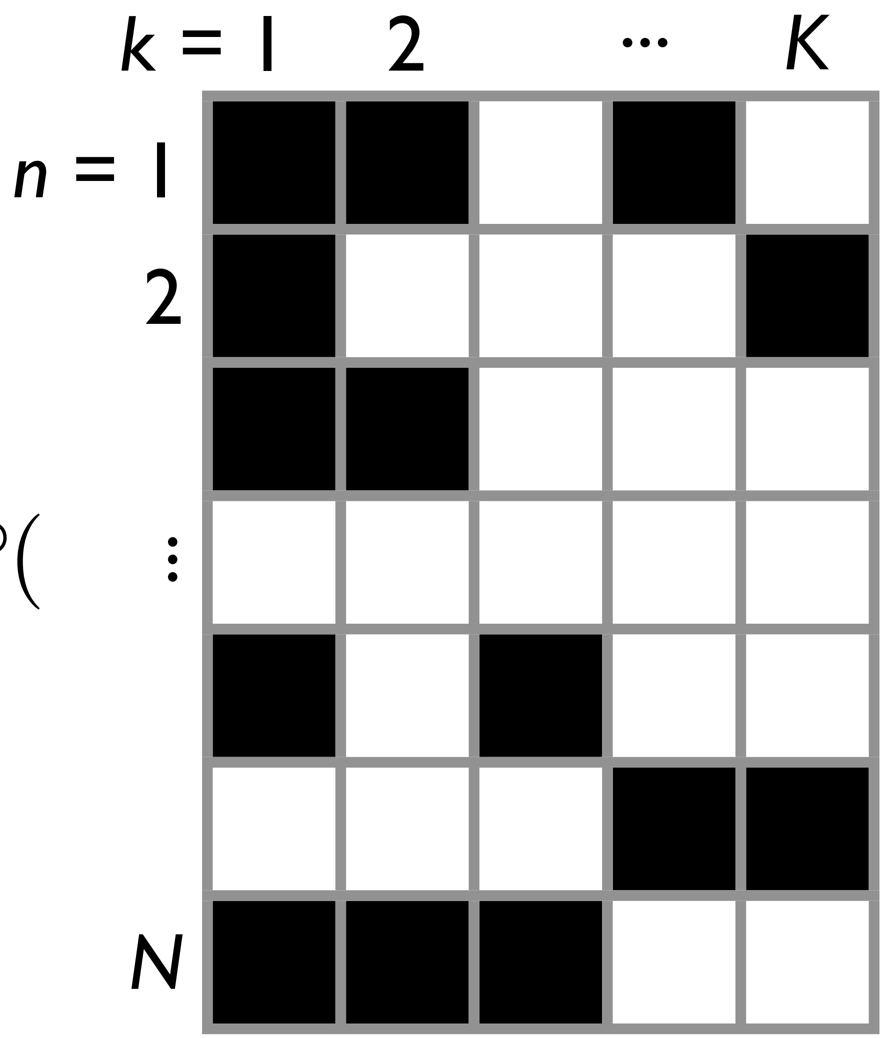
$$= \mathbb{E}\Big[\sum_{\text{distinct } i_k} \frac{1}{K!} \prod_{k=1}^{K} q_{i_k}^{S_{N,k}} (1 - q_{i_k})^{N - S_{N,k}} \cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N \Big]$$

Size of *k*th feature

[Broderick, Pitman, Jordan (submitted)]

Frequency models: EFPFs?

$\mathbb{P}($



$)$

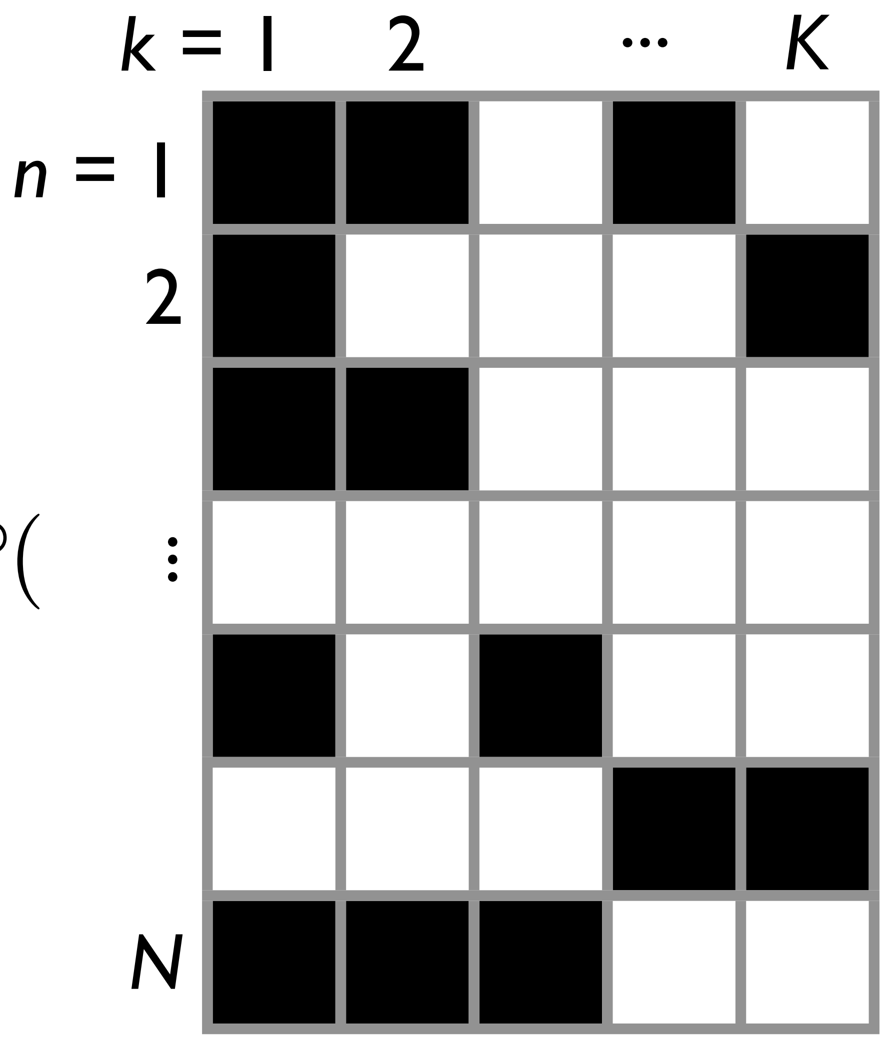
Number of features

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 (submitted)]

Frequency models: EFPFs?

$\mathbb{P}($


$)$

Number of features

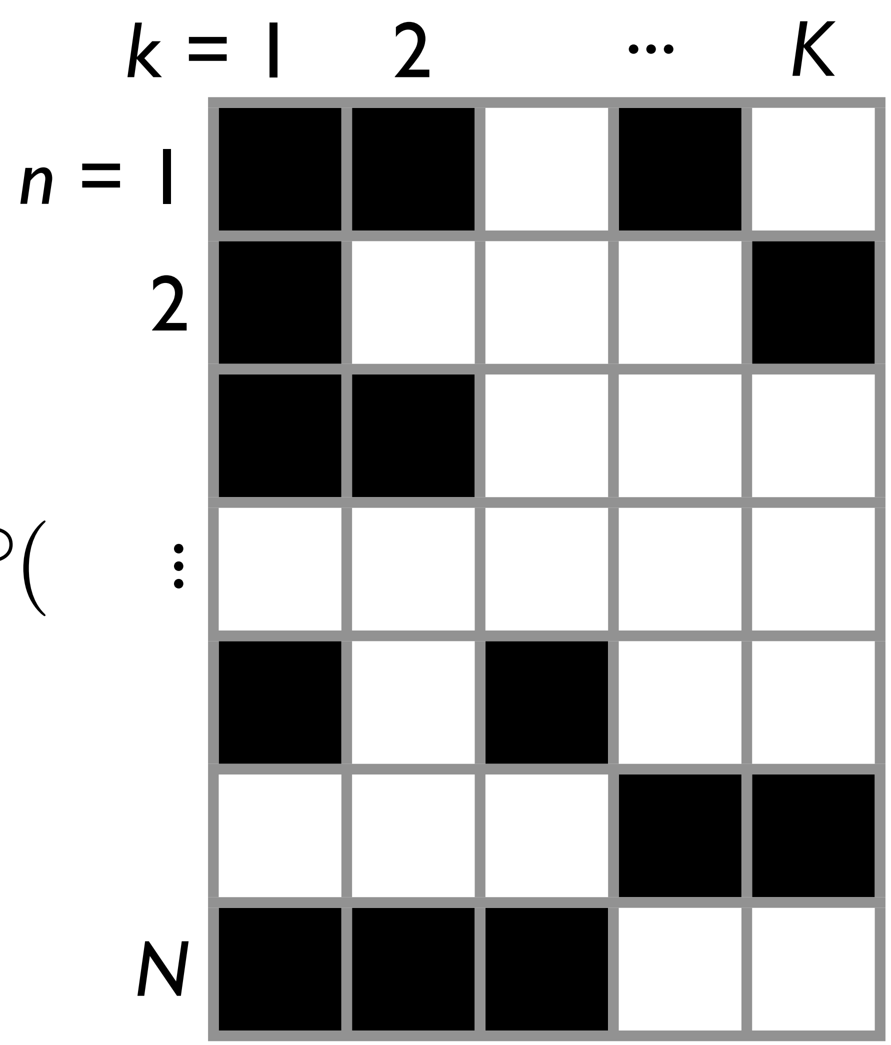
Number of data points

Size of *k*th 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 (submitted)]

Frequency models: EFPFs?

$\mathbb{P}($  $)$

Number of features

Number of data points

Size of *k*th feature

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

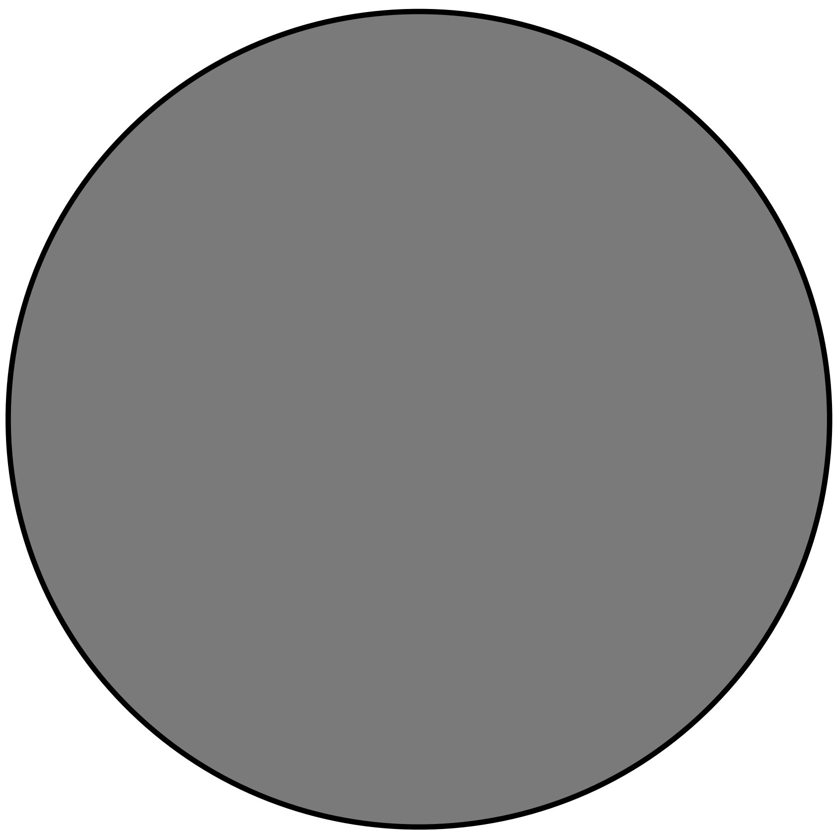
$$\cdot \prod_{j \notin \{i_k\}_{k=1}^{K}} (1 - q_j)^N \Big] = p(N; S_{N,1}, S_{N,2}, \ldots, S_{N,K})$$

EFPF

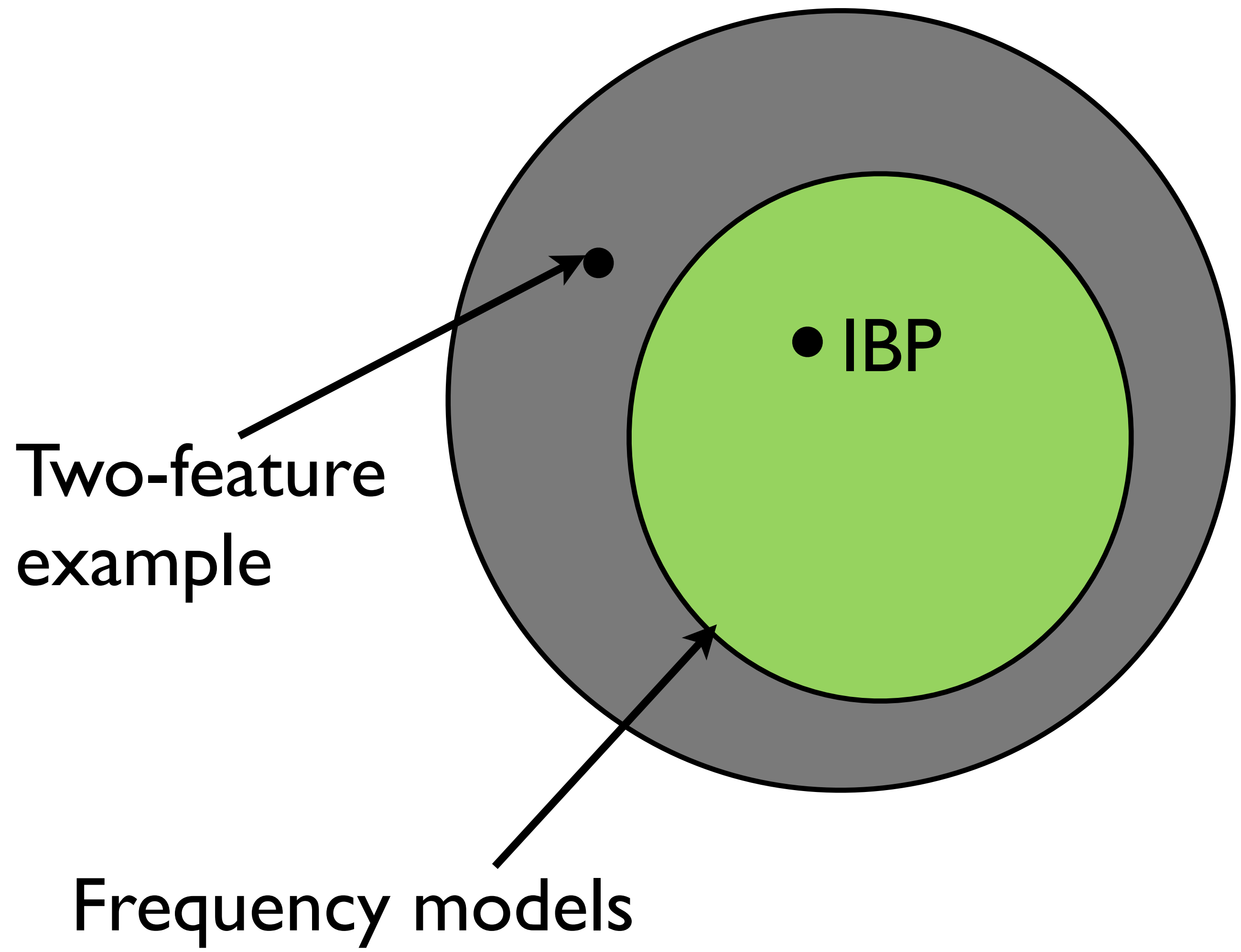
[Broderick, Pitman, Jordan (submitted)]

Frequency models: EFPFs?

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



Exchangeable feature distributions
= Feature paintbox allocations

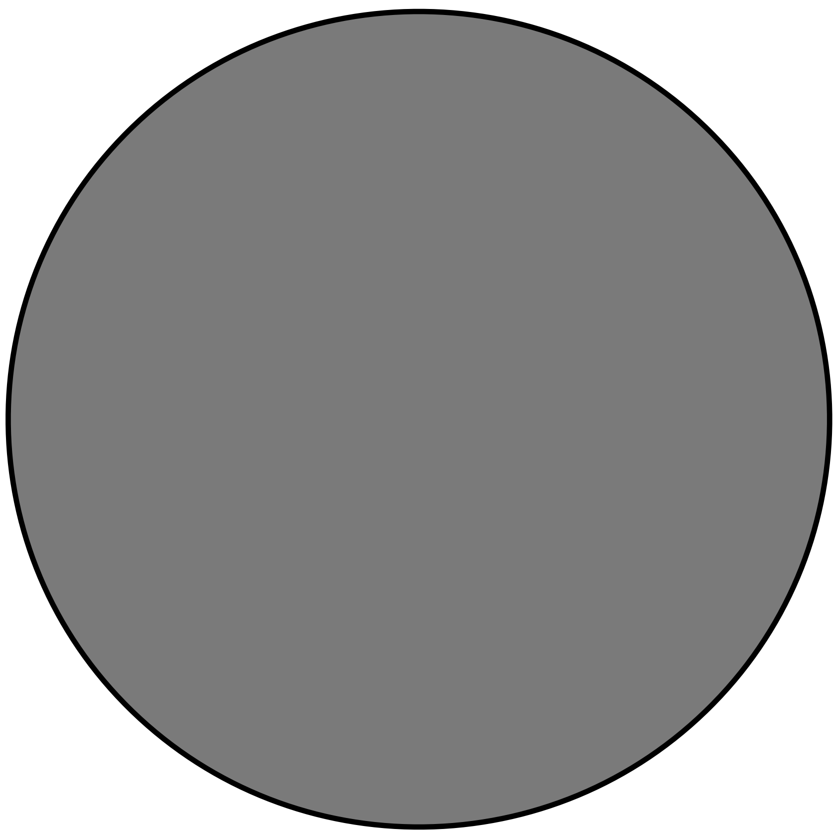


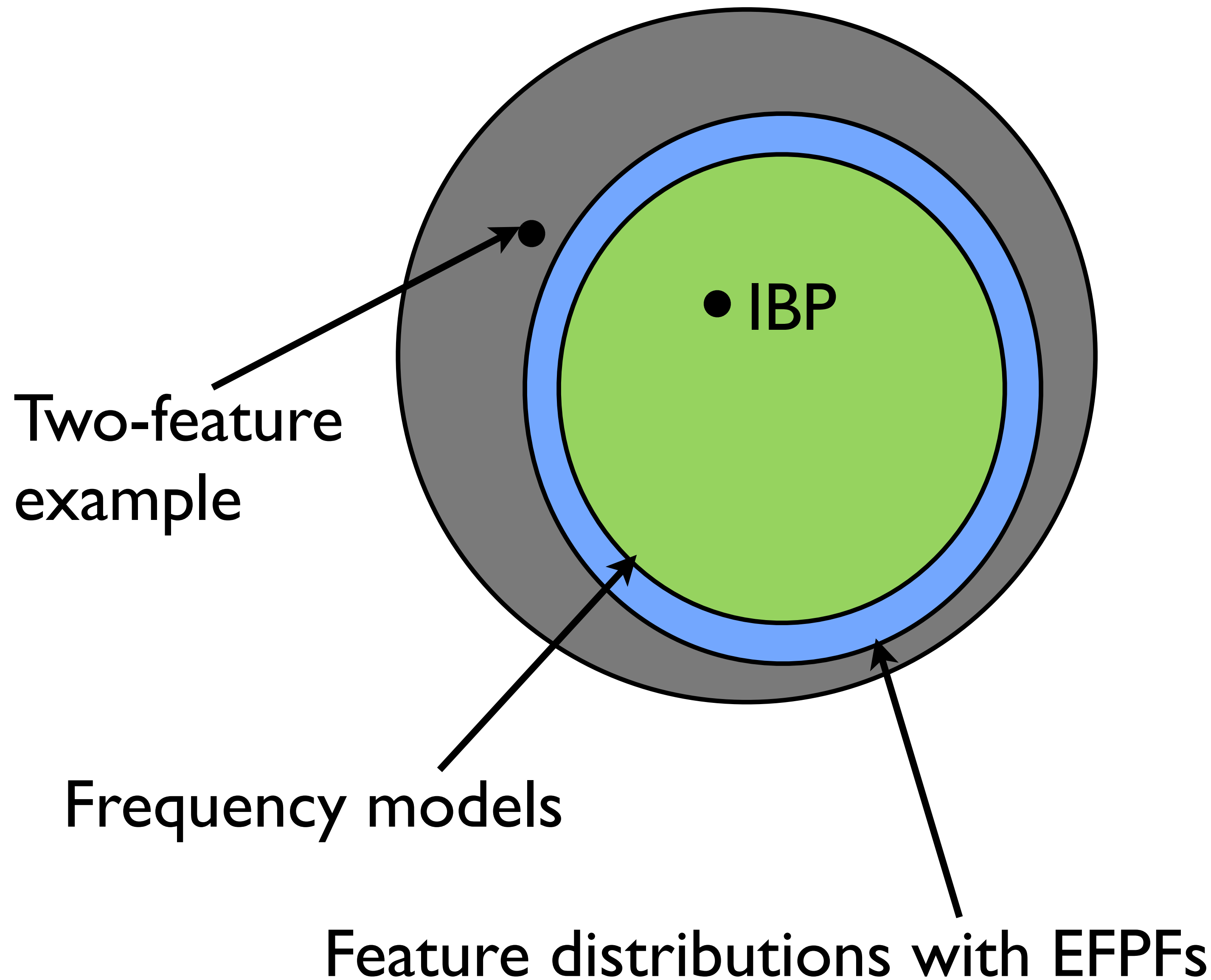
[Broderick, Pitman, Jordan (submitted)]

Frequency models: EFPFs?

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

Exchangeable feature distributions
= Feature paintbox allocations



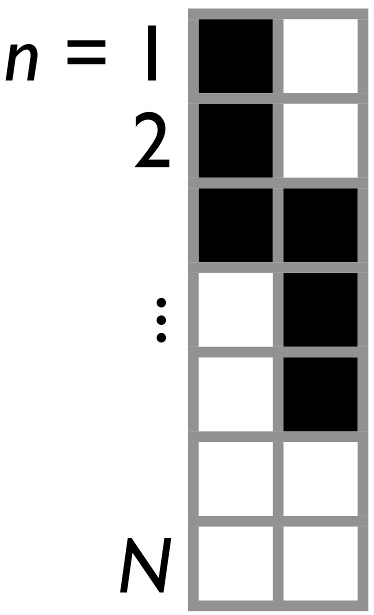


[Broderick, Pitman, Jordan (submitted)]

Distributions with EFPFs: frequencies?

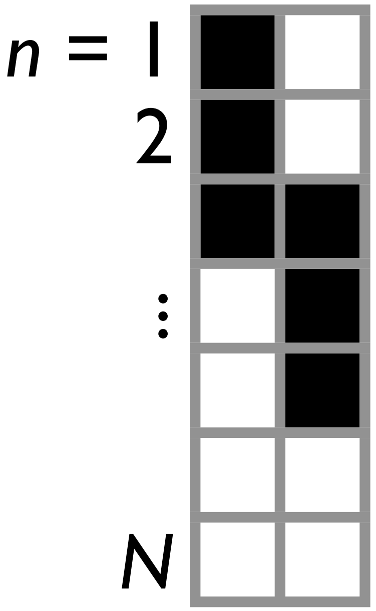
Distributions with EFPFs: frequencies?

Feature allocation



Distributions with EFPFs: frequencies?

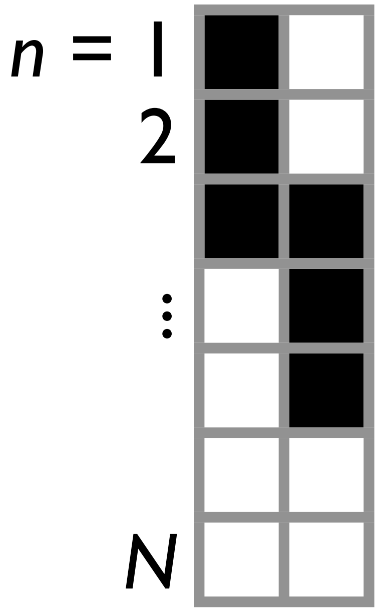
Feature allocation



K=2 for all N

Distributions with EFPFs: frequencies?

Feature allocation



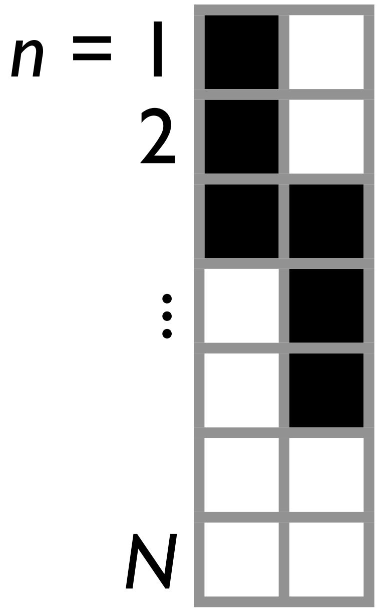
K=2 for all N

Assume EFPF

$$p(N; S_{N,1}, S_{N,2})$$

Distributions with EFPFs: frequencies?

Feature allocation



K=2 for all N

Assume EFPF

$$p(N; S_{N,1}, S_{N,2})$$

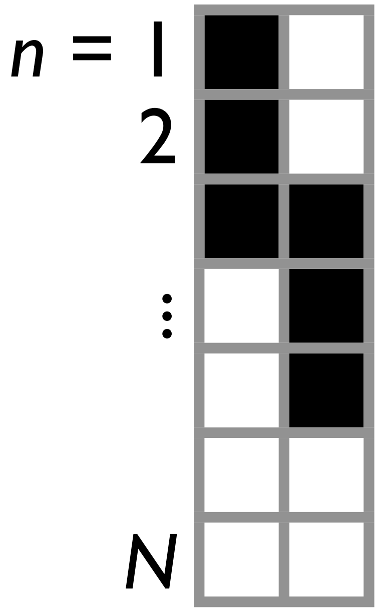
Want to show:

$$\exists q_1$$

$$\exists q_2$$

Distributions with EFPFs: frequencies?

Feature allocation

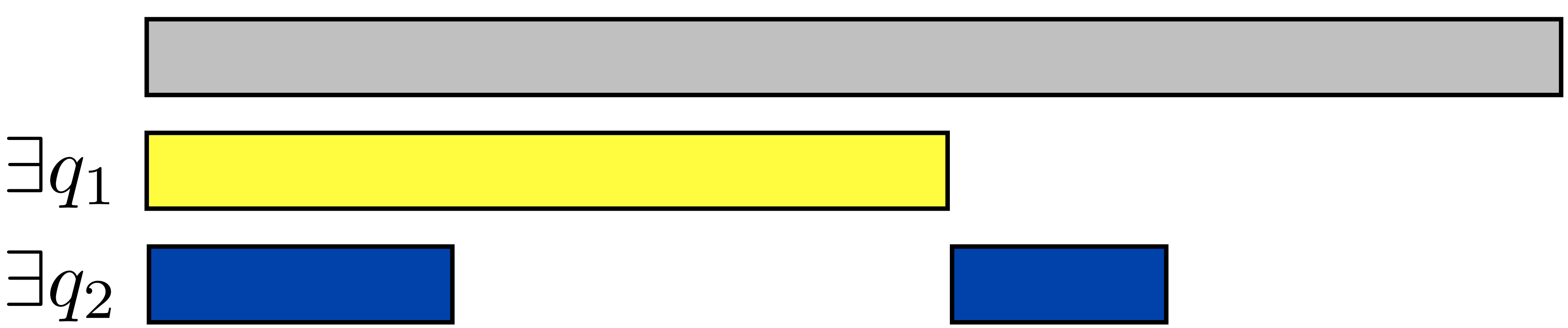


K=2 for all N

Assume EFPF

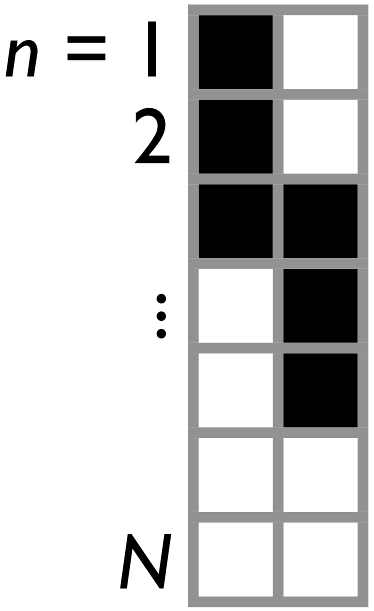
$p(N; S_{N,1}, S_{N,2})$

Want to show:



Distributions with EFPFs: frequencies?

Feature allocation



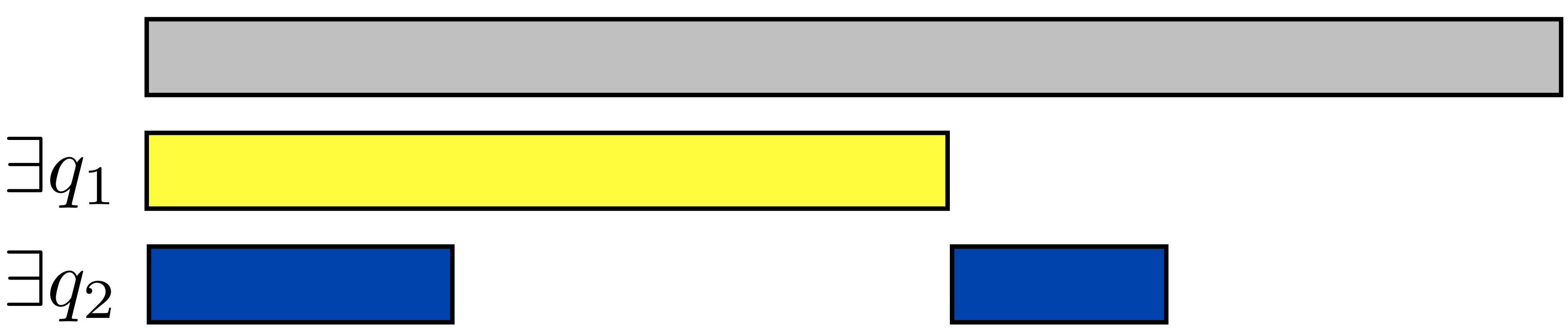
K=2 for all N

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

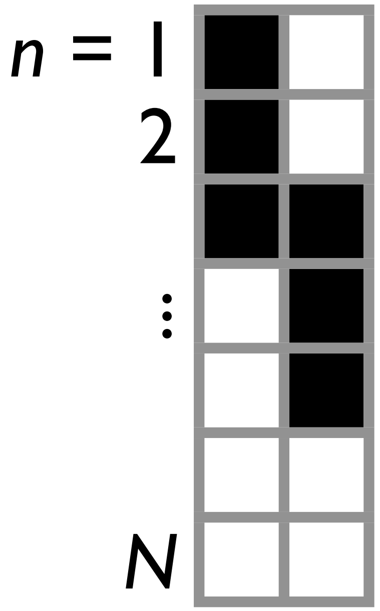
Want to show:

e.g. $\mathbb{P}(\text{row} = \quad | q_{1:2})$



Distributions with EFPFs: frequencies?

Feature allocation



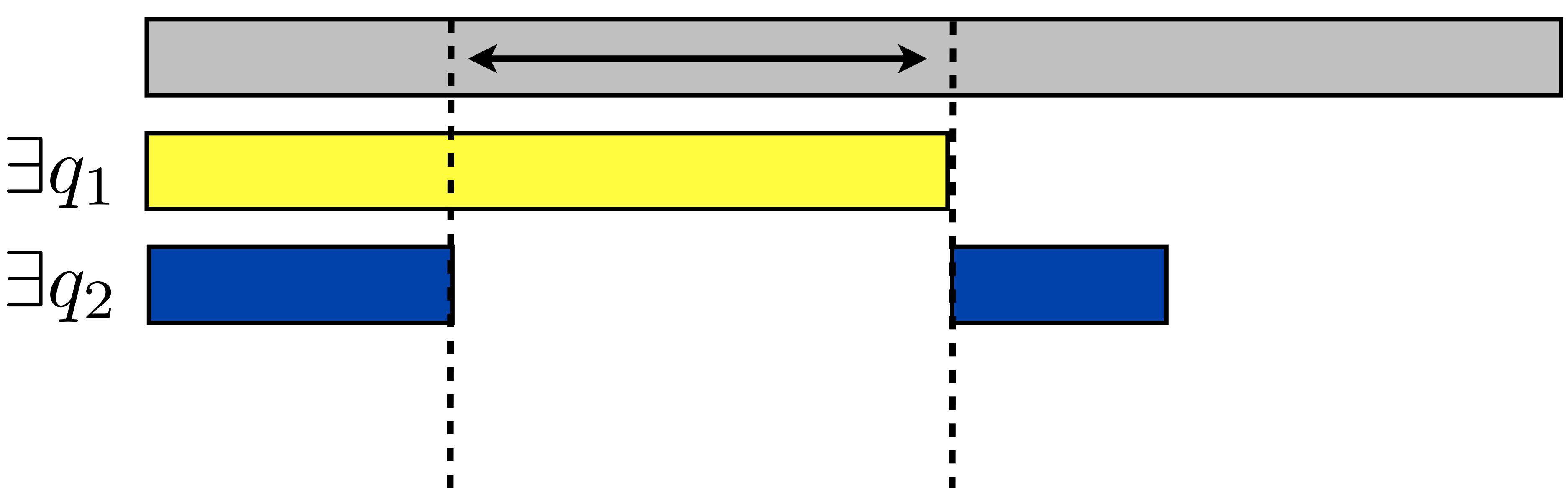
K=2 for all N

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

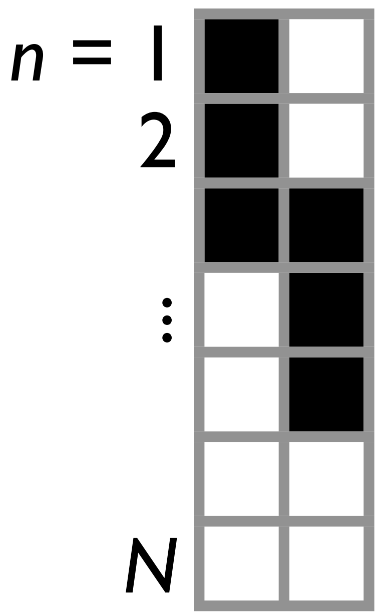
Want to show:

e.g. $\mathbb{P}(\mathrm{row} = \blacksquare\square \,|q_{1:2})$



Distributions with EFPFs: frequencies?

Feature allocation



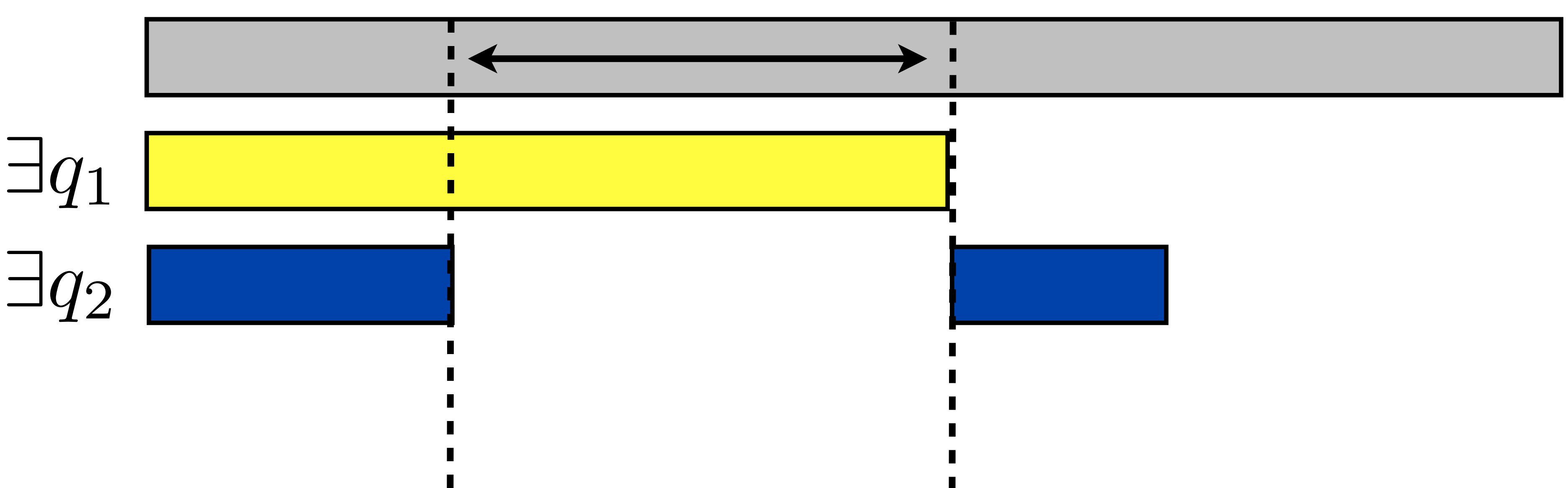
K=2 for all N

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

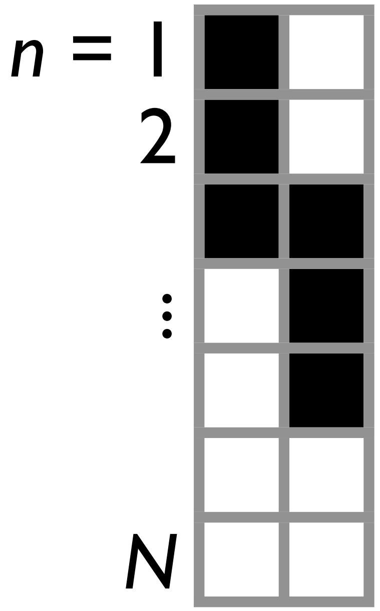
Want to show:

e.g. $\mathbb{P}(\text{row} = \blacksquare\square \,|q_{1:2}) = q_1(1-q_2)$



Distributions with EFPFs: frequencies?

Feature allocation

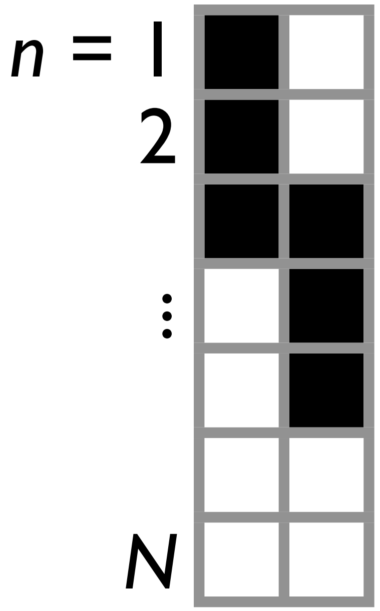


Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

Distributions with EFPFs: frequencies?

Feature allocation

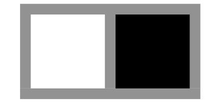


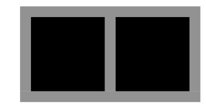
Assume EFPF

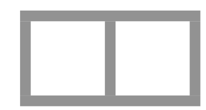
$p(N; S_{N,1}, S_{N,2})$

Feature paintbox

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

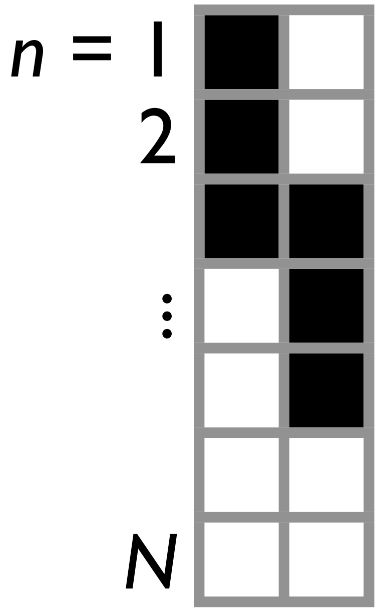
$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

Distributions with EFPFs: frequencies?

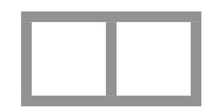
Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

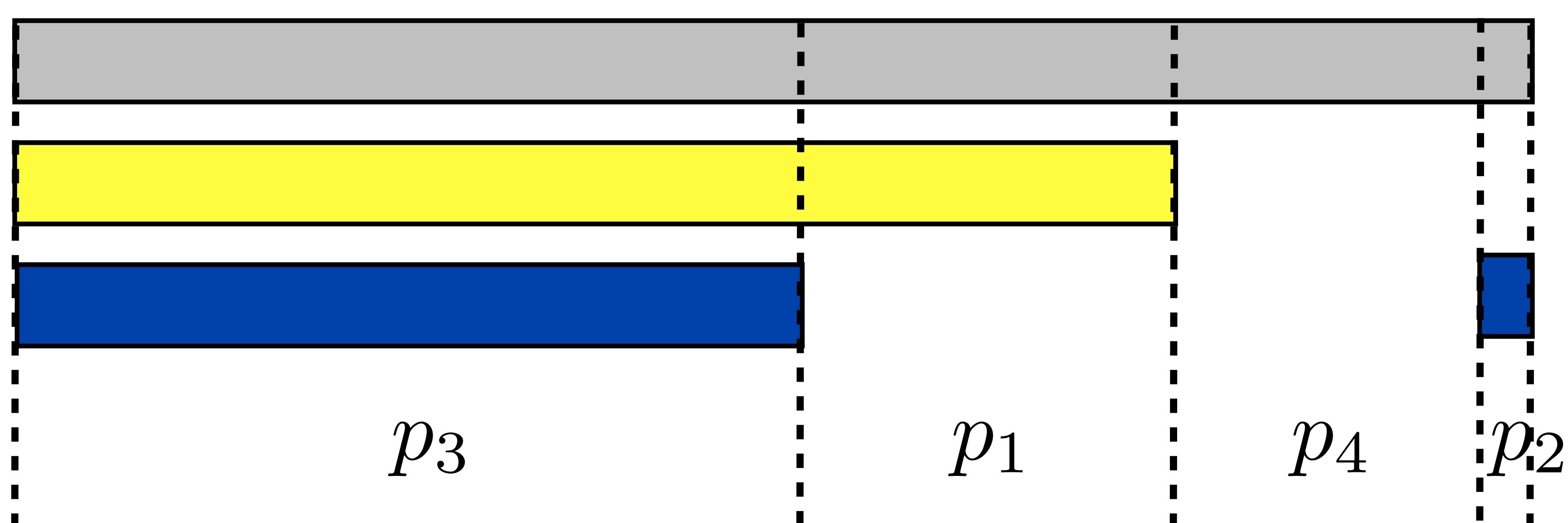
$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

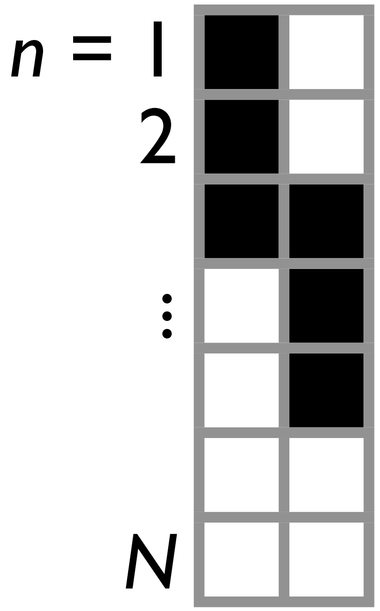
Assume EFPF

$p(N; S_{N,1}, S_{N,2})$



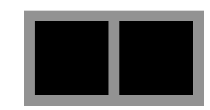
Distributions with EFPFs: frequencies?

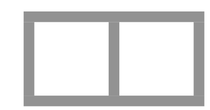
Feature allocation



$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_1$

$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_3$

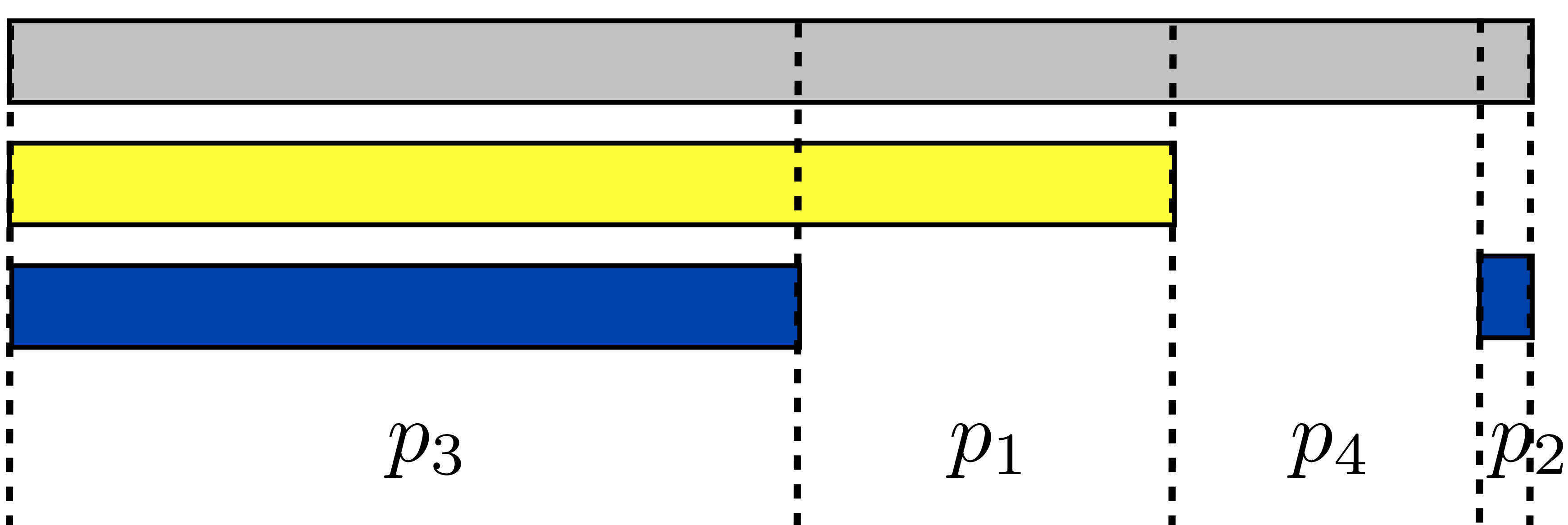
$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

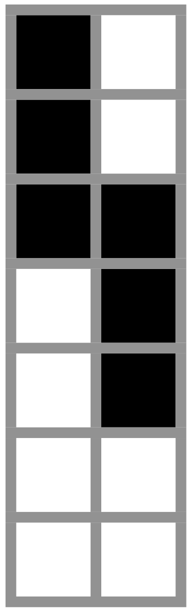
$q_1 = p_1 + p_3$

$q_2 = p_2 + p_3$



Distributions with EFPFs: frequencies?

Feature allocation



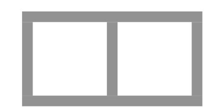
Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

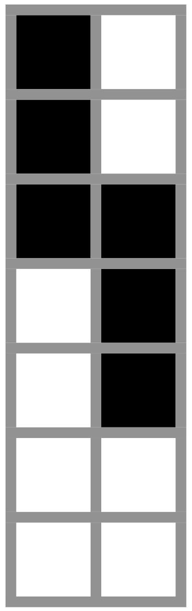
$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

Distributions with EFPFs: frequencies?

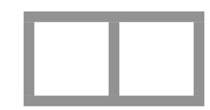
Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

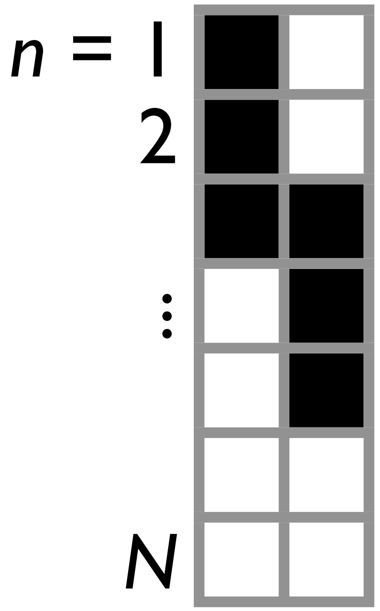
Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

$$\mathbb{P}(4; 2, 2)$$

Distributions with EFPFs: frequencies?

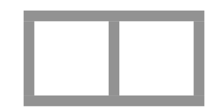
Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

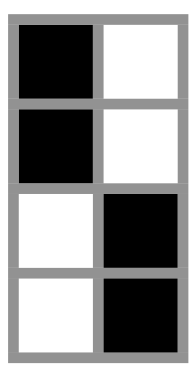
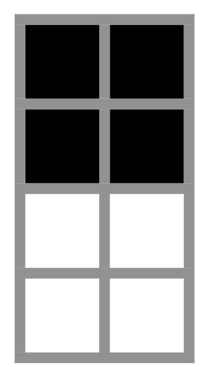
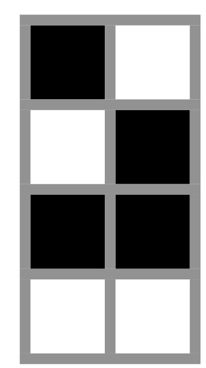
$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

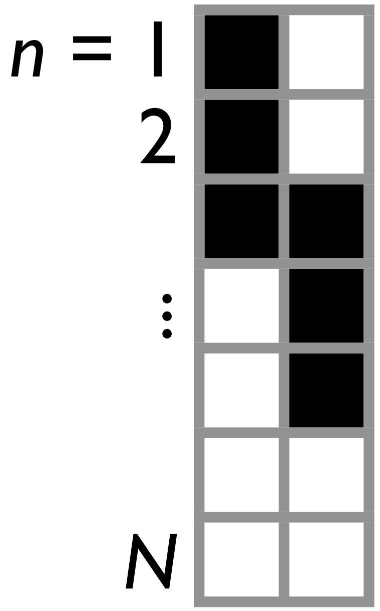
Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

$\mathbb{P}(4; 2, 2) = \mathbb{P}($  $) = \mathbb{P}($  $) = \mathbb{P}($  $)$

Distributions with EFPFs: frequencies?

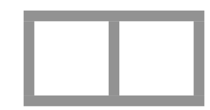
Feature allocation



$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_1$

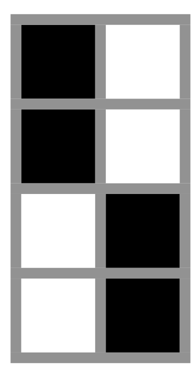
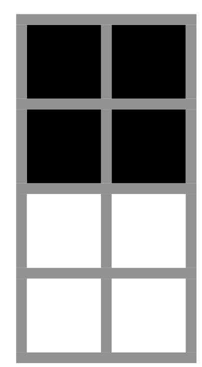
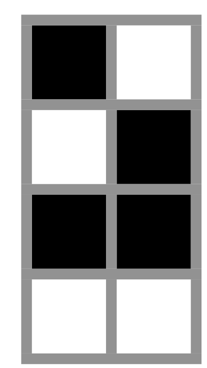
$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_4$

Assume EFPF

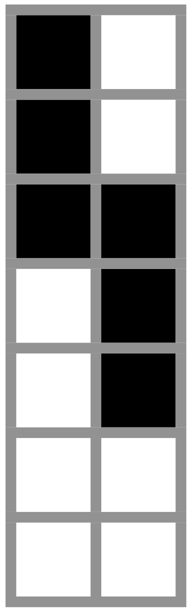
$p(N; S_{N,1}, S_{N,2})$

$\mathbb{P}(4; 2, 2) = \mathbb{P}($  $) = \mathbb{P}($  $) = \mathbb{P}($  $)$

$$\mathbb{E}[p_1^2 p_2^2] = \mathbb{E}[p_3^2 p_4^2] = \mathbb{E}[p_1 p_2 p_3 p_4]$$

Distributions with EFPFs: frequencies?

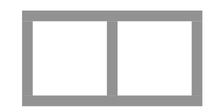
Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

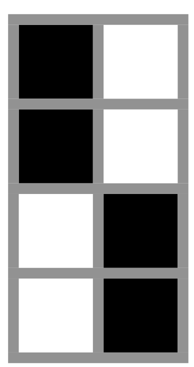
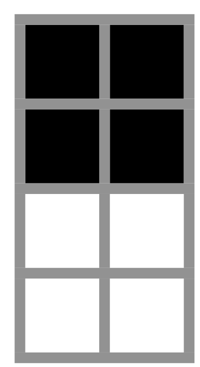
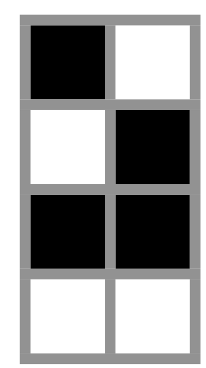
$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

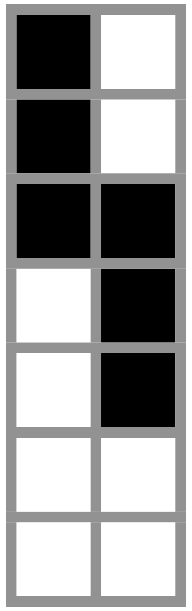
$\mathbb{P}(4; 2, 2) = \mathbb{P}($  $) = \mathbb{P}($  $) = \mathbb{P}($  $)$

$$\mathbb{E}[p_1^2 p_2^2] = \mathbb{E}[p_3^2 p_4^2] = \mathbb{E}[p_1 p_2 p_3 p_4]$$

$$\mathbb{E}[(p_1 p_2 - p_3 p_4)^2] = 0$$

Distributions with EFPFs: frequencies?

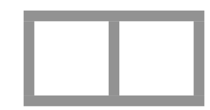
Feature allocation



$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_1$

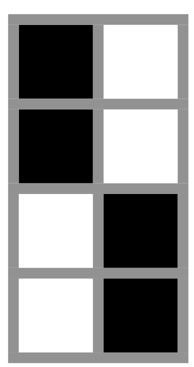
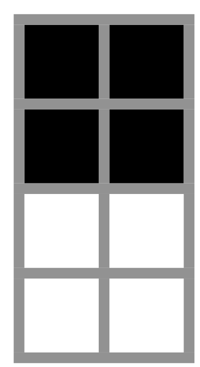
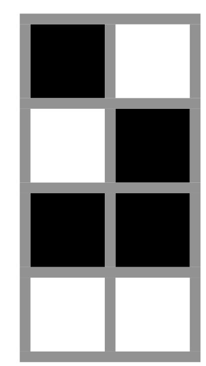
$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $|p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

$\mathbb{P}(4; 2, 2) = \mathbb{P}($  $) = \mathbb{P}($  $) = \mathbb{P}($  $)$

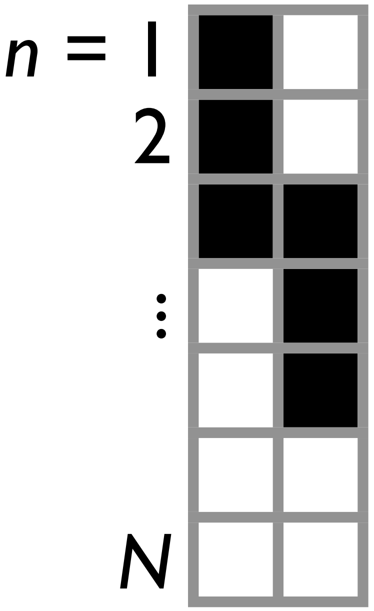
$$\mathbb{E}[p_1^2 p_2^2] = \mathbb{E}[p_3^2 p_4^2] = \mathbb{E}[p_1 p_2 p_3 p_4]$$

$$\mathbb{E}[(p_1 p_2 - p_3 p_4)^2] = 0$$

$$p_1 p_2 \overset{a.s.}{=} p_3 p_4$$

Distributions with EFPFs: frequencies?

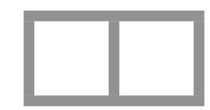
Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

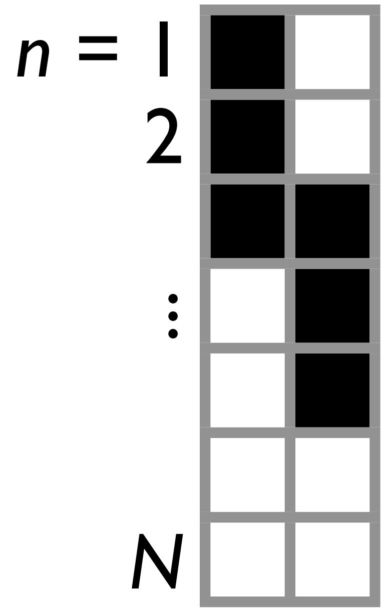
Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

$$p_1 p_2 \stackrel{a.s.}{=} p_3 p_4$$

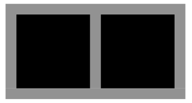
Distributions with EFPFs: frequencies?

Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

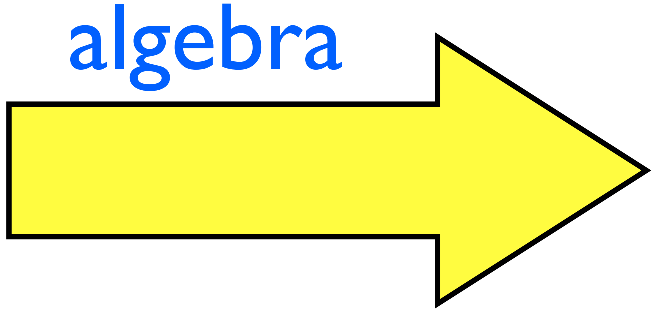
$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

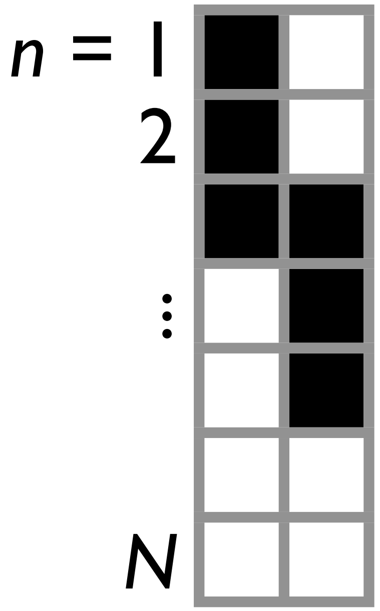
$$p_1 p_2 \overset{a.s.}{=} p_3 p_4$$

algebra



Distributions with EFPFs: frequencies?

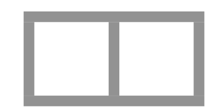
Feature allocation



$\mathbb{P}($row $=$  $| p_{1:4}) = p_1$

$\mathbb{P}($row $=$  $| p_{1:4}) = p_2$

$\mathbb{P}($row $=$  $| p_{1:4}) = p_3$

$\mathbb{P}($row $=$  $| p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

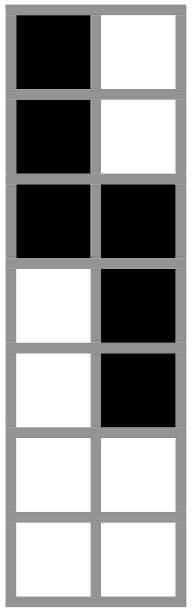
$$p_1 p_2 \overset{a.s.}{=} p_3 p_4$$

algebra

$$p_1 \overset{a.s.}{=} (p_1 + p_3)(1 - [p_2 + p_3])$$

Distributions with EFPFs: frequencies?

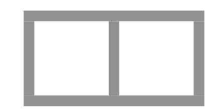
Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

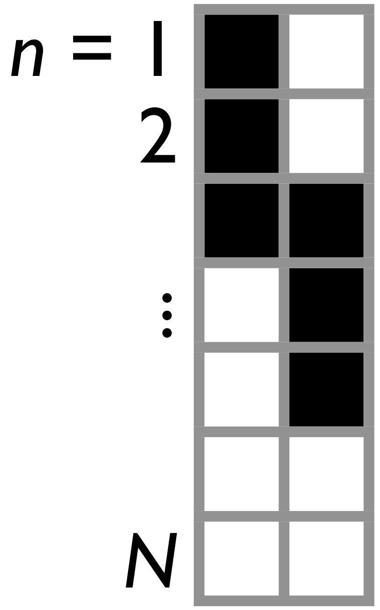
$$p_1 p_2 \stackrel{a.s.}{=} p_3 p_4$$

algebra

$$p_1 \stackrel{a.s.}{=} (p_1 + p_3)(1 - [p_2 + p_3])$$

Distributions with EFPFs: frequencies?

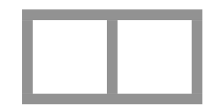
Feature allocation



$\mathbb{P}($row $=$  $| p_{1:4}) = p_1$

$\mathbb{P}($row $=$  $| p_{1:4}) = p_2$

$\mathbb{P}($row $=$  $| p_{1:4}) = p_3$

$\mathbb{P}($row $=$  $| p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

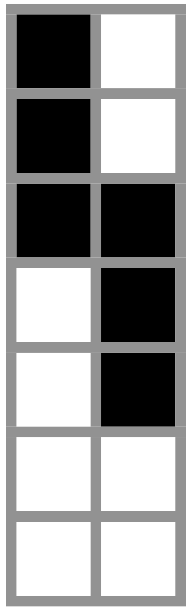
$$p_1 p_2 \overset{a.s.}{=} p_3 p_4$$

$$\xrightarrow{\text{algebra}}$$

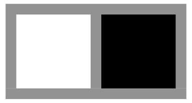
$$p_1 \overset{a.s.}{=} q_1(1 - [p_2 + p_3])$$

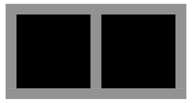
Distributions with EFPFs: frequencies?

Feature allocation



$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_1$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_2$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_3$

$\mathbb{P}(\text{row} =$  $| p_{1:4}) = p_4$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

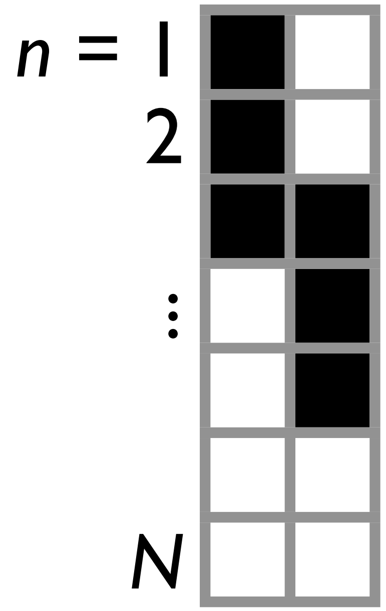
$$p_1 p_2 \stackrel{a.s.}{=} p_3 p_4$$

algebra

$$p_1 \stackrel{a.s.}{=} q_1(1 - [p_2 + p_3])$$

Distributions with EFPFs: frequencies?

Feature allocation



$$\mathbb{P}(\text{row} = \blacksquare\square \mid p_{1:4}) = p_1$$

$$\mathbb{P}(\text{row} = \square\blacksquare \mid p_{1:4}) = p_2$$

$$\mathbb{P}(\text{row} = \blacksquare\blacksquare \mid p_{1:4}) = p_3$$

$$\mathbb{P}(\text{row} = \square\square \mid p_{1:4}) = p_4$$

Assume EFPF

$p(N; S_{N,1}, S_{N,2})$

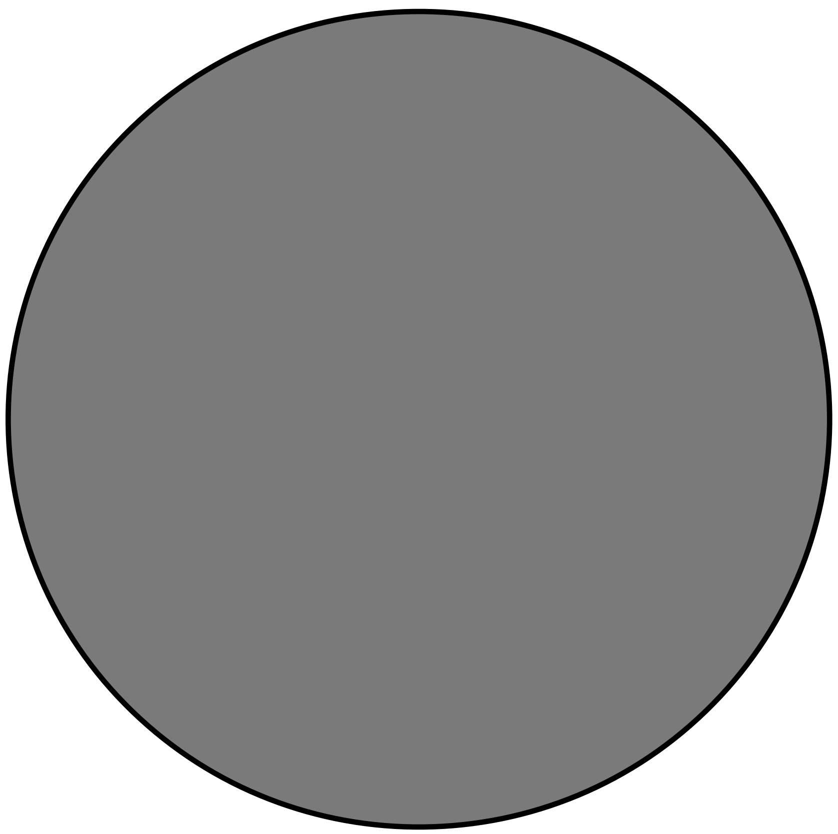
$$p_1 p_2 \stackrel{a.s.}{=} p_3 p_4$$

$$\xrightarrow{\text{algebra}}$$

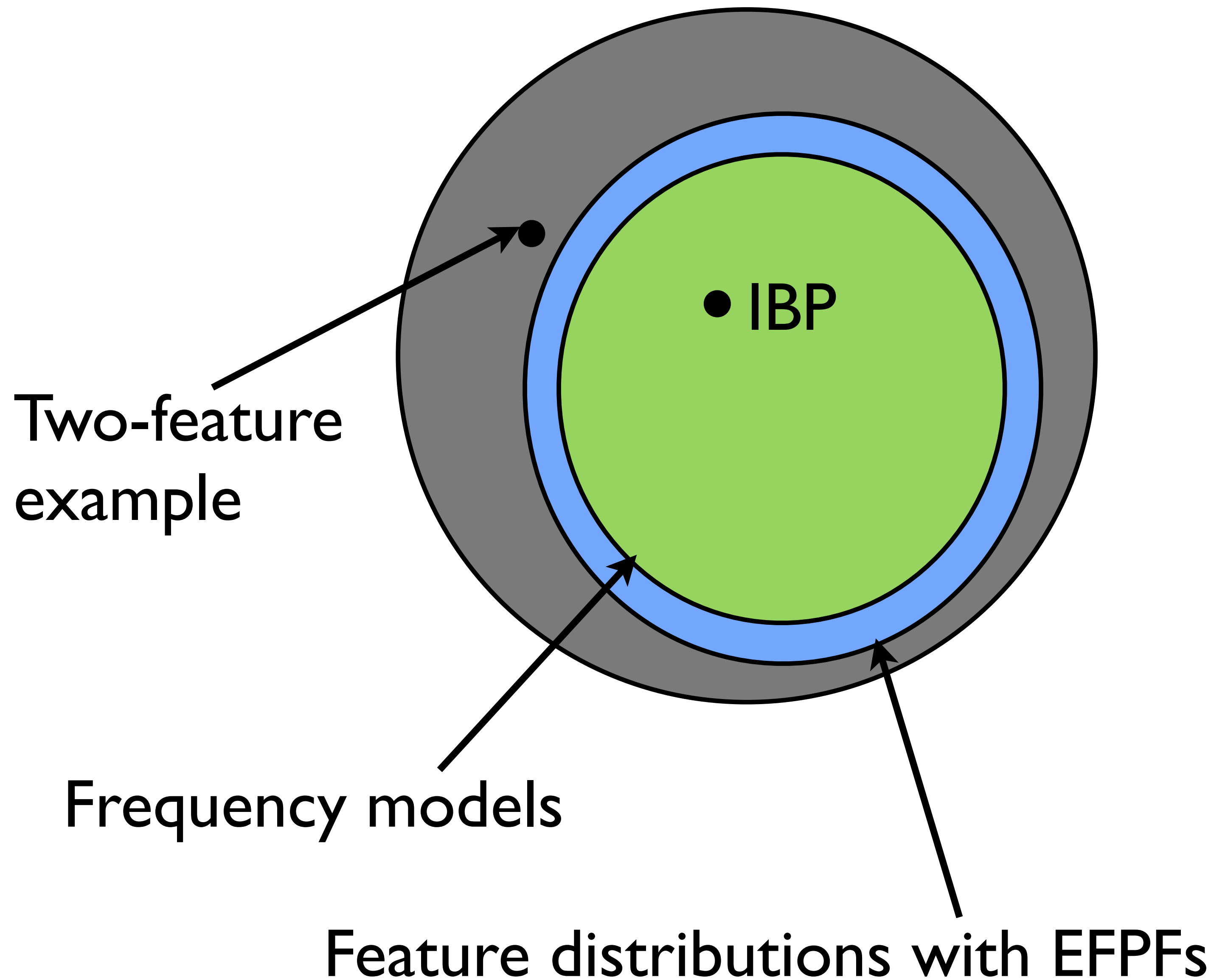
$$p_1 \stackrel{a.s.}{=} q_1(1 - q_2)$$

Distributions with EFPFs: frequencies?

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



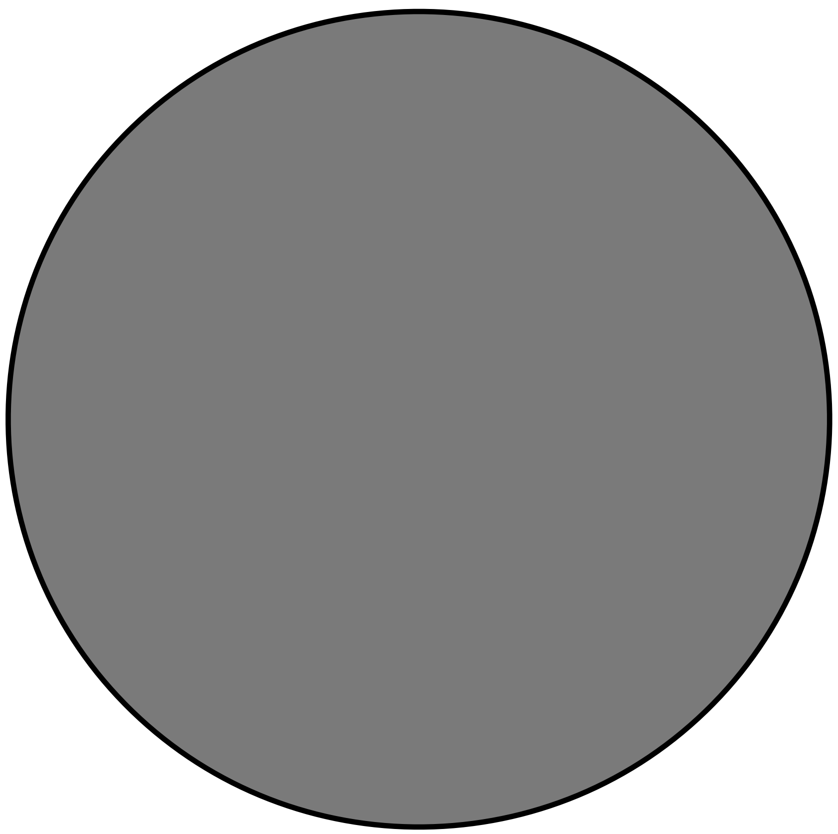
Exchangeable feature distributions
= Feature paintbox allocations



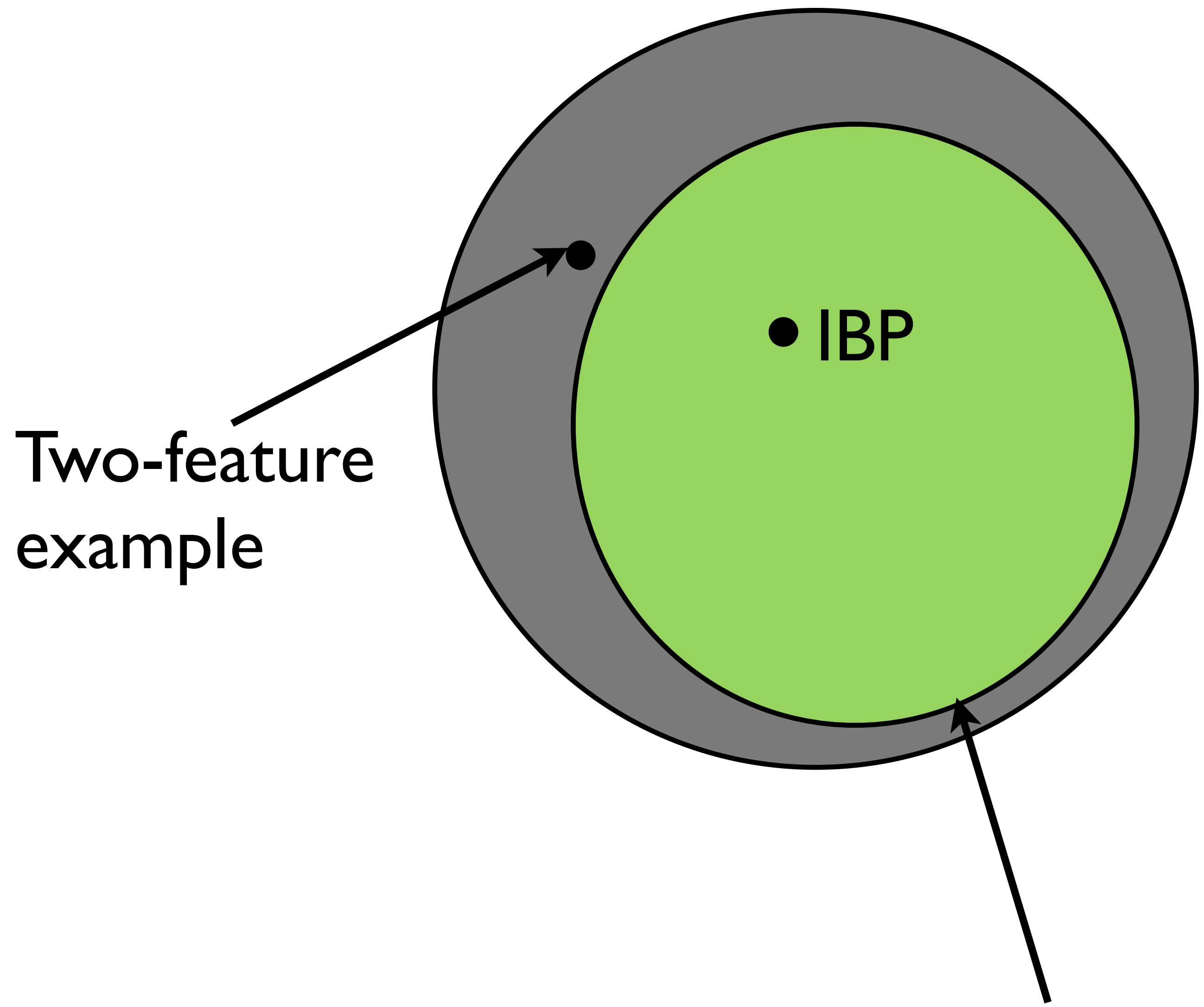
[Broderick, Pitman, Jordan (submitted)]

Distributions with EFPFs: frequencies?

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



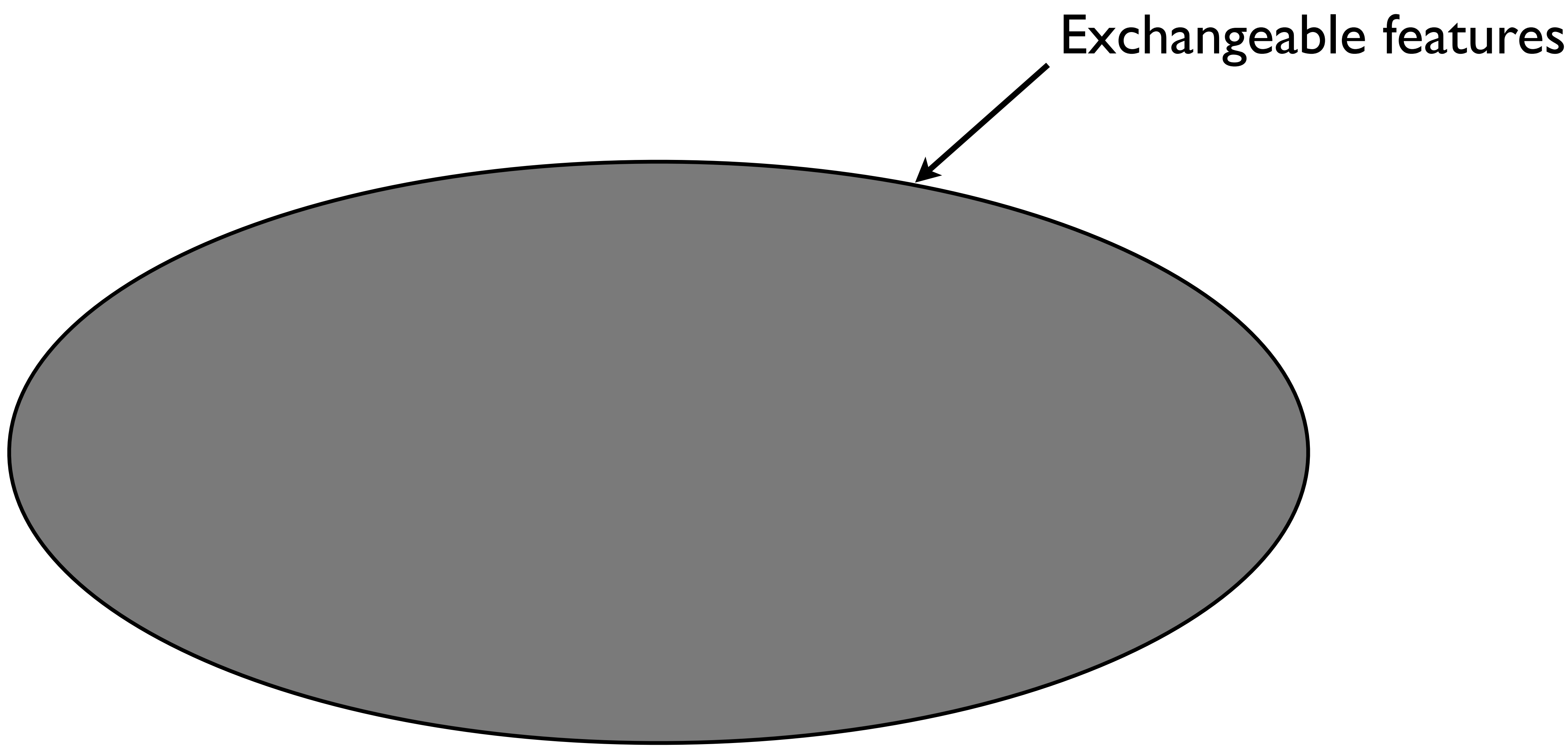
Exchangeable feature distributions
= Feature paintbox allocations



Feature distributions with EFPFs
= Frequency models

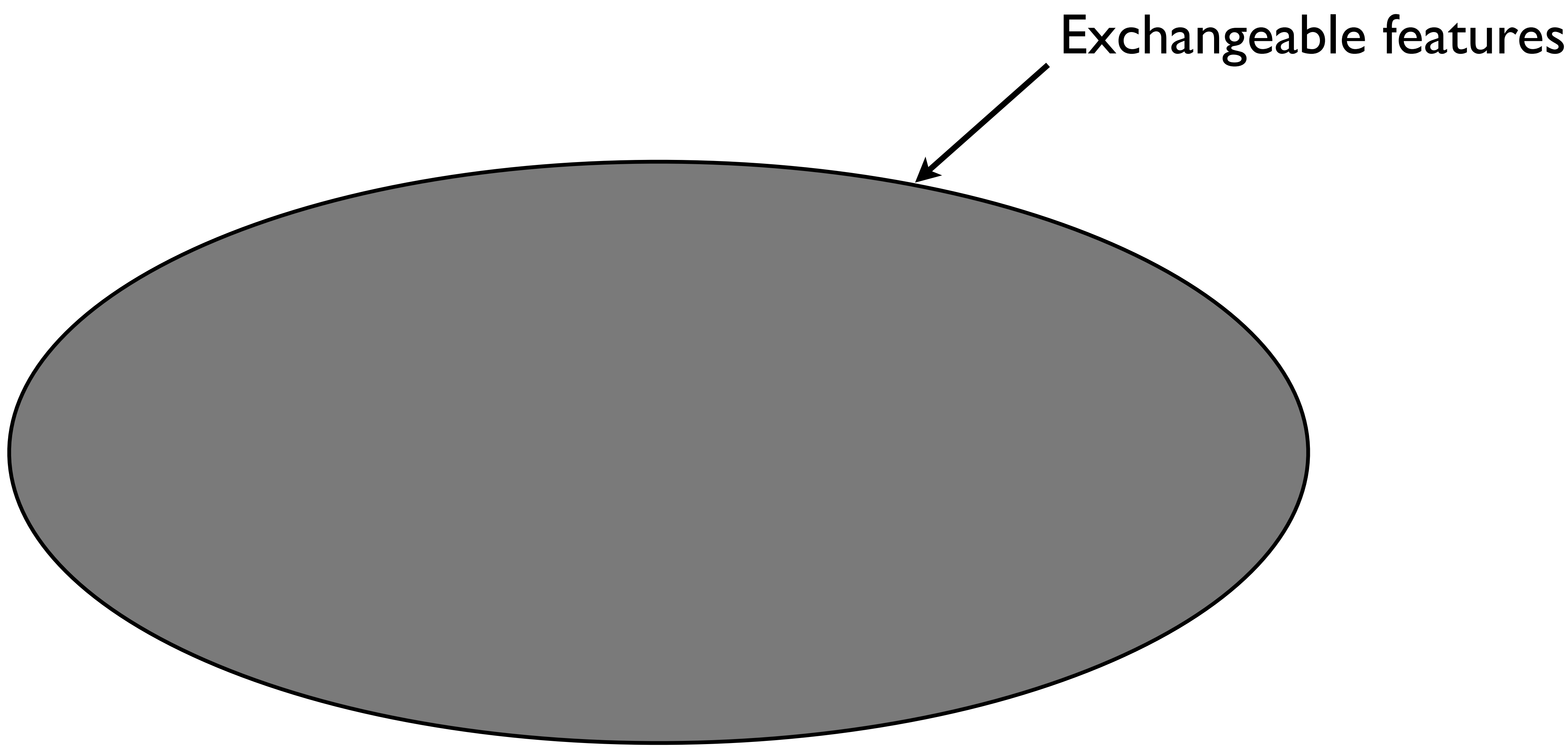
[Broderick, Pitman, Jordan (submitted)]

Conclusions



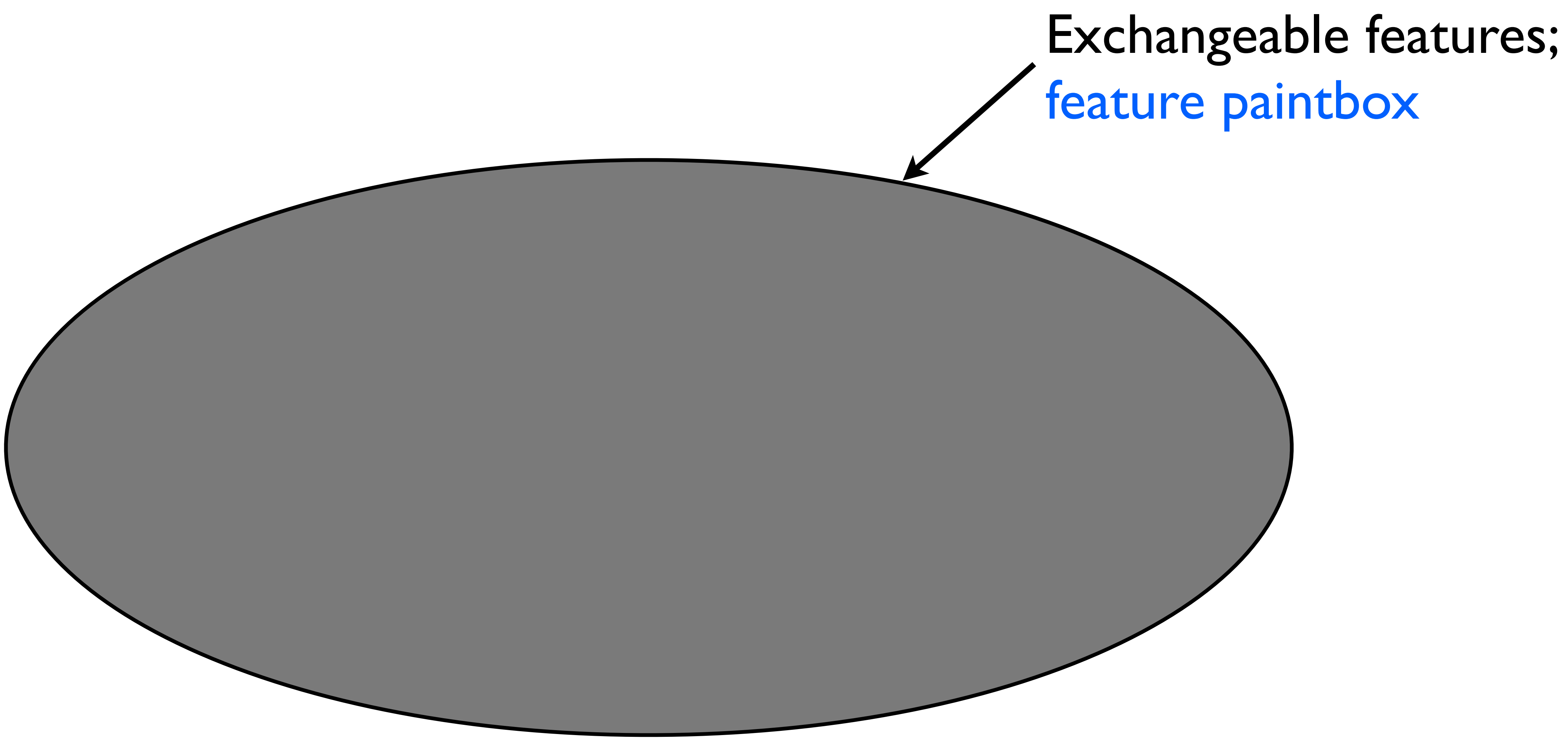
Conclusions

- Feature paintbox: characterization of exchangeable feature models



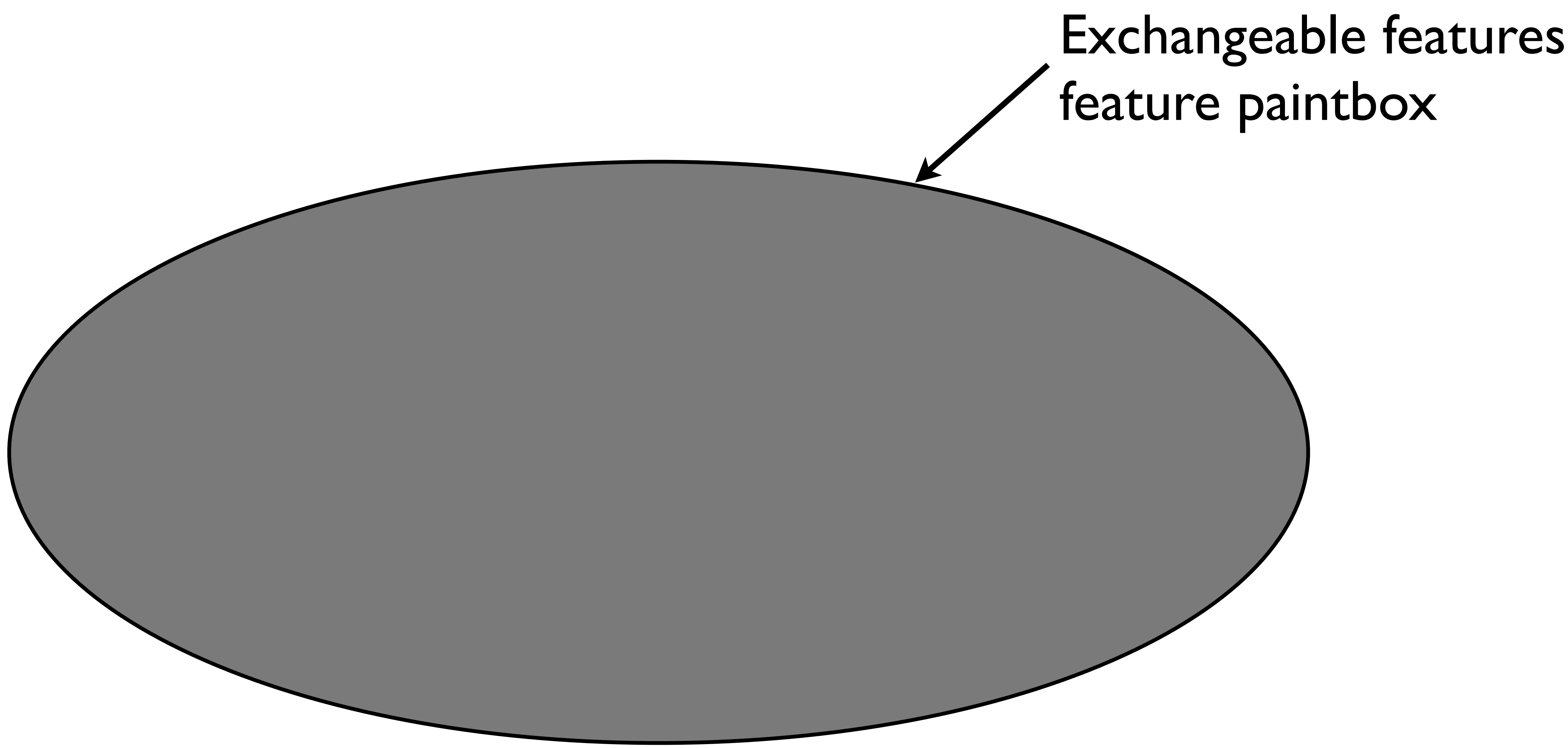
Conclusions

- Feature paintbox: characterization of exchangeable feature models



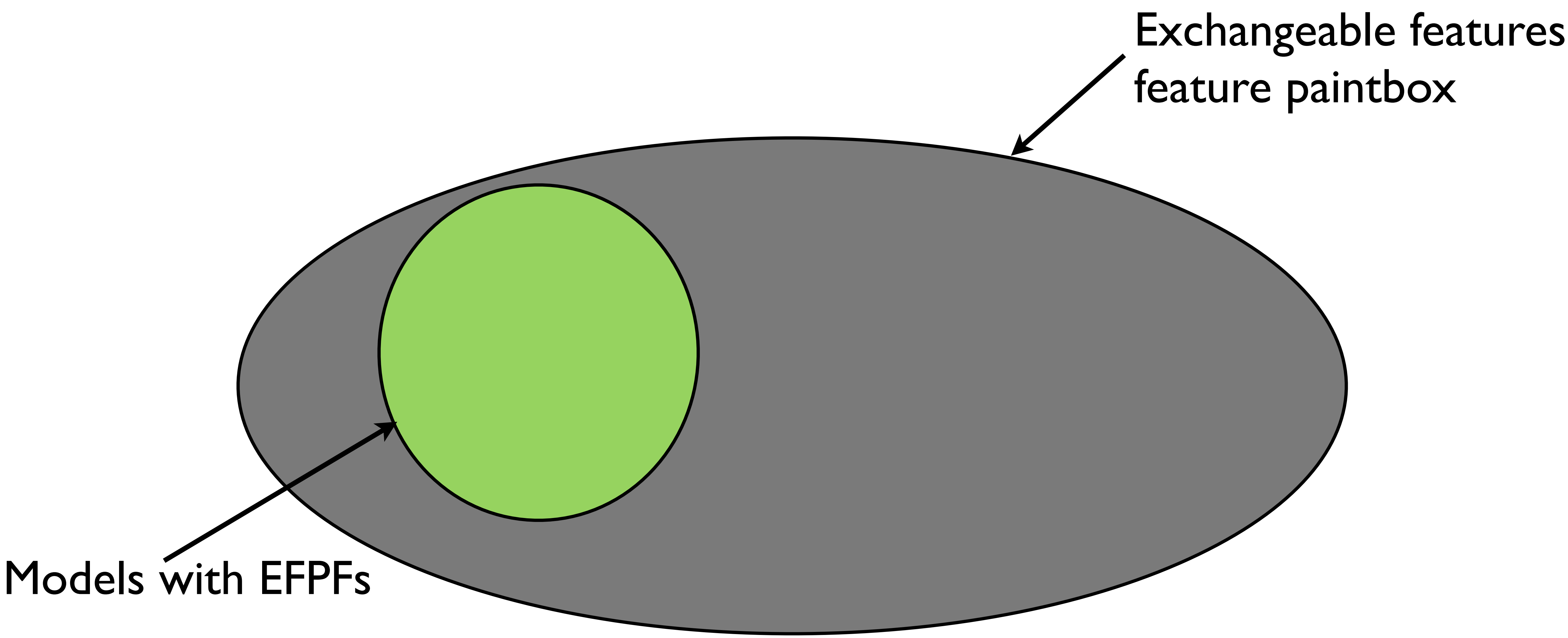
Conclusions

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



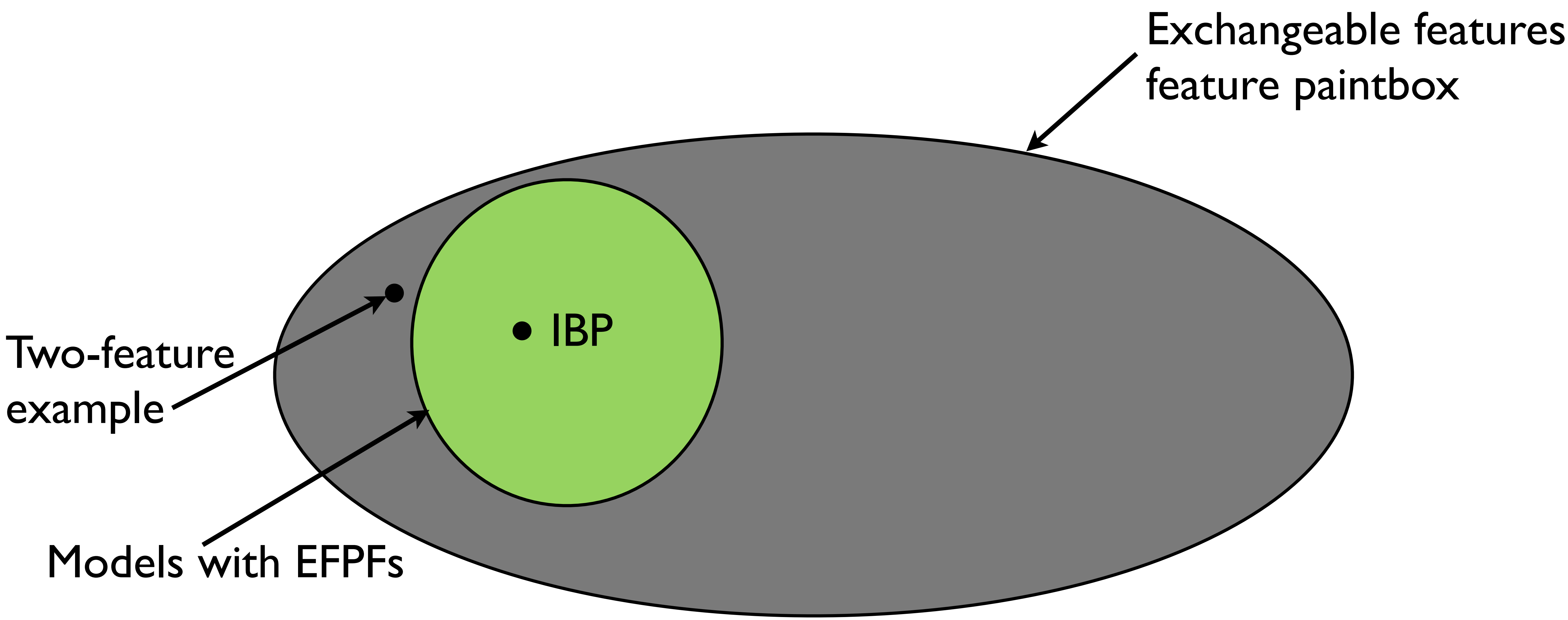
Conclusions

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



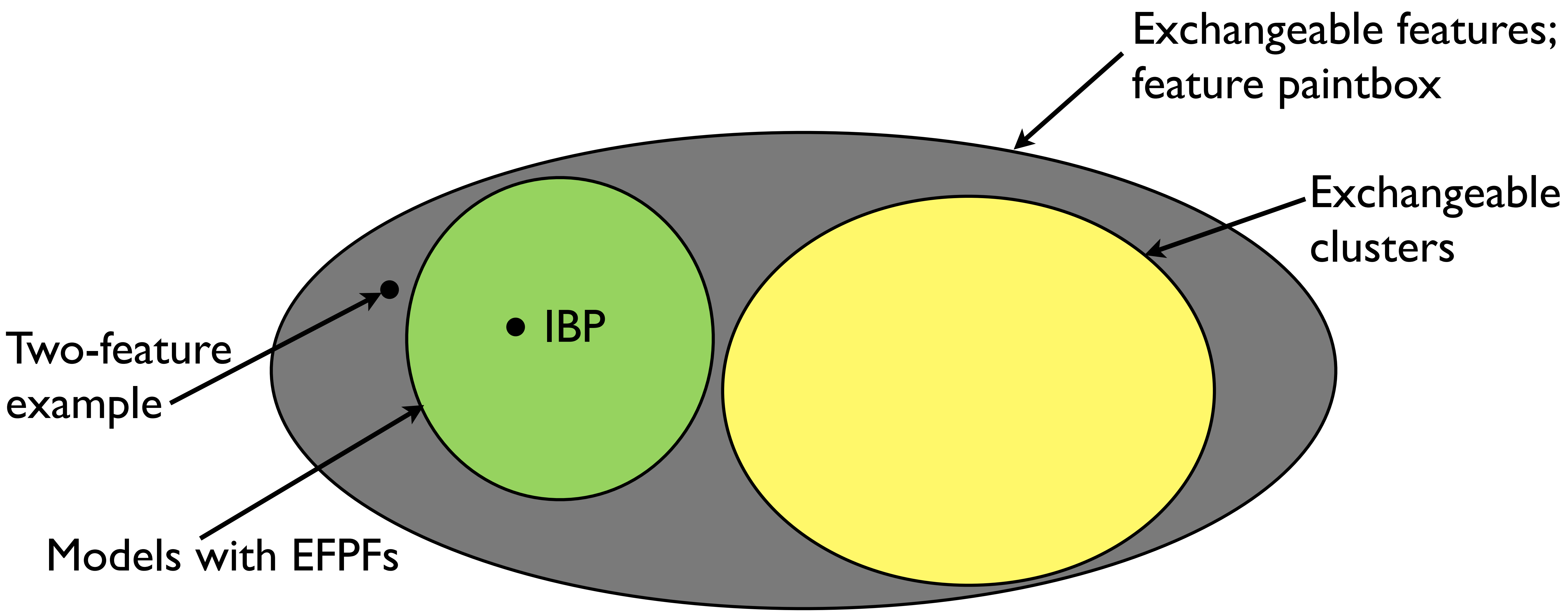
Conclusions

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



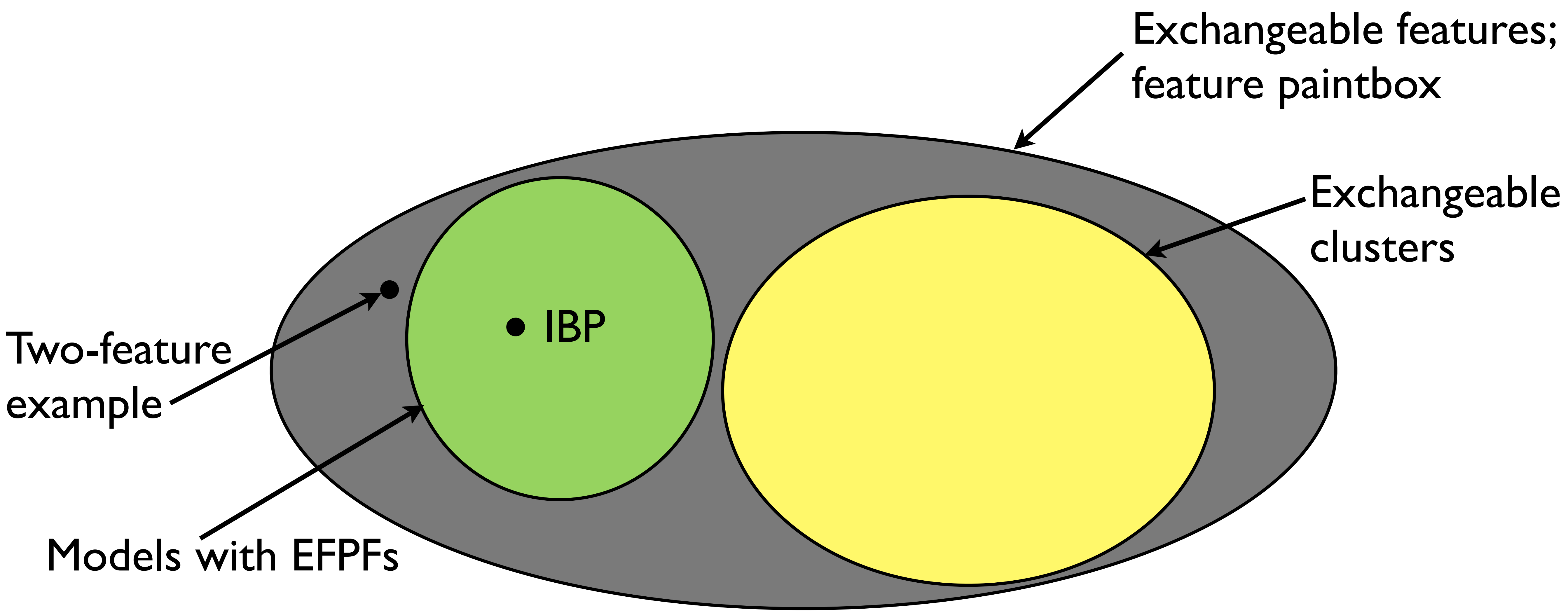
Conclusions

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



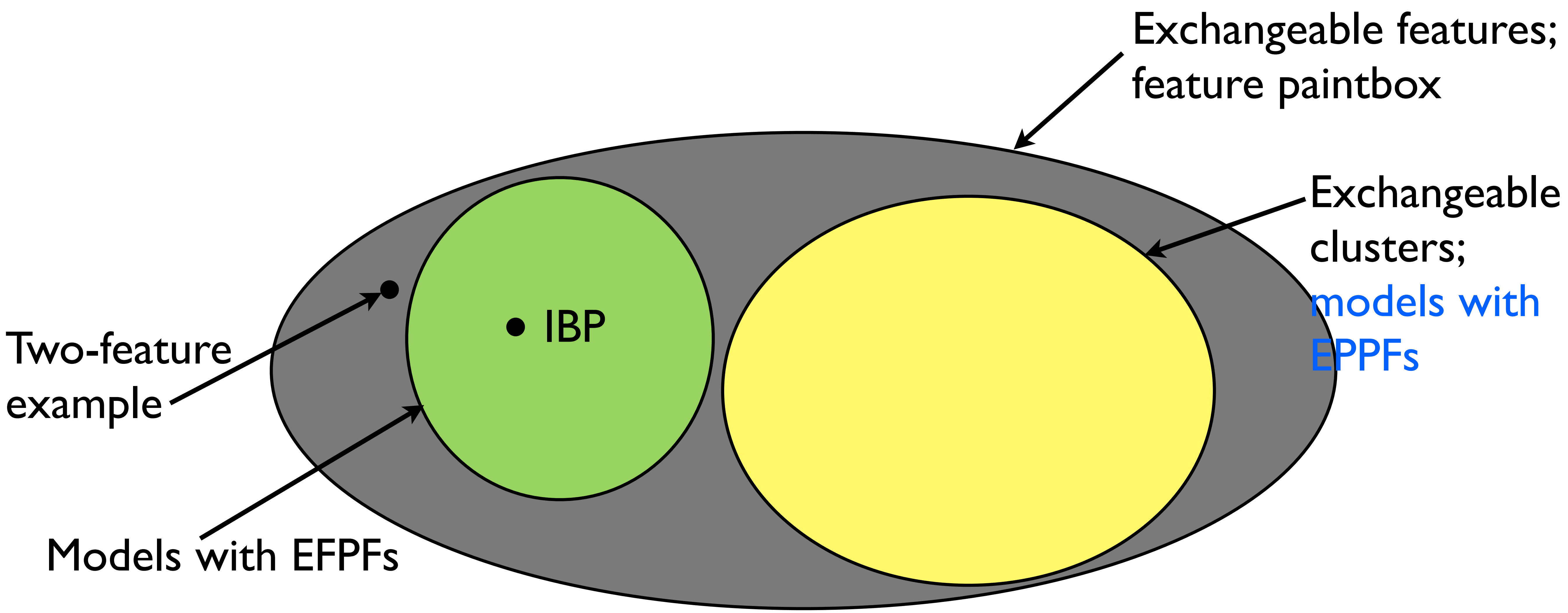
Conclusions

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



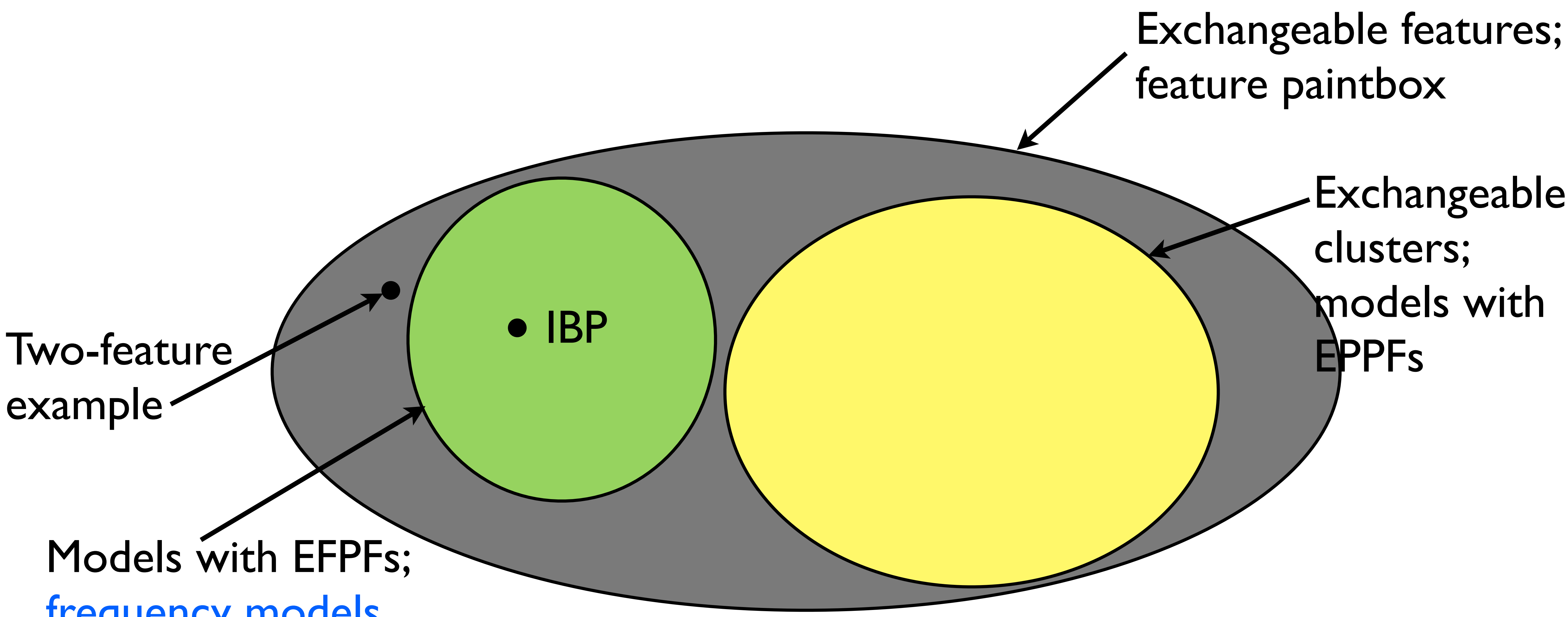
Conclusions

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



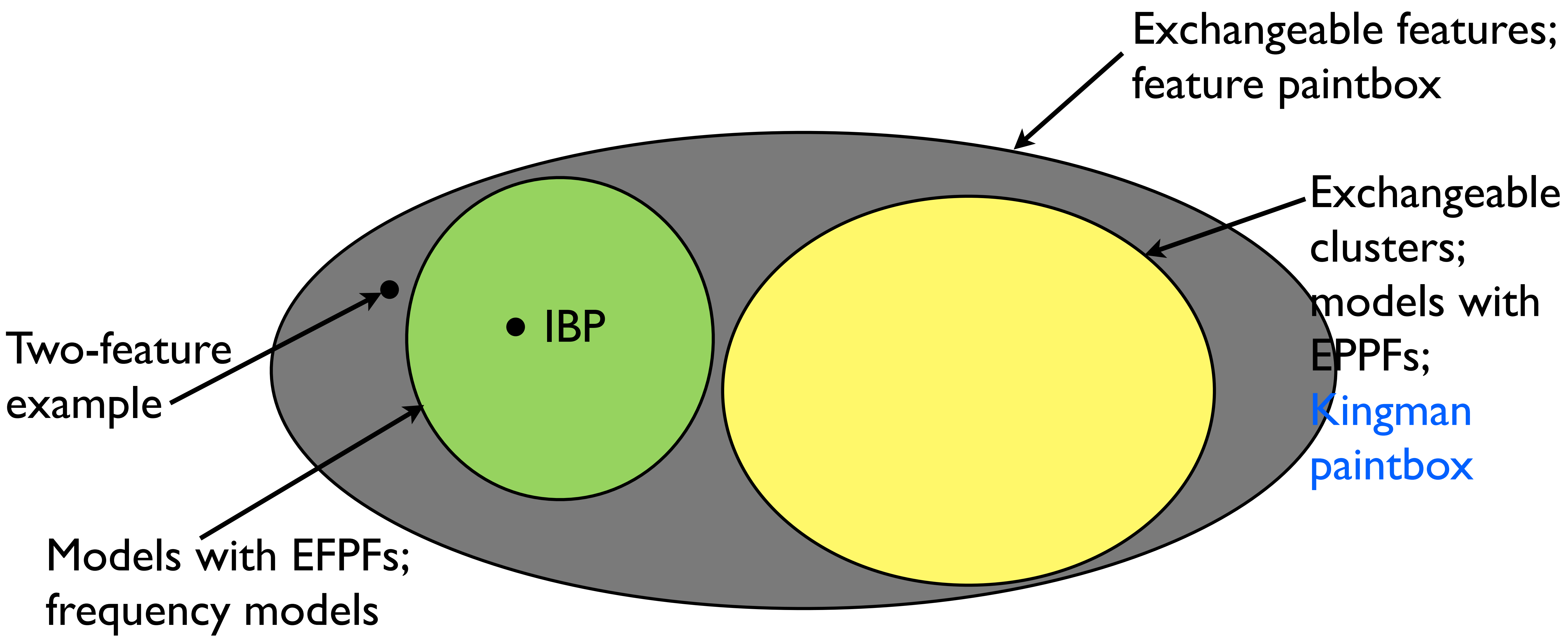
Conclusions

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



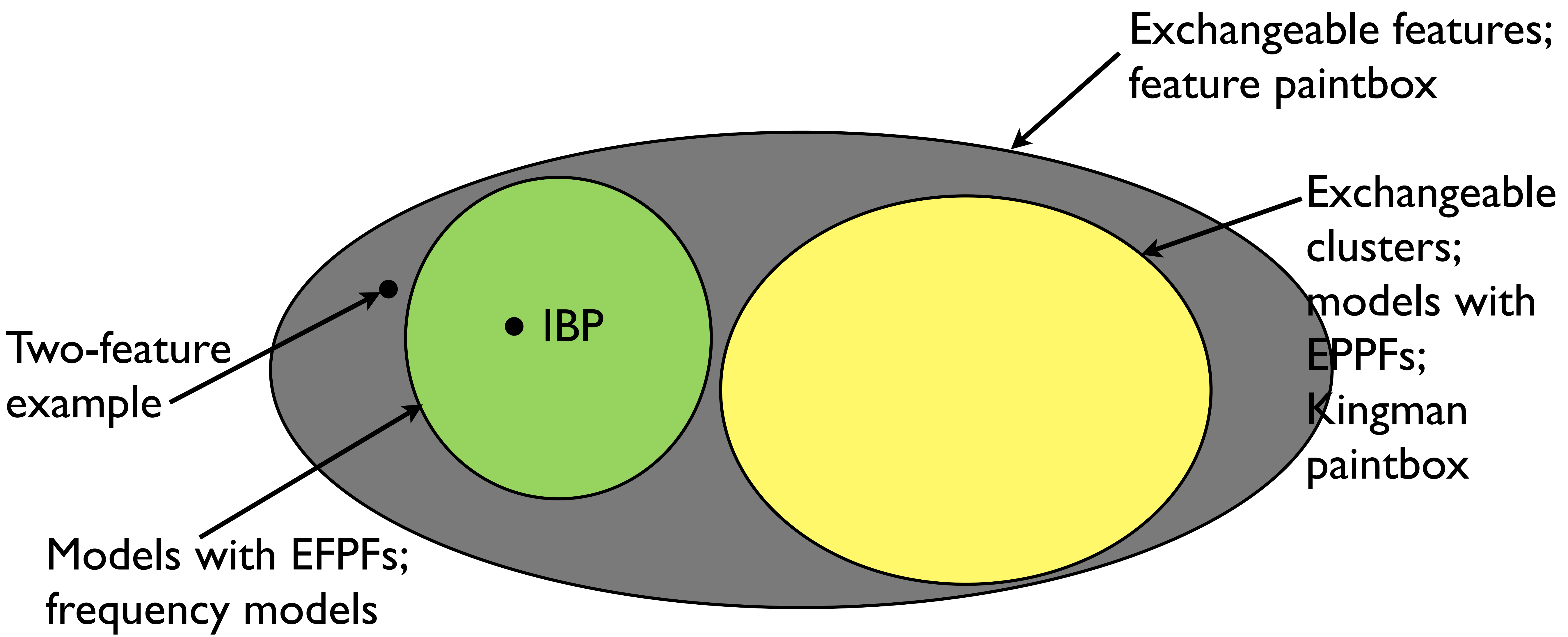
Conclusions

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



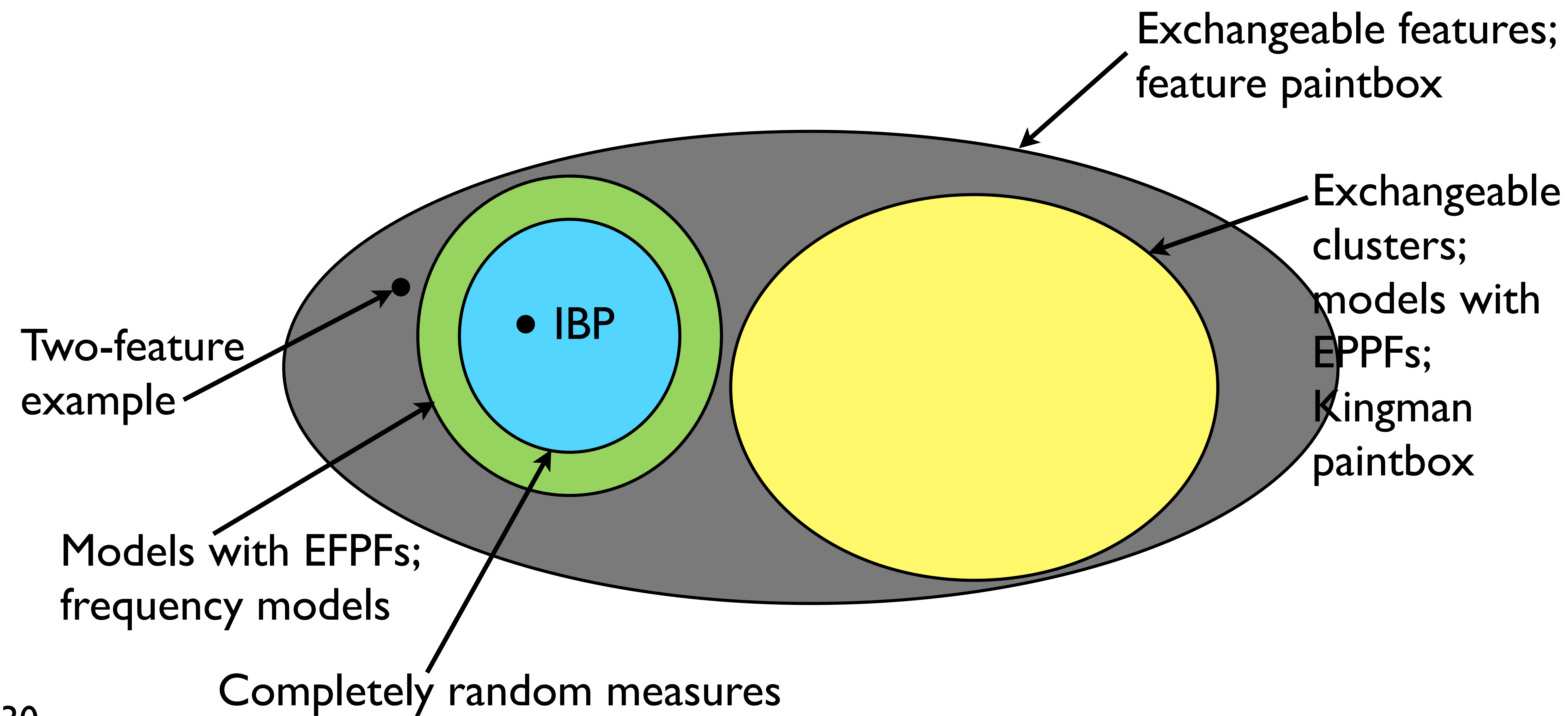
Conclusions

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



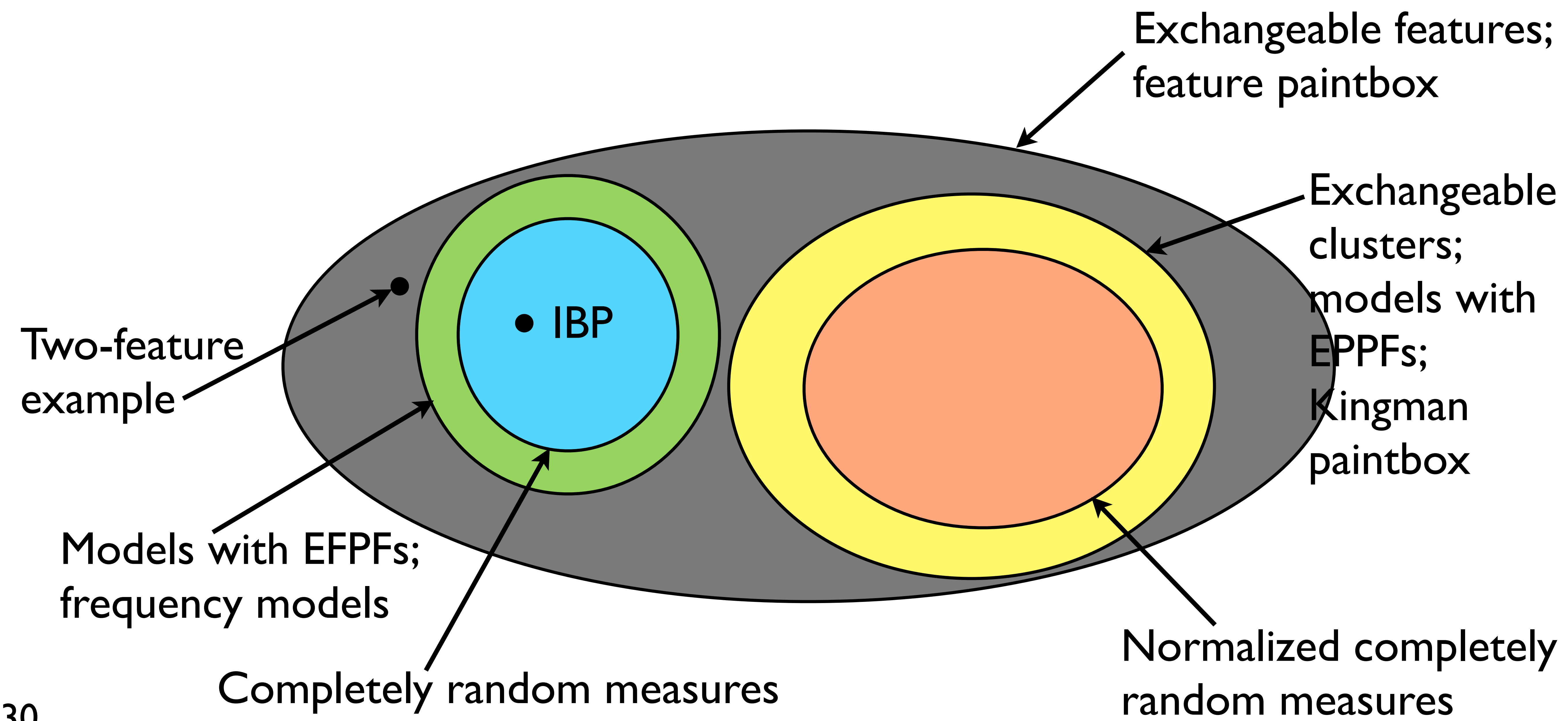
Conclusions

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



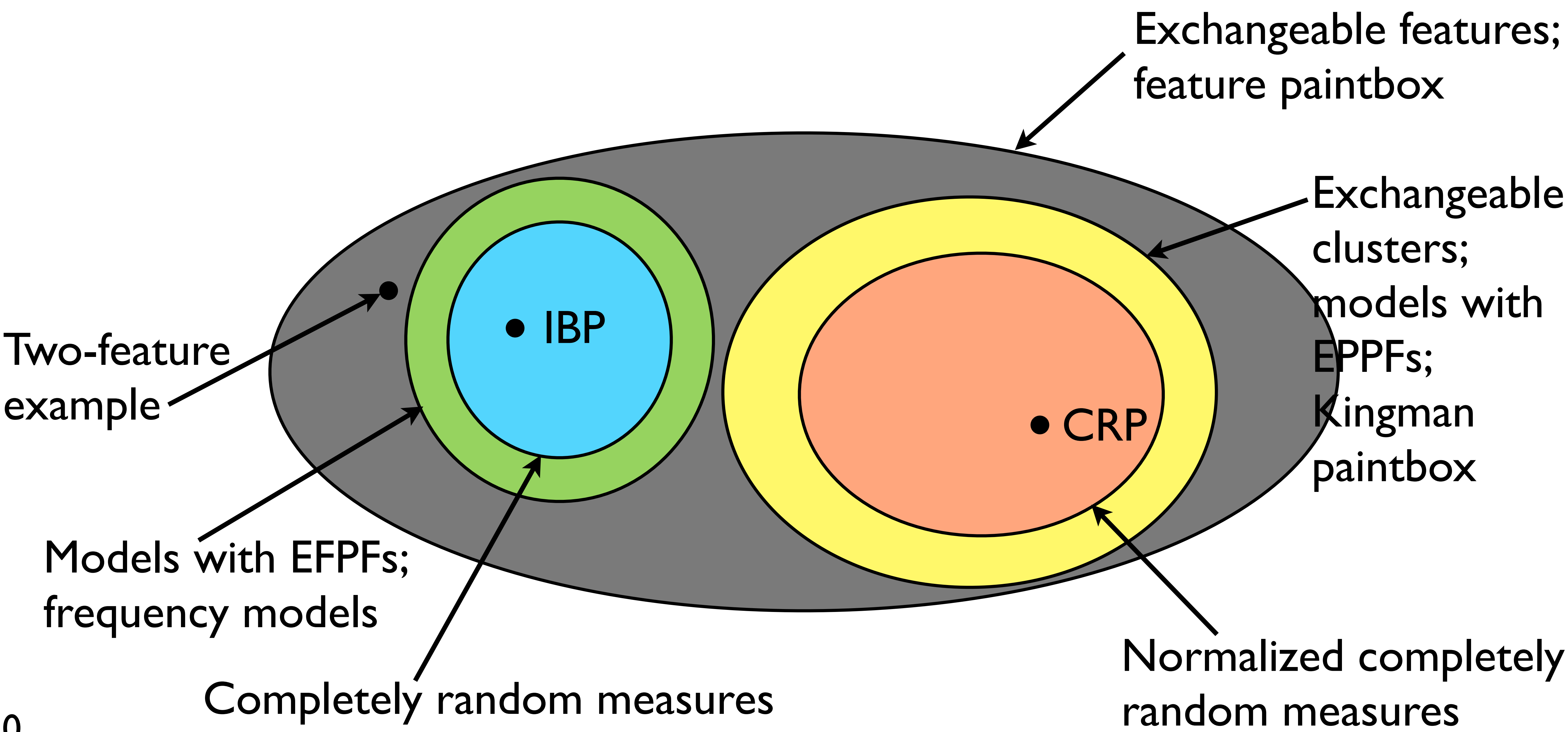
Conclusions

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



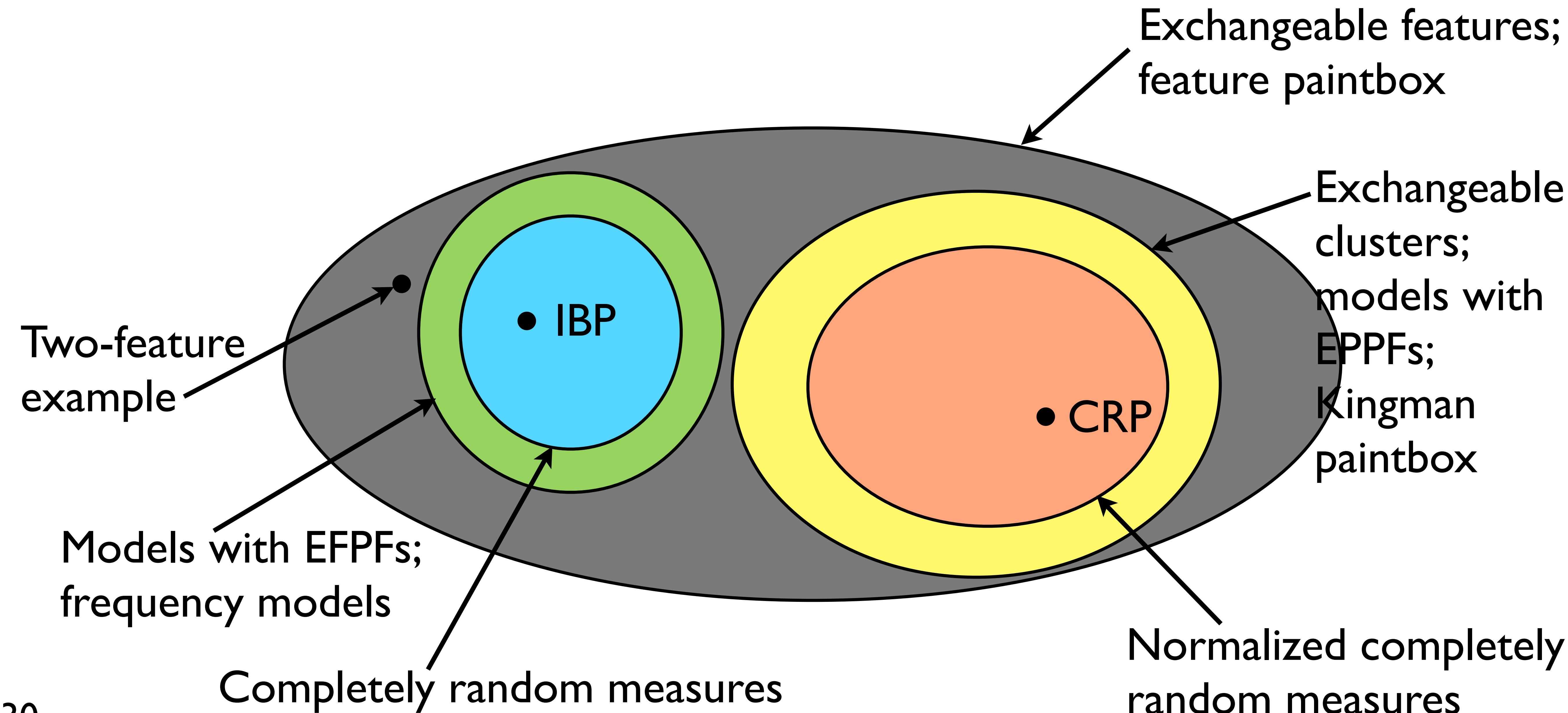
Conclusions

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



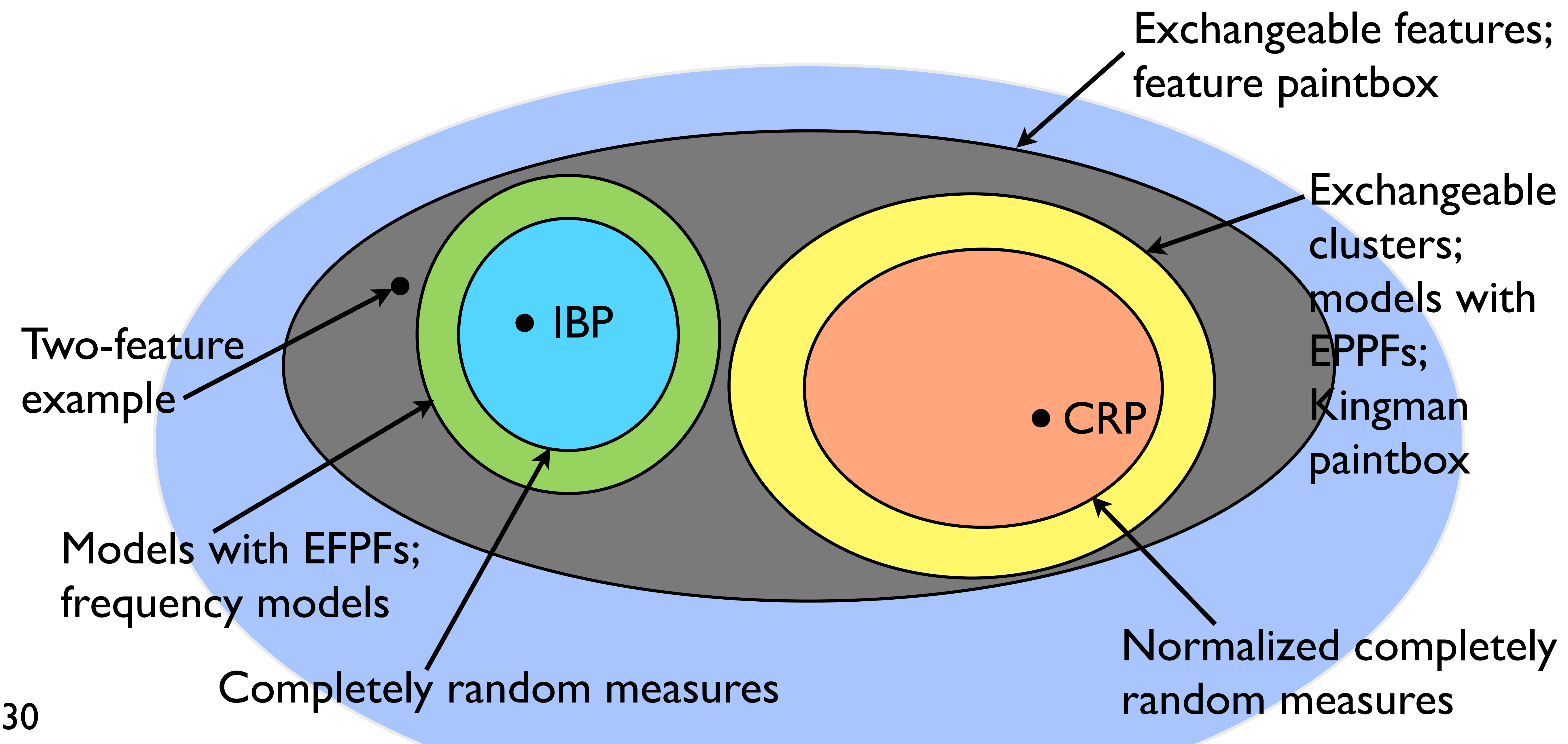
Conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections to fill in
- Other combinatorial structures



Conclusions

- Feature paintbox: characterization of exchangeable feature models
- Characterization of alternative correlation structure
- Remaining connections to fill in
- Other combinatorial structures



References

T. Broderick, M. I. Jordan, and J. Pitman. Clusters and features from combinatorial stochastic processes. *Arxiv preprint arXiv:1206.5862*, 2012.

T. Broderick, J. Pitman, and M. I. Jordan. Feature allocations, probability functions, and paintboxes. Submitted.

T. Broderick, L. Mackey, J. Paisley, and M. I. Jordan. Combinatorial clustering and the beta negative binomial process. *Arxiv preprint arXiv:1111.1802*, 2011.

T. Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In Y. Weiss, B. Scholkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems 18*, pages 475–482. MIT Press, Cambridge, MA, 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.

Y. Kim. Nonparametric Bayesian estimators for counting processes. *Annals of Statistics*, 27(2):562–588, 1999.

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

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 *Proceedings of the International Conference on Artificial Intelligence and Statistics*, volume 11, 2007.

M. Zhou, L. Hannah, D. Dunson, and L. Carin. Beta-negative binomial process and Poisson factor analysis. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, volume 15, 2012.