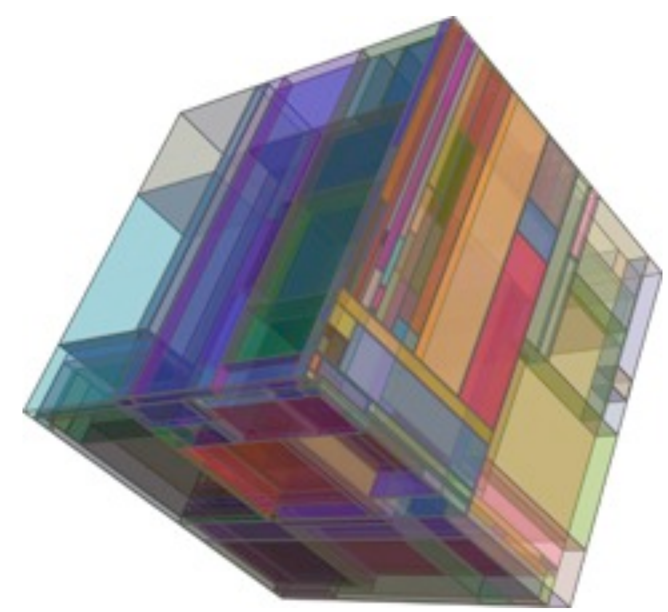
Clusters and features from combinatorial stochastic processes

Tamara Broderick
UC Berkeley

September 13, 2012



Nonparametric Bayesian statistics

Bayesian

- Specify a generative model
- Calculate posterior

Nonparametric (Bayesian)

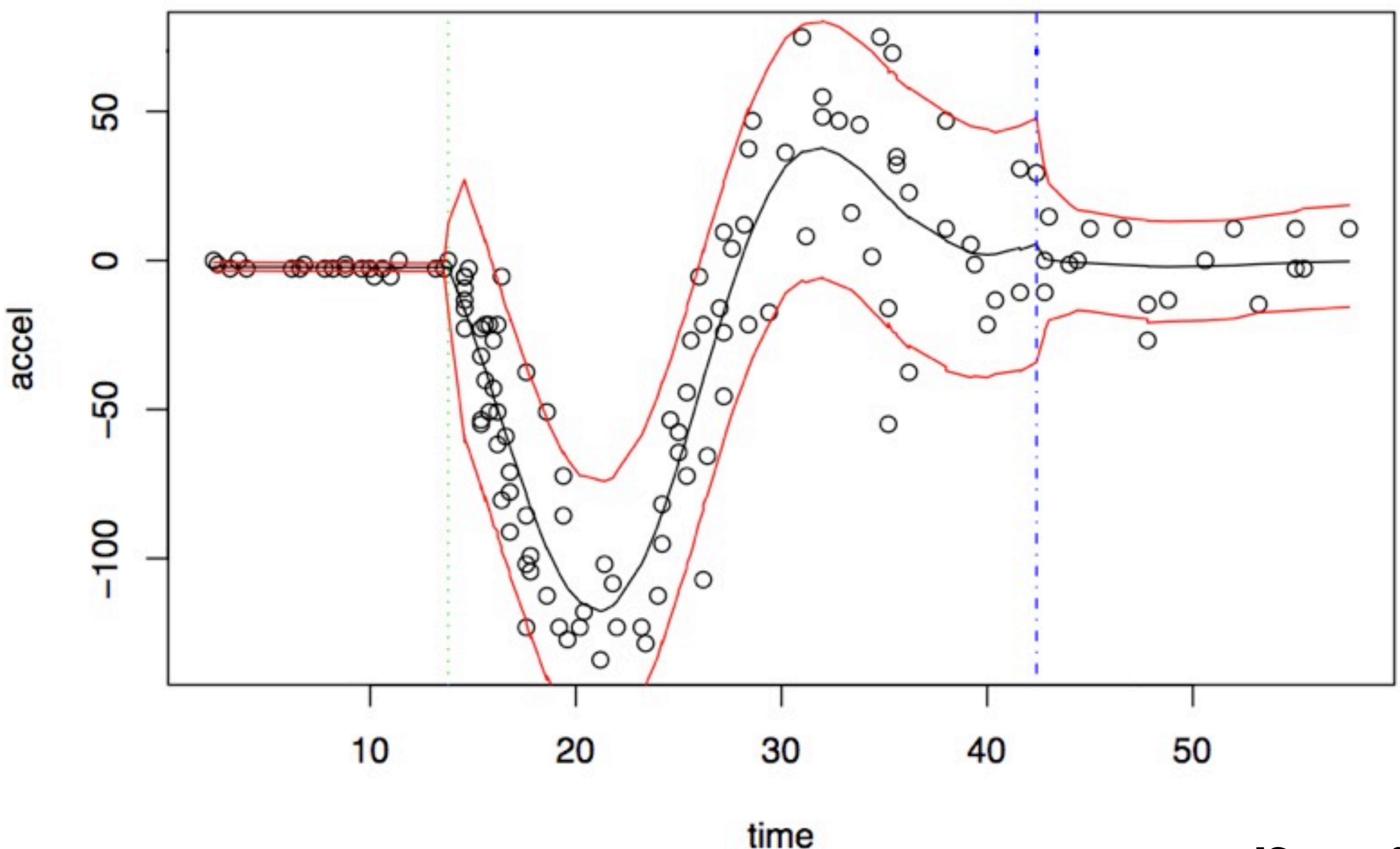
- Number of parameters grows with the size of the data

Nonparametric Bayesian statistics

Continuous/ordinal

- E.g. Gaussian process
- Supervised learning

smooth function



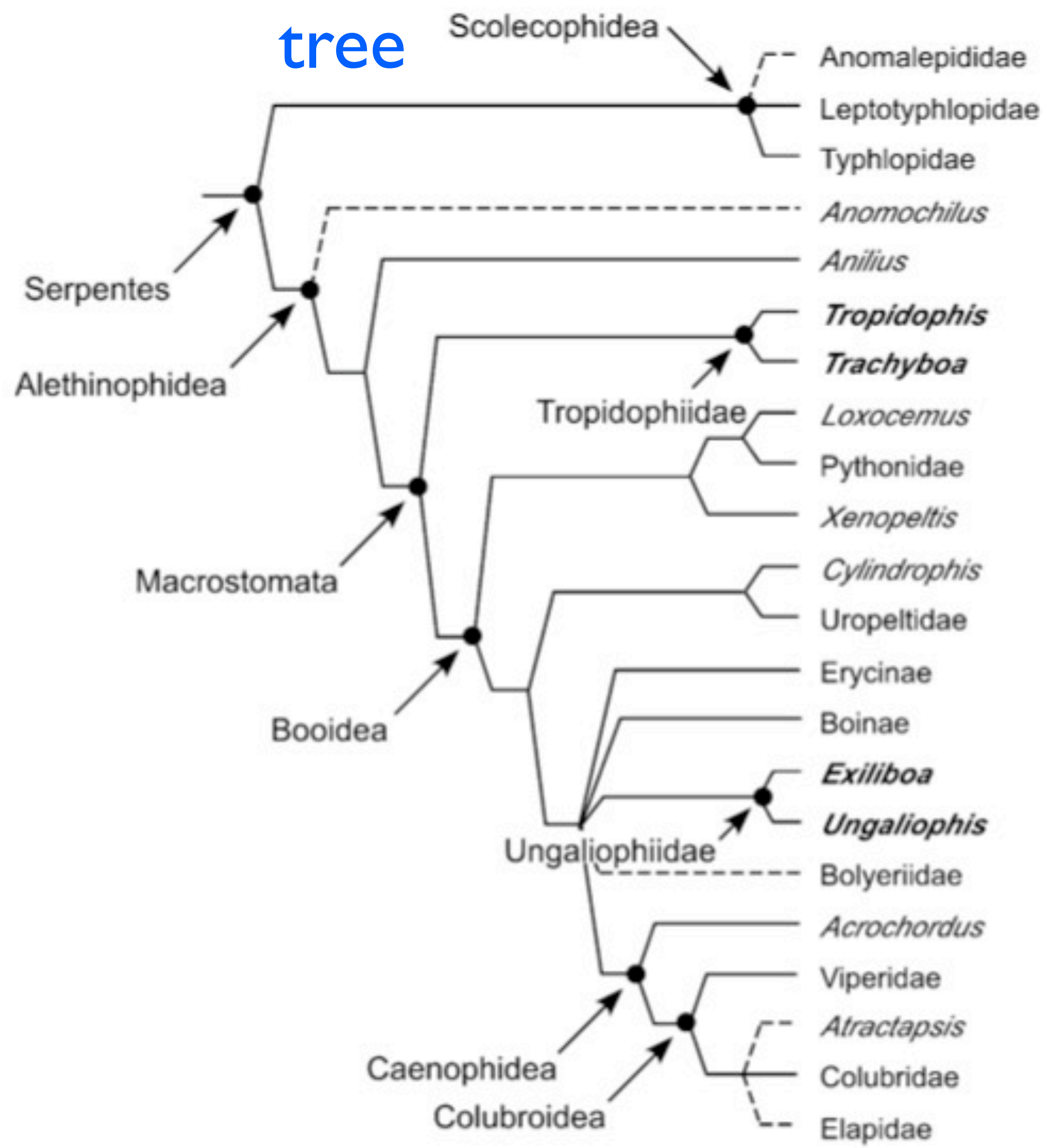
[Gramacy 2008]

Nonparametric Bayesian statistics

Discrete/combinatorial

- E.g. Dirichlet process
- Latent/unsupervised learning

tree



permutation

$$\sigma : 1 \to 5$$
$$2 \to 1$$
$$3 \to 4$$
$$4 \to 2$$
$$5 \to 3$$

[Wilcox et al 2002]

Outline

I. Clusters

Outline

I. Clusters

- Overview

Outline

I. Clusters

- Overview
- Distribution

Outline

I. Clusters

- Overview
- Distribution
- Proportions

Outline

I. Clusters

- Overview
- Distribution
- Proportions
- Random probability measure

Outline

I. Clusters

- Overview
- Distribution
- Proportions
- Random probability measure

II. Features

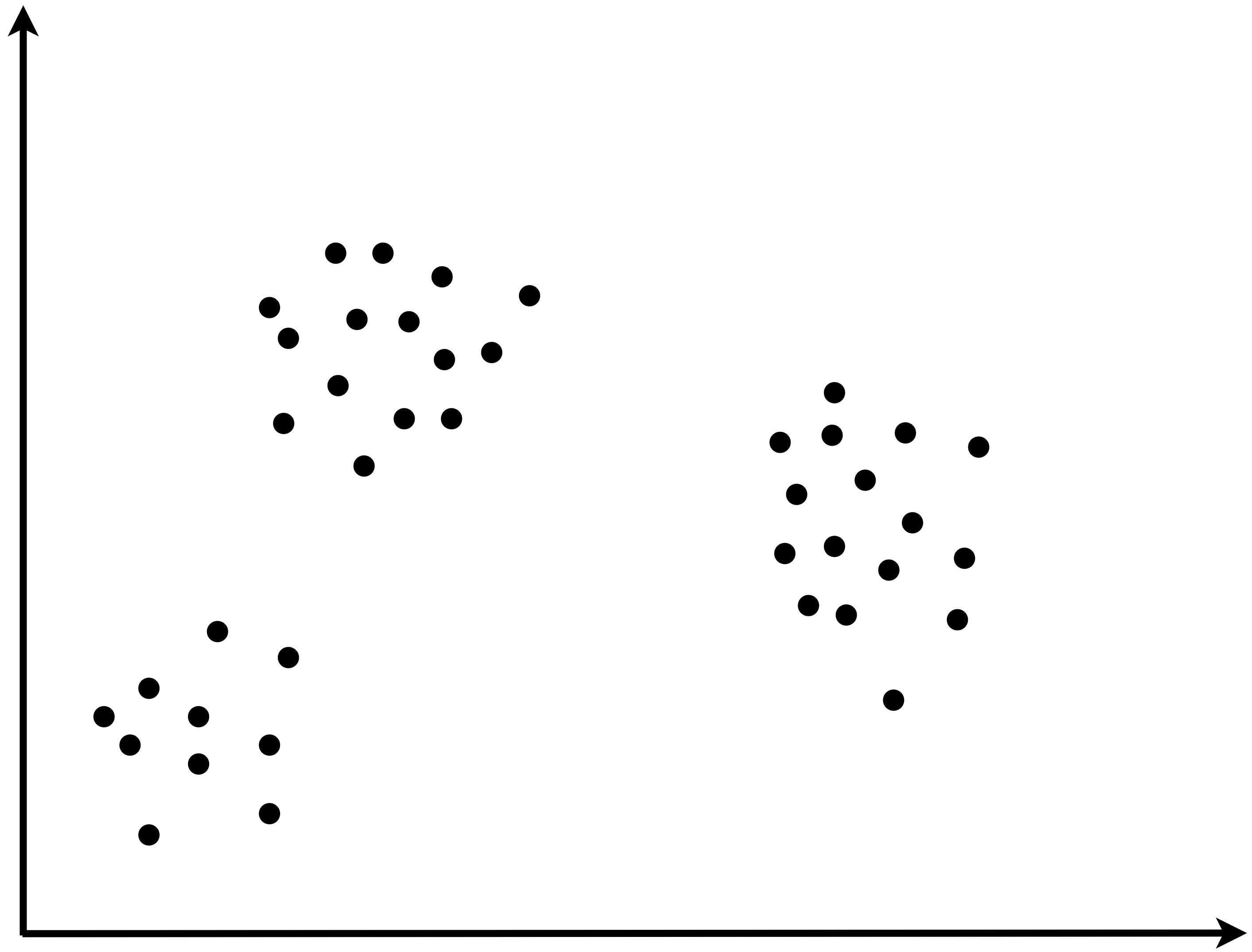
Outline

I. Clusters

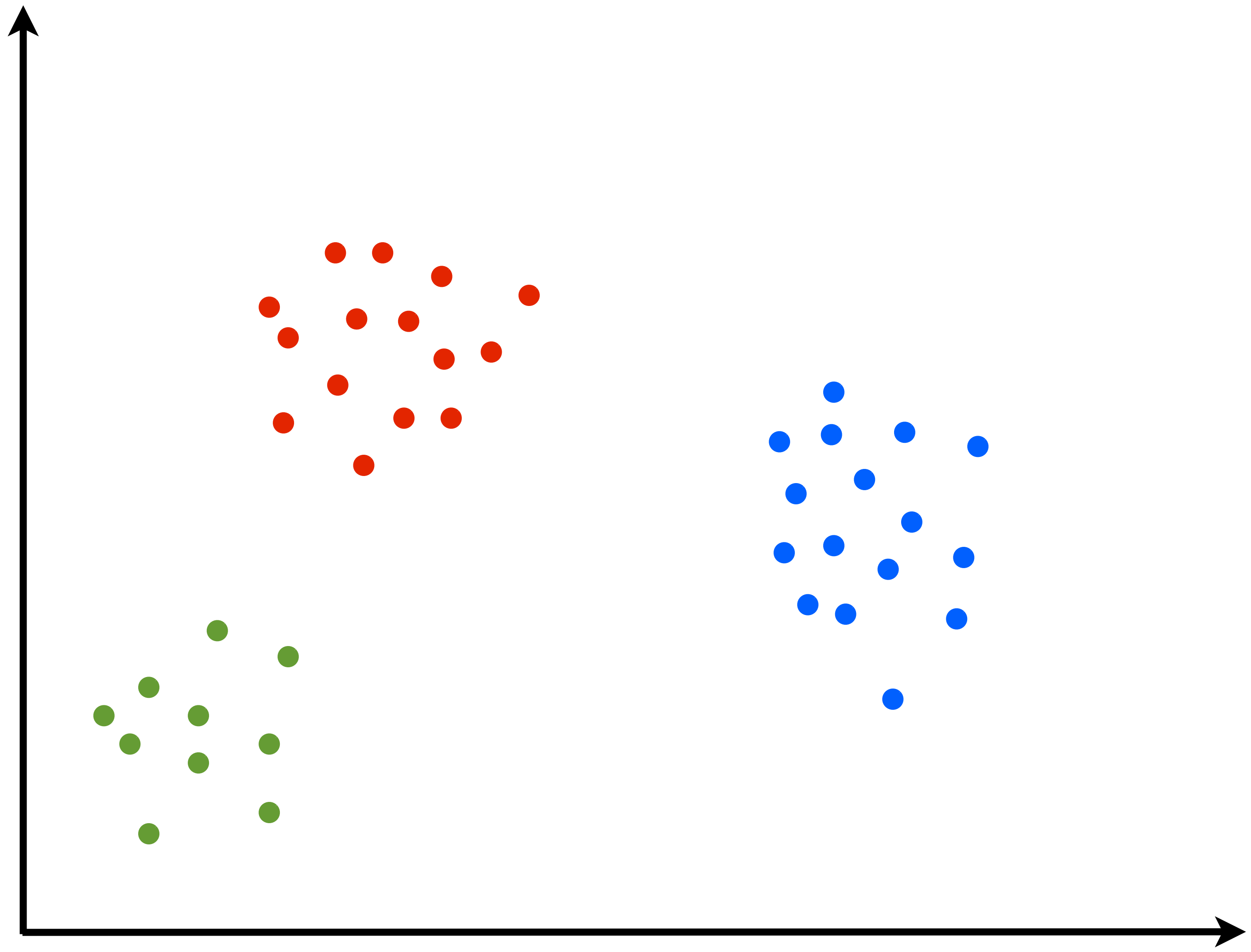
- Overview
- Distribution
- Proportions
- Random probability measure

II. Features

Clustering

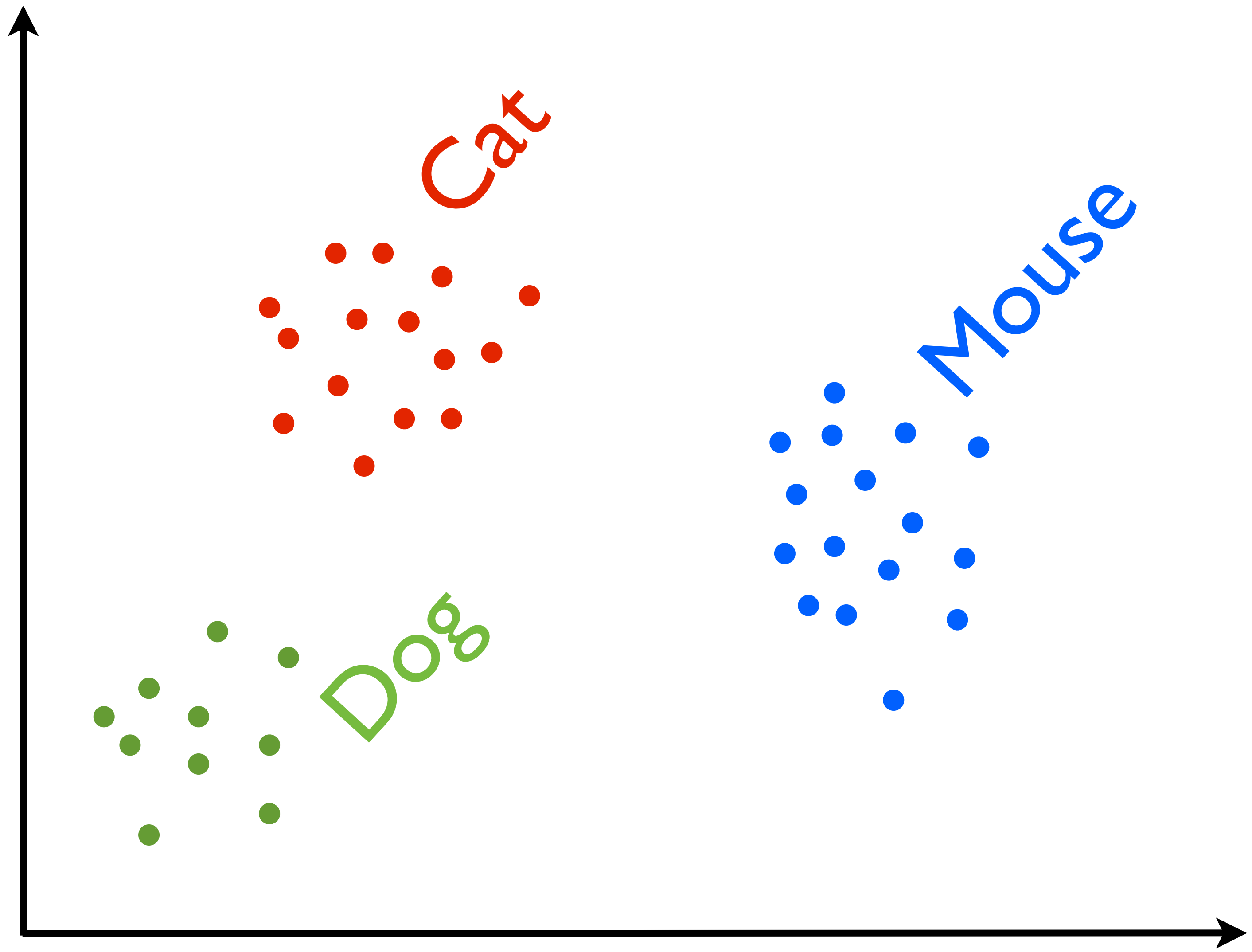


Clustering



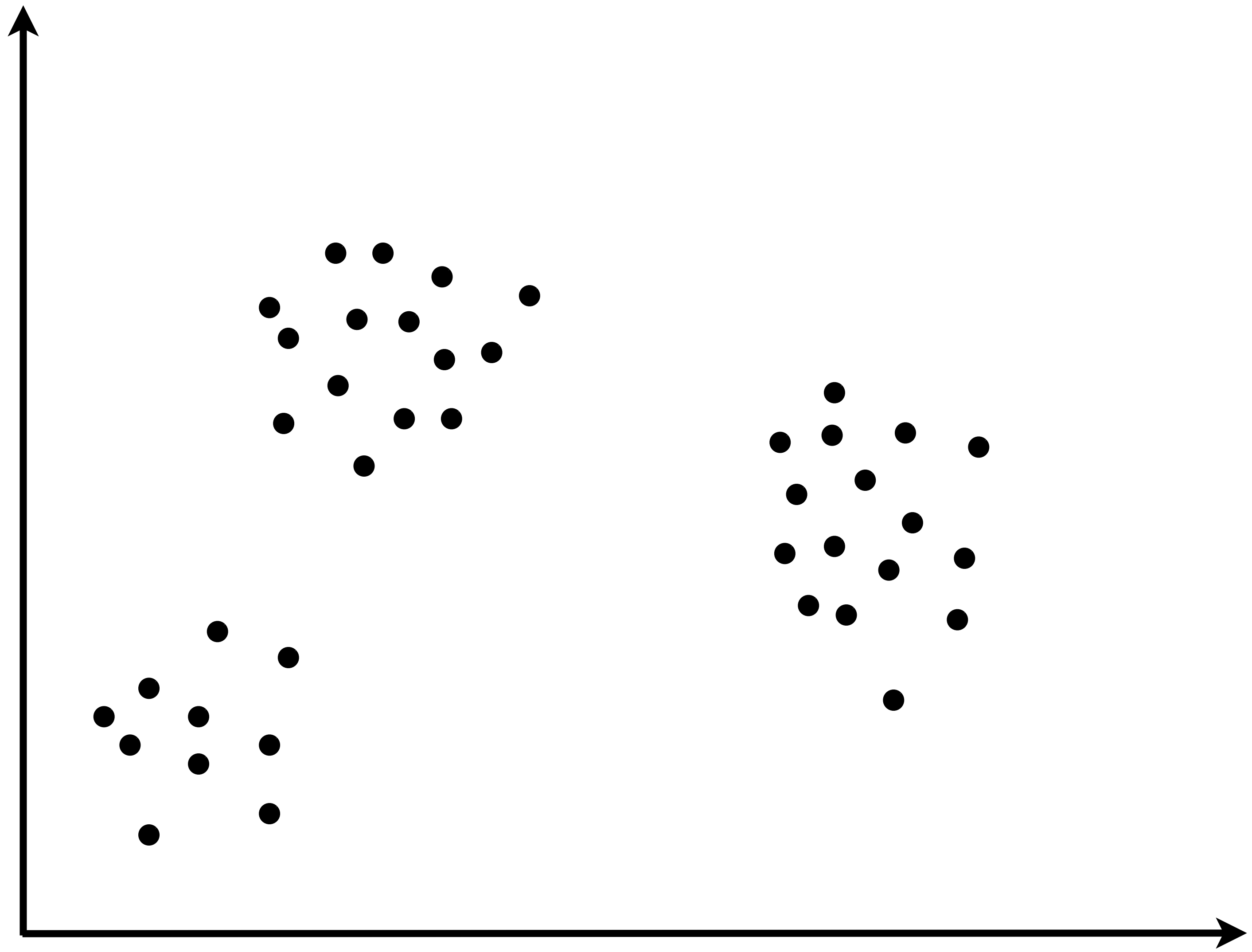
“clusters”,
“classes”,
“blocks (of a partition)”

Clustering



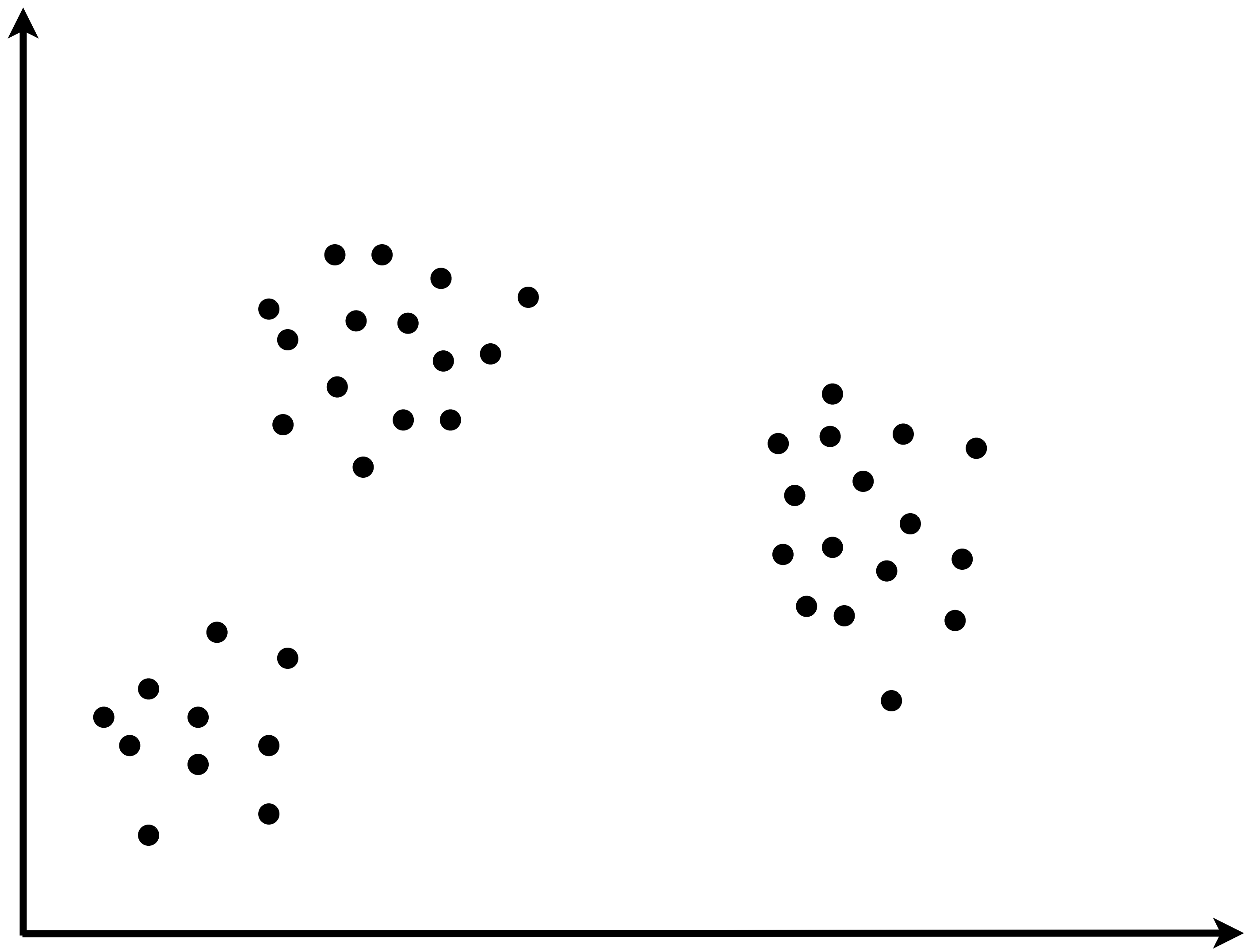
“clusters”,
“classes”,
“blocks (of a partition)”

Clustering

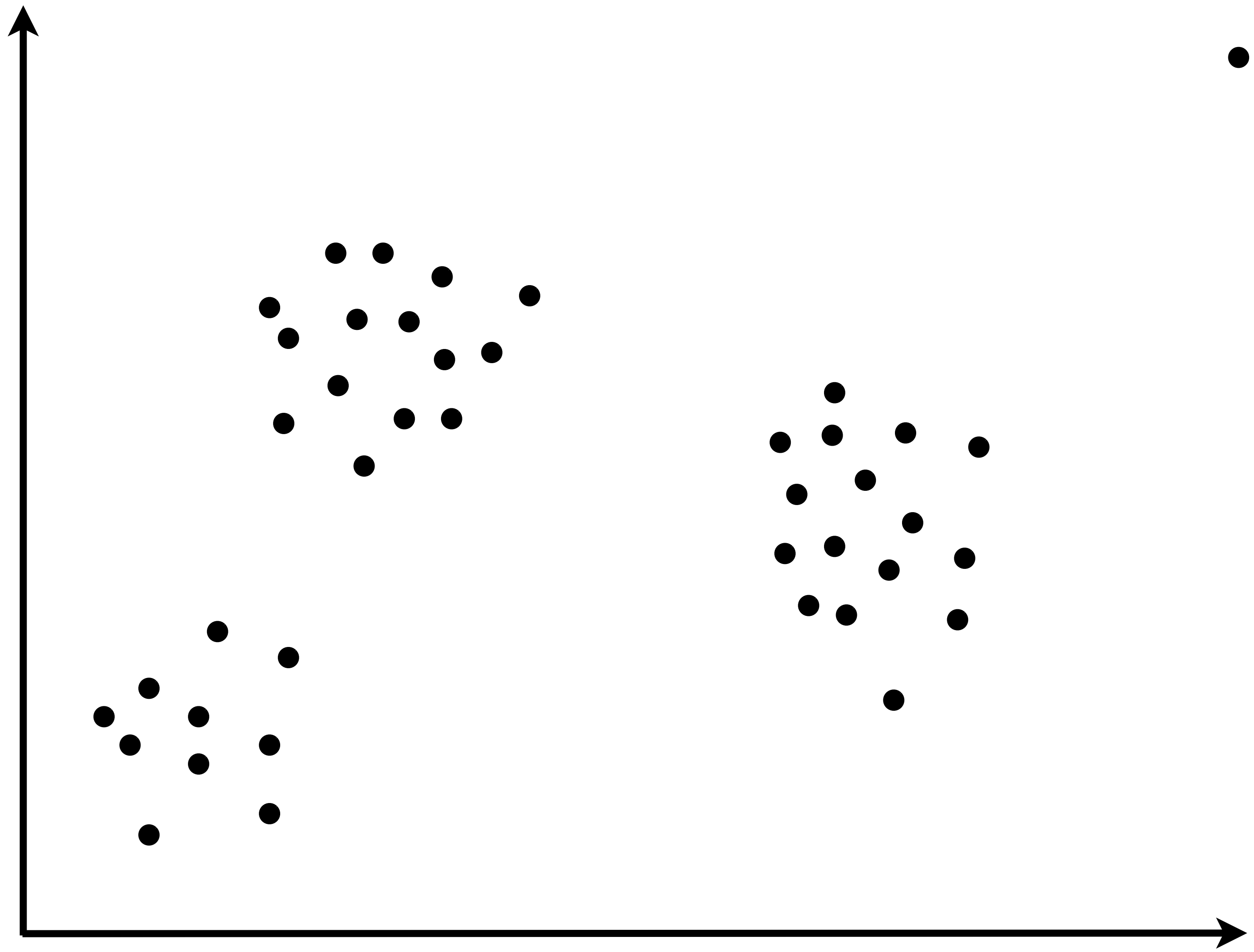


Clustering

...is hard



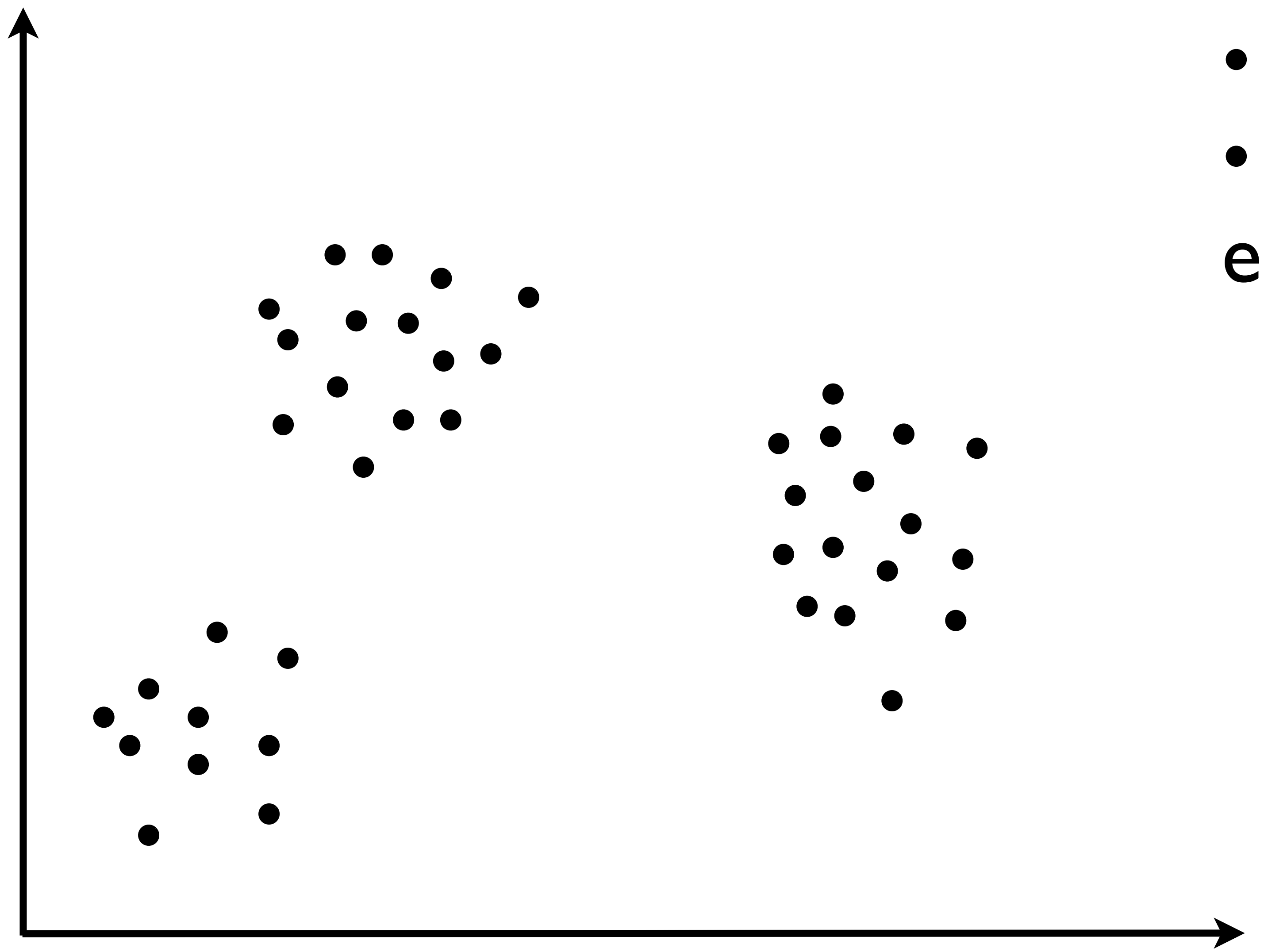
Clustering



...is hard

- Unsupervised

Clustering



...is hard

- Unsupervised
- Data dimensions not always easy to visualize

Clustering

...is useful

Clustering

...is useful

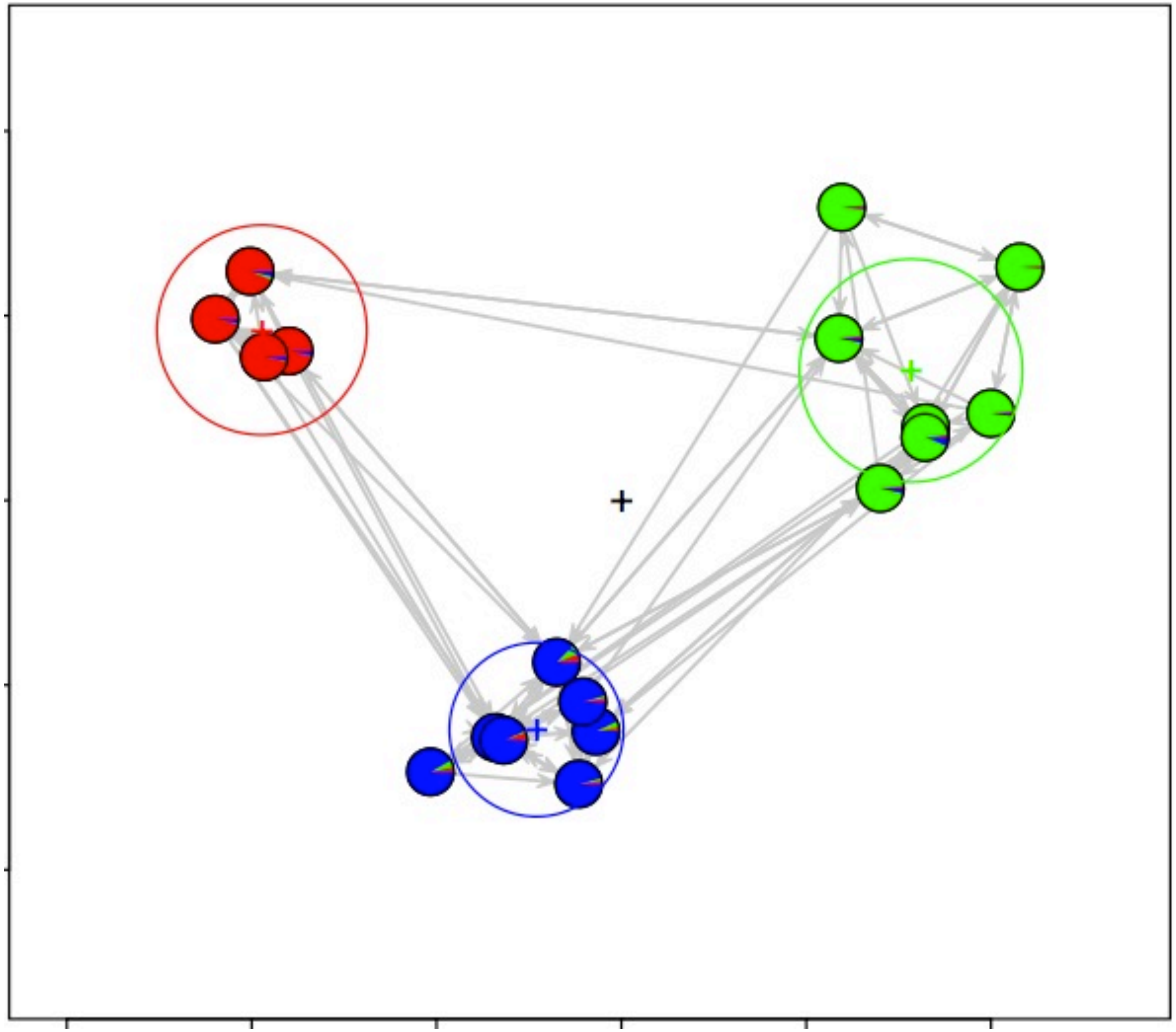
- Exploratory data analysis

Clustering

...is useful

- Exploratory data analysis

Network Analysis



[Krivitsky, Handcock 2008]

Clustering

...is useful

- Exploratory data analysis
- Classes are unspecified

(changing too quickly, expensive to label data, unknown, etc)

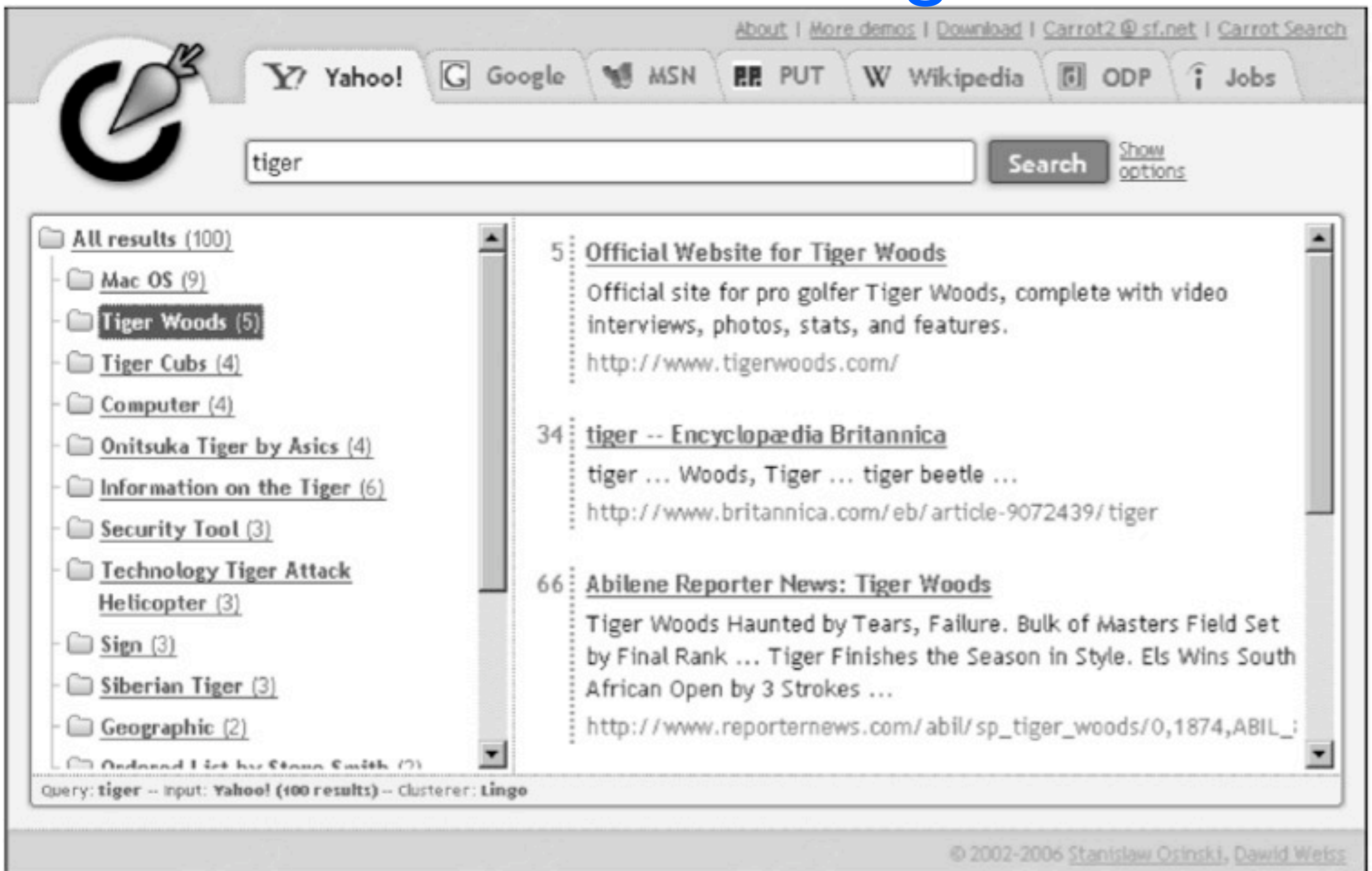
Clustering

...is useful

- Exploratory data analysis
- Classes are unspecified

(changing too quickly, expensive to label data, unknown, etc)

Document clustering



[Carpineto et al 2009]

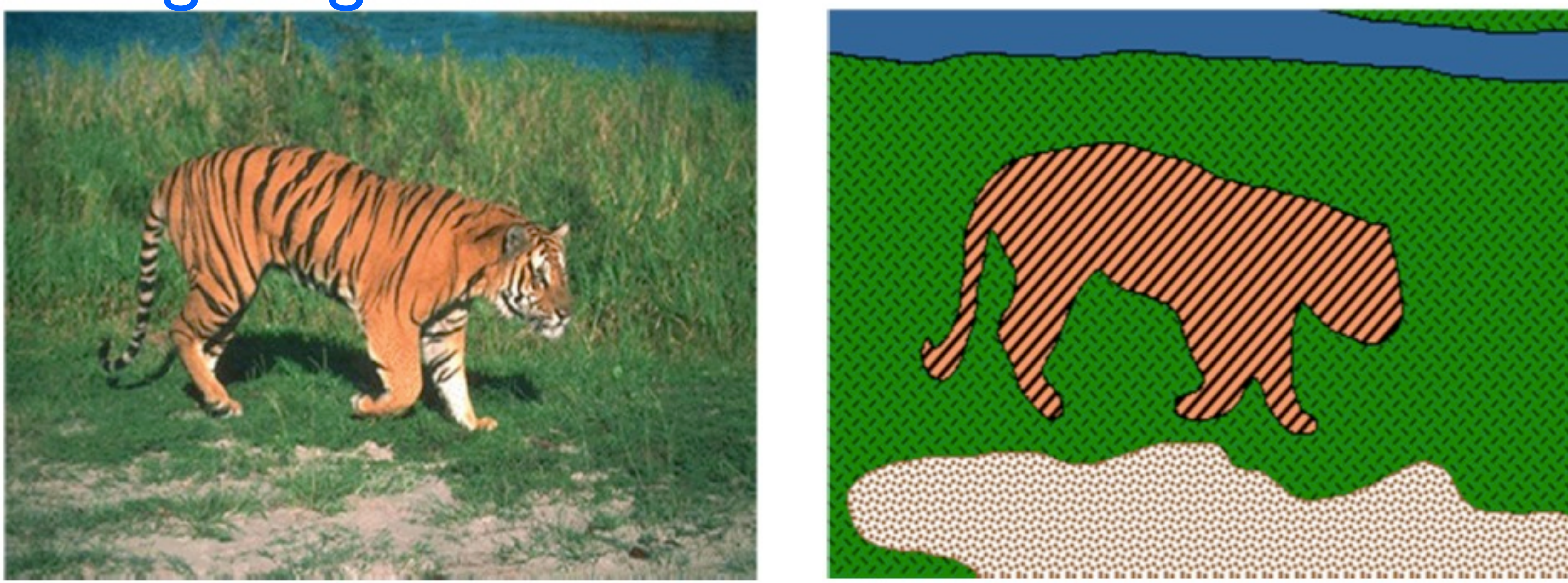
Clustering

...is useful

- Exploratory data analysis
- Classes are unspecified

(changing too quickly, expensive to label data, unknown, etc)

Image segmentation



[Fei-Fei 2011]

Clustering

...is useful

- Exploratory data analysis
- Classes are unspecified

(changing too quickly, expensive to label data, unknown, etc)

Topic Analysis

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

[...]arst Foundation will give $1.25 million to Lincoln Center, Metropoli-
[...]k Philharmonic and Juilliard School. "Our board felt that we had a
[...] a mark on the future of the performing arts with these grants an act
[...]ur traditional areas of support in health, medical research, education
[...] Hearst Foundation President Randolph A. Hearst said Monday in
[...]incoln Center's share will be $200,000 for its new building, which
[...]and provide new public facilities. The Metropolitan Opera Co. and
[...] will receive $400,000 each. The Juilliard School, where music and
the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

[Blei et al 2003]

Clustering

Why Bayesian?

Topic Analysis

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

earst Foundation will give $1.25 million to Lincoln Center, Metropoli-
k Philharmonic and Juilliard School. "Our board felt that we had a
a mark on the future of the performing arts with these grants an act
our traditional areas of support in health, medical research, education
Hearst Foundation President Randolph A. Hearst said Monday in
incoln Center's share will be $200,000 for its new building, which
and provide new public facilities. The Metropolitan Opera Co. and
will receive $400,000 each. The Juilliard School, where music and
the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

[Blei et al 2003]

Clustering

Why Bayesian?

- Flexibility to specify model

Topic Analysis

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

arst Foundation will give $1.25 million to Lincoln Center, Metropoli-
k Philharmonic and Juilliard School. "Our board felt that we had a
a mark on the future of the performing arts with these grants an act
our traditional areas of support in health, medical research, education
Hearst Foundation President Randolph A. Hearst said Monday in
incoln Center's share will be $200,000 for its new building, which
and provide new public facilities. The Metropolitan Opera Co. and
will receive $400,000 each. The Juilliard School, where music and
the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

[Blei et al 2003]

Clustering

Why Bayesian?

- Flexibility to specify model

Why nonparametric?

Topic Analysis

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

earst Foundation will give $1.25 million to Lincoln Center, Metropoli-
k Philharmonic and Juilliard School. "Our board felt that we had a
a mark on the future of the performing arts with these grants an act
our traditional areas of support in health, medical research, education
Hearst Foundation President Randolph A. Hearst said Monday in
incoln Center's share will be $200,000 for its new building, which
and provide new public facilities. The Metropolitan Opera Co. and
will receive $400,000 each. The Juilliard School, where music and
the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter
of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000
donation, too.

[Blei et al 2003]

Clustering

Why Bayesian?

- Flexibility to specify model

Why nonparametric?

- Don't know the number of clusters in advance

Topic Analysis

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

earst Foundation will give $1.25 million to Lincoln Center, Metropoli-
k Philharmonic and Juilliard School. "Our board felt that we had a
a mark on the future of the performing arts with these grants an act
our traditional areas of support in health, medical research, education
Hearst Foundation President Randolph A. Hearst said Monday in
Lincoln Center's share will be $200,000 for its new building, which
and provide new public facilities. The Metropolitan Opera Co. and
will receive $400,000 each. The Juilliard School, where music and
the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

[Blei et al 2003]

Outline

I. Clusters

- Overview
- Distribution
- Proportions
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
- Proportions
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
    - ◇ Clusters
    - ◇ Data given clusters
    - ◇ Posterior
- Proportions
- Random probability measure

II. Features

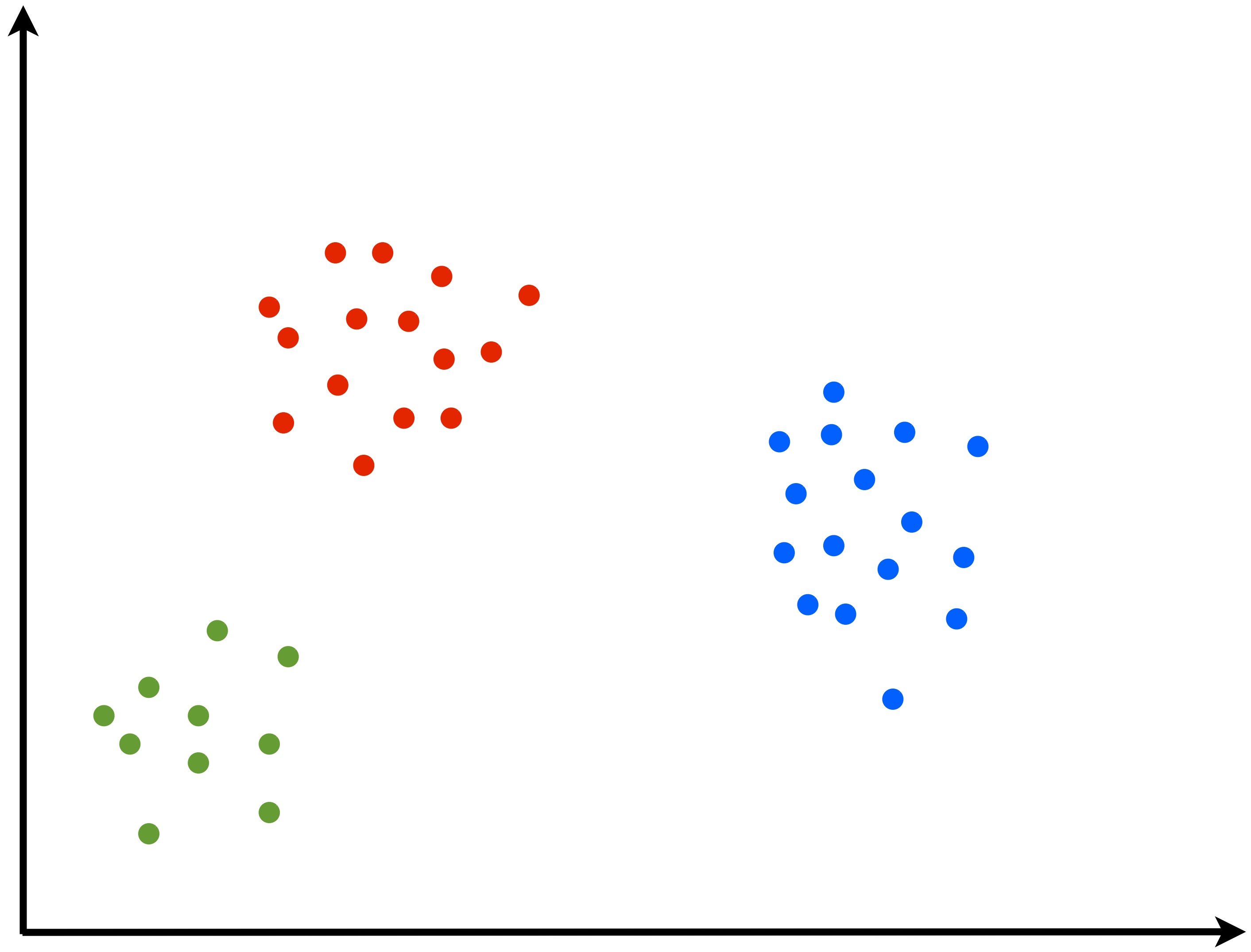
Outline

I. Clusters

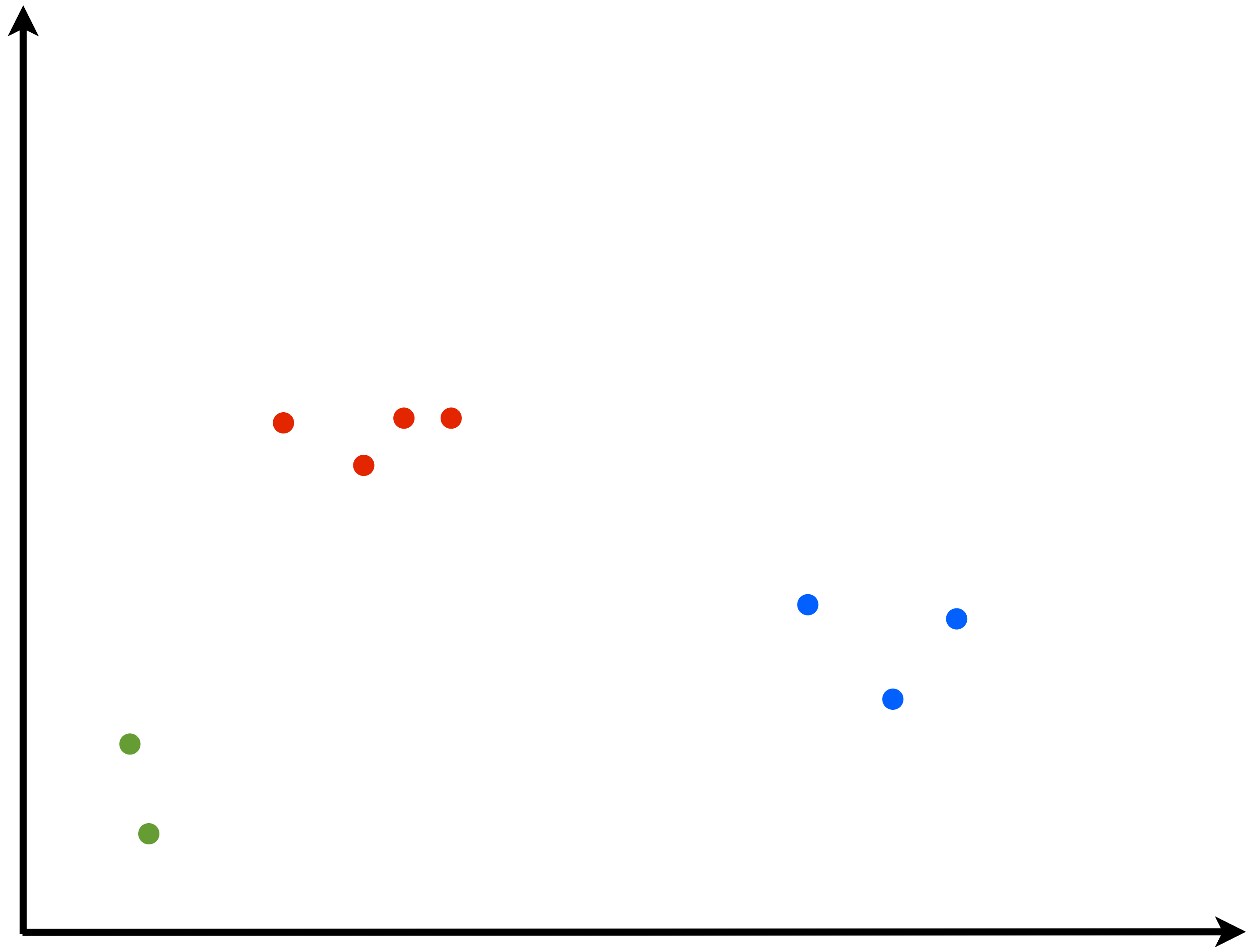
- Overview
- Distribution
    - ◇ Clusters
    - ◇ Data given clusters
    - ◇ Posterior
- Proportions
- Random probability measure

II. Features

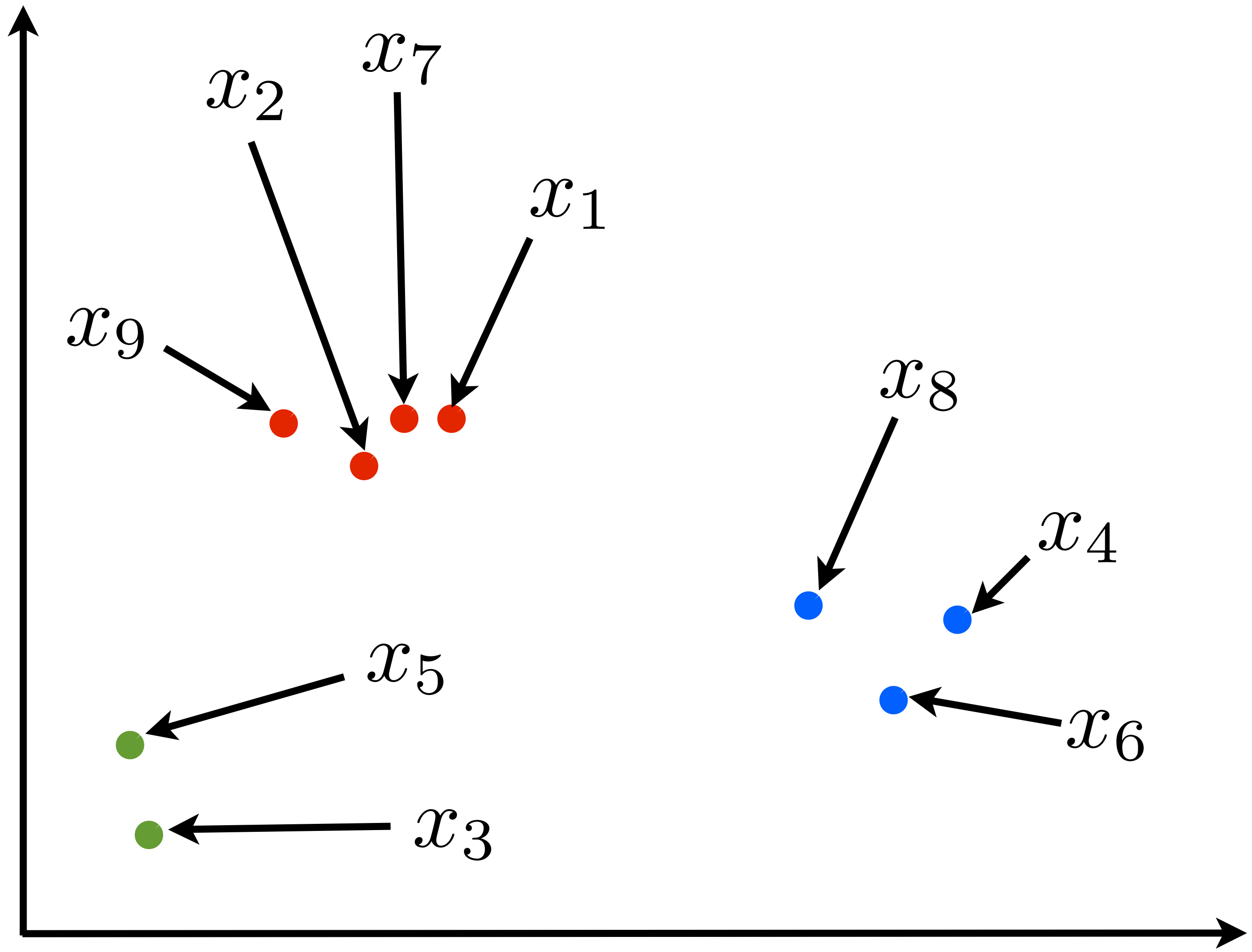
Clustering



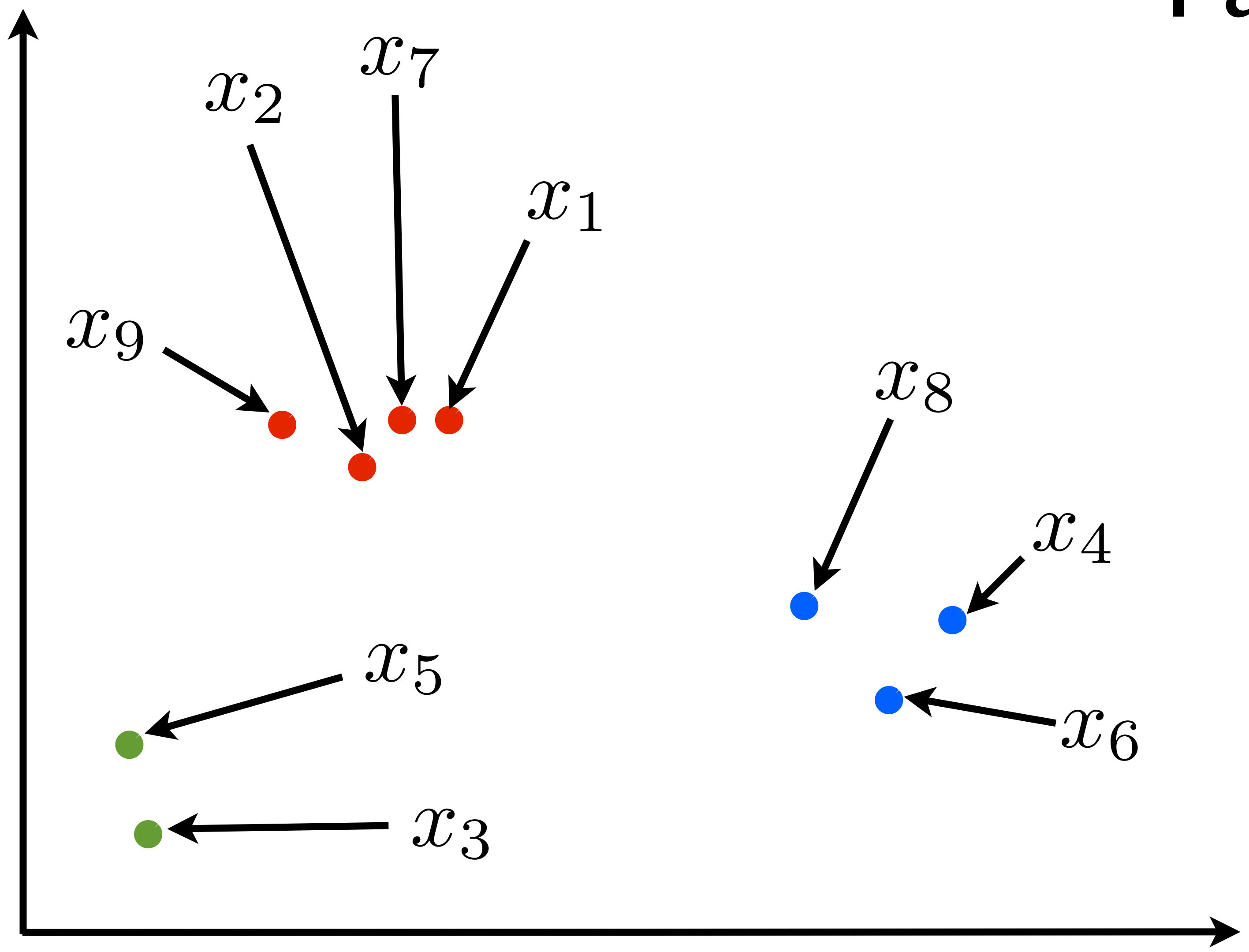
Clustering



Clustering

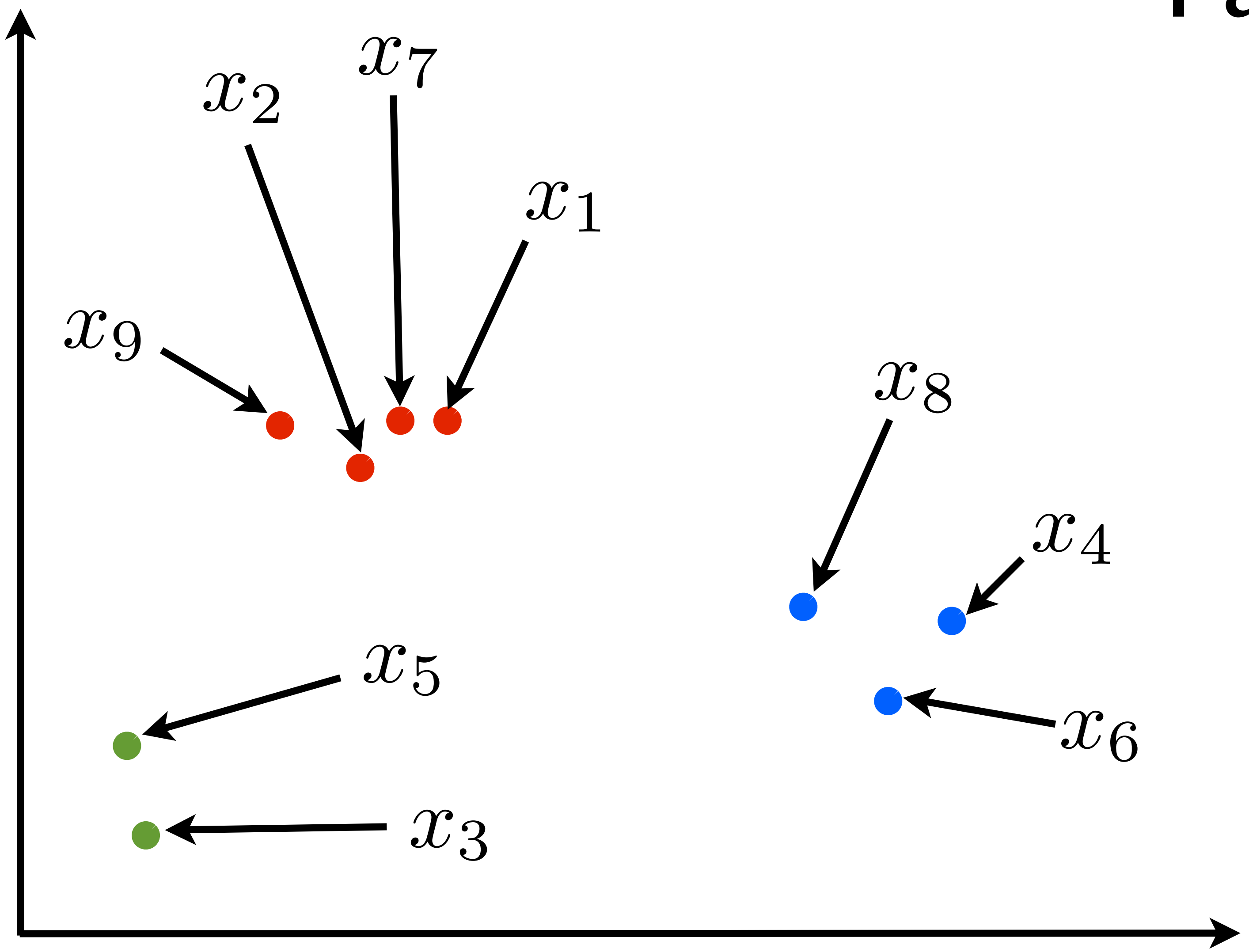


Clustering



Partition of 1, 2, ..., 9

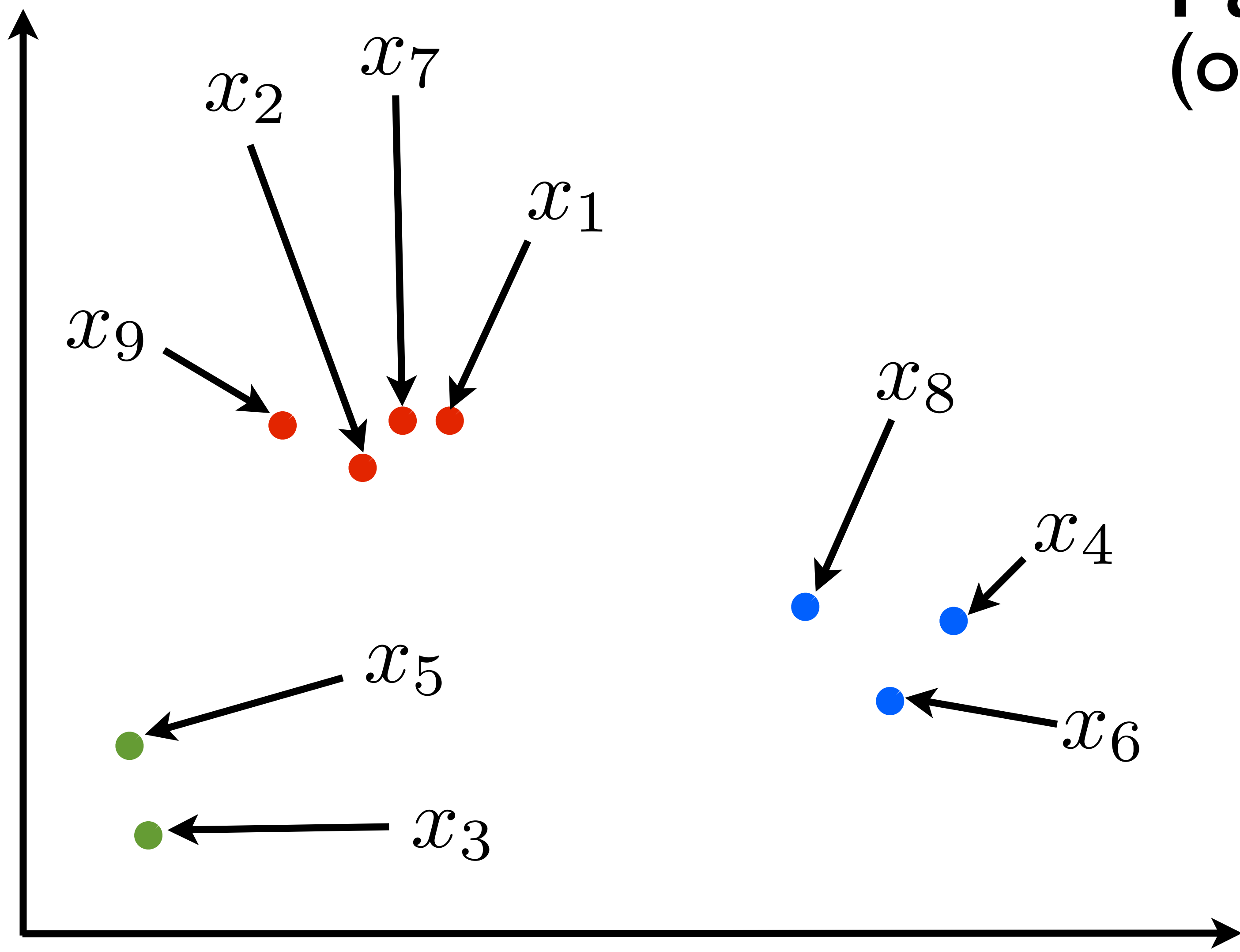
Clustering



Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

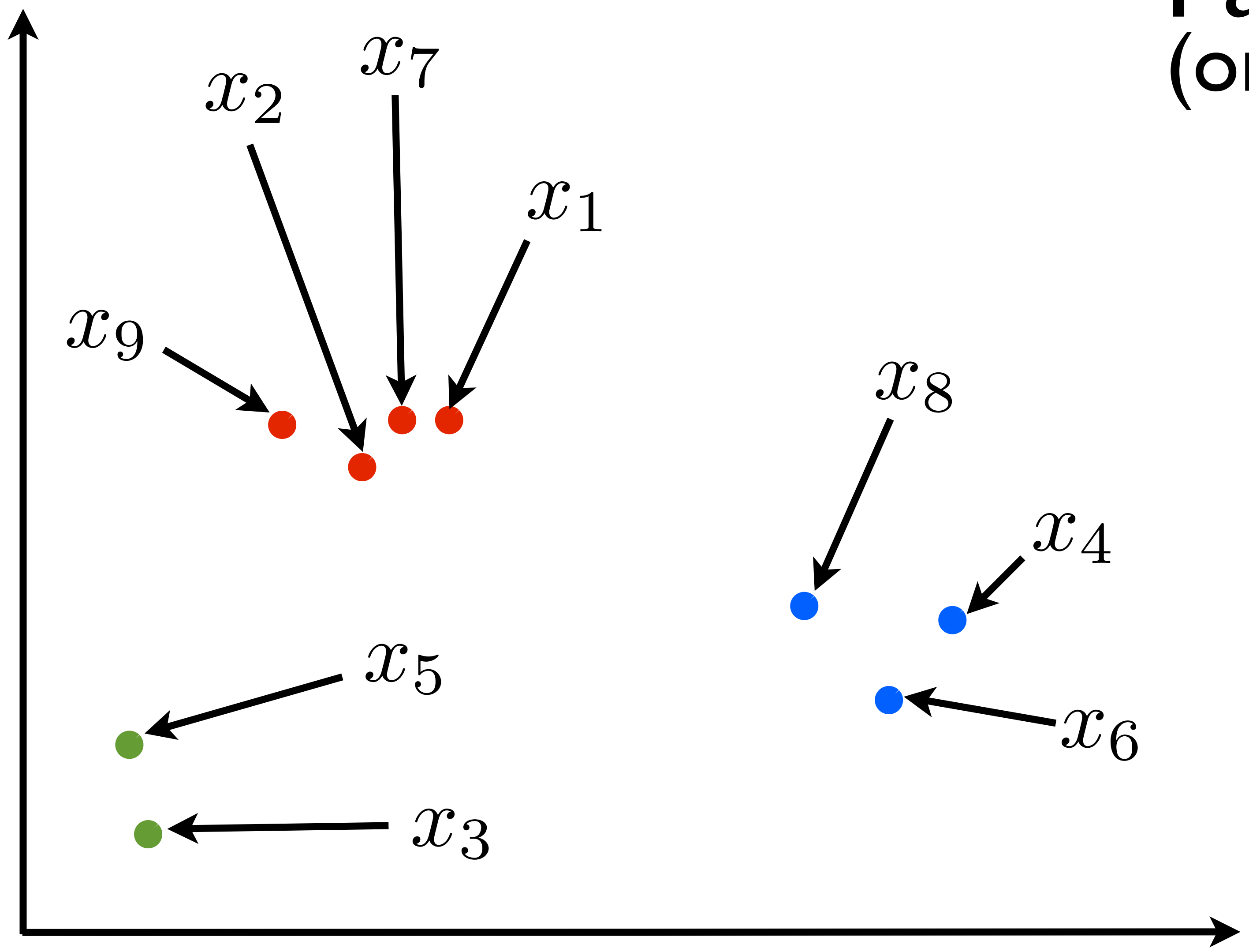
Clustering



Partition of 1, 2, ..., 9
(or clustering)

$$\pi_9 = \{\{9,2,7,1\}, \{8,4,6\}, \{5,3\}\}$$

Clustering

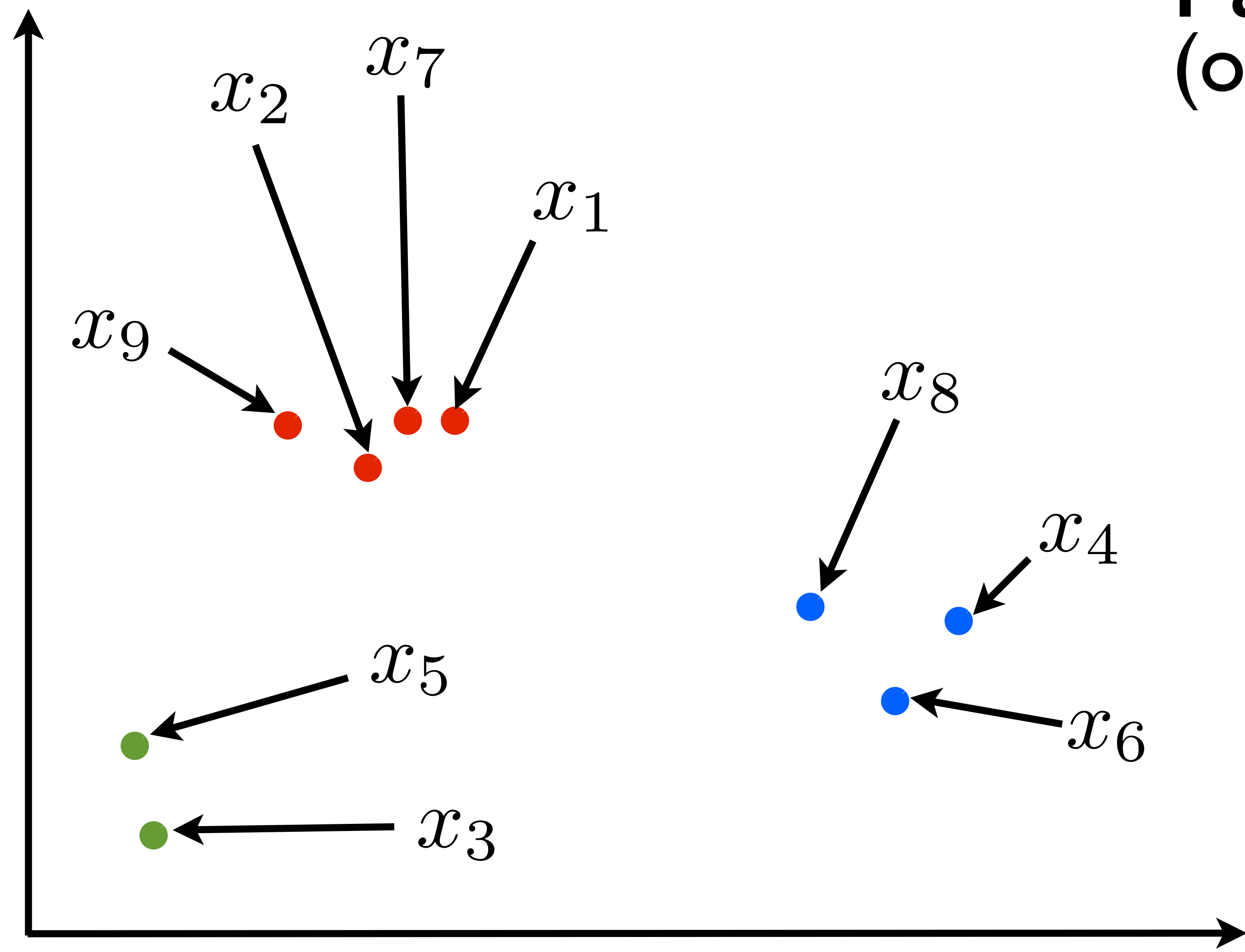


Partition of 1, 2, ..., 9
(or clustering)

$$\pi_9 = \{\{9, 2, 7, 1\},\ \{8, 4, 6\},\ \{5, 3\}\}$$

cluster

Clustering



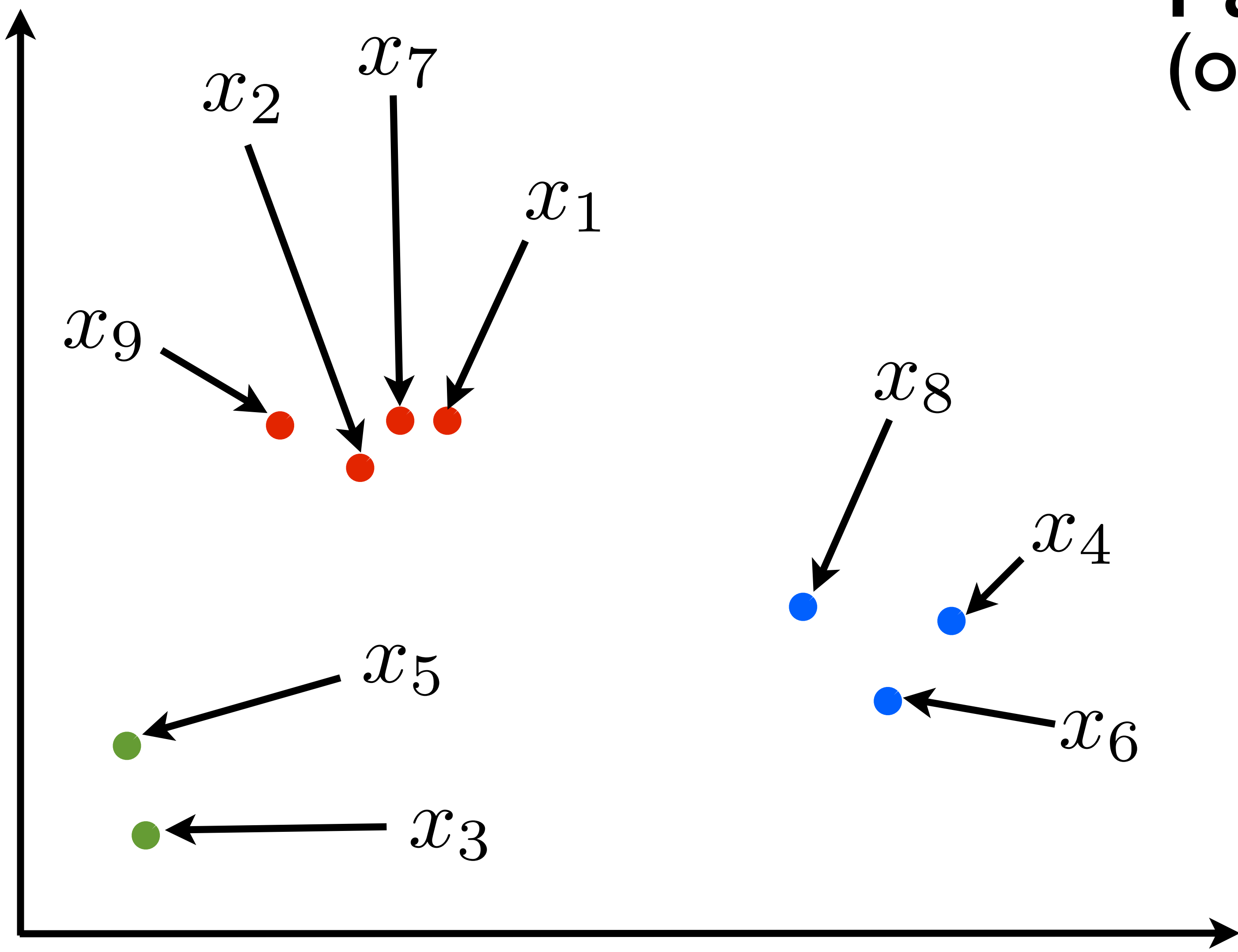
Partition of 1, 2, ..., 9
(or clustering)

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

cluster

N: Number of data points

Clustering



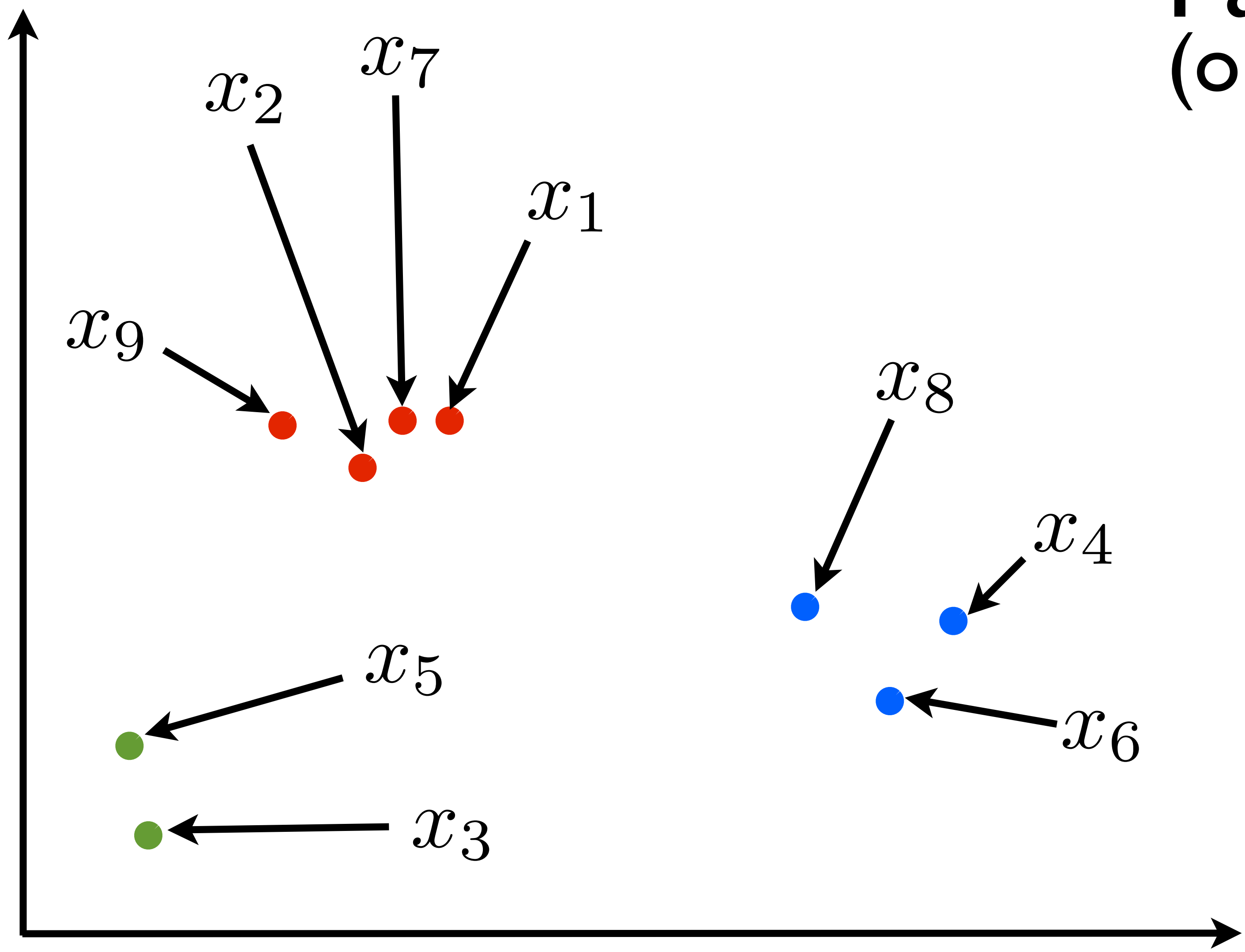
Partition of 1, 2, ..., 9
(or clustering)

$$\pi_9 = \{\{9,2,7,1\},\ \{8,4,6\},\ \{5,3\}\}$$

cluster

N: Number of data points
K: Number of clusters

Clustering



Partition of 1, 2, ..., 9
(or clustering)

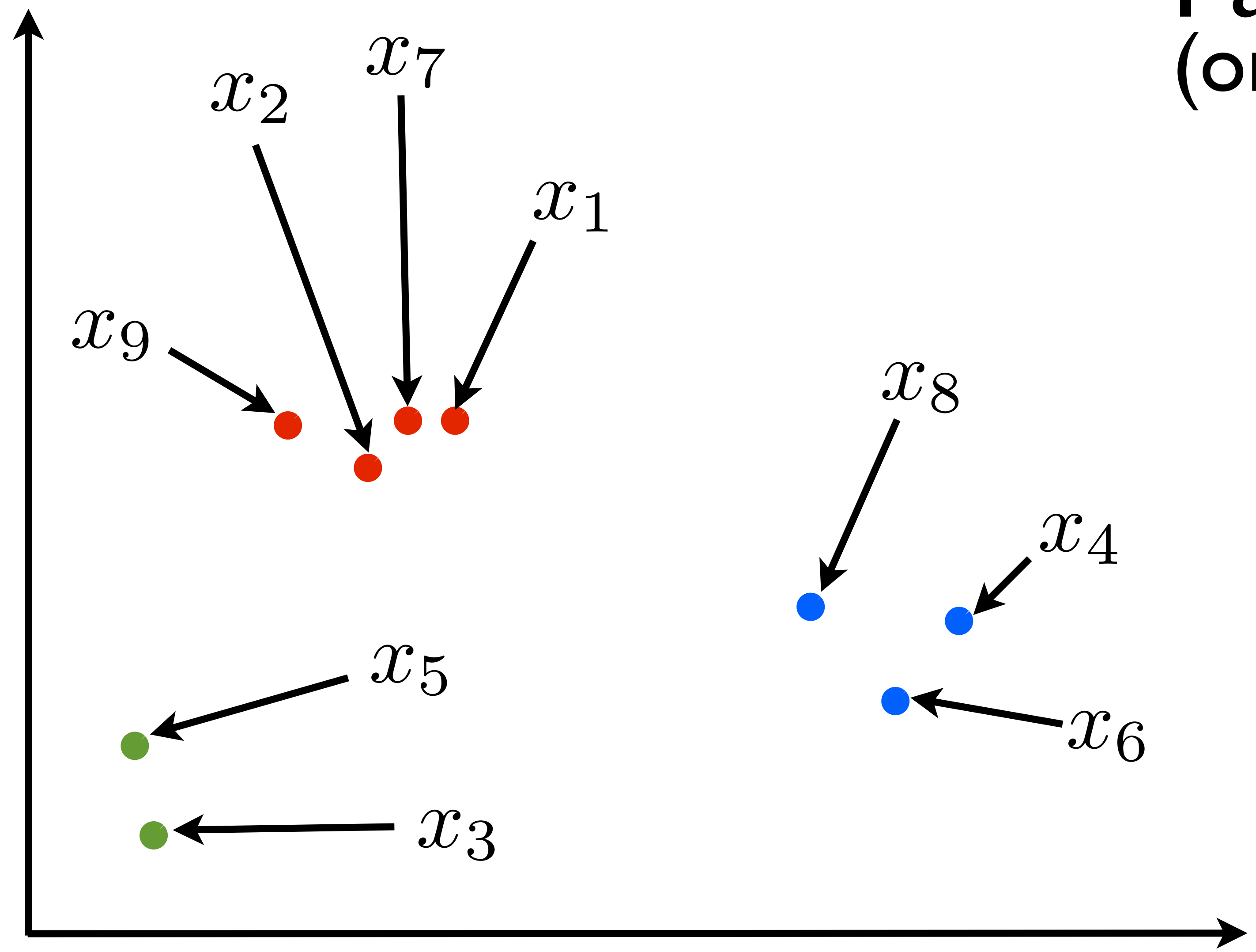
$$\pi_9 = \{\{9,2,7,1\},\ \{8,4,6\},\ \{5,3\}\}$$

cluster

N: Number of data points (N = 9)

K: Number of clusters

Clustering



Partition of 1, 2, ..., 9
(or clustering)

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

cluster

N: Number of data points (N = 9)

K: Number of clusters (K = 3)

Clustering

Random partition

Clustering

Random partition

Partition of 1, 2, ..., 9

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

Partition of 1, 2, ..., 9

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

- Exchangeable

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

- Exchangeable

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9,2,7,1\}, \{8,4,6\}, \{5,3\}\}$$

$$\pi'_9 = \{\{1,3,8,2\}, \{9,5,7\}, \{6,4\}\}$$

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

- Exchangeable

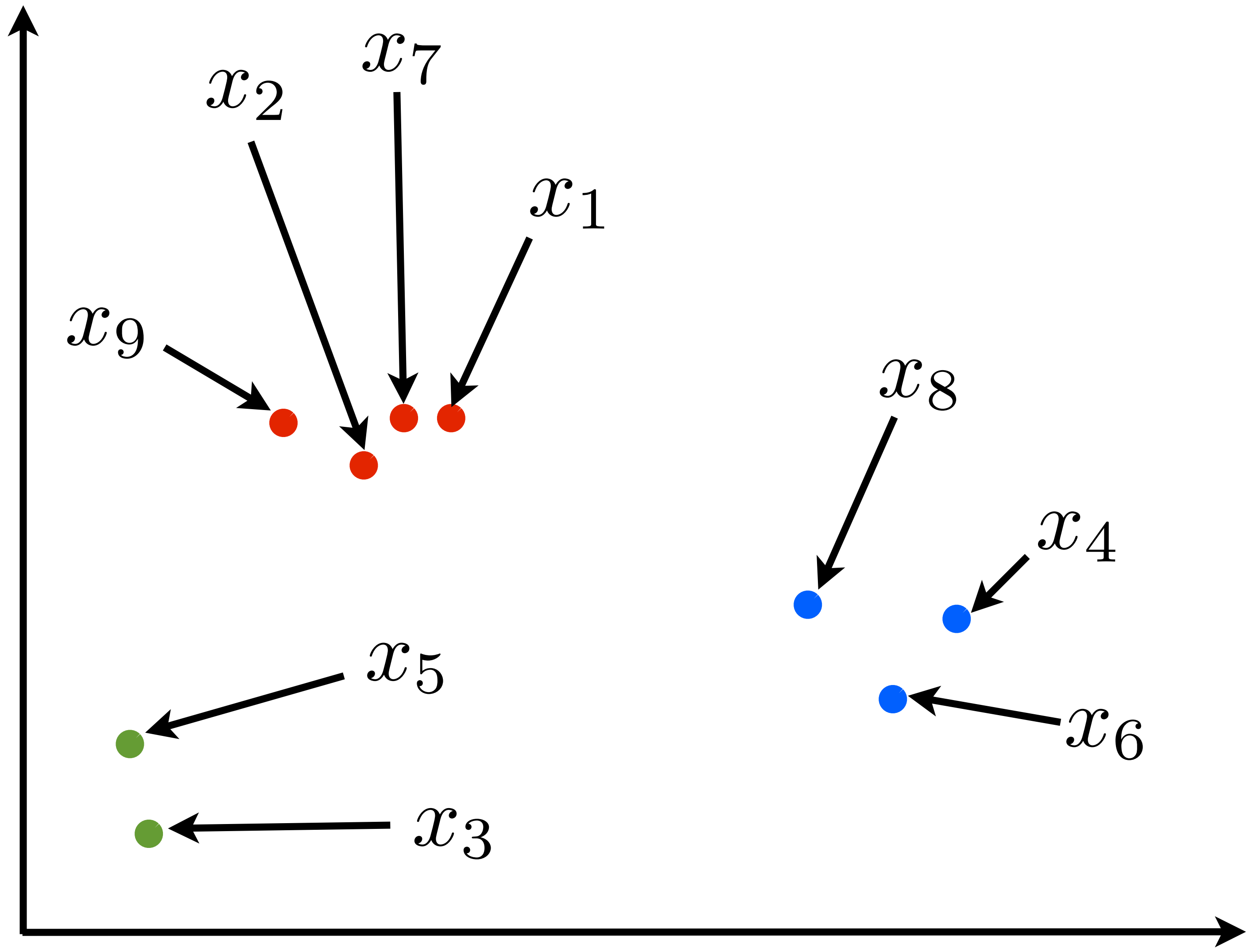
Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

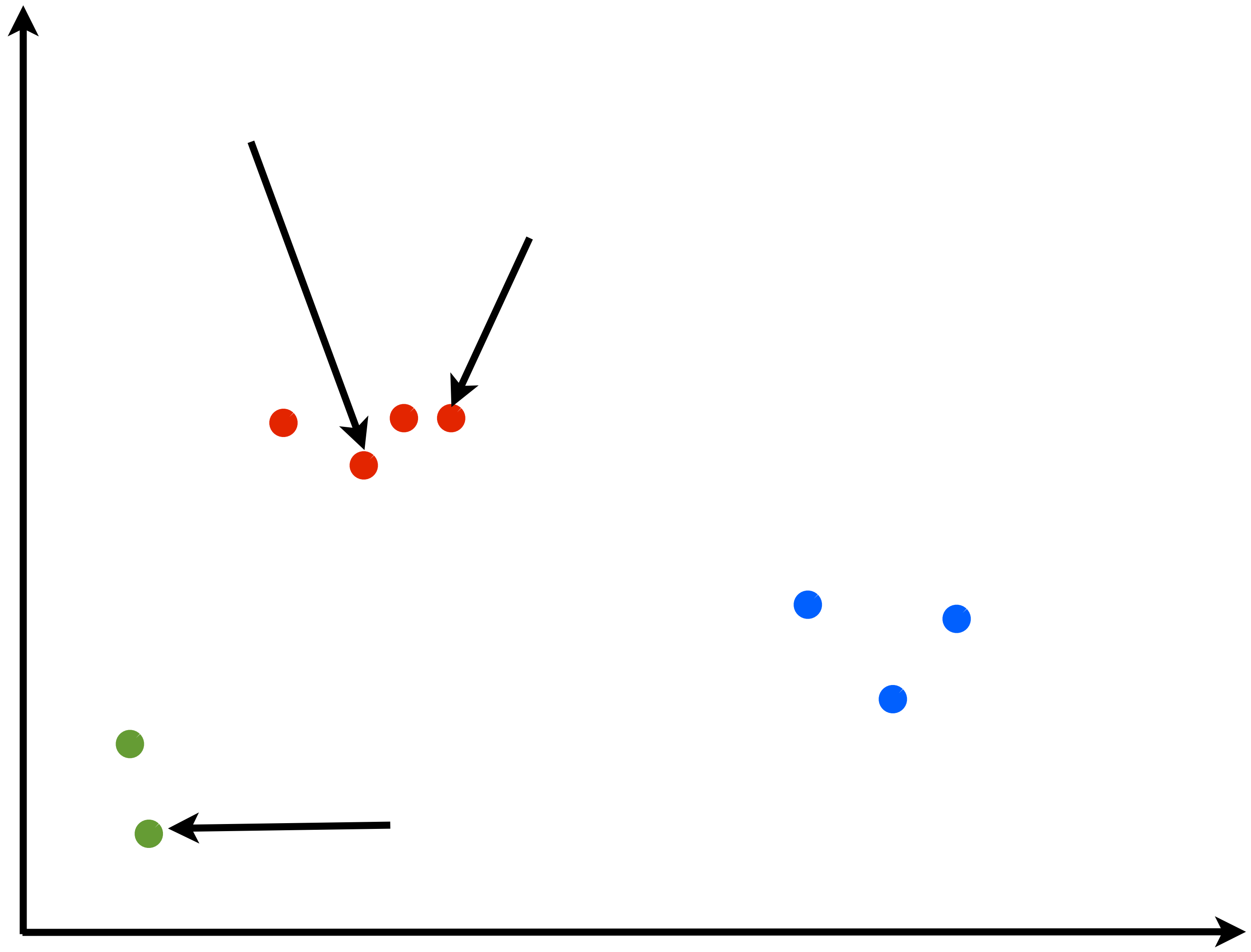
$$\mathbb{P}(\Pi_9 = \pi_9) = \mathbb{P}(\Pi_9 = \pi_9')$$

$$\pi_9' = \{\{1, 3, 8, 2\}, \{9, 5, 7\}, \{6, 4\}\}$$

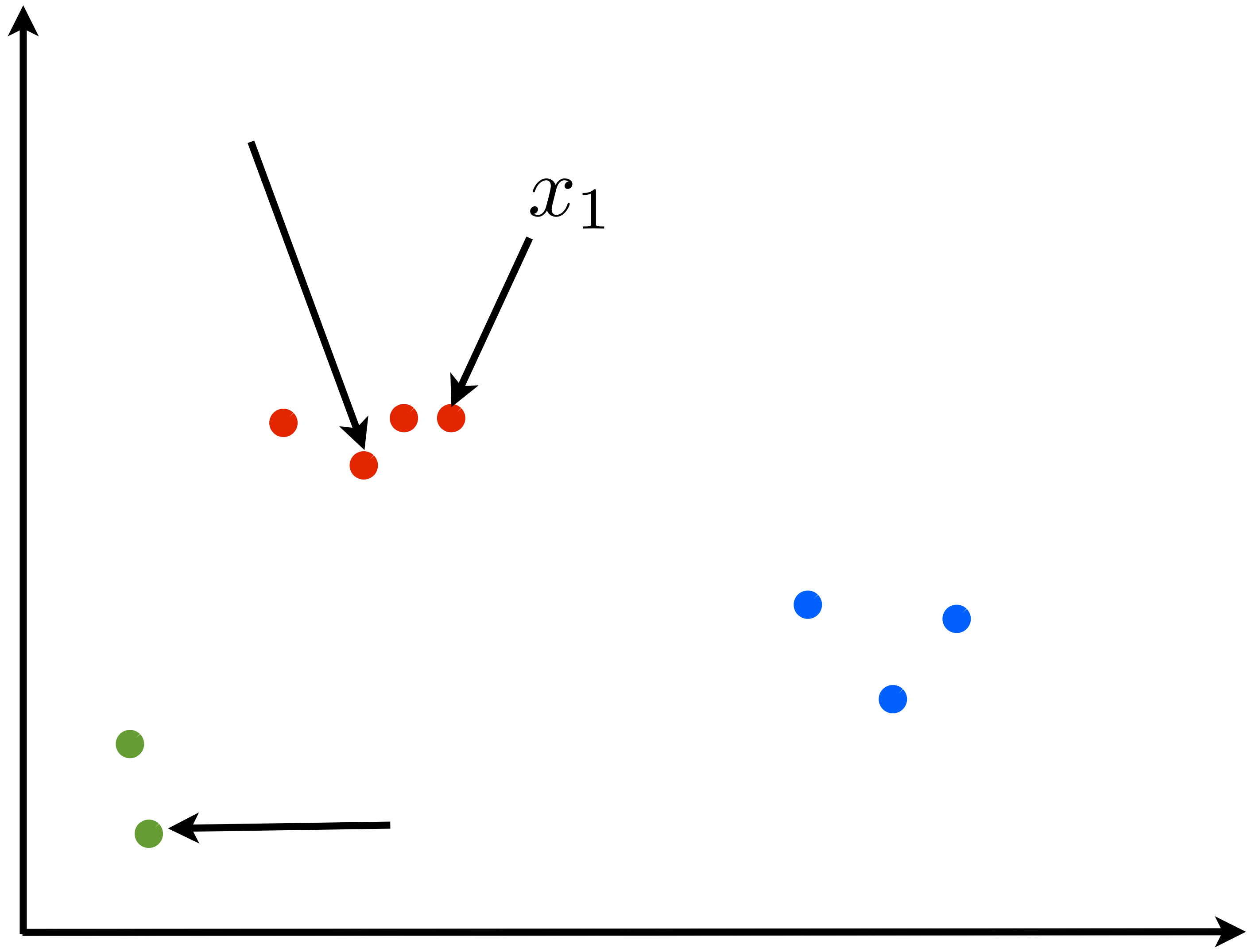
Exchangeability



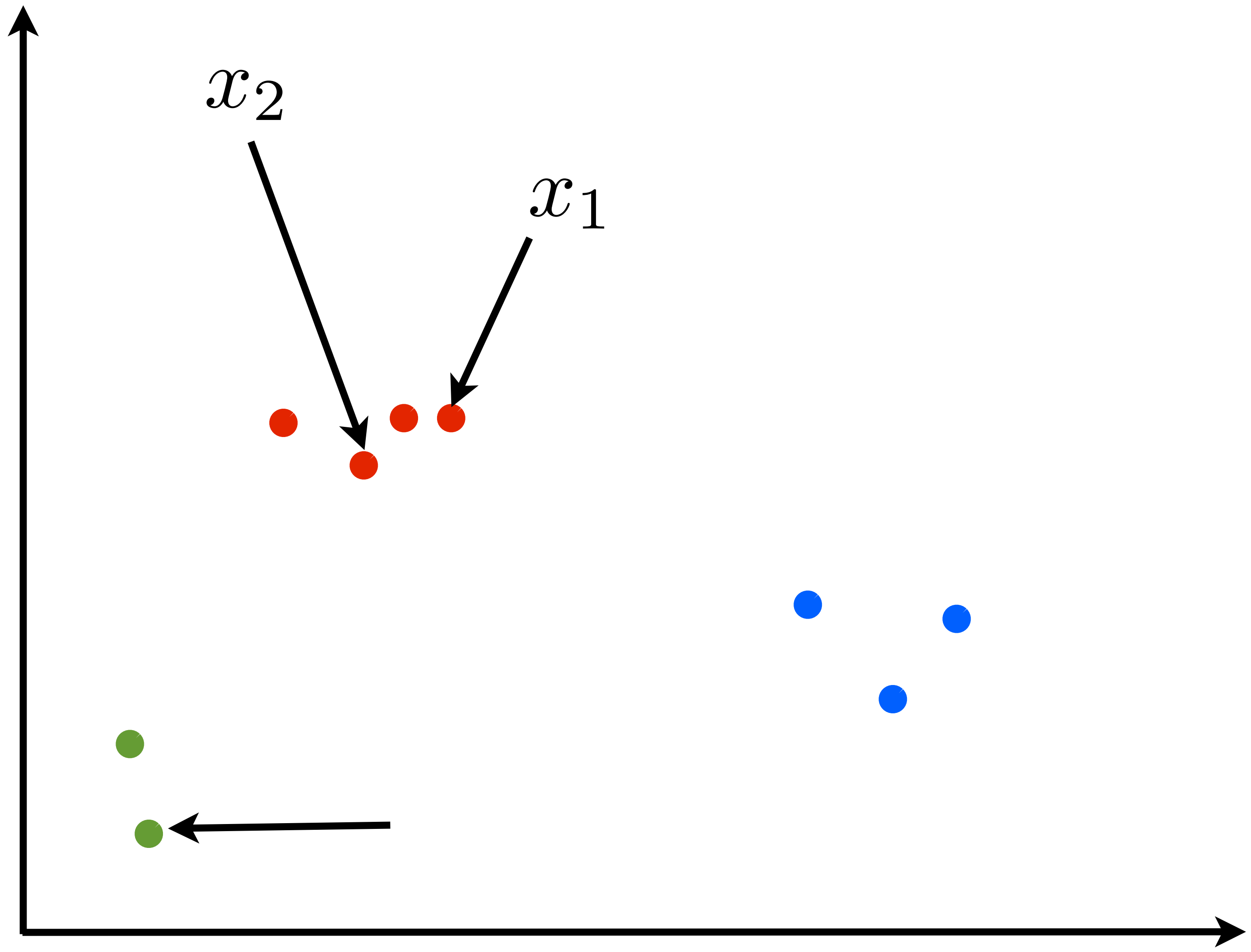
Exchangeability



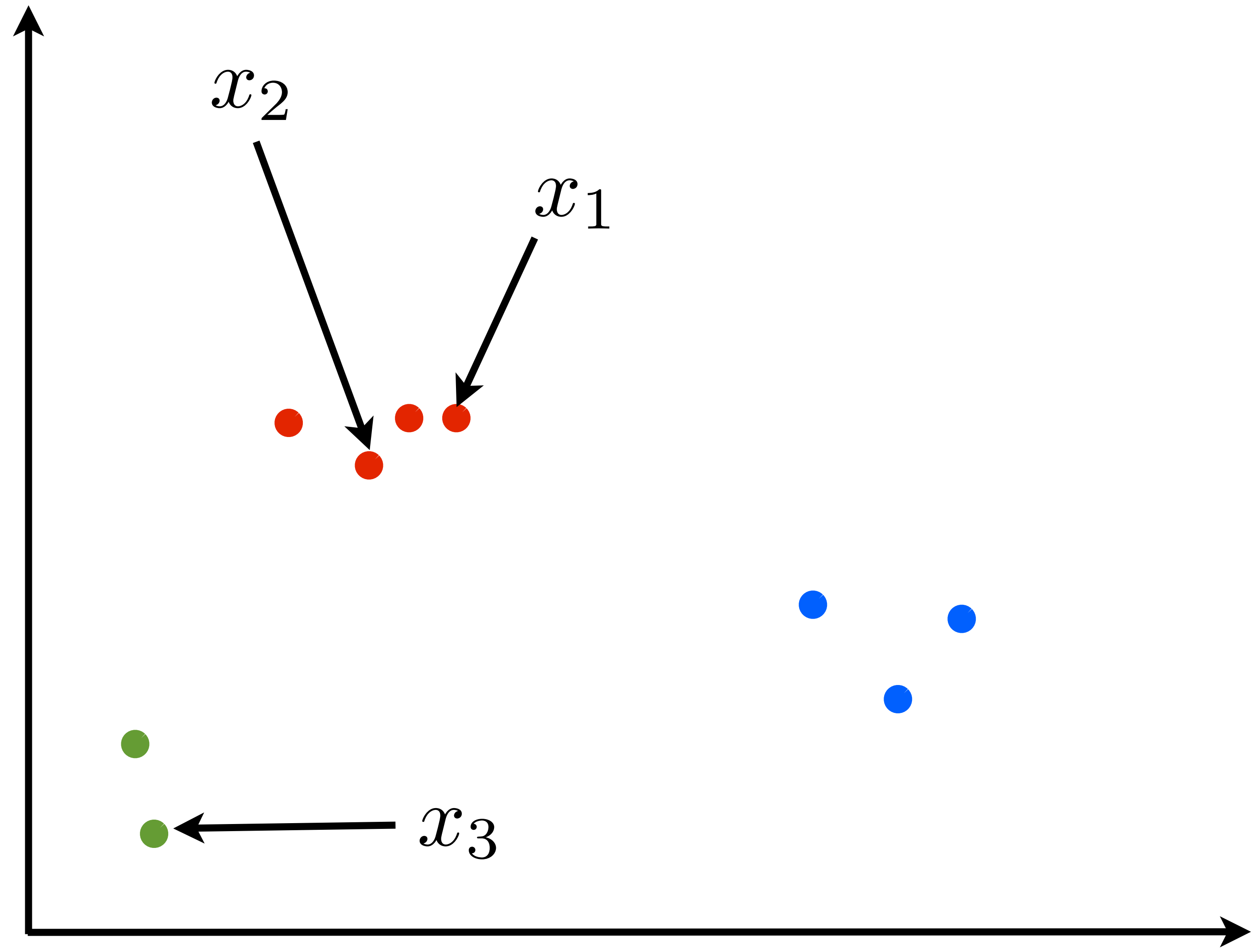
Exchangeability



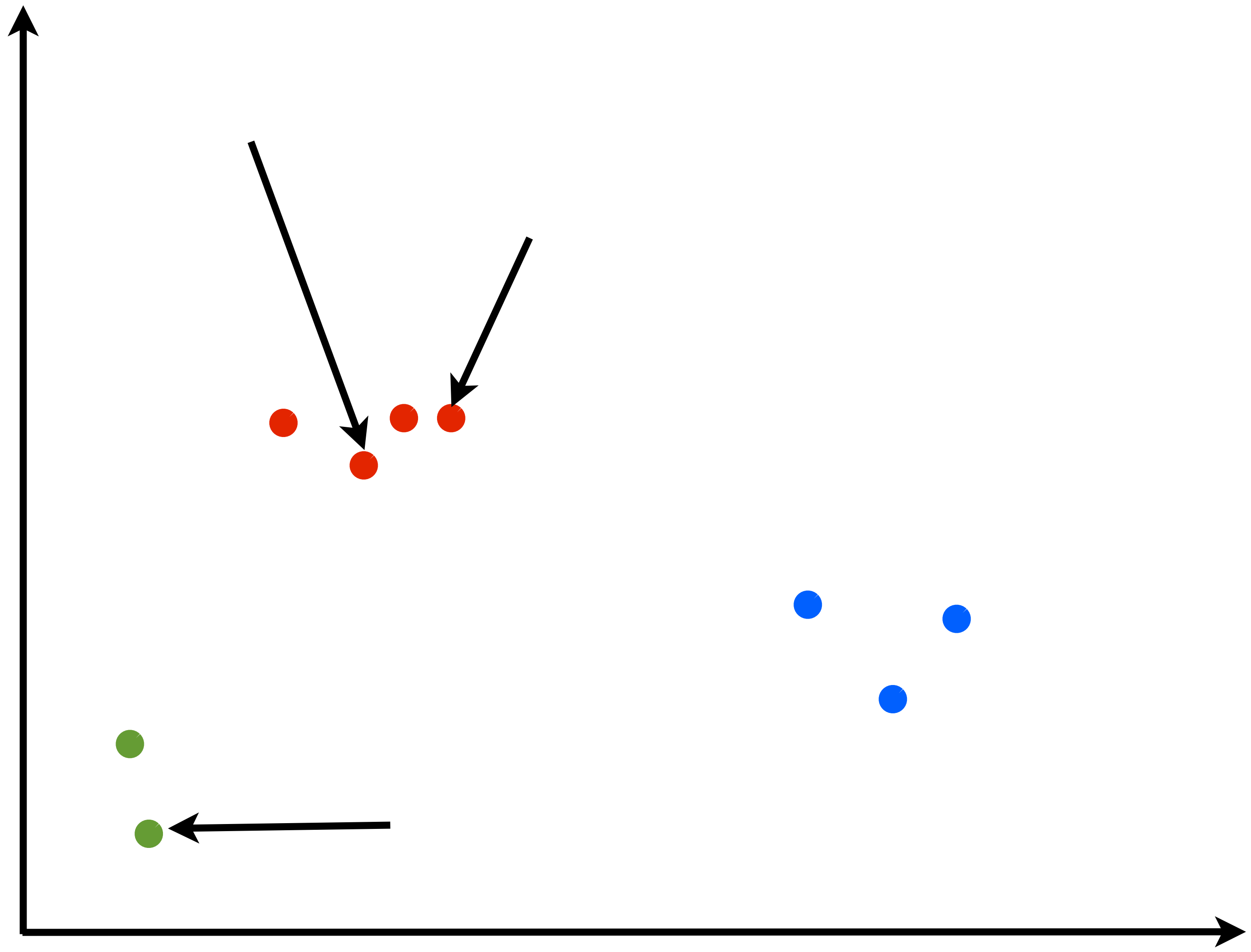
Exchangeability



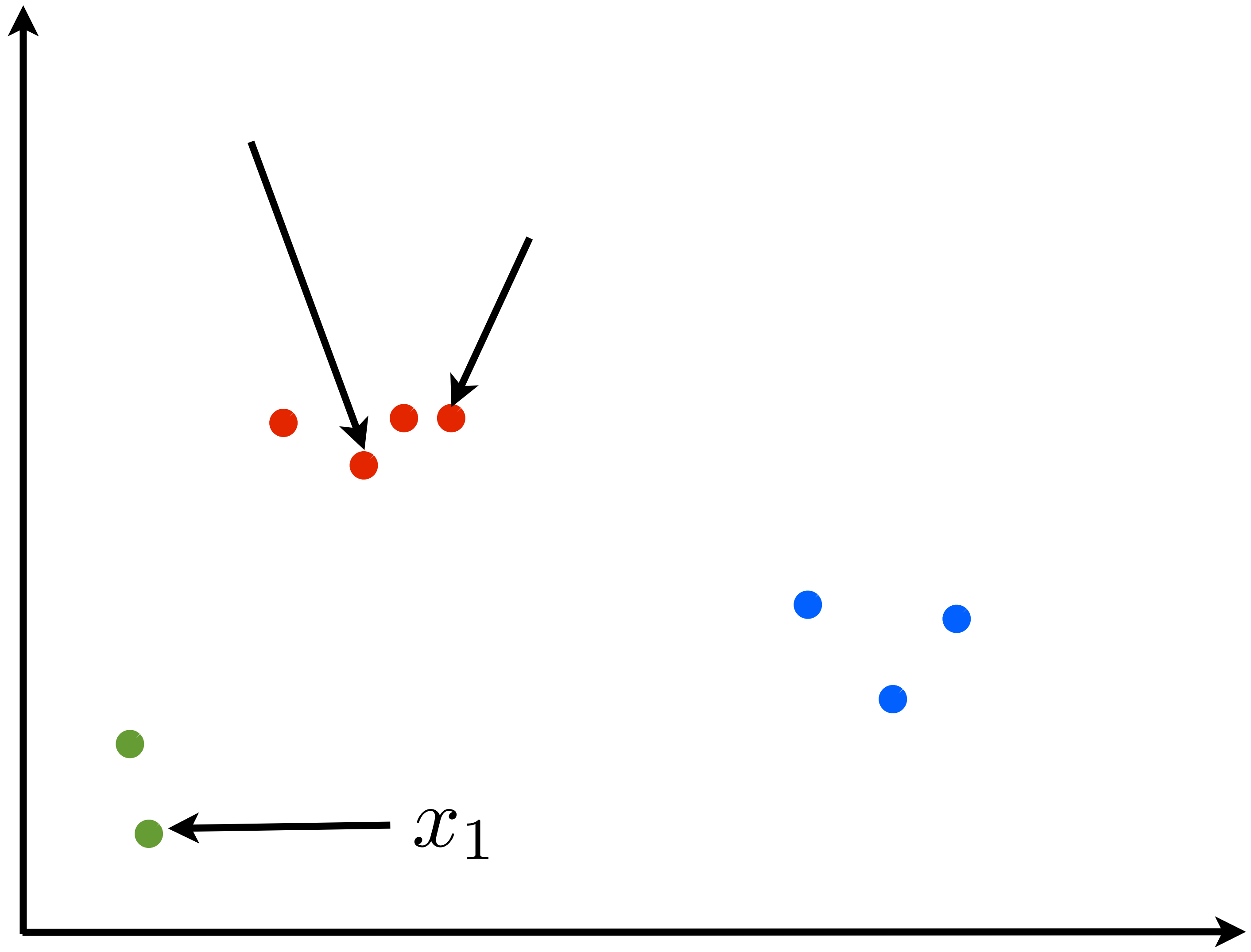
Exchangeability



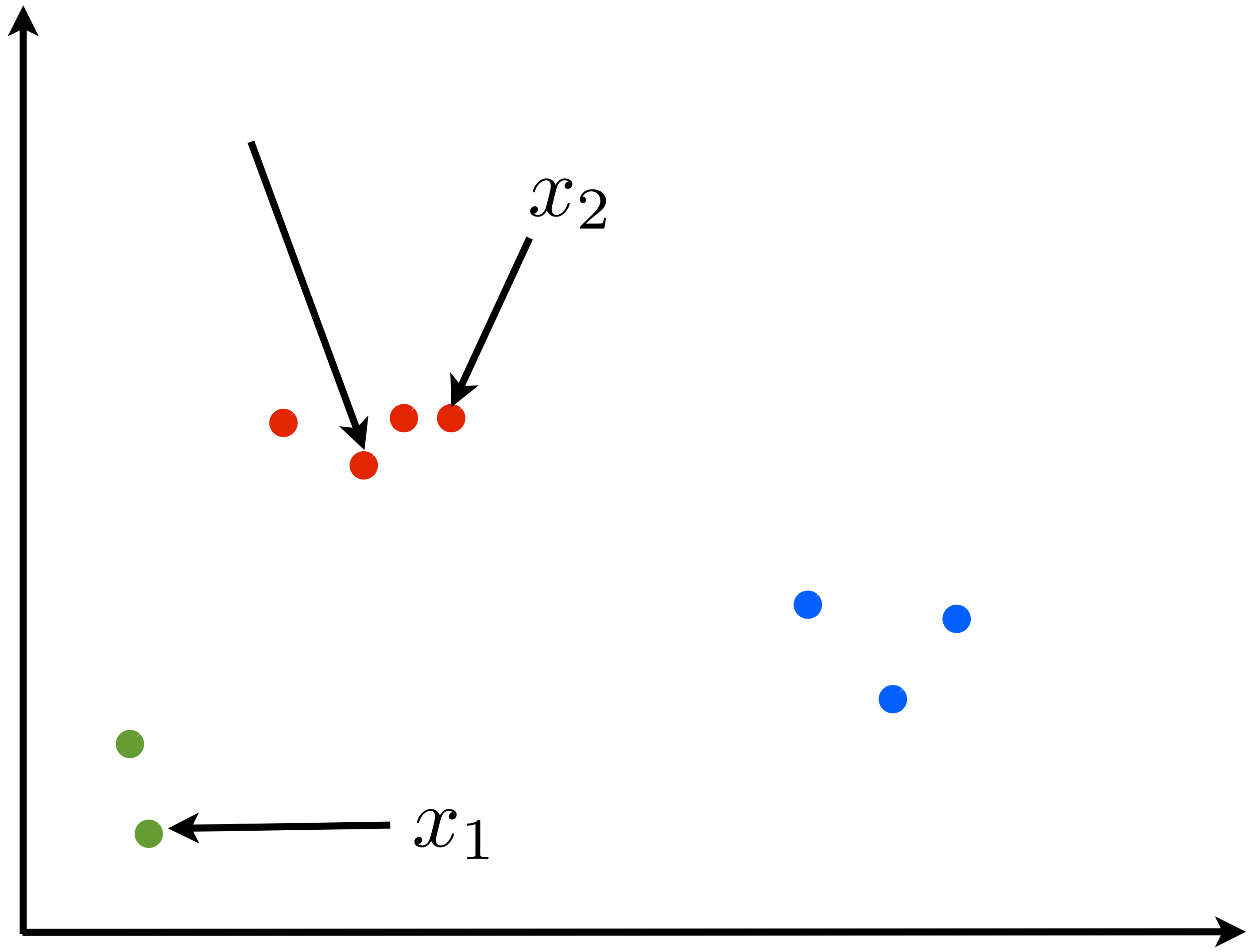
Exchangeability



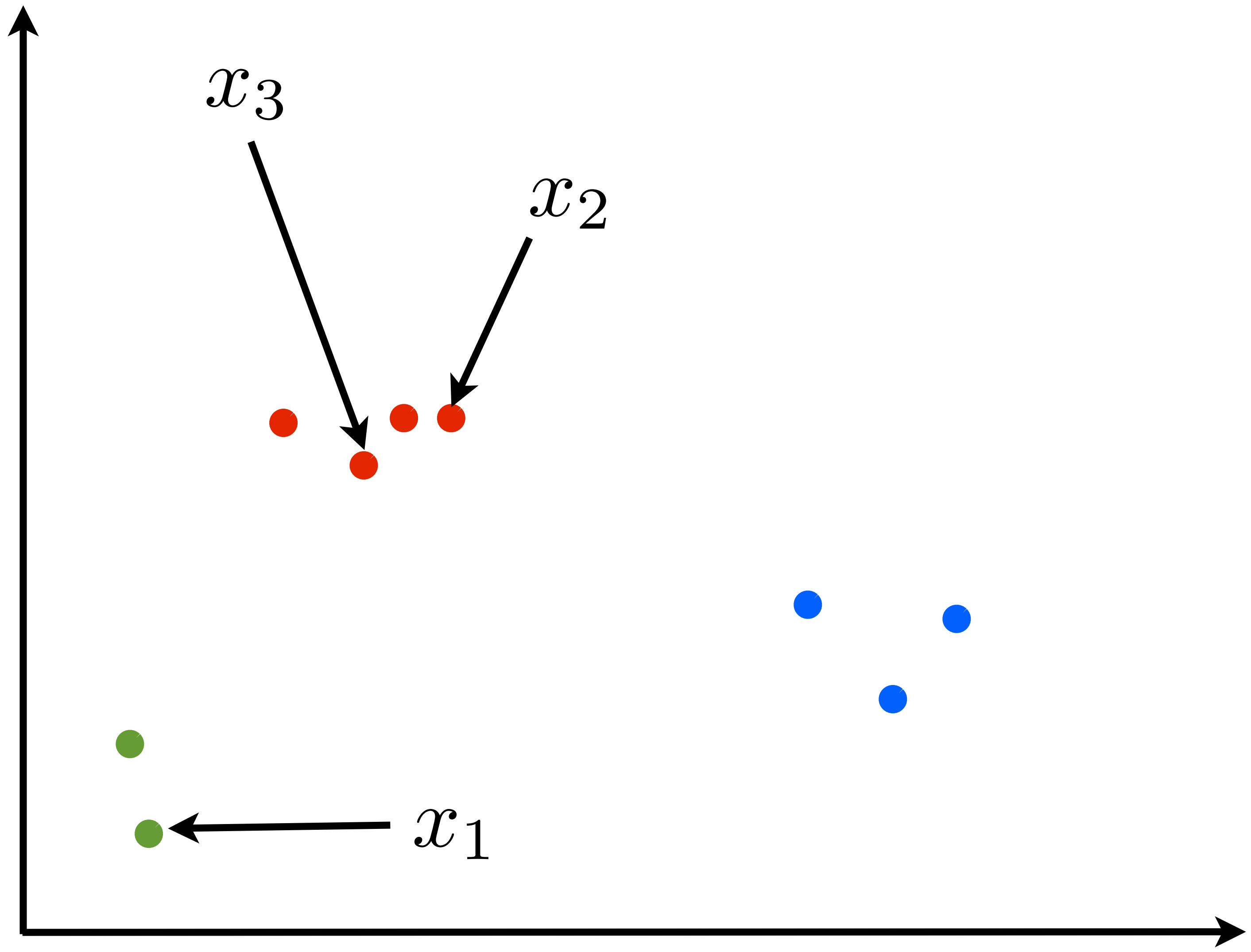
Exchangeability



Exchangeability



Exchangeability



Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

- Exchangeable

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9,2,7,1\}, \{8,4,6\}, \{5,3\}\}$$

$$\mathbb{P}(\Pi_9 = \pi_9) = \mathbb{P}(\Pi_9 = \pi_9')$$

$$\pi_9' = \{\{1,3,8,2\}, \{9,5,7\}, \{6,4\}\}$$

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9,2,7,1\},\{8,4,6\},\{5,3\}\}$$

- Exchangeable

$$\mathbb{P}(\Pi_9 = \pi_9) = \mathbb{P}(\Pi_9 = \pi_9')$$

$$\pi_9' = \{\{1,3,8,2\},\{9,5,7\},\{6,4\}\}$$

- (Almost surely) consistent sequence of partitions

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

- Exchangeable

$$\mathbb{P}(\Pi_9 = \pi_9) = \mathbb{P}(\Pi_9 = \pi_9')$$

$$\pi_9' = \{\{1, 3, 8, 2\}, \{9, 5, 7\}, \{6, 4\}\}$$

- (Almost surely) consistent sequence of partitions

$$\pi_{10} = \{\{9, 2, 7, 1\}, \{8, 4, 6, 10\}, \{5, 3\}\}$$

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

- Exchangeable

$$\mathbb{P}(\Pi_9 = \pi_9) = \mathbb{P}(\Pi_9 = \pi_9')$$

$$\pi_9' = \{\{1, 3, 8, 2\}, \{9, 5, 7\}, \{6, 4\}\}$$

- (Almost surely) consistent sequence of partitions

$$\pi_{10} = \{\{9\}, \{2\}, \{7\}, \{1, 10\} \{8\}, \{4, 5\}, \{6, 3\}\}$$

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

- Exchangeable

- (Almost surely) consistent sequence of partitions

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

$$\mathbb{P}(\Pi_9 = \pi_9) = \mathbb{P}(\Pi_9 = \pi_9')$$

$$\pi_9' = \{\{1, 3, 8, 2\}, \{9, 5, 7\}, \{6, 4\}\}$$

$$\pi_{10} = \{\{9\}, \{2\}, \{7\}, \{1, 10\} \{8\}, \{4, 5\}, \{6, 3\}\}$$

Clustering

Random partition

$$\mathbb{P}(\Pi_N = \pi_N)$$

Partition of 1, 2, ..., 9

$$\pi_9 = \{\{9,2,7,1\}, \{8,4,6\}, \{5,3\}\}$$

- Exchangeable

$$\mathbb{P}(\Pi_9 = \pi_9) = \mathbb{P}(\Pi_9 = \pi_9')$$

$$\pi_9' = \{\{1,3,8,2\}, \{9,5,7\}, \{6,4\}\}$$

- (Almost surely) consistent sequence of partitions

$$\pi_{10} = \{\{9,2,7,10,1\}, \{8,4,6\}, \{5,3\}\}$$

Clustering

- What does $\mathbb{P}(\Pi_N = \pi_N)$ look like?

Clustering

- What does $\mathbb{P}(\Pi_N = \pi_N)$ look like?
- Take any partition $\pi_N = \{A_1, A_2, \ldots, A_K\}$

Clustering

- What does $\mathbb{P}(\Pi_N = \pi_N)$ look like?
- Take any partition $\pi_N = \{A_1, A_2, \ldots, A_K\}$

$$\mathbb{P}(\Pi_N = \pi_N) = p(|A_1|, |A_2|, \ldots, |A_K|)$$

[Pitman 1995]

Clustering

- What does $\mathbb{P}(\Pi_N = \pi_N)$ look like?
- Take any partition $\pi_N = \{A_1, A_2, \ldots, A_K\}$

$$\mathbb{P}(\Pi_N = \pi_N) = p(|A_1|, |A_2|, \ldots, |A_K|)$$

p: symmetric in its arguments

[Pitman 1995]

Clustering

- What does $\mathbb{P}(\Pi_N = \pi_N)$ look like?
- Take any partition $\pi_N = \{A_1, A_2, \ldots, A_K\}$

$$\mathbb{P}(\Pi_N = \pi_N) = p(|A_1|, |A_2|, \ldots, |A_K|)$$

p: symmetric in its arguments

"Exchangeable partition probability function" (EPPF)

[Pitman 1995]

Outline

I. Clusters

- Overview
- Distribution
  - ◇ Clusters
  - ◇ Data given clusters
  - ◇ Posterior
- Proportions
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
    - ◇ Clusters (Example: Chinese restaurant process)
    - ◇ Data given clusters
    - ◇ Posterior
- Proportions
- Random probability measure

II. Features

EPPF Example

EPPF Example

Chinese restaurant process

[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition

[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition

[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

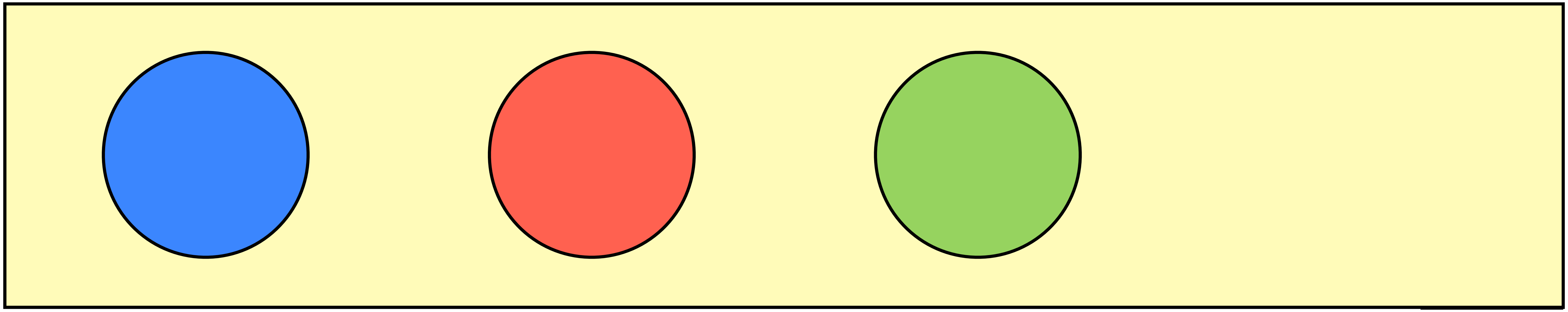
- Restaurant ⇔ partition
  - Table ⇔ cluster

[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition
  - Table ⇔ cluster

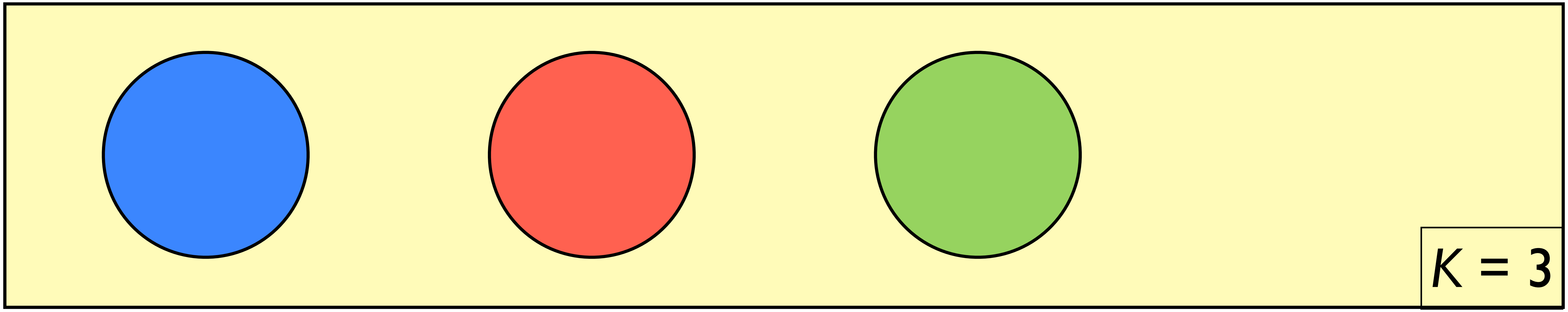


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition
    - Table ⇔ cluster

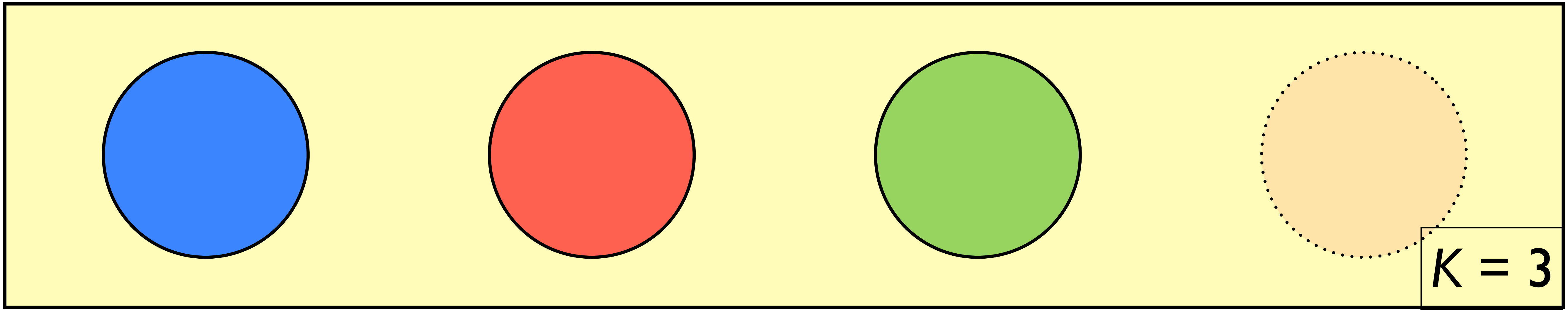


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition
    - Table ⇔ cluster

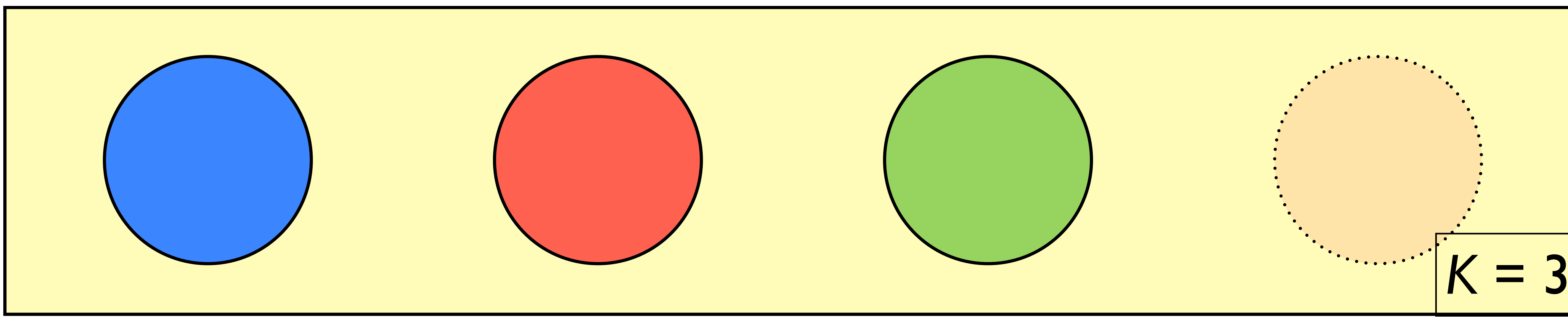


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition
  - Table ⇔ cluster
  - Customer ⇔ index

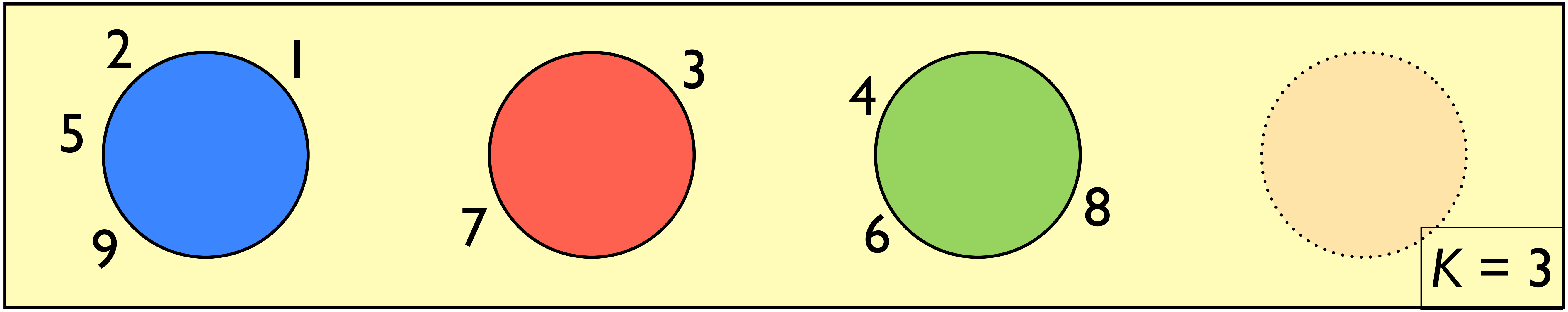


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition
  - Table ⇔ cluster
  - Customer ⇔ index

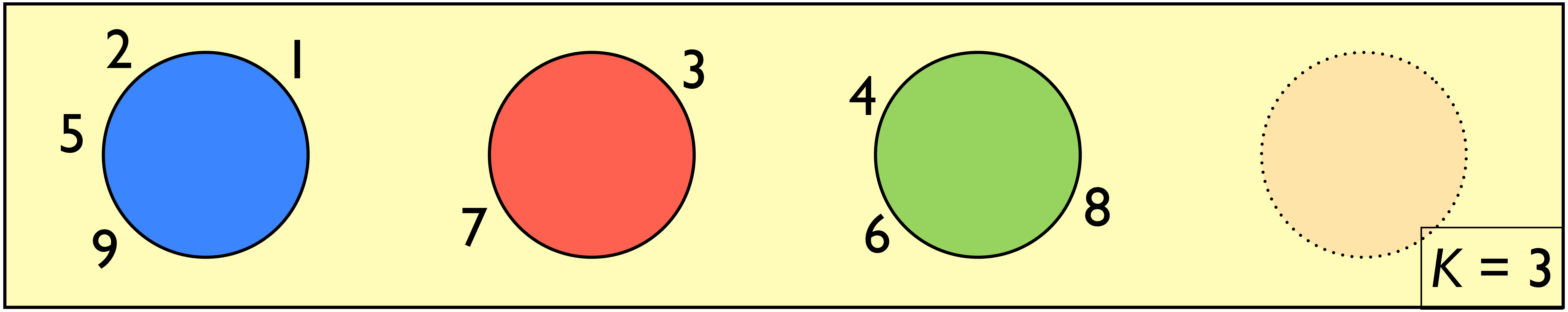


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Restaurant ⇔ partition
  - Table ⇔ cluster
  - Customer ⇔ index

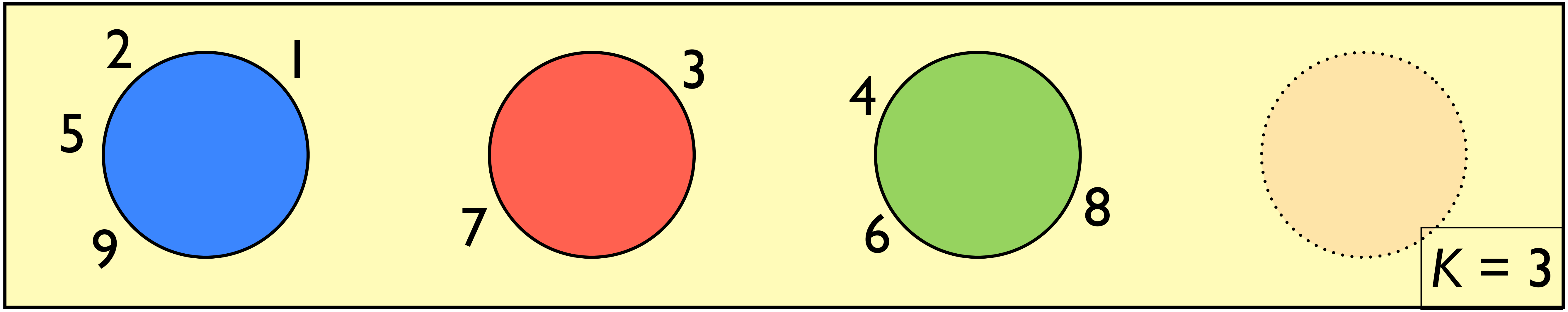


$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

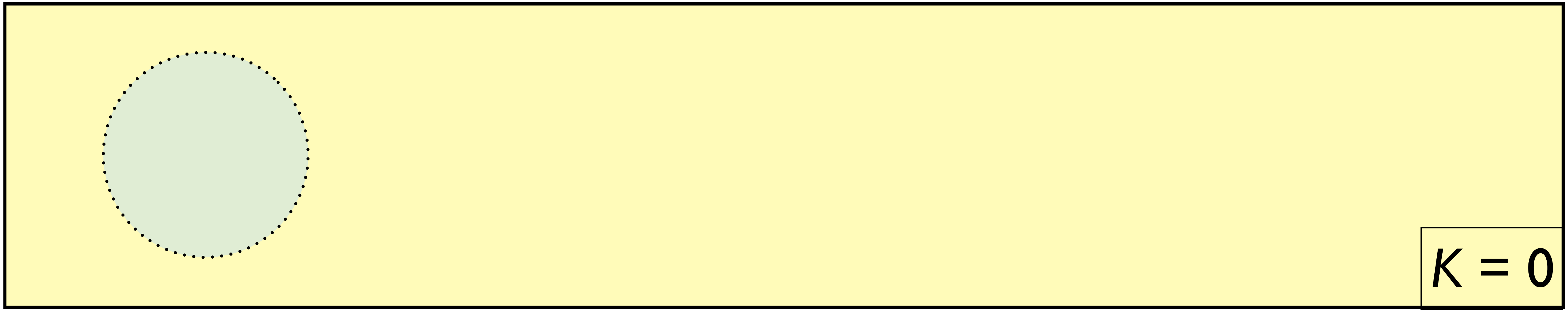
Chinese restaurant process



[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

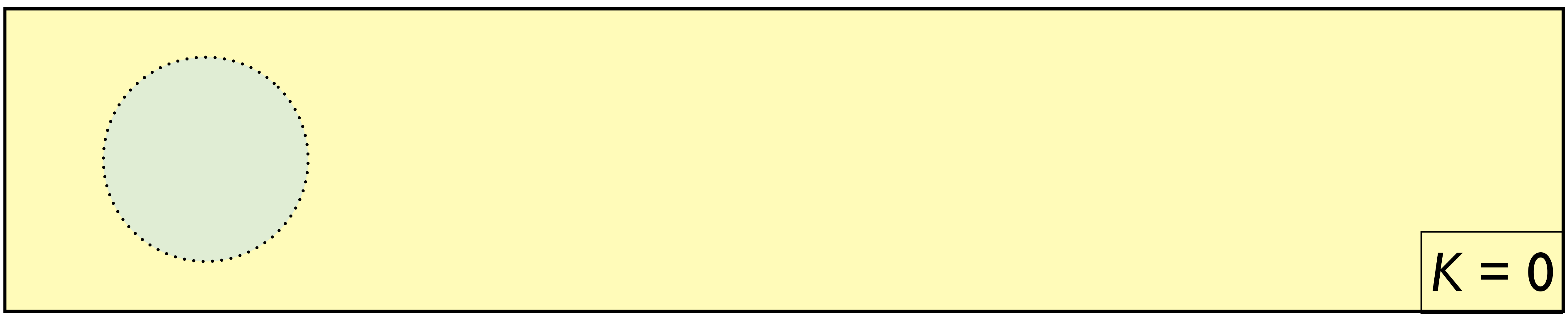


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Customers prefer popular tables

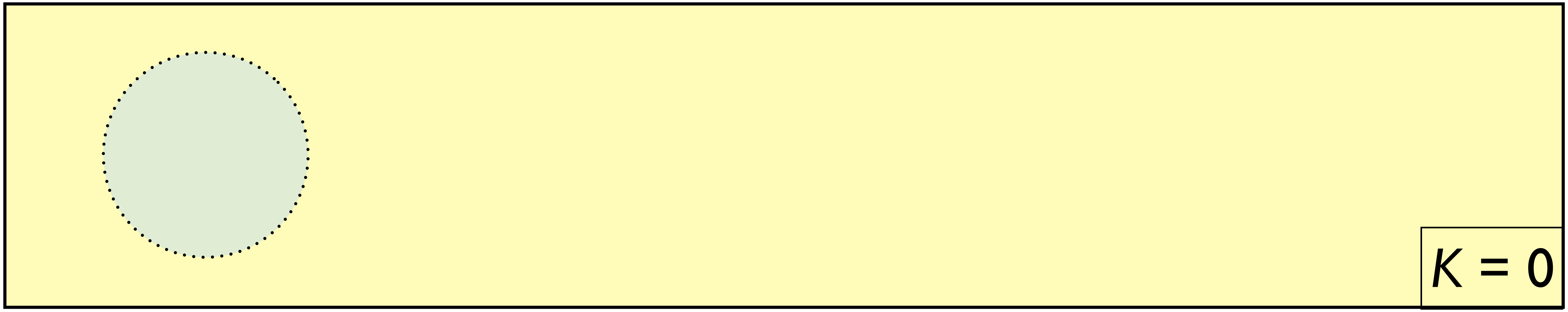


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits

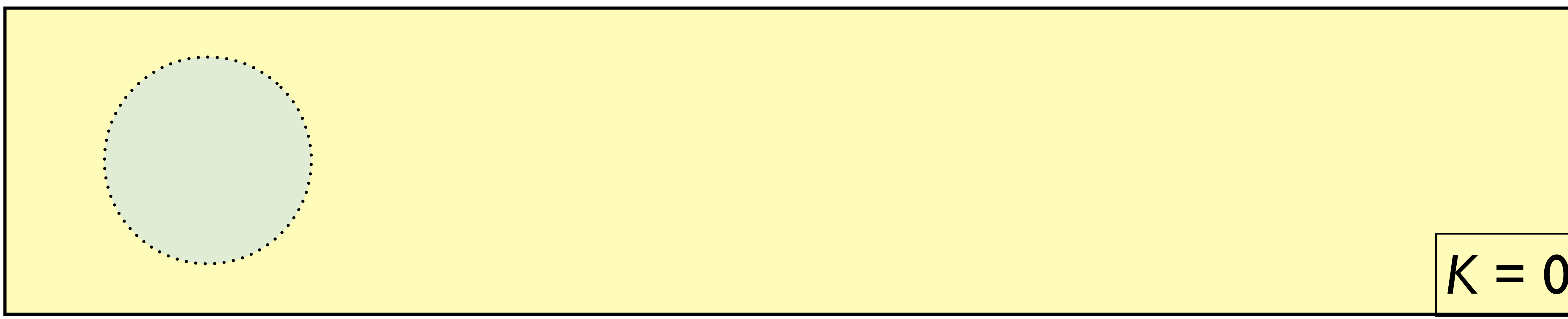


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$

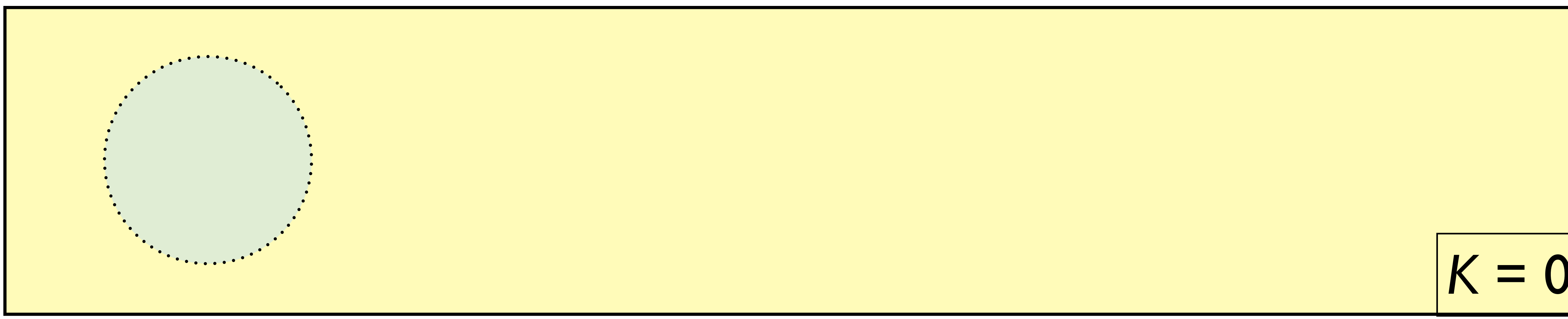


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$



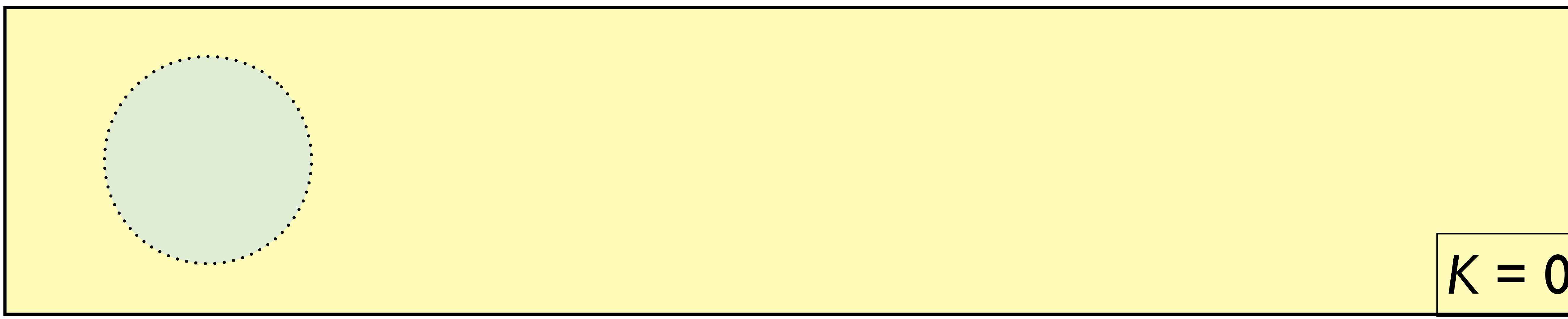
[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

concentration parameter

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+1 with probability $\propto \theta$

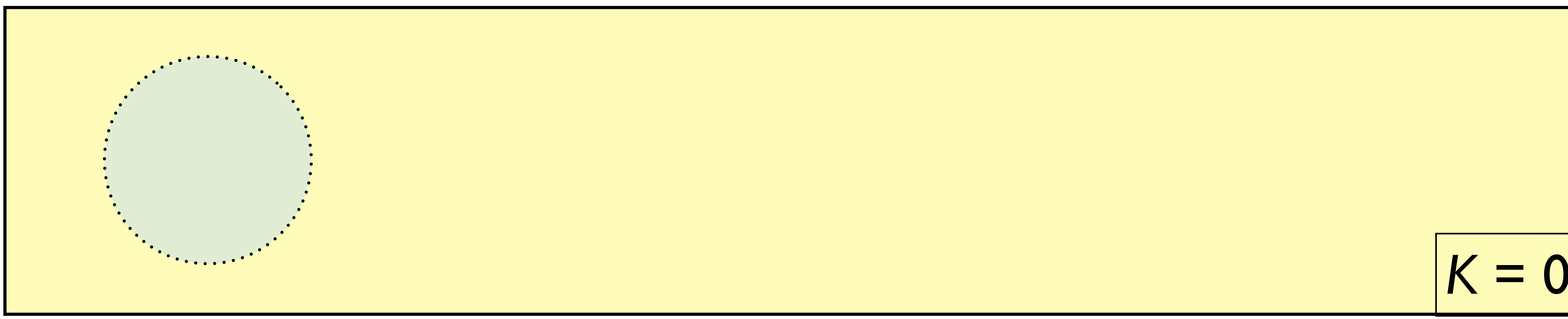


[Blackwell, MacQueen 1973; Aldous 1985]

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

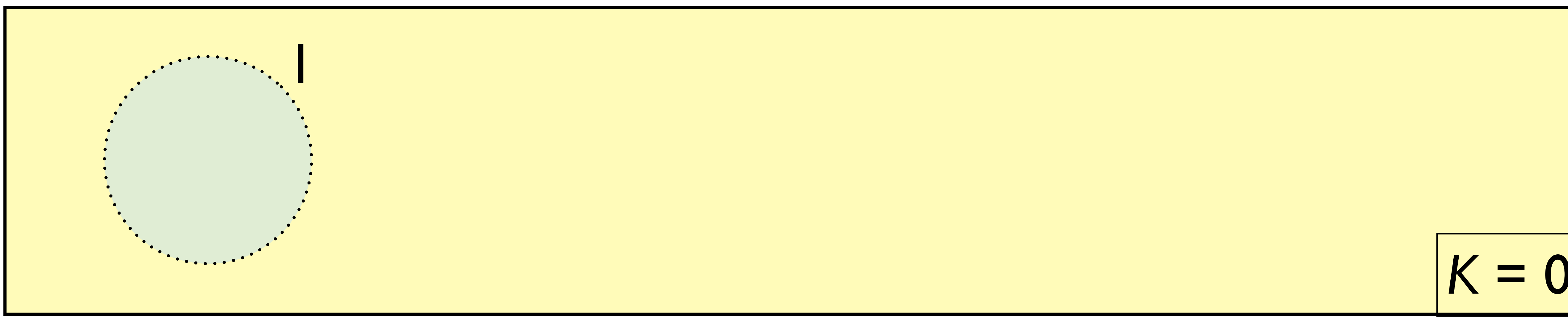


$$\frac{\theta}{\theta}$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

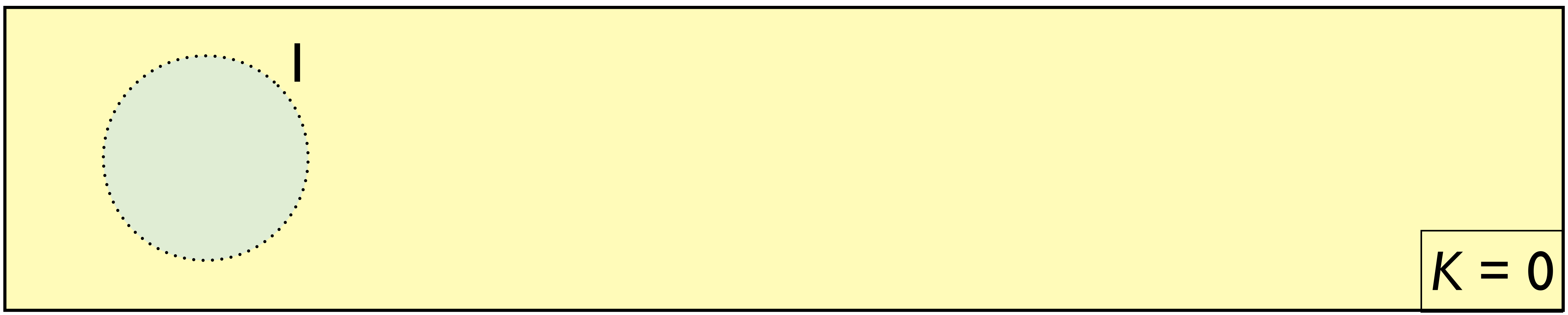


$$\frac{\theta}{\theta}$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$



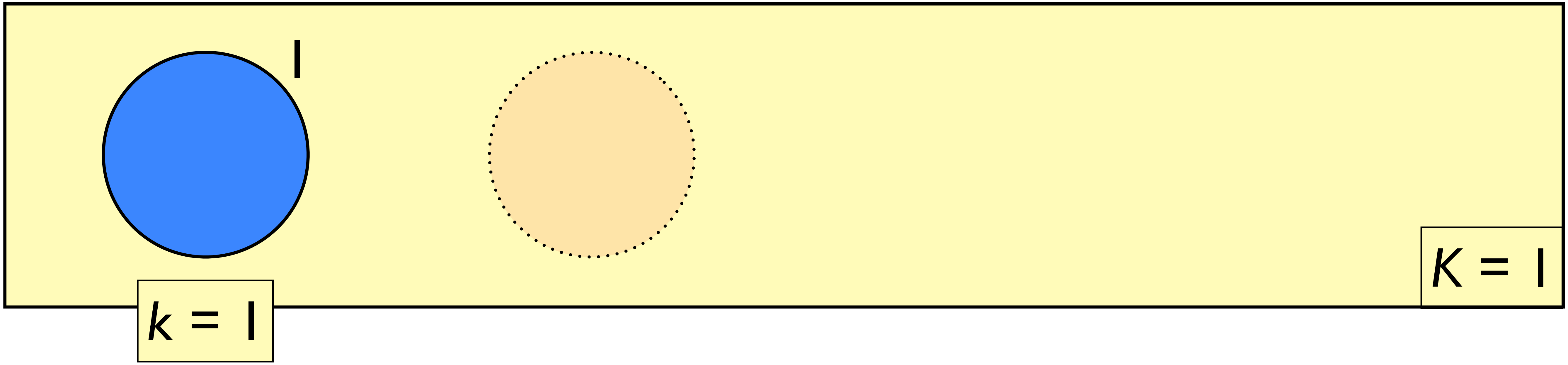
$$\frac{\theta}{\theta}$$

$$\mathbb{P}(\Pi_1 = \pi_1) = 1$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

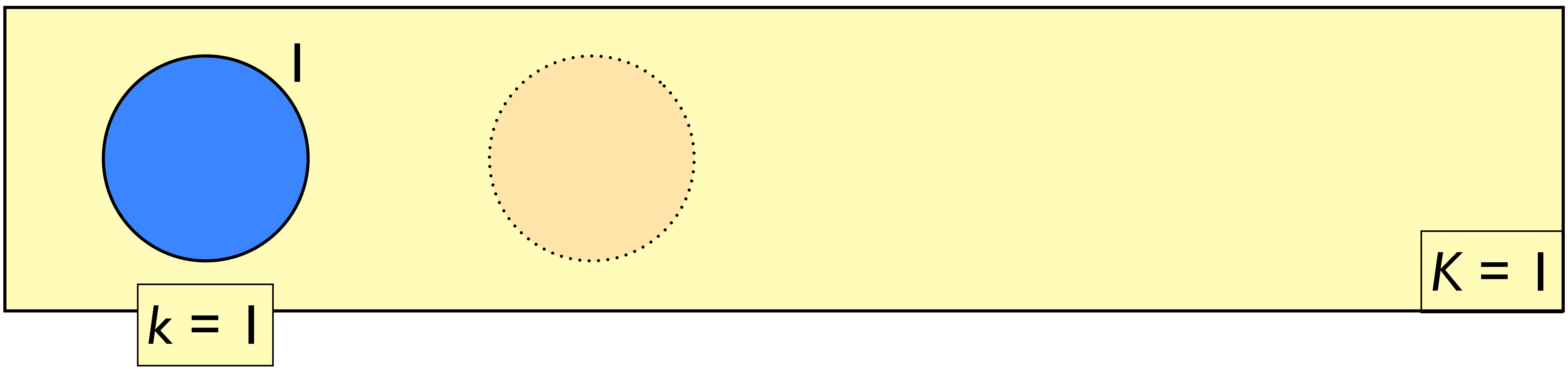


$$\mathbb{P}(\Pi_1 = \pi_1) = 1$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$



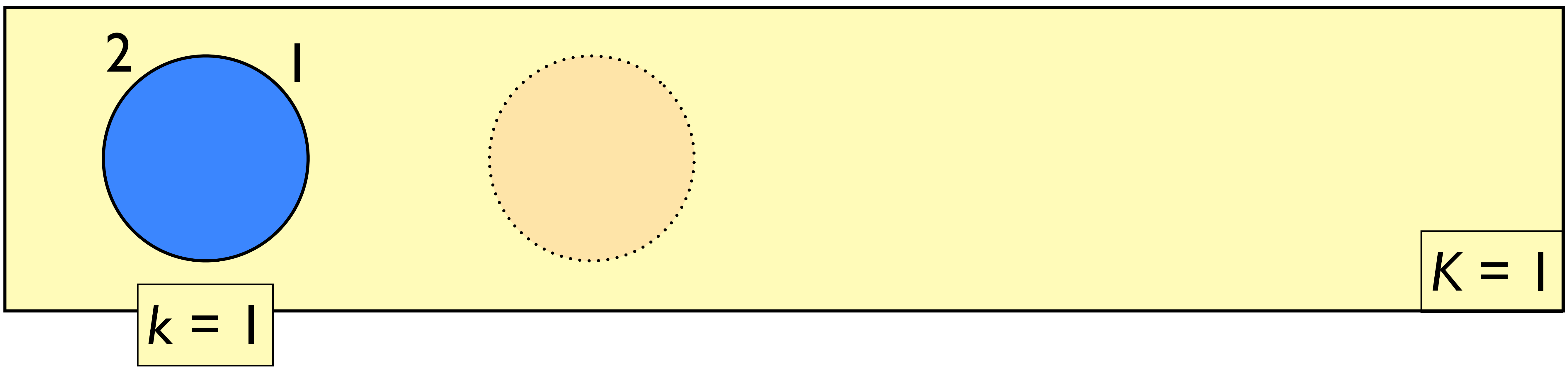
$$\frac{1}{1+\theta} \qquad \frac{\theta}{1+\theta}$$

$$\mathbb{P}(\Pi_1 = \pi_1) = 1$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$



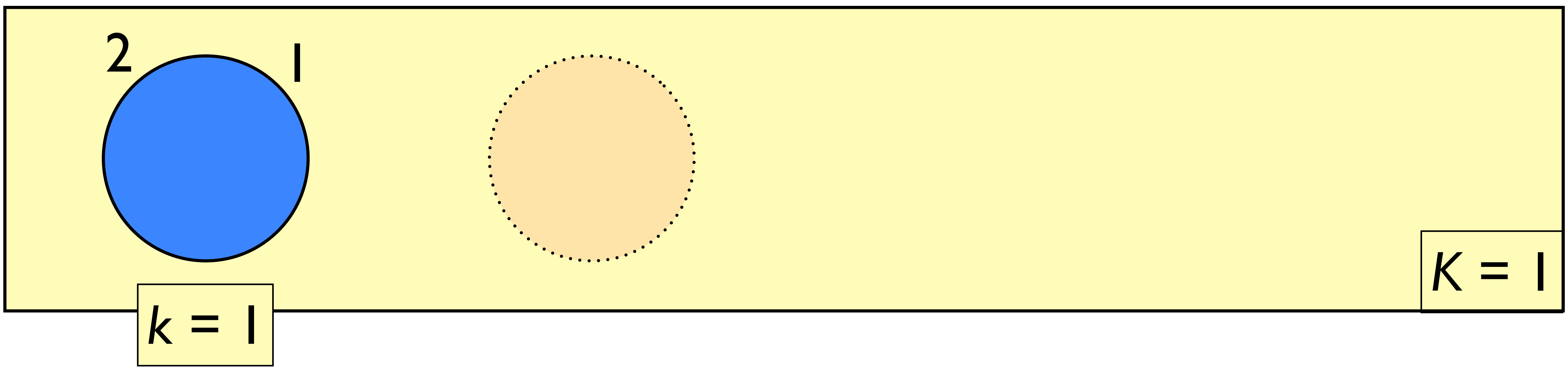
$\frac{1}{1+\theta}$ $\frac{\theta}{1+\theta}$

$\mathbb{P}(\Pi_1 = \pi_1) = 1$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$



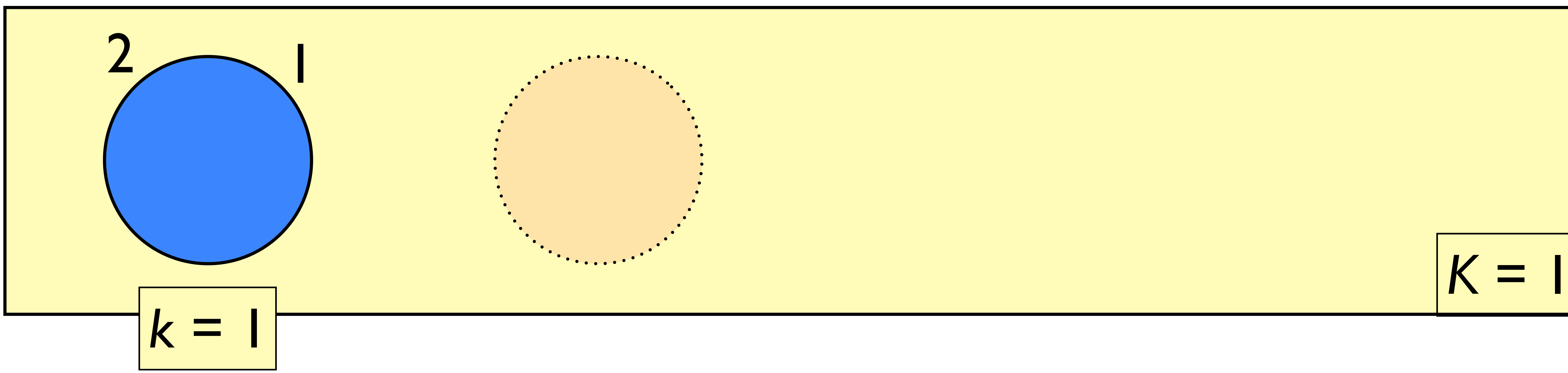
$$\frac{1}{1+\theta} \qquad \frac{\theta}{1+\theta}$$

$$\mathbb{P}(\Pi_2 = \pi_2) = \frac{1}{1+\theta}$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$



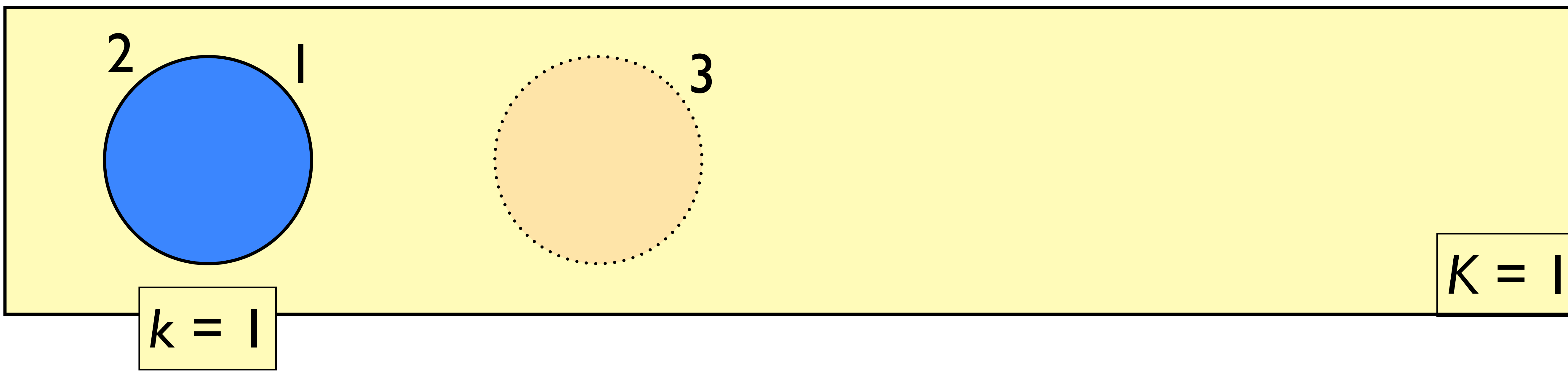
$$\frac{2}{2+\theta} \qquad \frac{\theta}{2+\theta}$$

$$\mathbb{P}(\Pi_2 = \pi_2) = \frac{1}{1+\theta}$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$



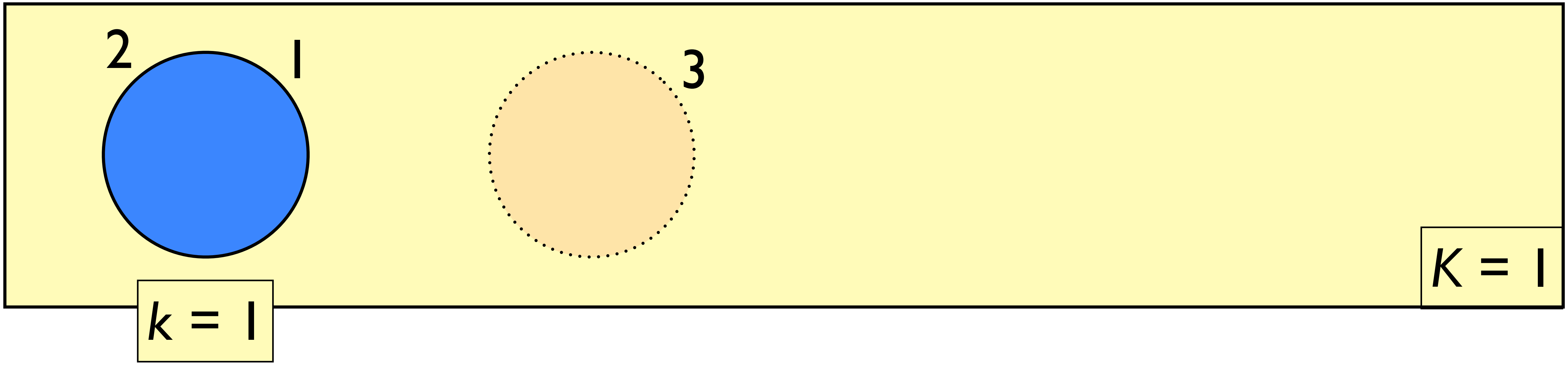
$$\frac{2}{2+\theta} \qquad \frac{\theta}{2+\theta}$$

$$\mathbb{P}(\Pi_2 = \pi_2) = \frac{1}{1+\theta}$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$



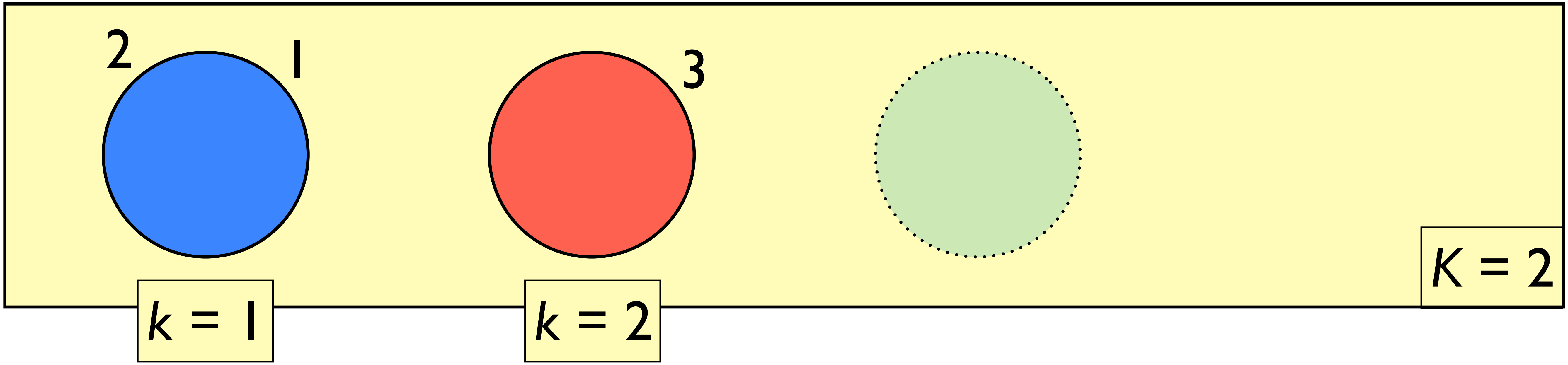
$$\frac{2}{2+\theta} \qquad\qquad \frac{\theta}{2+\theta}$$

$$\mathbb{P}(\Pi_3 = \pi_3) = \frac{1}{(1+\theta)(2+\theta)} \cdot \theta$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$

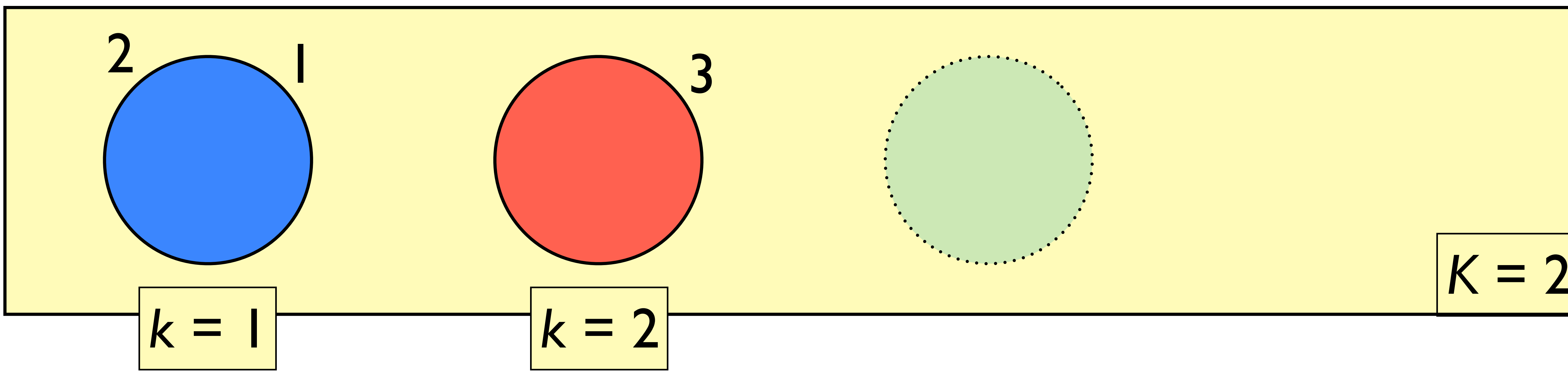


$$\mathbb{P}(\Pi_3 = \pi_3) = \frac{1}{(1+\theta)(2+\theta)} \cdot \theta$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$



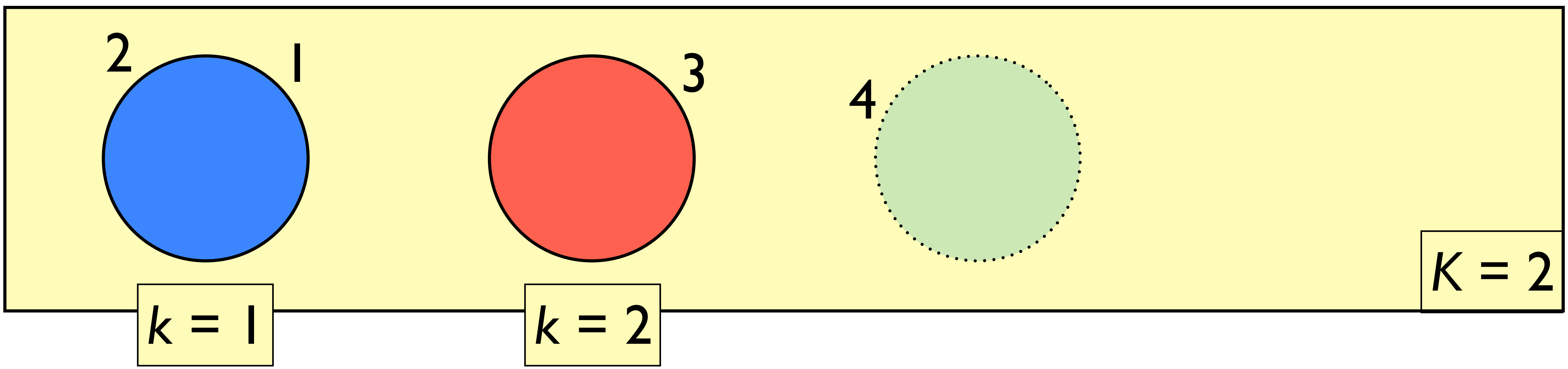
$$\frac{2}{3+\theta} \qquad \frac{1}{3+\theta} \qquad \frac{\theta}{3+\theta}$$

$$\mathbb{P}(\Pi_3 = \pi_3) = \frac{1}{(1+\theta)(2+\theta)} \cdot \theta$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$



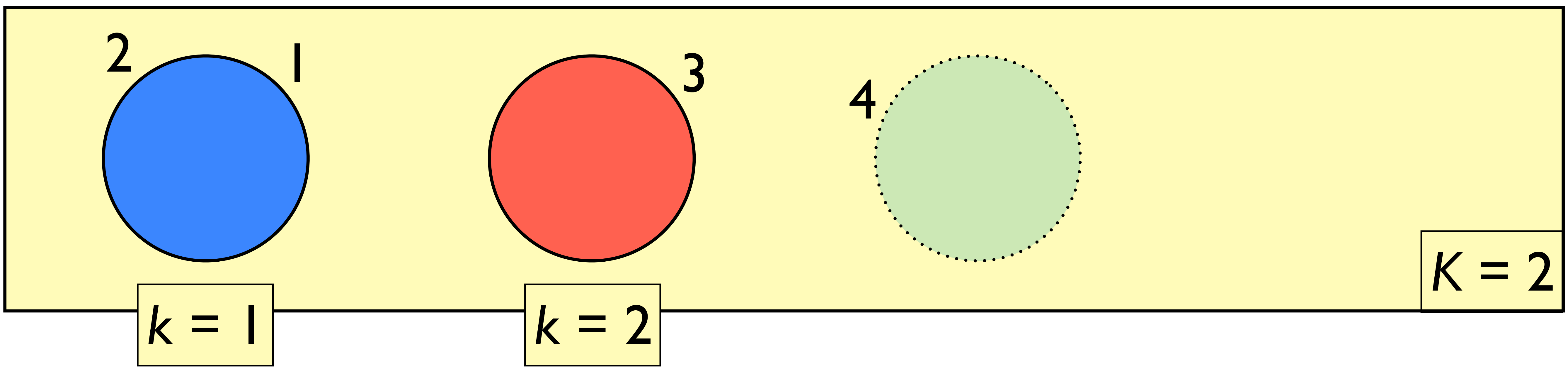
$$\frac{2}{3+\theta} \qquad \frac{1}{3+\theta} \qquad \frac{\theta}{3+\theta}$$

$$\mathbb{P}(\Pi_4 = \pi_4) = \frac{1}{(1+\theta)(2+\theta)(3+\theta)} \cdot \theta^2$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+1 with probability $\propto \theta$



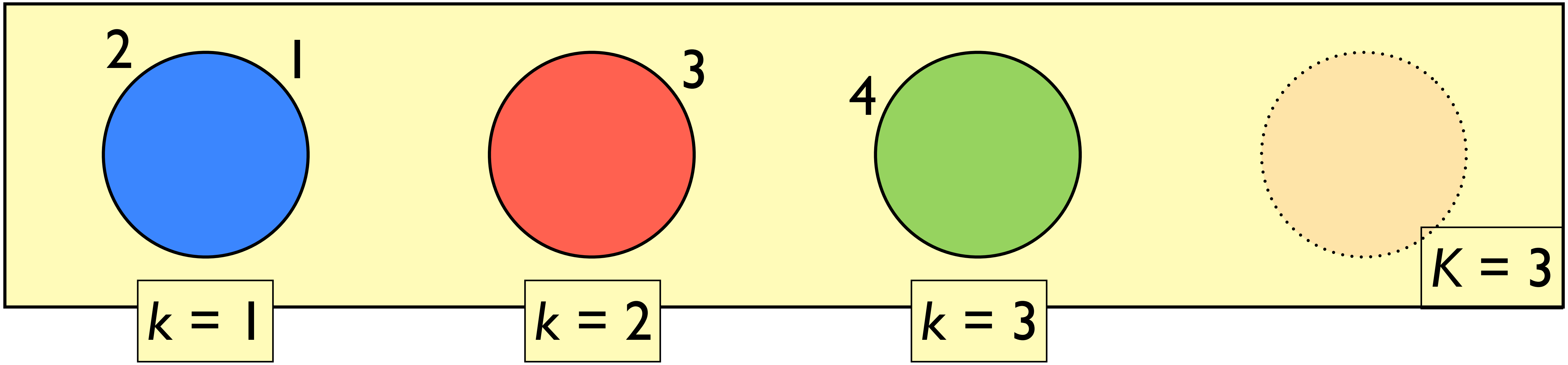
$\frac{2}{3+\theta}$ $\frac{1}{3+\theta}$ $\frac{\theta}{3+\theta}$

$$\mathbb{P}(\Pi_4 = \pi_4) = \frac{1}{\prod_{n=1}^{3}(n+\theta)} \cdot \theta^2$$

EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

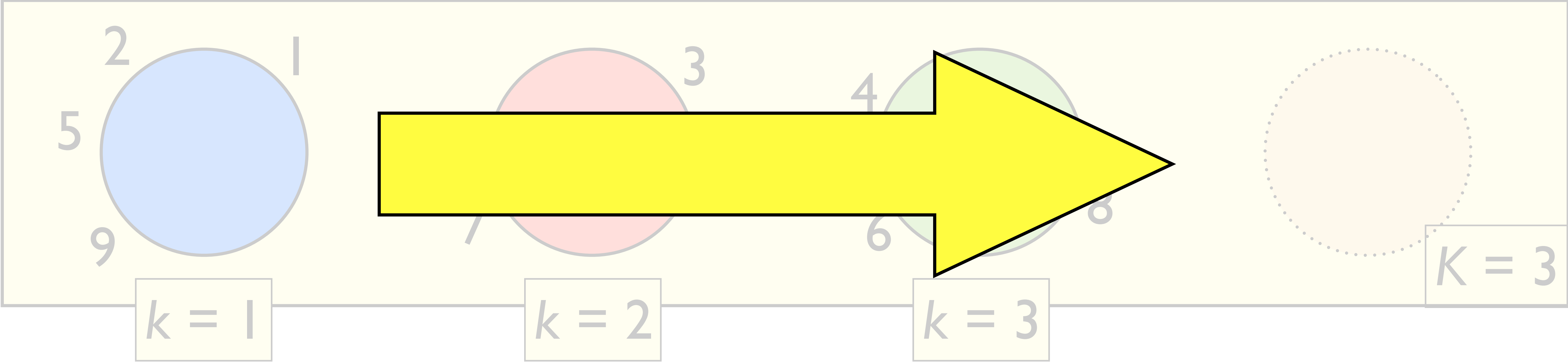


$$\mathbb{P}(\Pi_4 = \pi_4) = \frac{1}{\prod_{n=1}^{3}(n+\theta)} \cdot \theta^2$$

EPPF Example

Chinese restaurant process

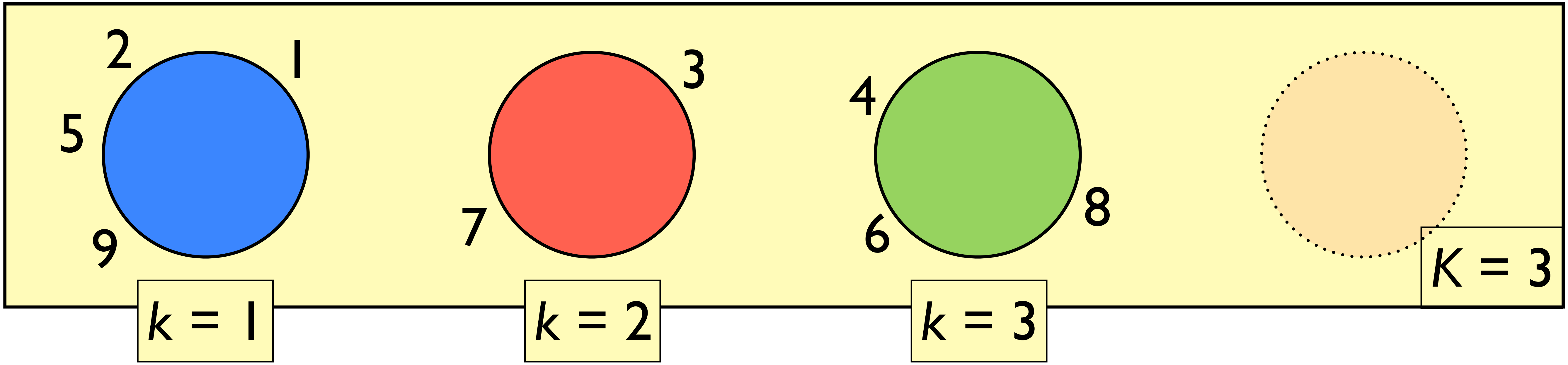
- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\ \text{people there})$
  - at new table *K*+*1* with probability $\propto \theta$



EPPF Example

Chinese restaurant process

- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$



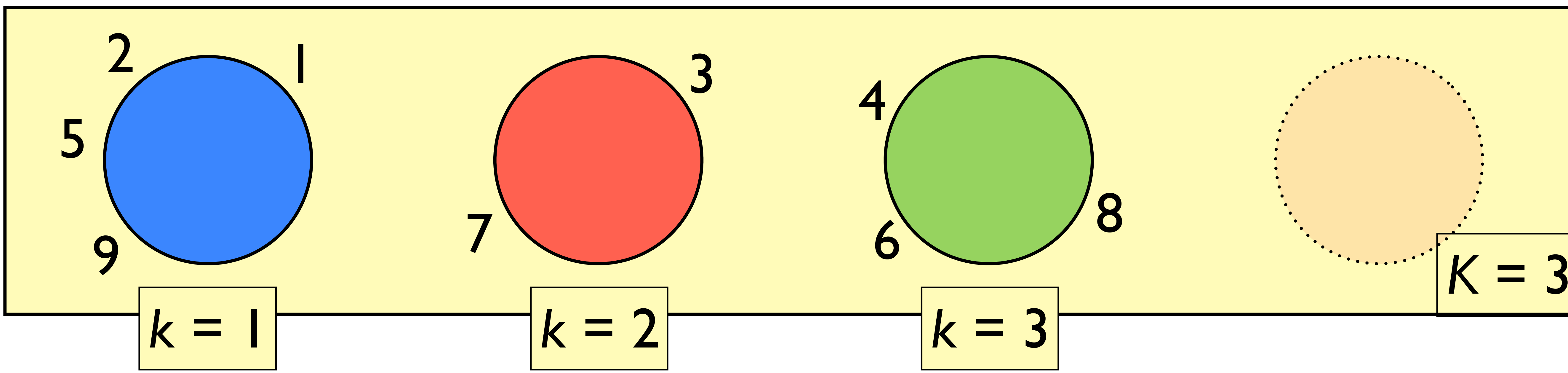
$$\mathbb{P}(\Pi_9 = \pi_9) = \frac{1}{\prod_{n=1}^{9}(n+\theta)} \cdot \theta^2 \cdot (3! \cdot 1! \cdot 2!)$$

EPPF Example

related to number of clusters

Chinese restaurant process

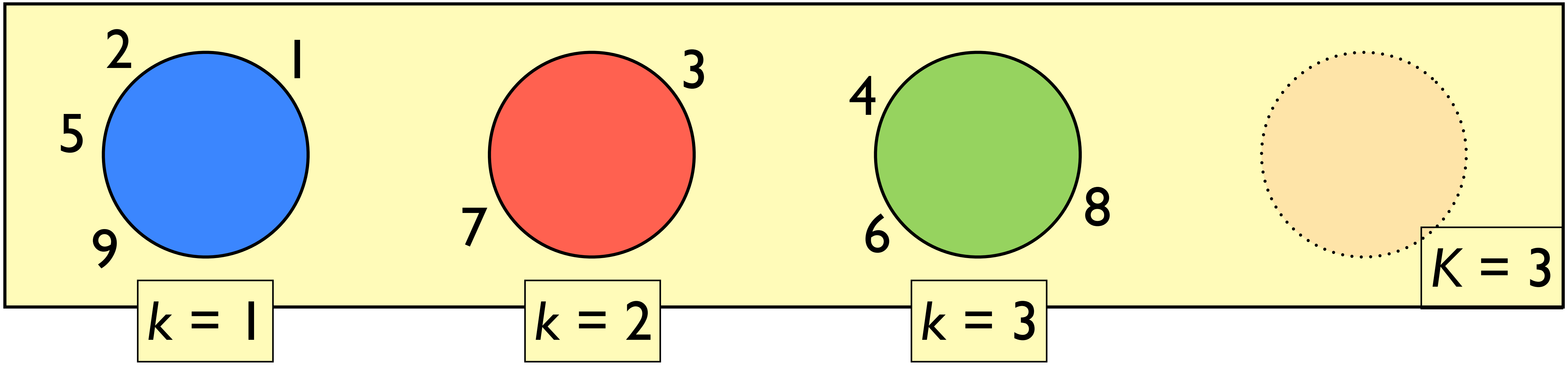
- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$



$$\mathbb{P}(\Pi_9 = \pi_9) = \frac{1}{\prod_{n=1}^{9}(n+\theta)} \cdot \theta^2 \cdot (3! \cdot 1! \cdot 2!)$$

EPPF Example

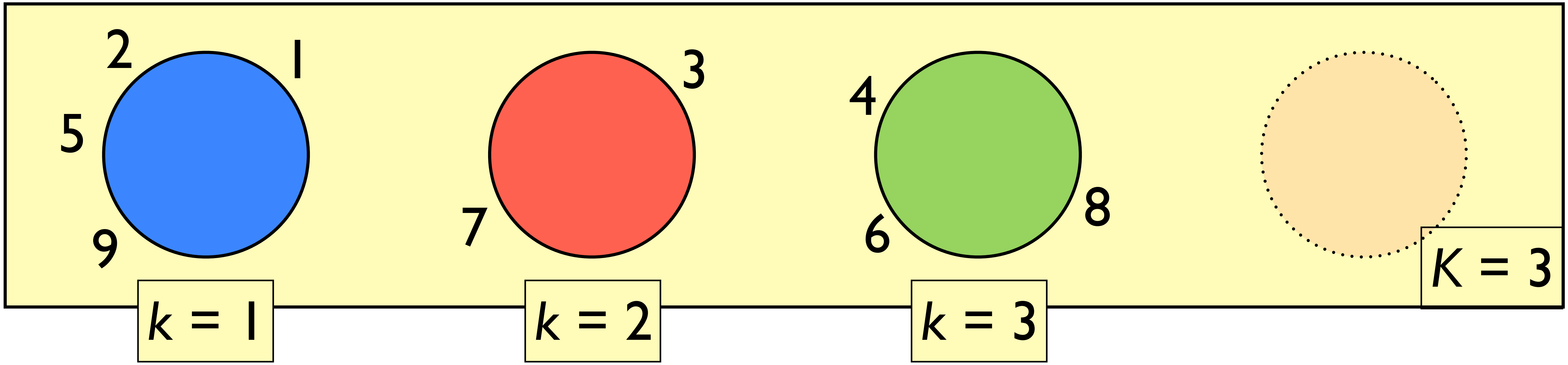
Chinese restaurant process



$$\mathbb{P}(\Pi_9 = \pi_9) = \frac{1}{\prod_{n=1}^{9}(n+\theta)} \cdot \theta^2 \cdot (3! \cdot 1! \cdot 2!)$$

EPPF Example

Chinese restaurant process

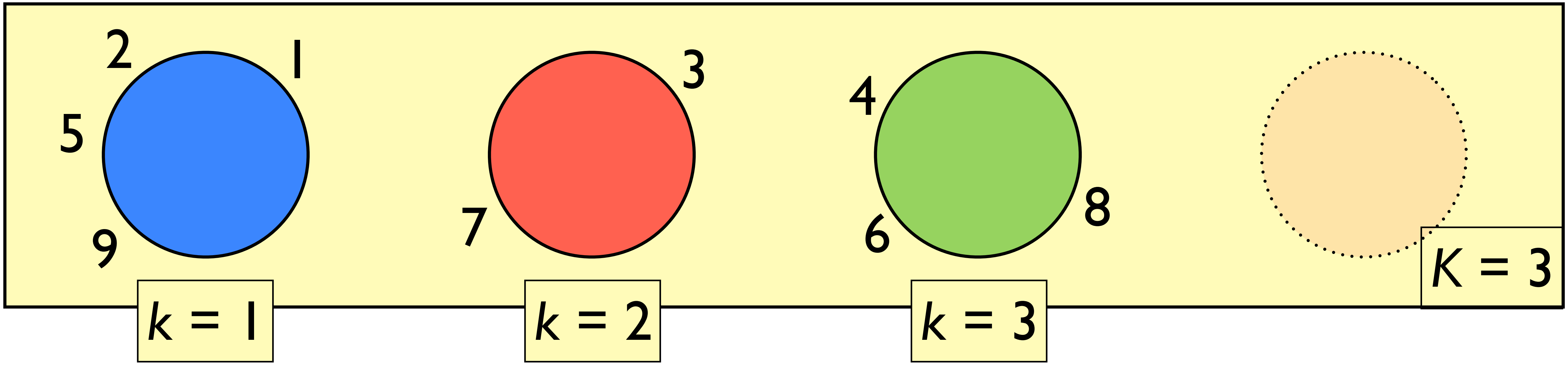


$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

$$\mathbb{P}(\Pi_9 = \pi_9) = \frac{1}{\prod_{n=1}^{9}(n+\theta)} \cdot \theta^2 \cdot (3! \cdot 1! \cdot 2!)$$

EPPF Example

Chinese restaurant process

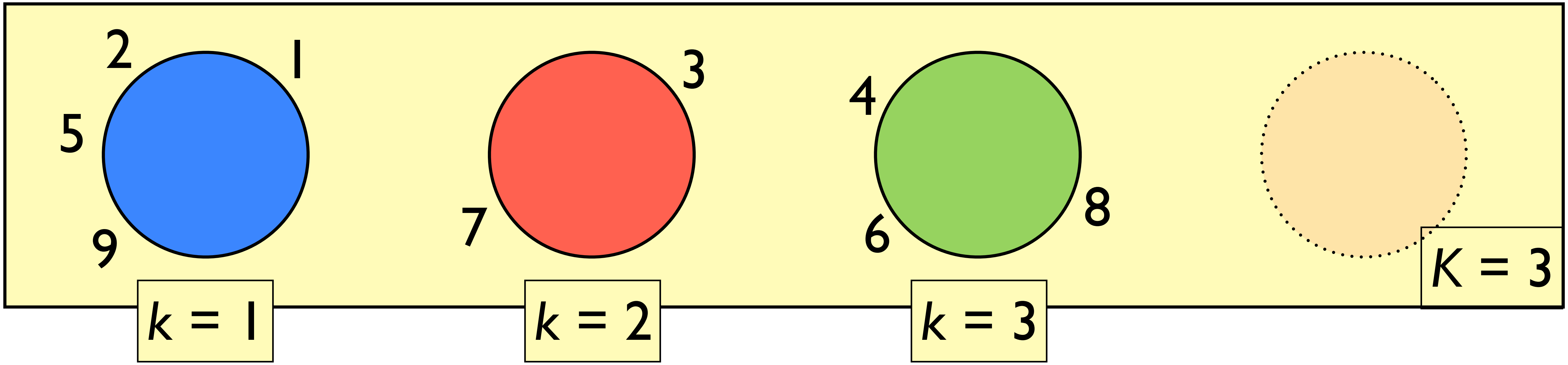


$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_9 = \pi_9) = \frac{1}{\prod_{n=1}^{9}(n+\theta)} \cdot \theta^2 \cdot (3! \cdot 1! \cdot 2!)$$

EPPF Example

Chinese restaurant process

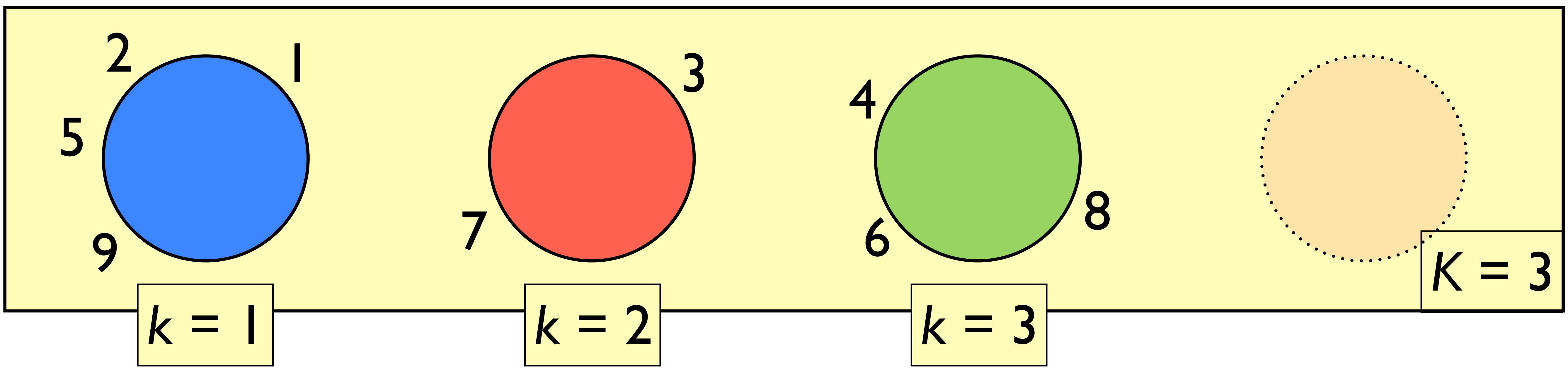


$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

EPPF Example

Chinese restaurant process



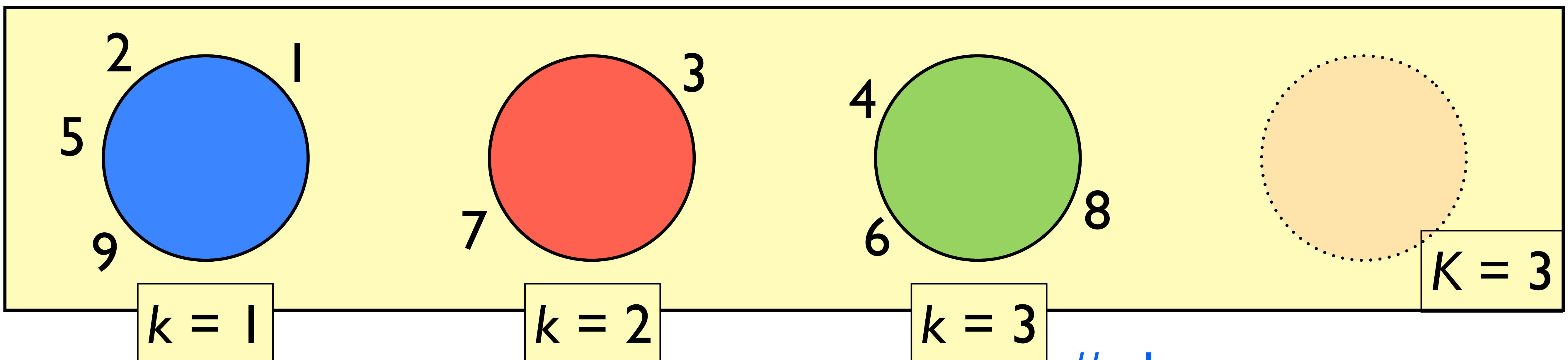
$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

size of kth cluster

EPPF Example

Chinese restaurant process



clusters

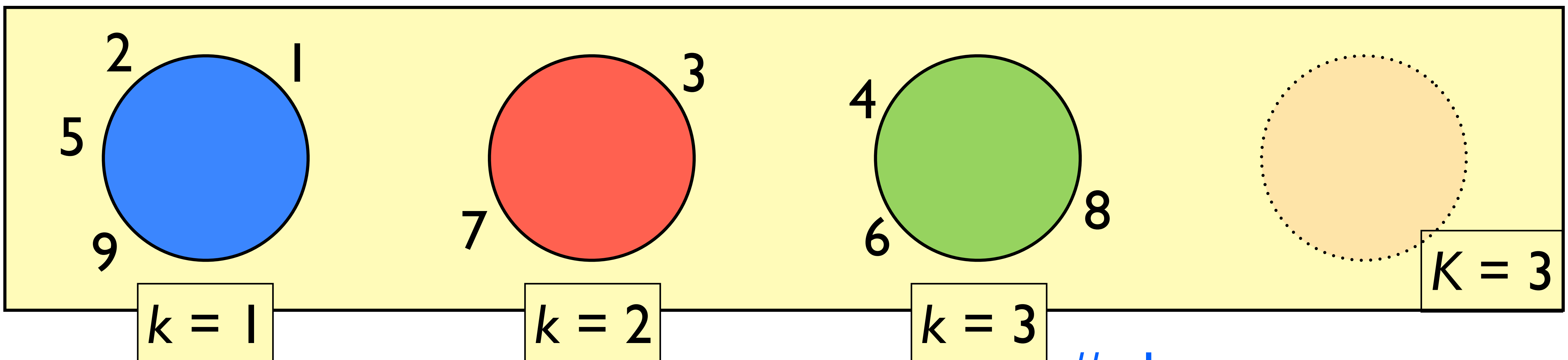
size of kth cluster

$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

EPPF Example

Chinese restaurant process



clusters

data points

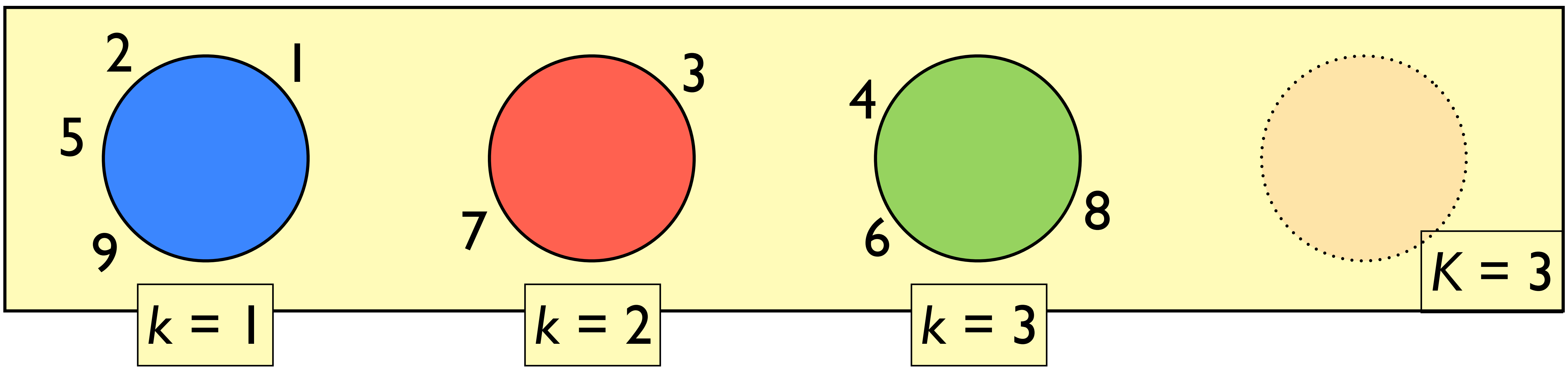
size of kth cluster

$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

EPPF Example

Chinese restaurant process

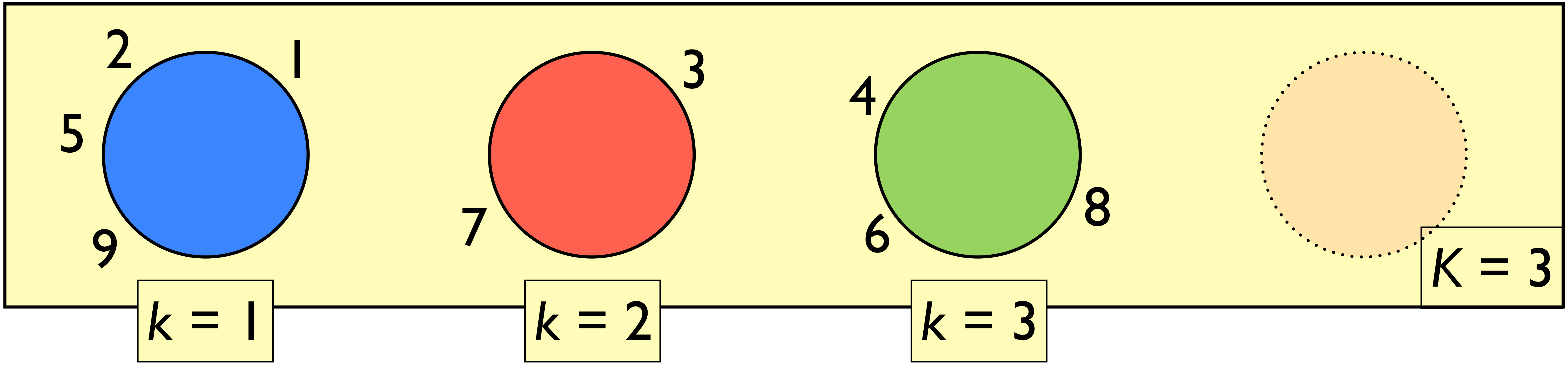


$$\pi_N = \{A_1, A_2, \ldots, A_K\} \quad \mathbb{P}(\Pi_N = \pi_N) = p(|A_1|, |A_2|, \ldots, |A_K|)$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K}(|A_k| - 1)! \quad \text{(EPPF)}$$

EPPF Example

Chinese restaurant process



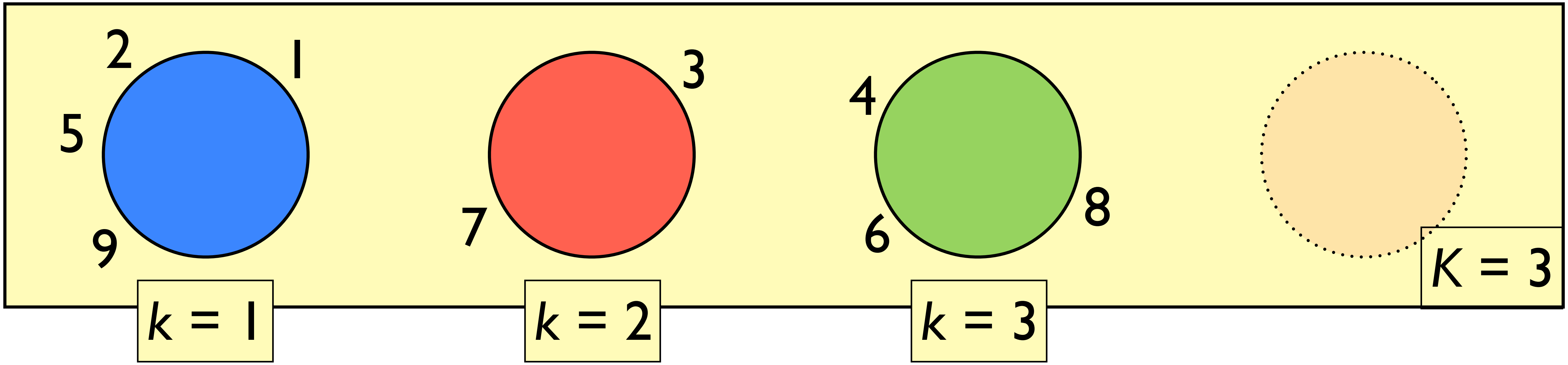
$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

EPPF Example

Chinese restaurant process

- Exchangeable



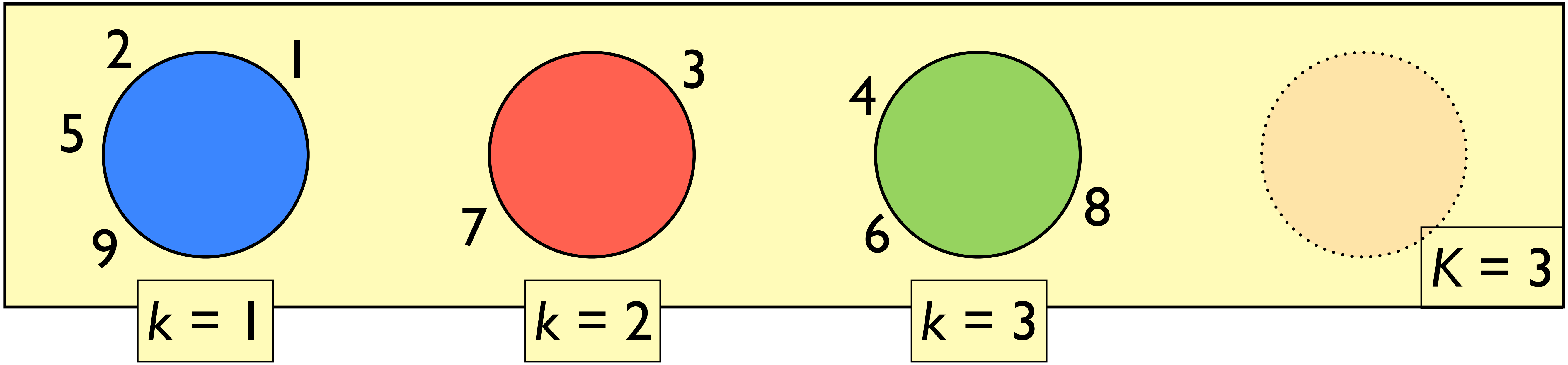
$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

EPPF Example

Chinese restaurant process

- Exchangeable
- Consistent



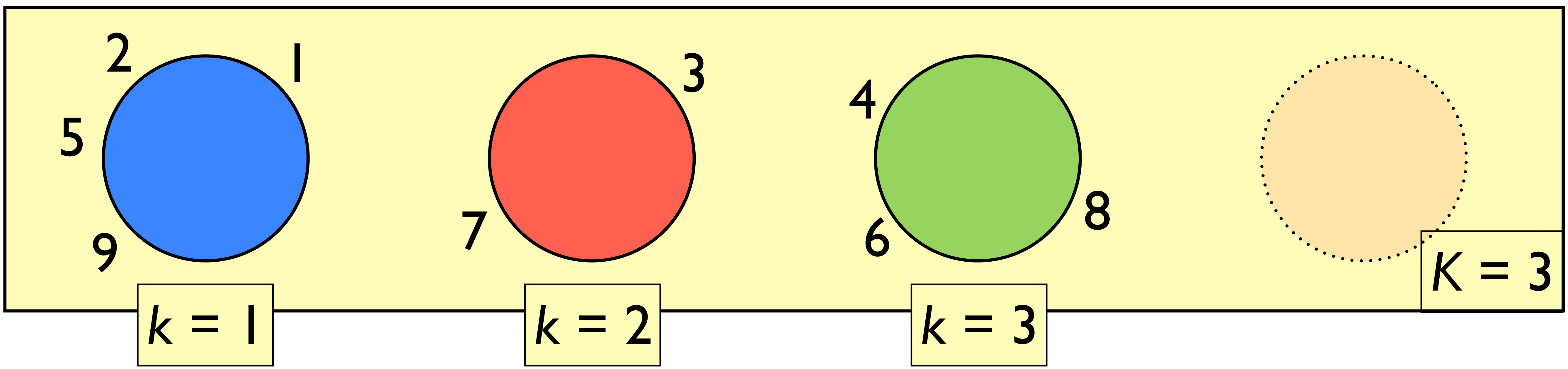
$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

EPPF Example

Chinese restaurant process

- Exchangeable
- Consistent
- Random number of clusters



$$\pi_N = \{A_1, A_2, \ldots, A_K\}$$

$$\mathbb{P}(\Pi_N = \pi_N) = \frac{1}{\prod_{n=1}^{N-1}(n+\theta)} \cdot \theta^{K-1} \cdot \prod_{k=1}^{K} (|A_k| - 1)!$$

Outline

I. Clusters

- Overview
- Distribution
    - ◇ Clusters (Example: Chinese restaurant process)
    - ◇ Data given clusters
    - ◇ Posterior
- Proportions
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
  - ◇ Clusters (Example: Chinese restaurant process)
  - ◇ Data given clusters
  - ◇ Posterior
- Proportions
- Random probability measure

II. Features

Outline

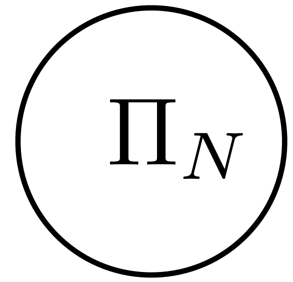
I. Clusters

- Overview
- Distribution
  - ◇ Clusters (Example: Chinese restaurant process)
  - ◇ Data given clusters (Example: Gaussian mixture)
  - ◇ Posterior
- Proportions
- Random probability measure

II. Features

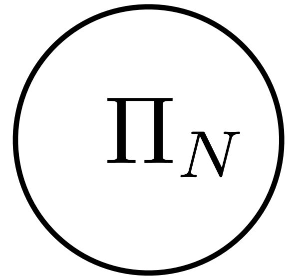
EPPF: Part of full generative model

EPPF: Part of full generative model



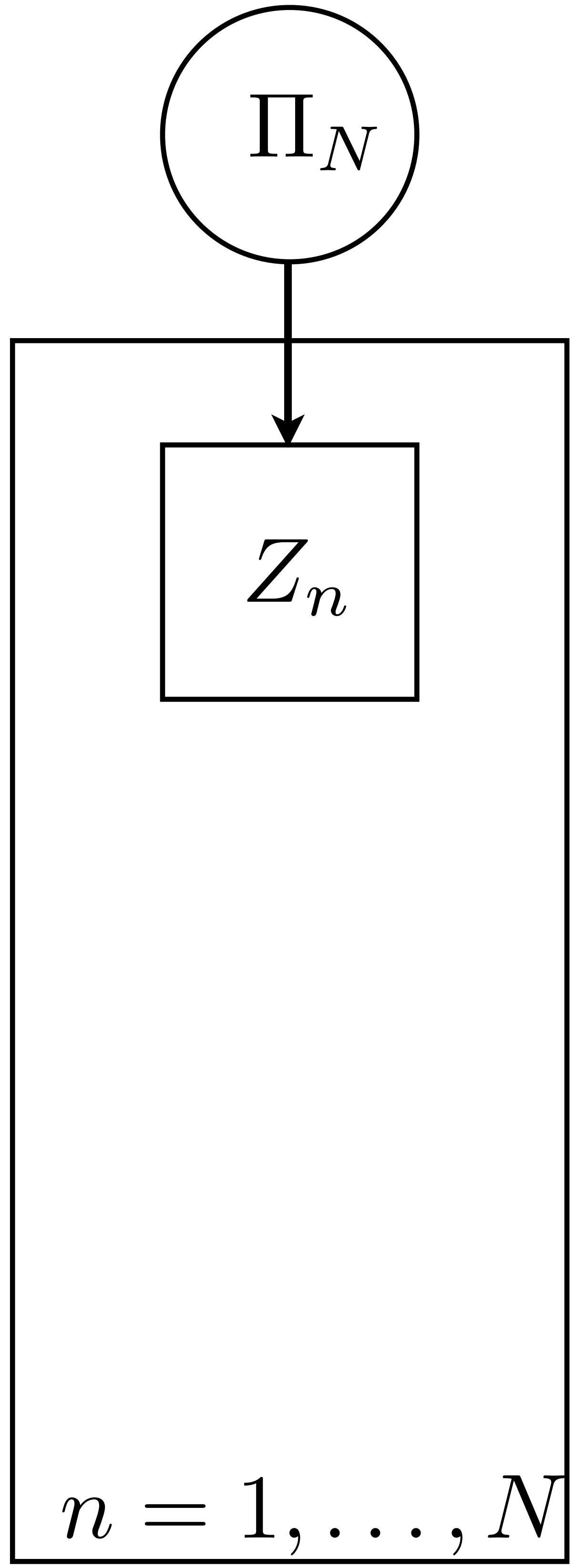
$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

EPPF: Part of full generative model



$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

EPPF: Part of full generative model



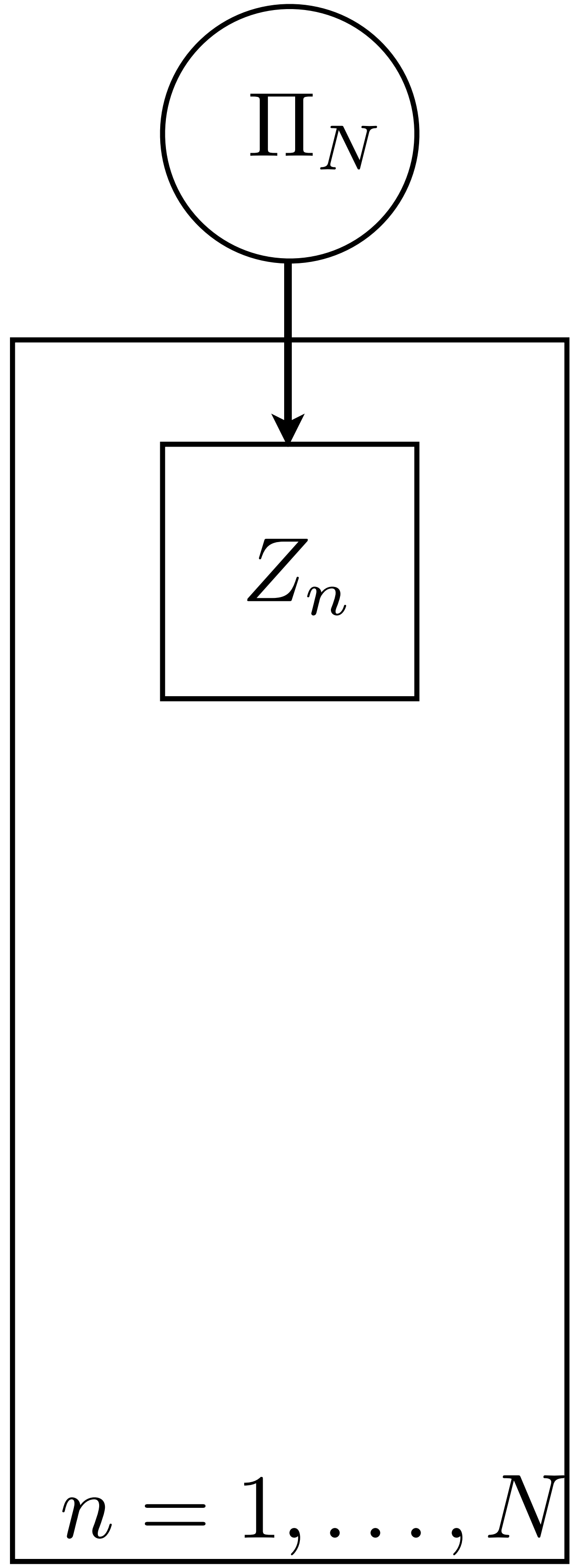
$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

$$z_9 = z_2 = z_5 = z_1 =$$

$$z_7 = z_3 =$$

$$z_8 = z_4 = z_6 =$$

EPPF: Part of full generative model



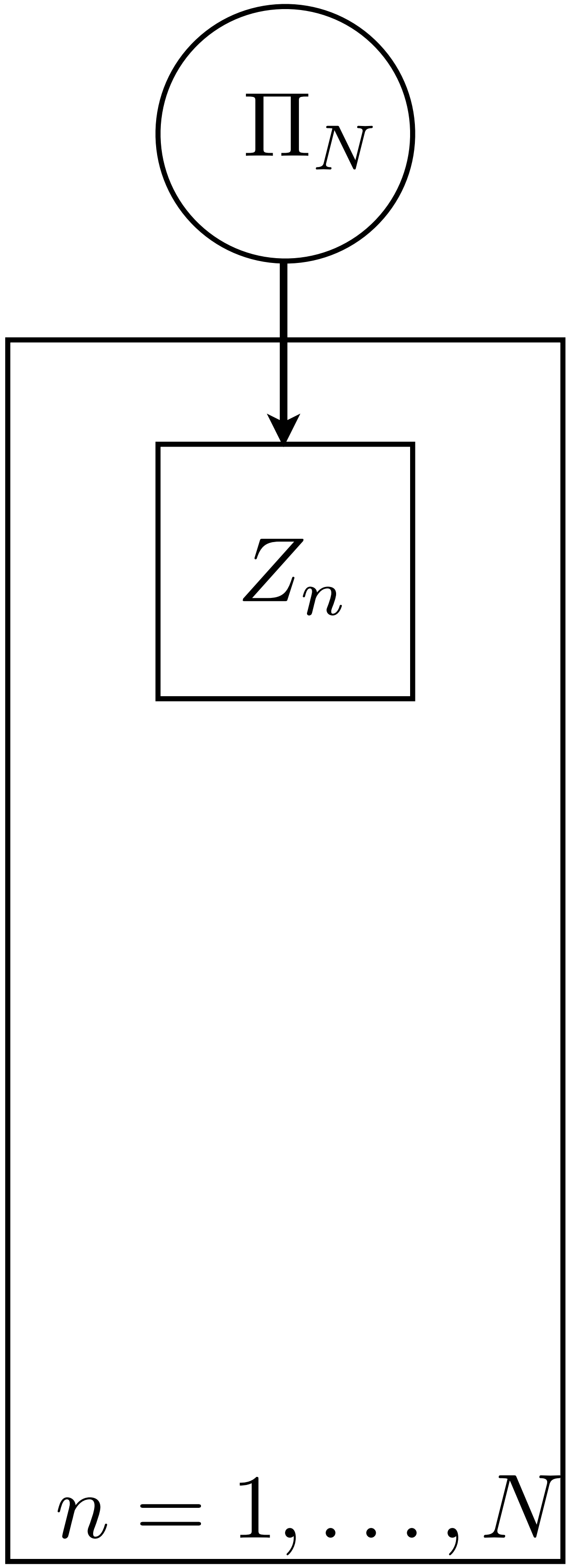
$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

$$z_9 = z_2 = z_5 = z_1 = 1$$

$$z_7 = z_3 = 2$$

$$z_8 = z_4 = z_6 = 3$$

EPPF: Part of full generative model

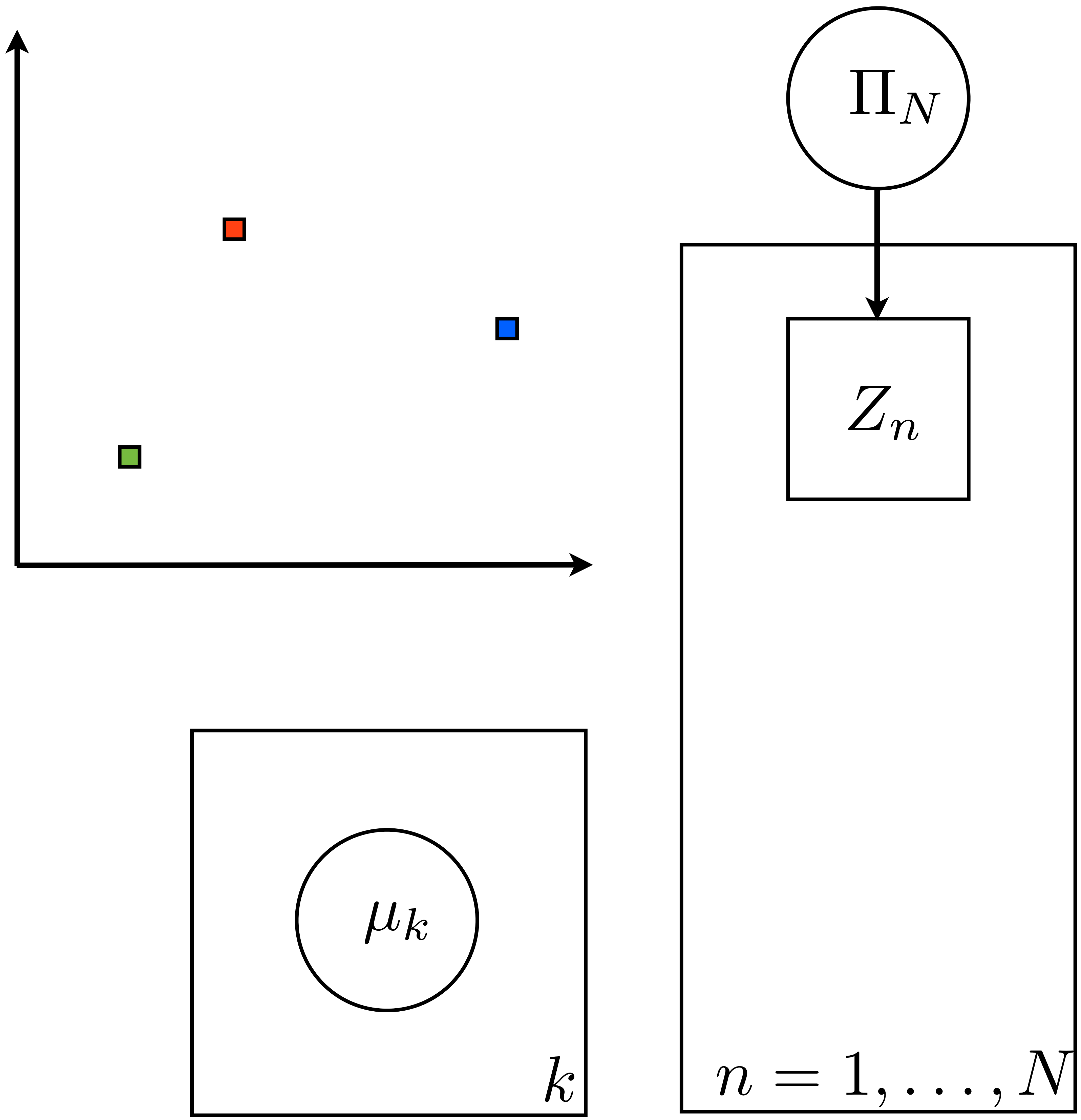


$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

$$z_9 = z_2 = z_5 = z_1 = 1$$
$$z_7 = z_3 = 2$$
$$z_8 = z_4 = z_6 = 3$$

"cluster indicators"

EPPF: Part of full generative model



$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

$$z_9 = z_2 = z_5 = z_1 = 1$$

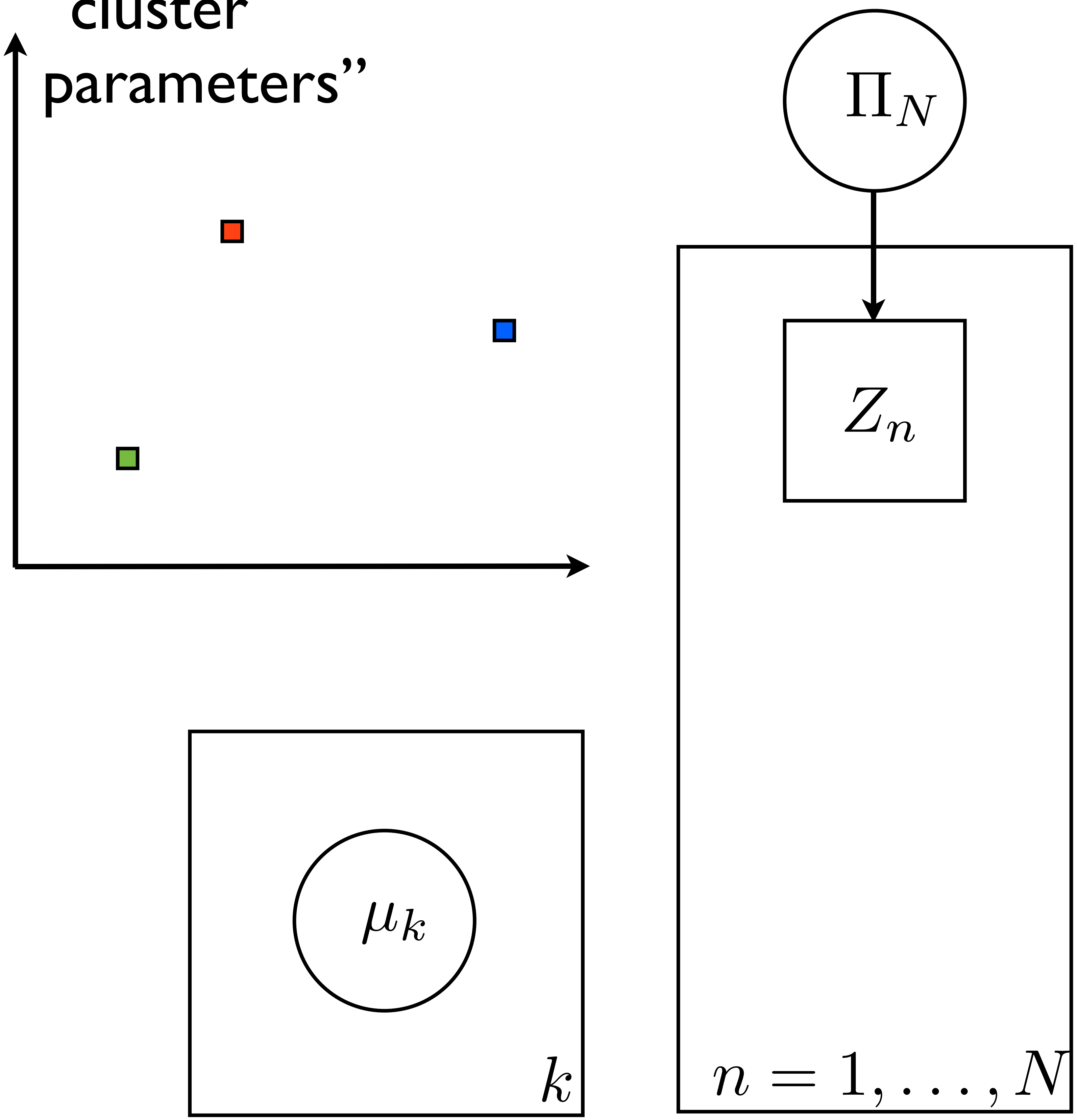
$$z_7 = z_3 = 2$$

$$z_8 = z_4 = z_6 = 3$$

"cluster indicators"

EPPF: Part of full generative model

"cluster parameters"



$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

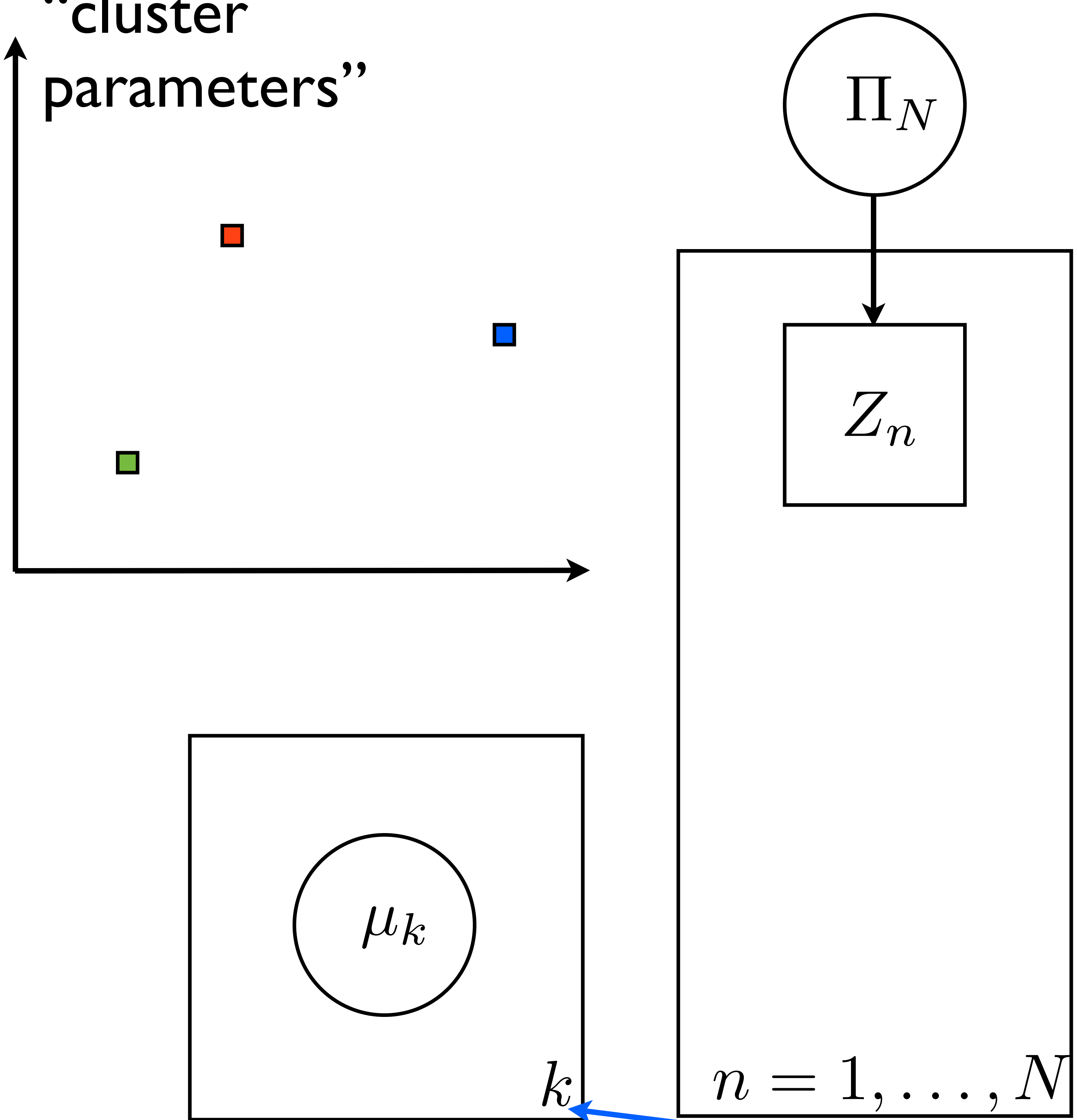
$$z_9 = z_2 = z_5 = z_1 = 1$$

$$z_7 = z_3 = 2$$

$$z_8 = z_4 = z_6 = 3$$

"cluster indicators"

EPPF: Part of full generative model



$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

$$z_9 = z_2 = z_5 = z_1 = 1$$

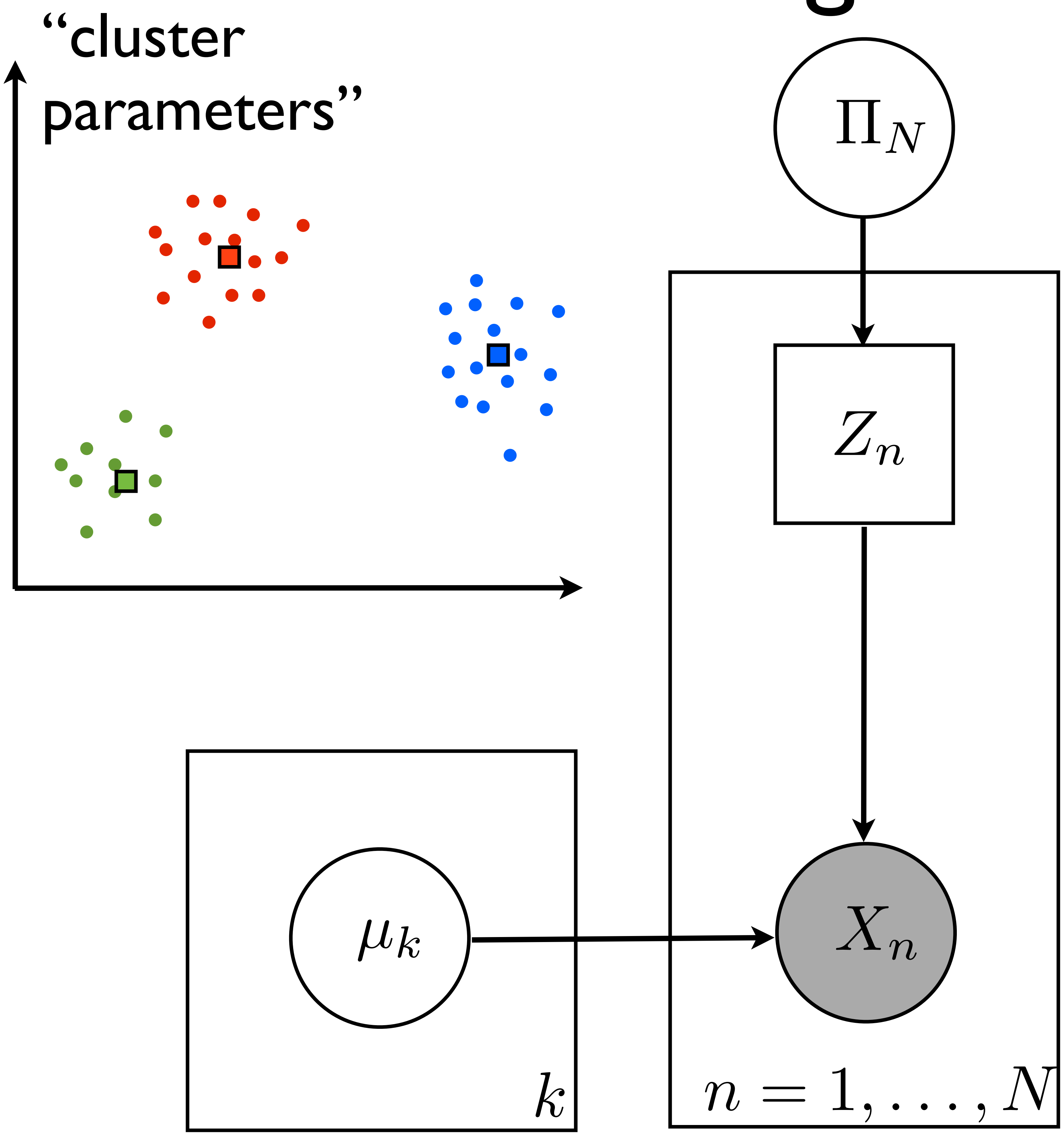
$$z_7 = z_3 = 2$$

$$z_8 = z_4 = z_6 = 3$$

"cluster indicators"

Can think of
k = 1, 2, ... , but only
use finitely many

EPPF: Part of full generative model



$$\pi_9 = \{\{9, 2, 5, 1\}, \{7, 3\}, \{8, 4, 6\}\}$$

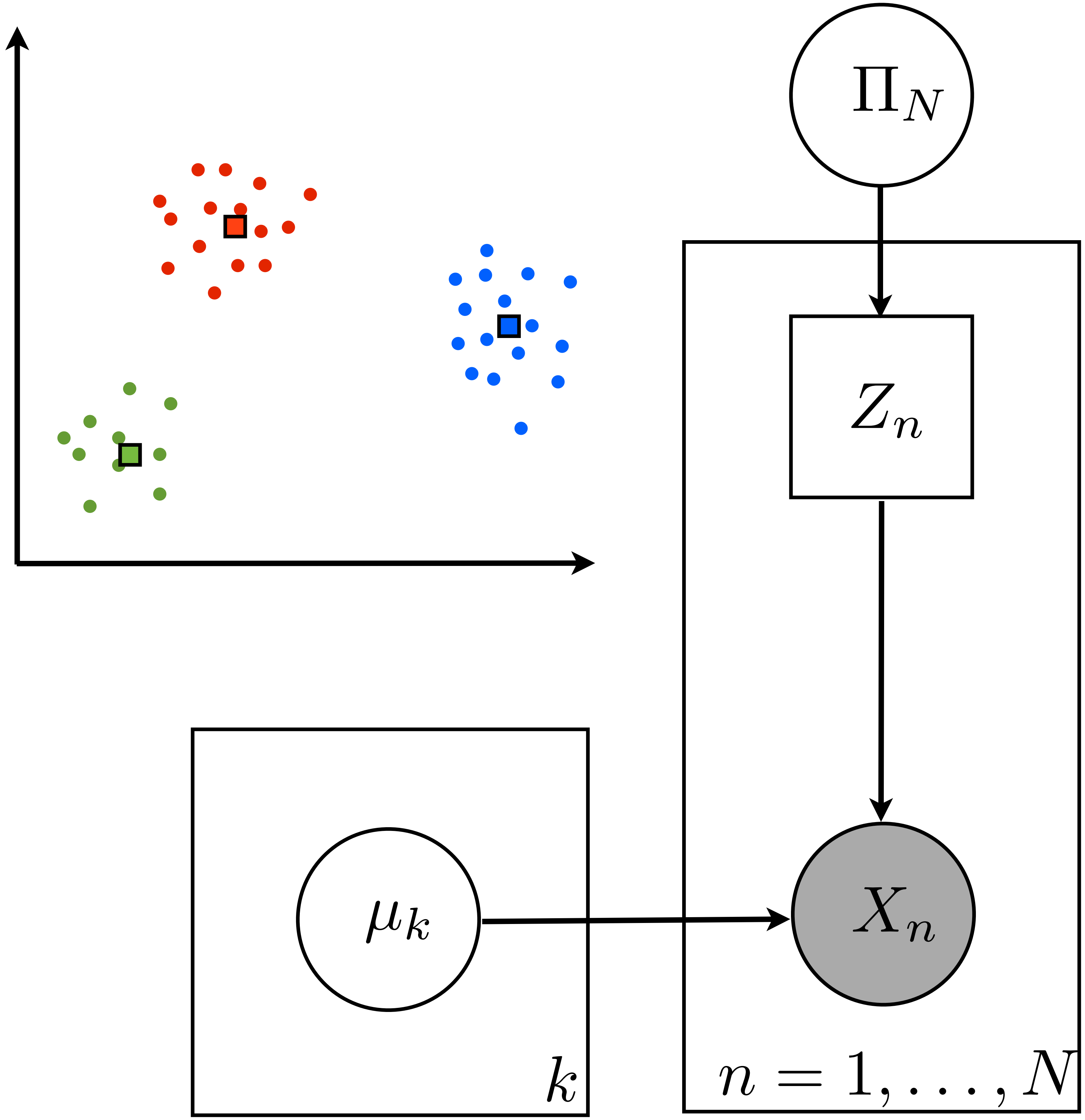
$$z_9 = z_2 = z_5 = z_1 = 1$$

$$z_7 = z_3 = 2$$

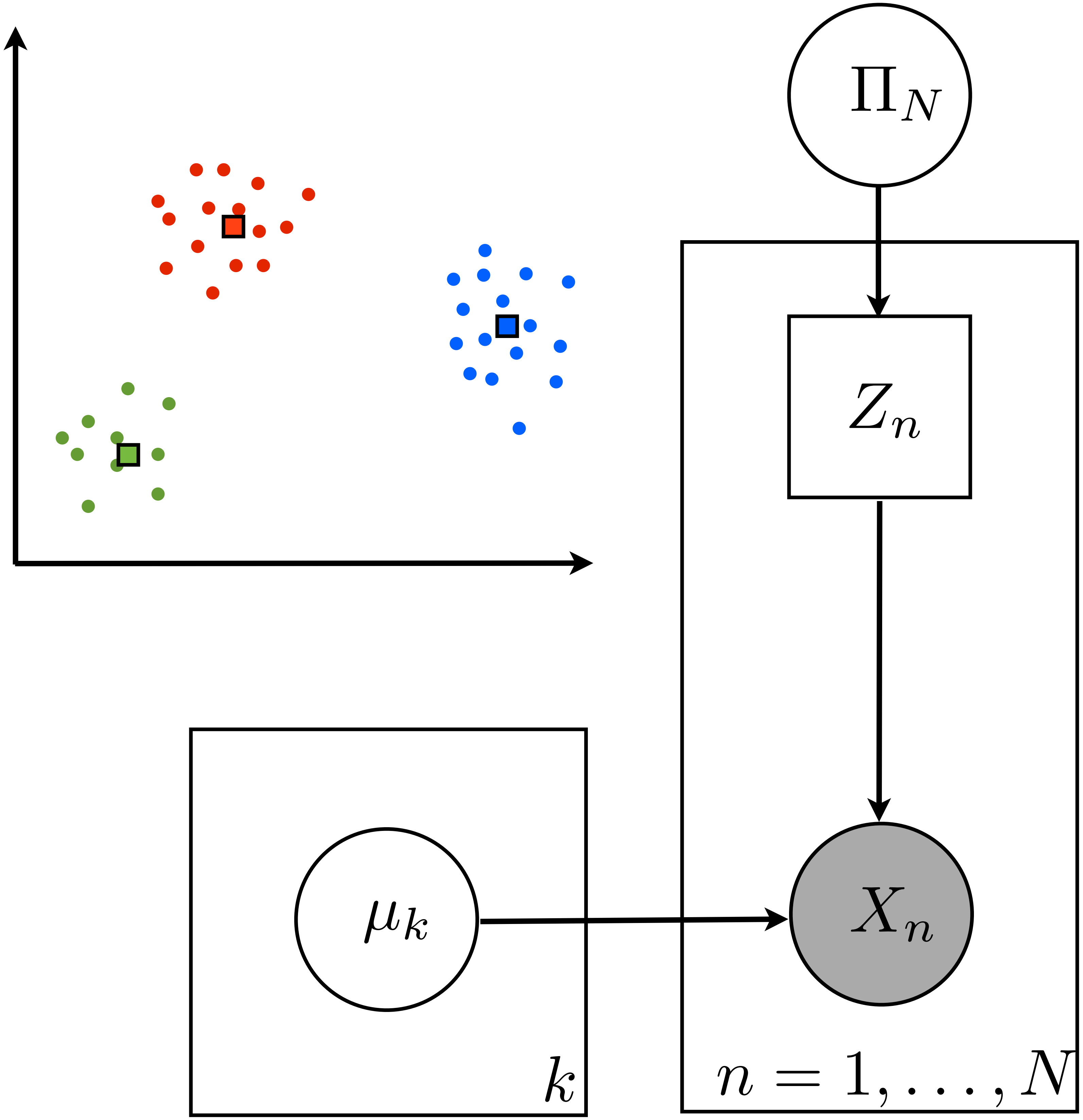
$$z_8 = z_4 = z_6 = 3$$

"cluster indicators"

EPPF: Part of full generative model

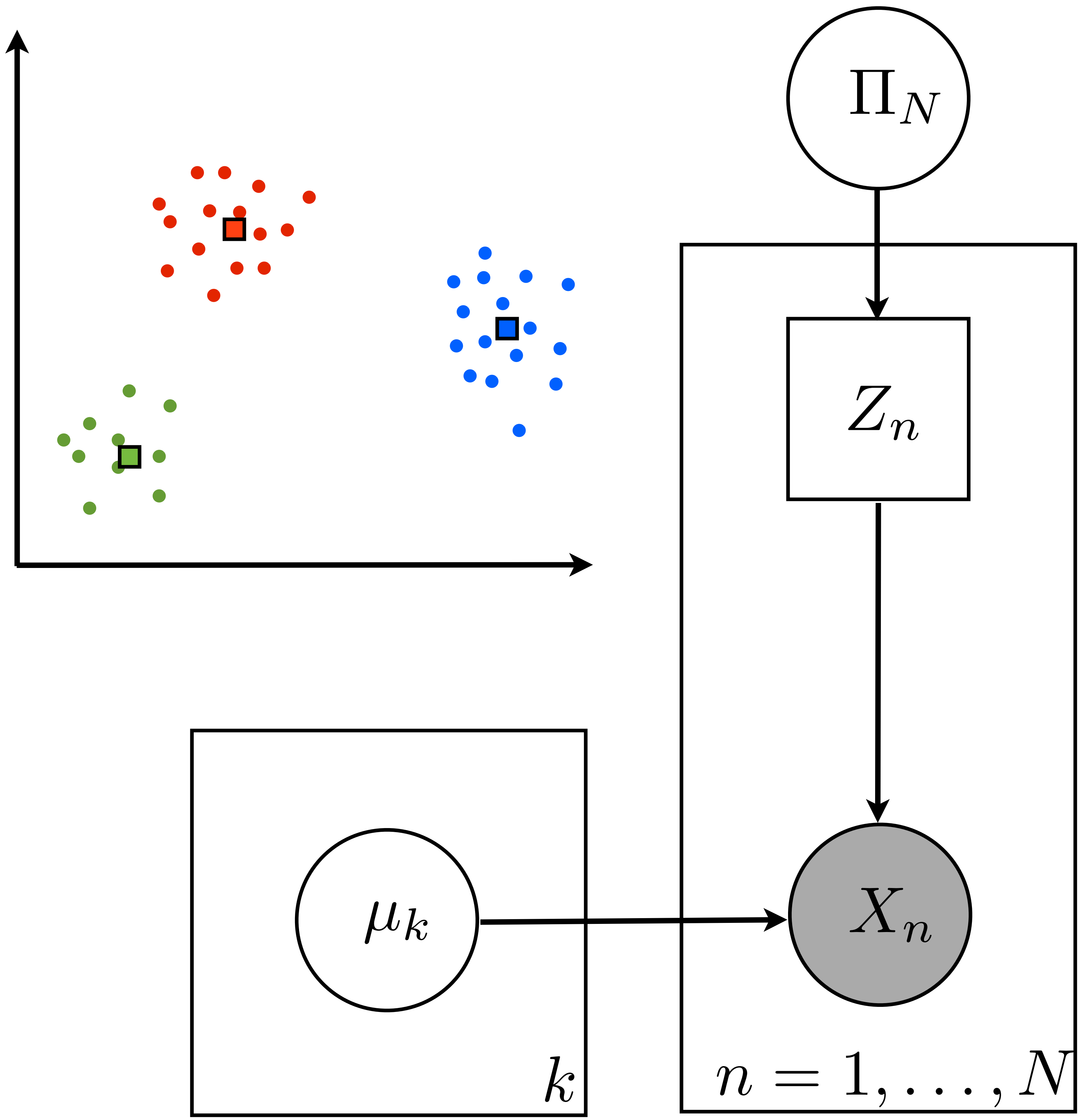


EPPF: Part of full generative model



$$\Pi_N \sim p$$

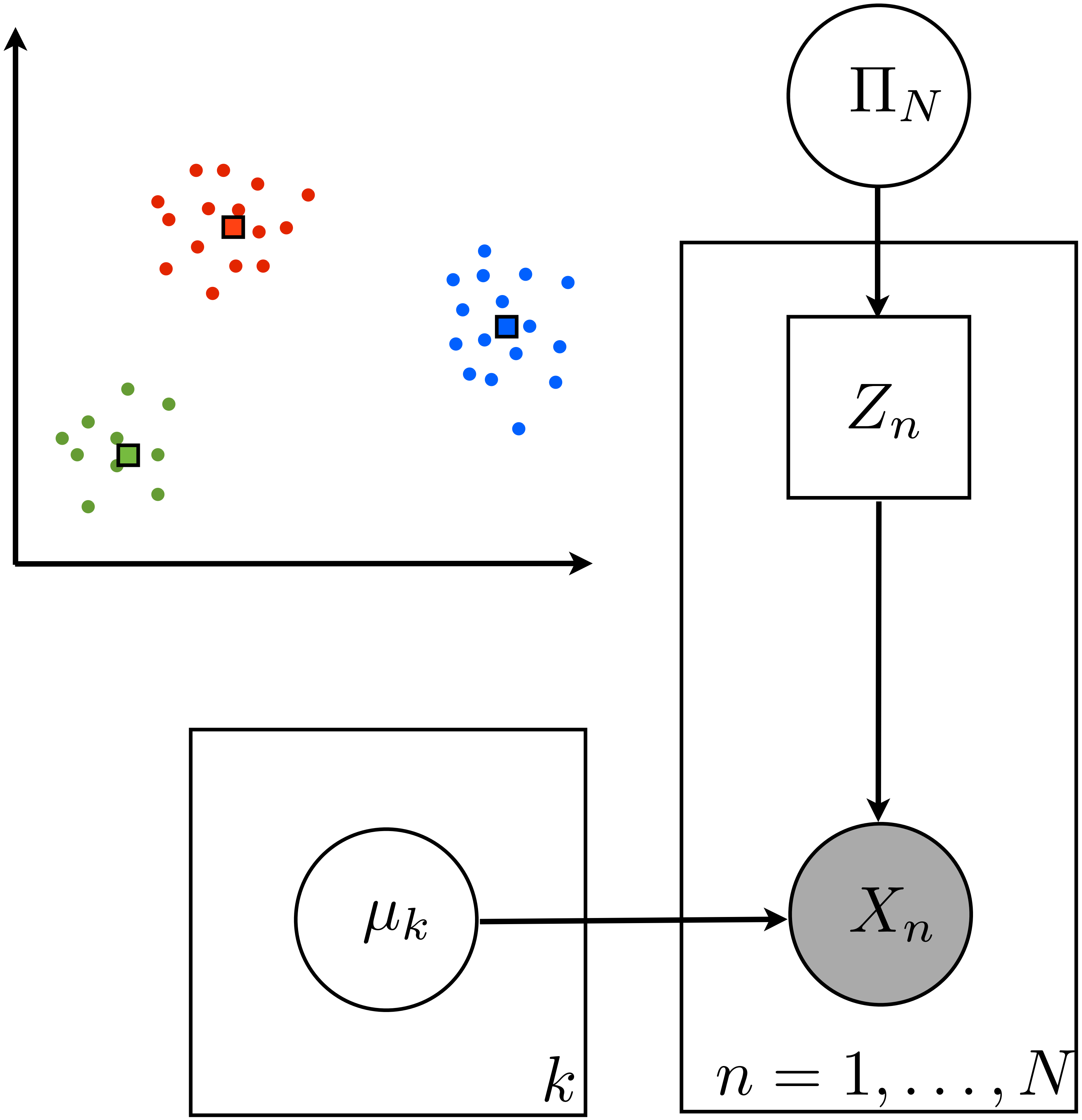
EPPF: Part of full generative model



$$\Pi_N \sim p$$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(0, \rho^2 I_D)$$

EPPF: Part of full generative model

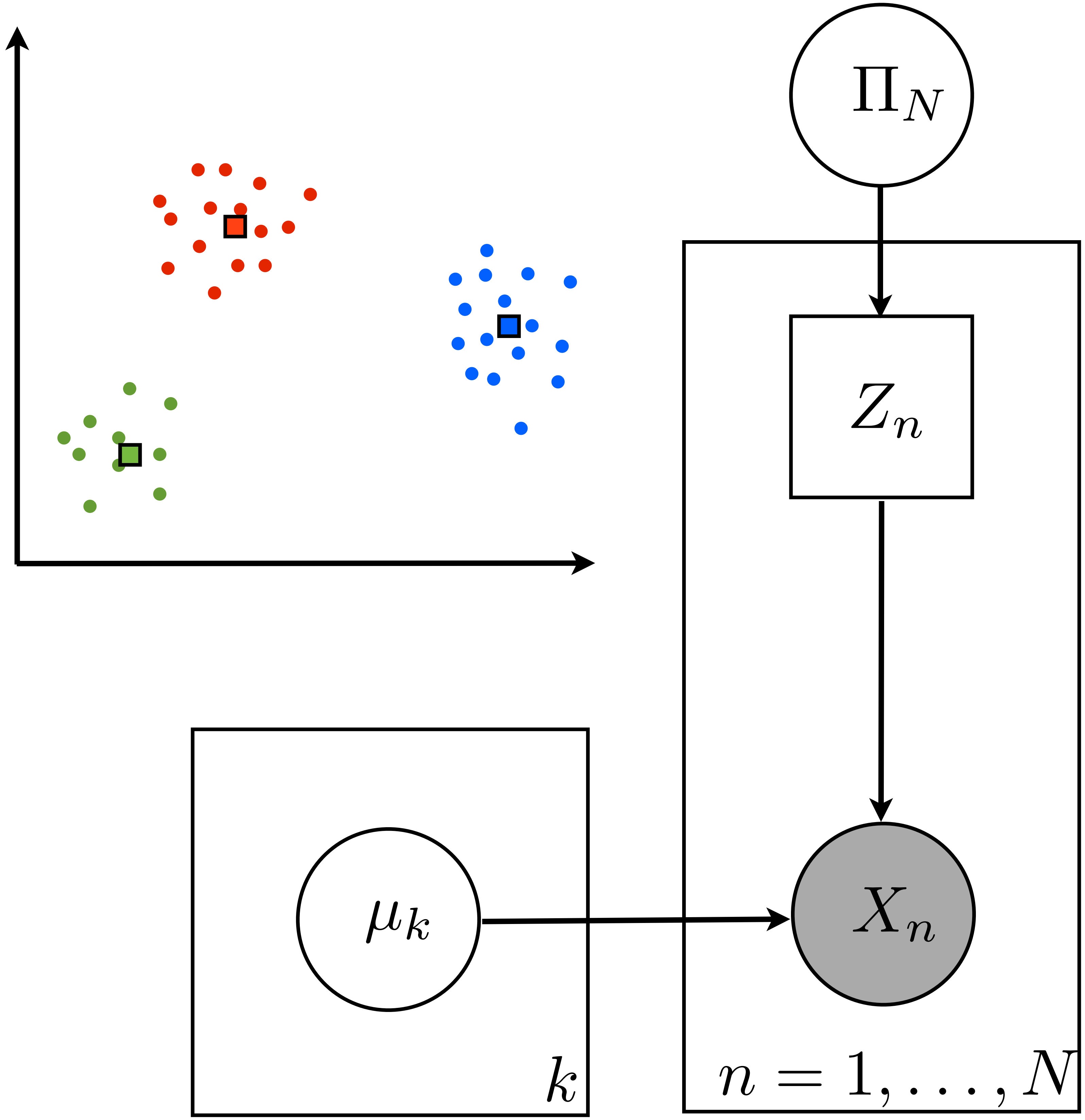


$$\Pi_N \sim p$$

dimension of a data point

$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(0, \rho^2 I_D)$$

EPPF: Part of full generative model

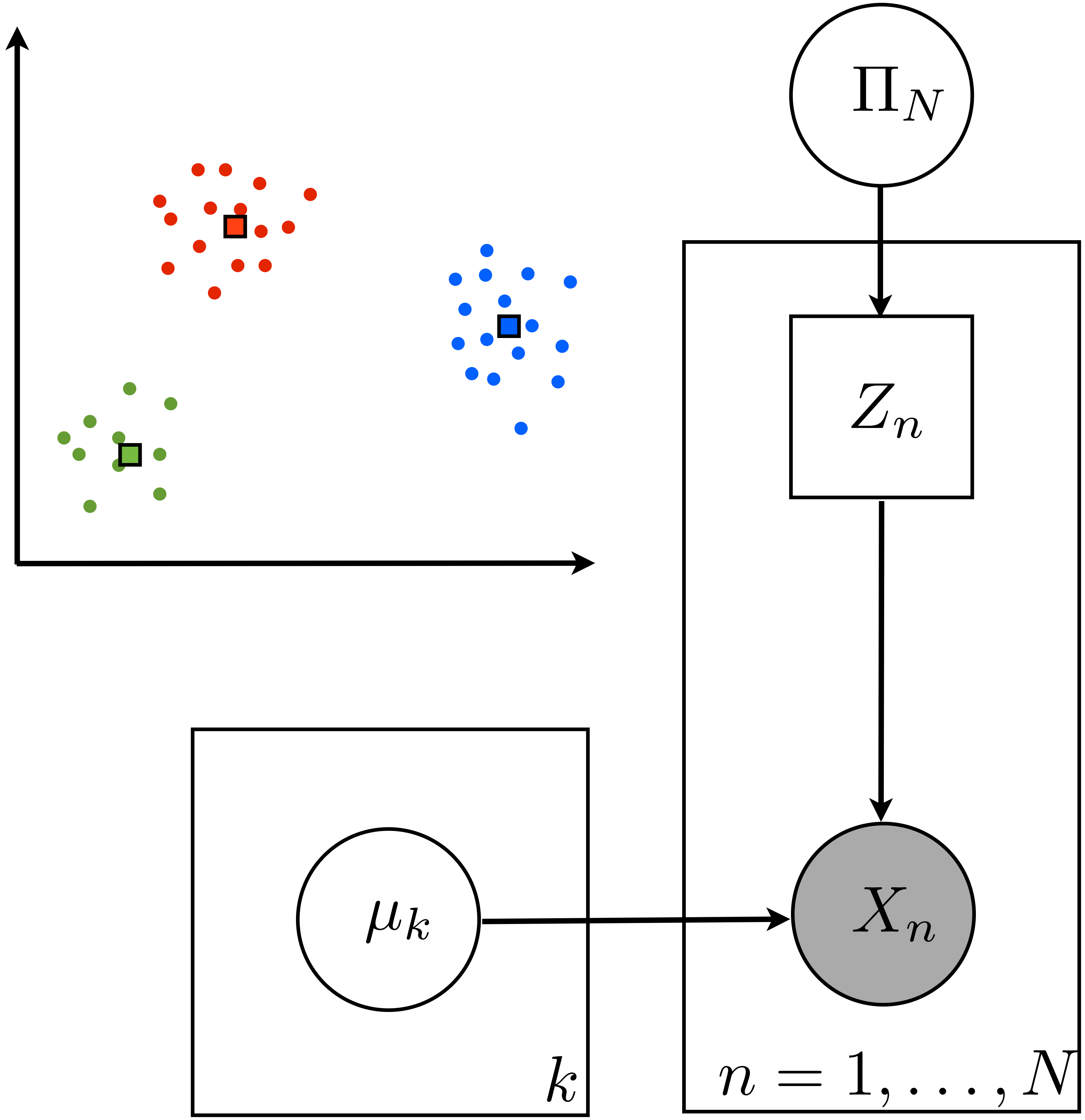


$$\Pi_N \sim p$$

$$\mu_k \stackrel{iid}{\sim} \mathcal{N}(0, \rho^2 I_D)$$

$$X_n \stackrel{indep}{\sim} \mathcal{N}(\mu_{Z_n}, \sigma^2 I_D)$$

EPPF: Part of full generative model



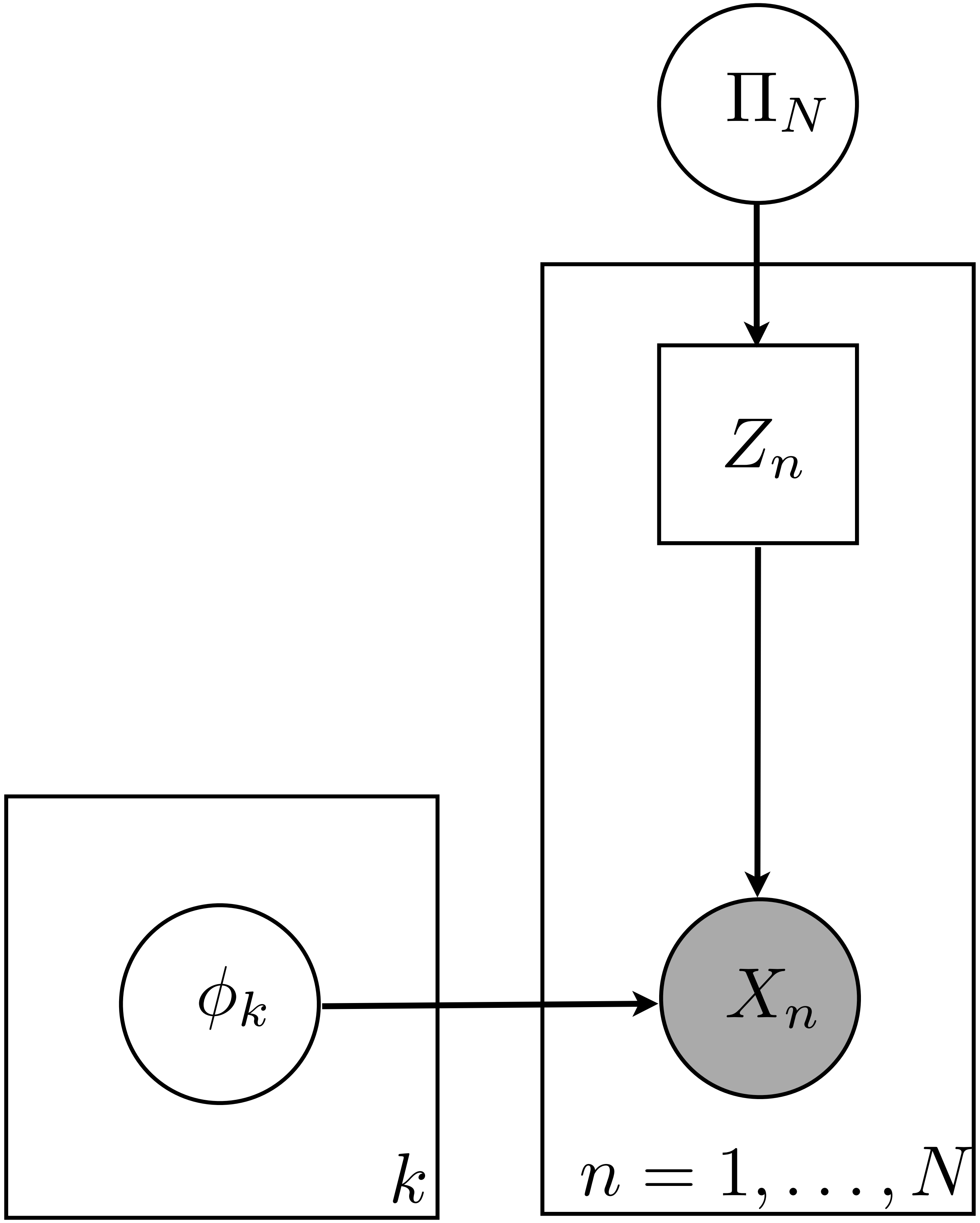
$\Pi_N \sim p$

$$\mu_k \overset{iid}{\sim} \mathcal{N}(0, \rho^2 I_D)$$

$$X_n \overset{indep}{\sim} \mathcal{N}(\mu_{Z_n}, \sigma^2 I_D)$$

"Gaussian mixture model"

EPPF: Part of full generative model



$$\Pi_N \sim p$$

$$\phi_k \overset{iid}{\sim} H$$

$$X_n \overset{indep}{\sim} F(\phi_{Z_n})$$

Outline

I. Clusters

- Overview
- Distribution
  - ◇ Clusters (Example: Chinese restaurant process)
  - ◇ Data given clusters (Example: Gaussian mixture)
  - ◇ Posterior
- Proportions
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
  - ◇ Clusters (Example: Chinese restaurant process)
  - ◇ Data given clusters (Example: Gaussian mixture)
  - ◇ Posterior
- Proportions
- Random probability measure

II. Features

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

D: data dimension
N: number data points
K: (random) number of clusters

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

all data points
(N vectors of length D)

D: data dimension
N: number data points
K: (random) number of clusters

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

all cluster indicators
(N integers)

all data points
(N vectors of length D)

D: data dimension
N: number data points
K: (random) number of clusters

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

all cluster indicators
(N integers)

all cluster means
(K vectors of length D)

all data points
(N vectors of length D)

D: data dimension
N: number data points
K: (random) number of clusters

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

- Usually can't do exact calculation

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

- Usually can't do exact calculation
- Approximation (MCMC, variational methods)

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

- Usually can't do exact calculation
- Approximation (MCMC, variational methods)

Gibbs sampling

[Geman, Geman 1984]

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

- Usually can't do exact calculation
- Approximation (MCMC, variational methods)

Gibbs sampling ← Type of MCMC method

[Geman, Geman 1984]

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

- Usually can't do exact calculation
- Approximation (MCMC, variational methods)

Gibbs sampling

- Sample each variable conditioned on the rest

[Geman, Geman 1984]

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

- Usually can't do exact calculation
- Approximation (MCMC, variational methods)

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n | X, \mu, Z_{-n}), \quad n = 1, \ldots, N$$

$$\mathbb{P}(\mu_k | X, Z, \mu_{-k}), \quad k = 1, \ldots, K$$

EPPF: Calculating posterior

Calculating posterior: $\mathbb{P}(Z, \mu | X)$

- Usually can't do exact calculation
- Approximation (MCMC, variational methods)

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n | X, \mu, Z_{-n}), \quad n = 1, \dots, N$$

$$\mathbb{P}(\mu_k | X, Z, \mu_{-k}), \quad k = 1, \dots, K$$

function of Z

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n | X, \mu, Z_{-n})$$

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n|X,\mu,Z_{-n}) = \frac{\mathbb{P}(X,Z,\mu)}{\mathbb{P}(X,Z_{-n},\mu)}$$

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n|X,\mu,Z_{-n}) = \frac{\mathbb{P}(X,Z,\mu)}{\mathbb{P}(X,Z_{-n},\mu)}$$

$$\propto \frac{\mathbb{P}(\Pi_N)\mathbb{P}(X_n|Z_n,\mu)}{\mathbb{P}(\Pi_{N-1})}$$

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n|X,\mu,Z_{-n}) = \frac{\mathbb{P}(X,Z,\mu)}{\mathbb{P}(X,Z_{-n},\mu)}$$

$$\propto \frac{\mathbb{P}(\Pi_N)\mathbb{P}(X_n|Z_n,\mu)}{\mathbb{P}(\Pi_{N-1})}$$

use exchangeability

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n|X,\mu,Z_{-n}) = \frac{\mathbb{P}(X,Z,\mu)}{\mathbb{P}(X,Z_{-n},\mu)}$$

$$\propto \frac{\mathbb{P}(\Pi_N)\mathbb{P}(X_n|Z_n,\mu)}{\mathbb{P}(\Pi_{N-1})}$$

e.g. Chinese restaurant process for clusters;
Gaussian mixture for
data given clusters

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n | X, \mu, Z_{-n}) = \frac{\mathbb{P}(X, Z, \mu)}{\mathbb{P}(X, Z_{-n}, \mu)}$$

$$\propto \frac{\mathbb{P}(\Pi_N)\mathbb{P}(X_n | Z_n, \mu)}{\mathbb{P}(\Pi_{N-1})}$$

e.g. CRP for clusters;
Gaussian mixture for
data given clusters

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n|X,\mu,Z_{-n}) = \frac{\mathbb{P}(X,Z,\mu)}{\mathbb{P}(X,Z_{-n},\mu)}$$

$$\propto \frac{\mathbb{P}(\Pi_N)\mathbb{P}(X_n|Z_n,\mu)}{\mathbb{P}(\Pi_{N-1})}$$

e.g. CRP for clusters; Gaussian mixture for data given clusters

$$= \begin{cases} \mathcal{N}(X_n|\mu_k,\sigma^2 I_D)\frac{|A_{-n,k}|}{N-1+\theta} & Z_n = k \\ \mathcal{N}(X_n|0,(\rho^2+\sigma^2)I_D)\frac{\theta}{N-1+\theta} & Z_n \text{ new} \end{cases}$$

[Escobar 1994; West, Muller, Escobar 1994; Escobar, West 1995; Bush, MacEachern 1996]

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n | X, \mu, Z_{-n}) = \frac{\mathbb{P}(X, Z, \mu)}{\mathbb{P}(X, Z_{-n}, \mu)}$$

$$\propto \frac{\mathbb{P}(\Pi_N)\mathbb{P}(X_n | Z_n, \mu)}{\mathbb{P}(\Pi_{N-1})}$$

e.g. CRP for clusters; Gaussian mixture for data given clusters

$$= \begin{cases} \mathcal{N}(X_n | \mu_k, \sigma^2 I_D) \frac{|A_{-n,k}|}{N-1+\theta} & Z_n = k \\ \mathcal{N}(X_n | 0, (\rho^2 + \sigma^2) I_D) \frac{\theta}{N-1+\theta} & Z_n \text{ new} \end{cases}$$

[Escobar 1994; West, Muller, Escobar 1994; Escobar, West 1995; Bush, MacEachern 1996]

EPPF: Calculating posterior

Gibbs sampling

- Sample each variable conditioned on the rest

$$\mathbb{P}(Z_n | X, \mu, Z_{-n}) = \frac{\mathbb{P}(X, Z, \mu)}{\mathbb{P}(X, Z_{-n}, \mu)}$$

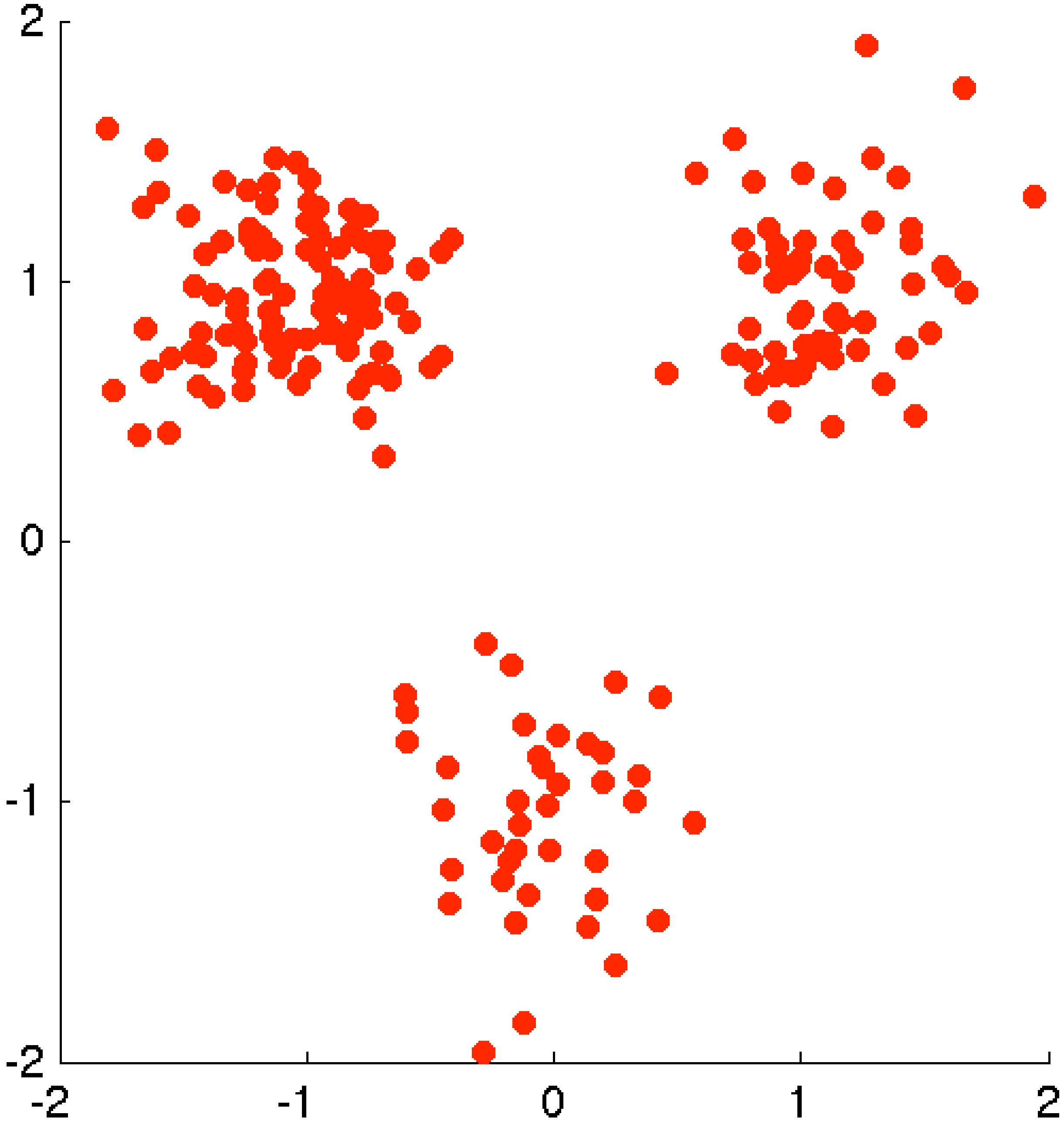
$$\propto \frac{\mathbb{P}(\Pi_N)\mathbb{P}(X_n | Z_n, \mu)}{\mathbb{P}(\Pi_{N-1})}$$

e.g. CRP for clusters; Gaussian mixture for data given clusters

$$= \begin{cases} \mathcal{N}(X_n | \mu_k, \sigma^2 I_D) \frac{|A_{-n,k}|}{N-1+\theta} & Z_n = k \\ \mathcal{N}(X_n | 0, (\rho^2 + \sigma^2) I_D) \frac{\theta}{N-1+\theta} & Z_n \text{ new} \end{cases}$$

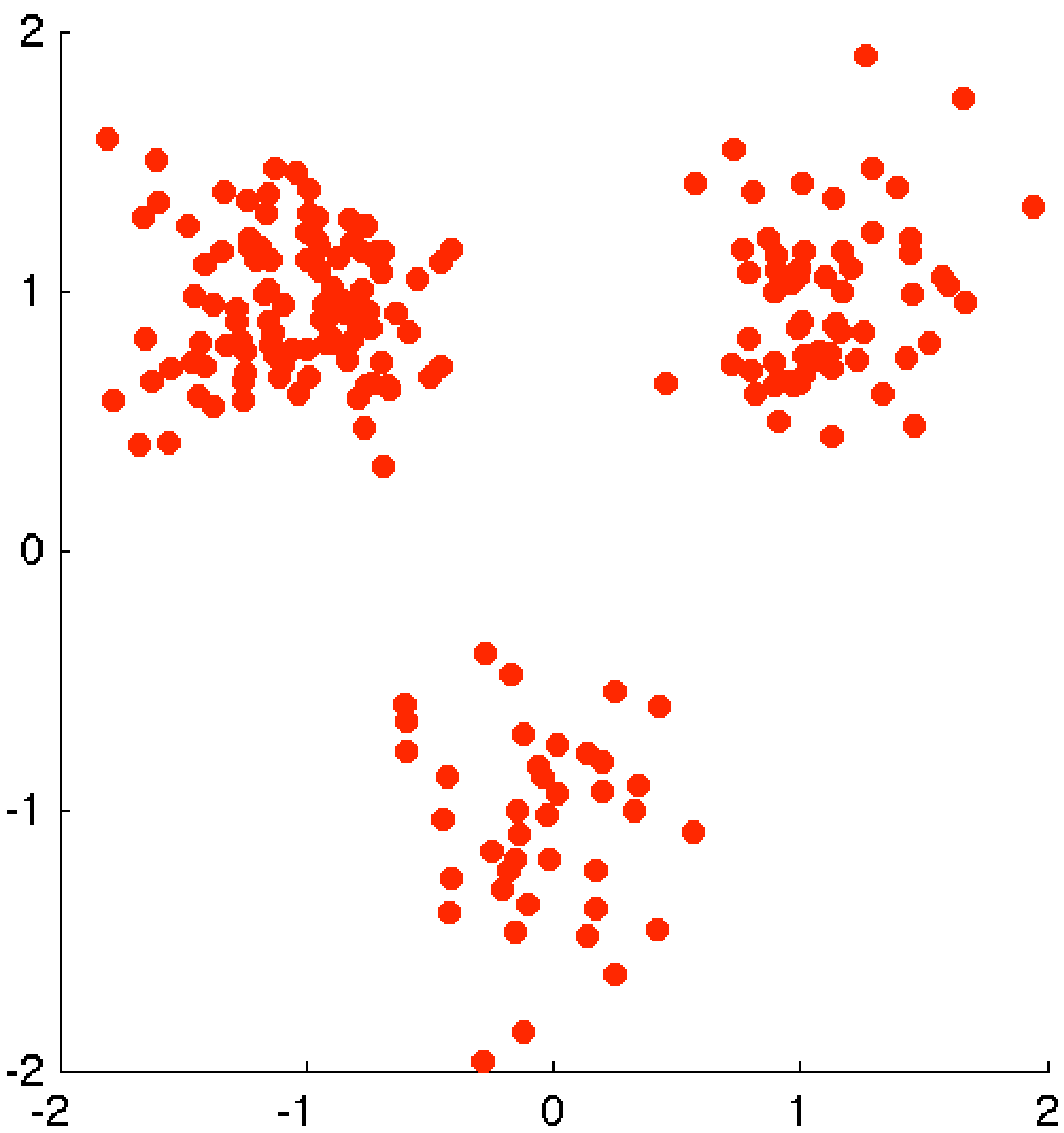
[Escobar 1994; West, Muller, Escobar 1994; Escobar, West 1995; Bush, MacEachern 1996]

EPPF: Calculating posterior



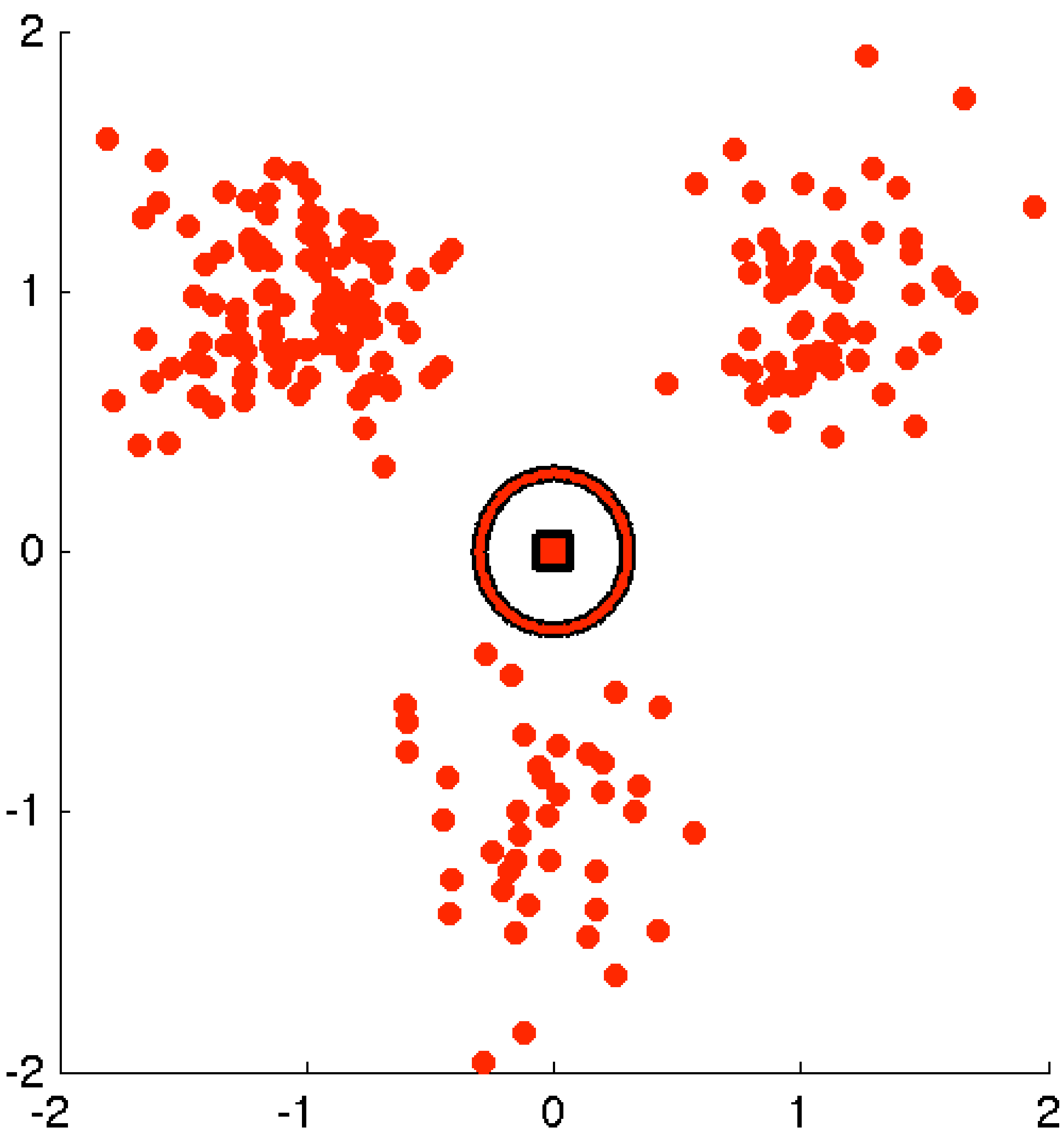
- Initialize
- Repeat
  - ◇ Sample cluster indicators
  - ◇ Sample cluster parameters

EPPF: Calculating posterior



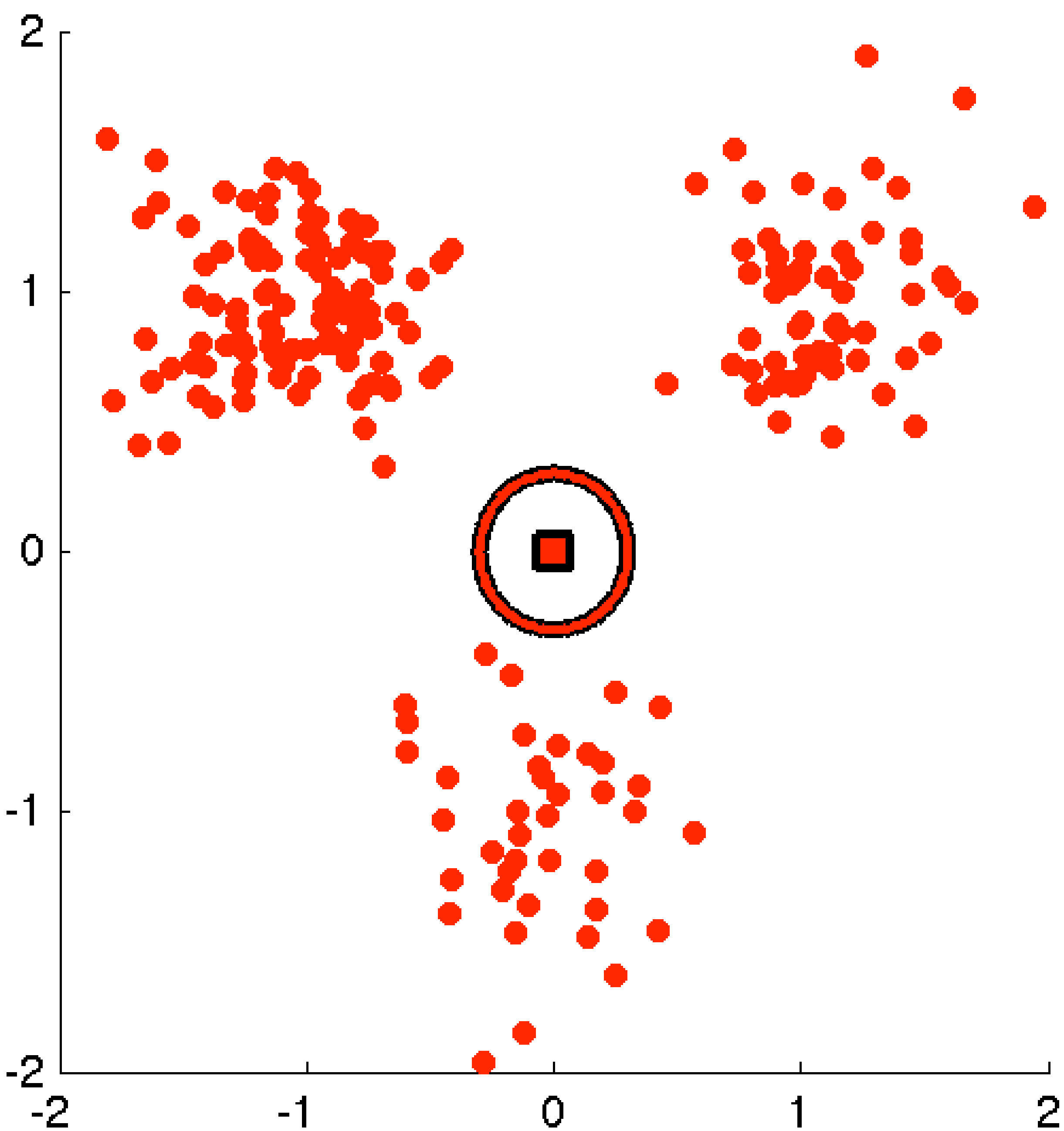
- Initialize
- Repeat
  - ◇ Sample cluster indicators
  - ◇ Sample cluster parameters

EPPF: Calculating posterior



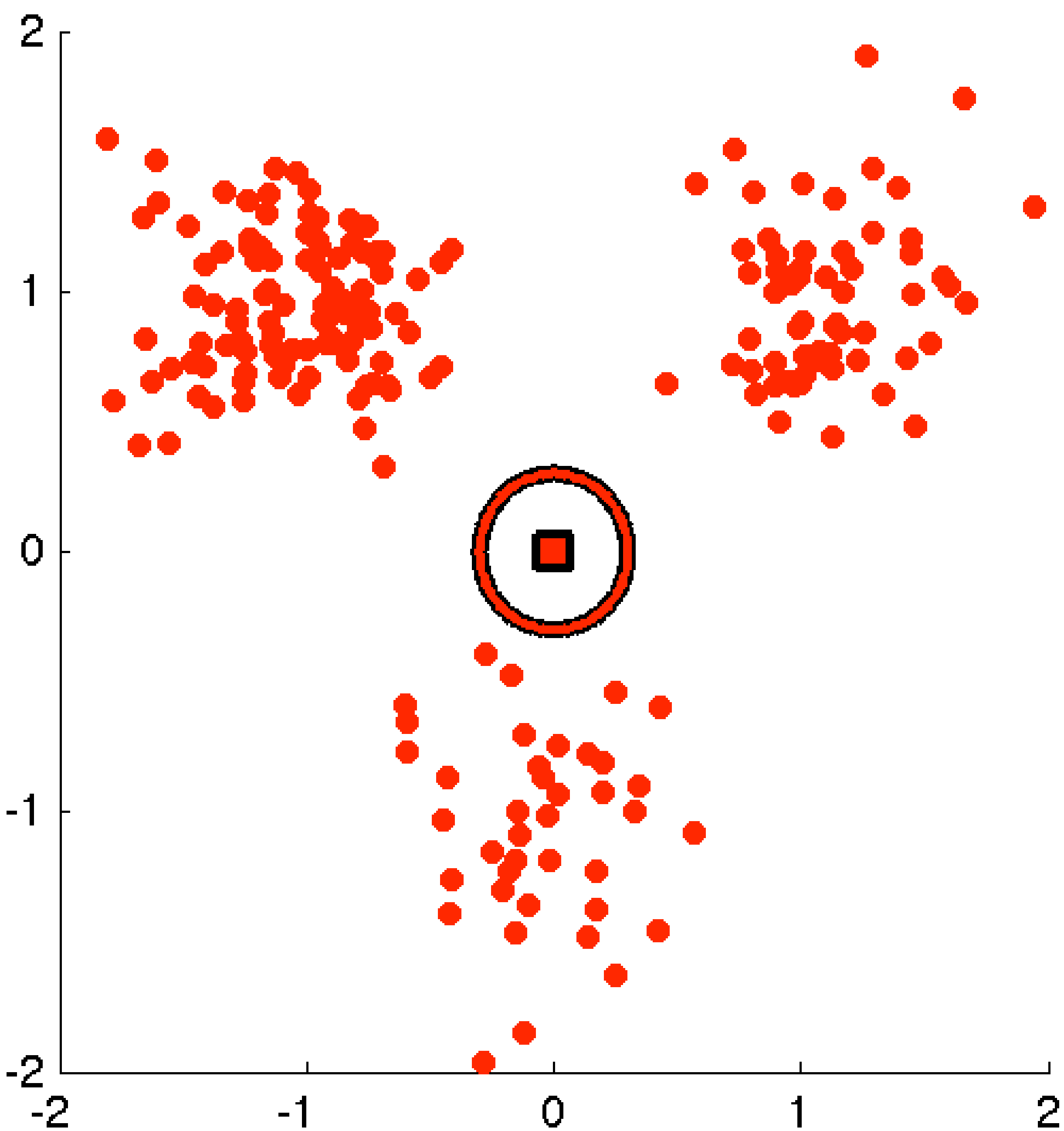
- Assign all points to one cluster
- Repeat
  - ◇ Sample cluster indicators
  - ◇ Sample cluster parameters

EPPF: Calculating posterior



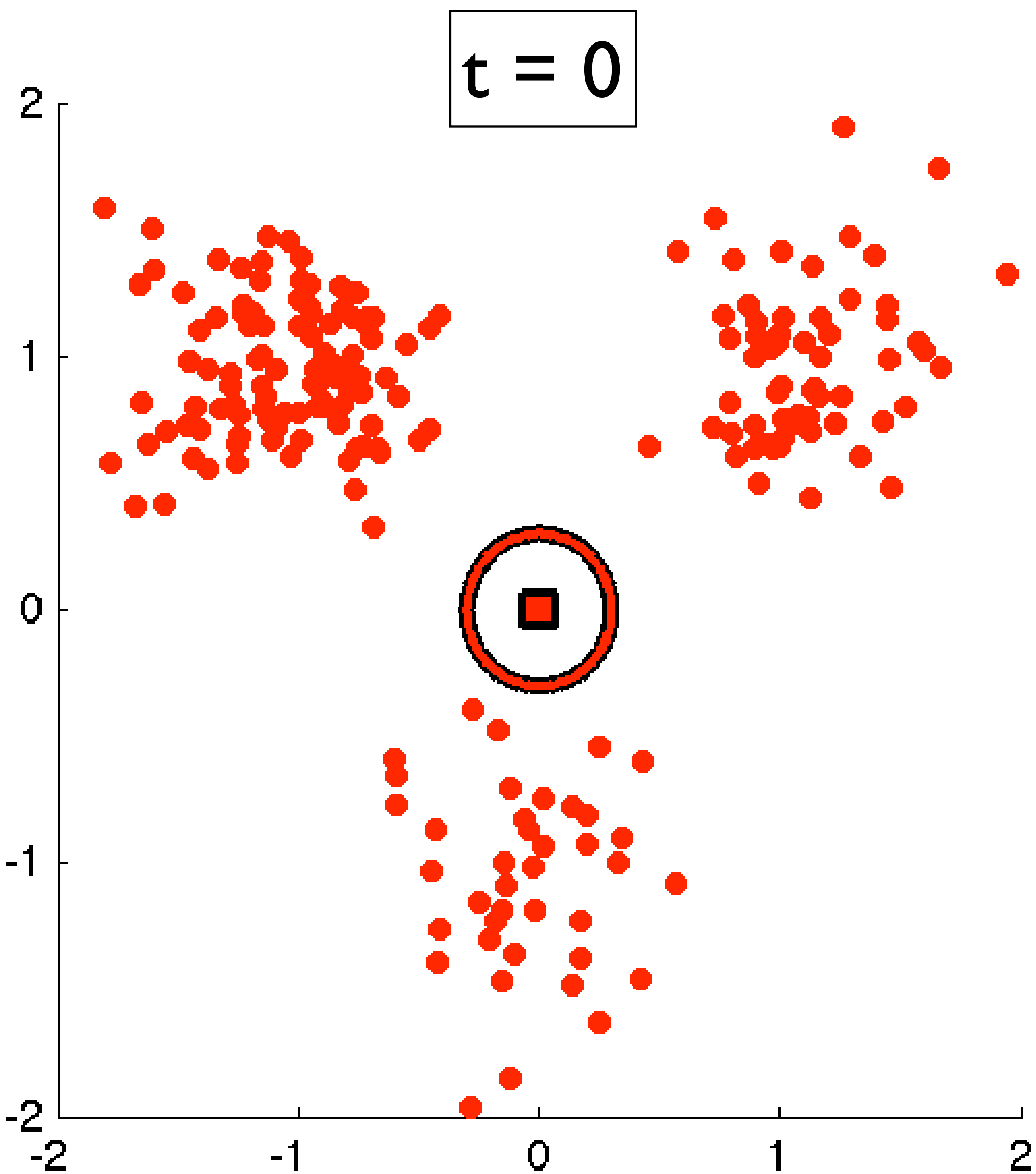
- Assign all points to one cluster
- Repeat
  - ◇ Sample cluster indicators
  - ◇ Sample cluster parameters

EPPF: Calculating posterior



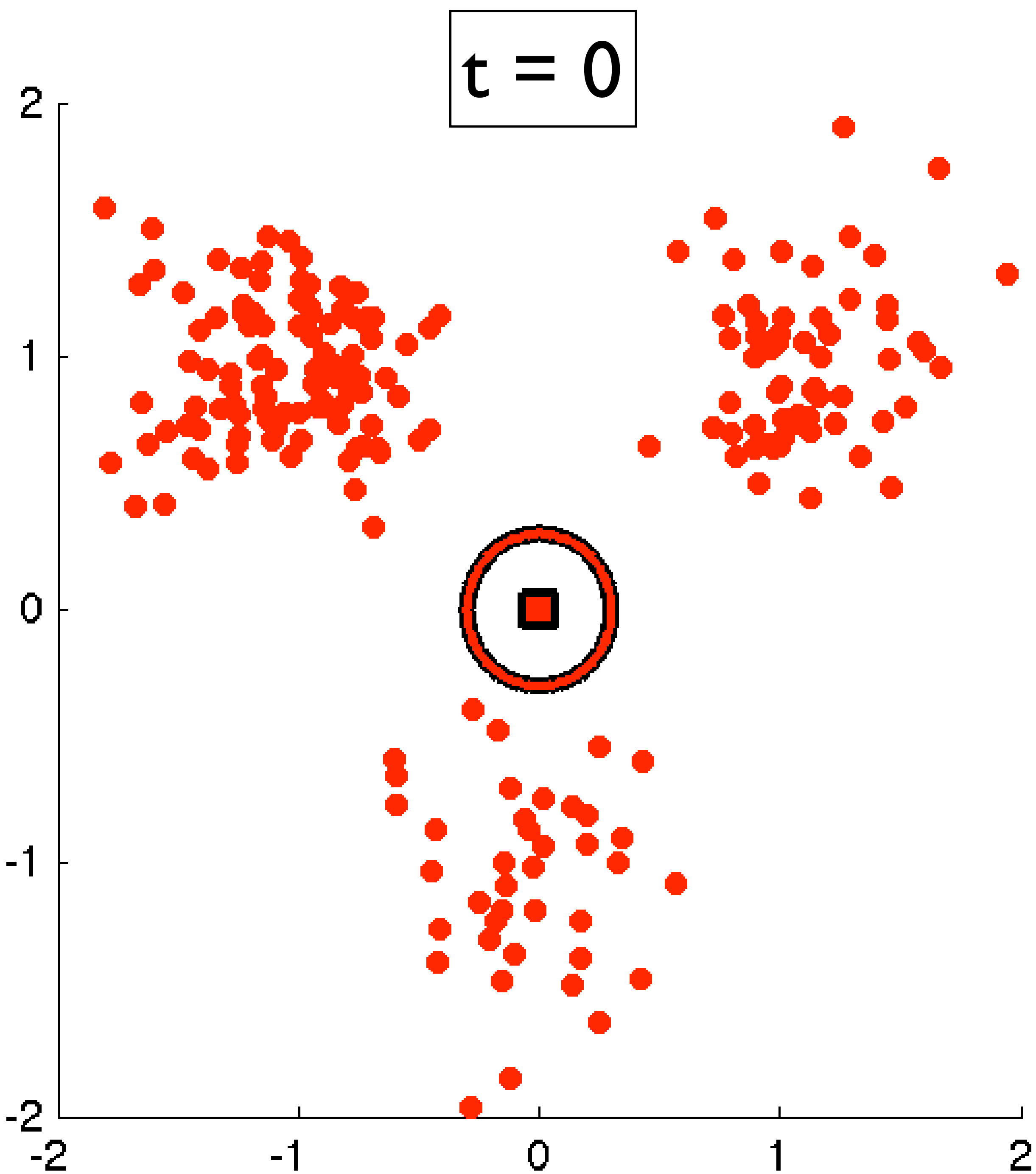
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ Sample cluster indicators
  - ◇ Sample cluster parameters

EPPF: Calculating posterior



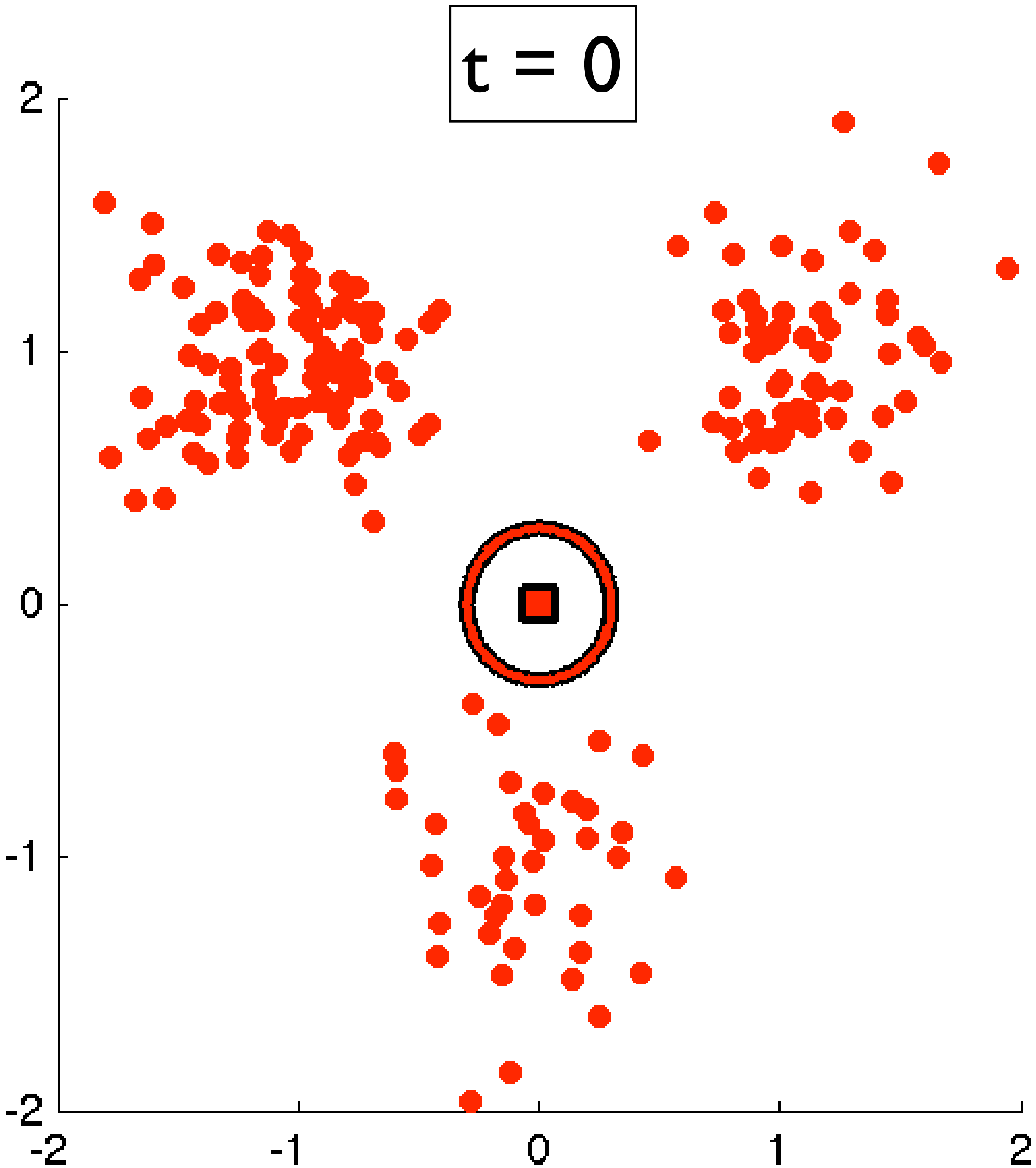
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ Sample cluster indicators
  - ◇ Sample cluster parameters

EPPF: Calculating posterior



- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ Sample cluster indicators
  - ◇ Sample cluster parameters

EPPF: Calculating posterior

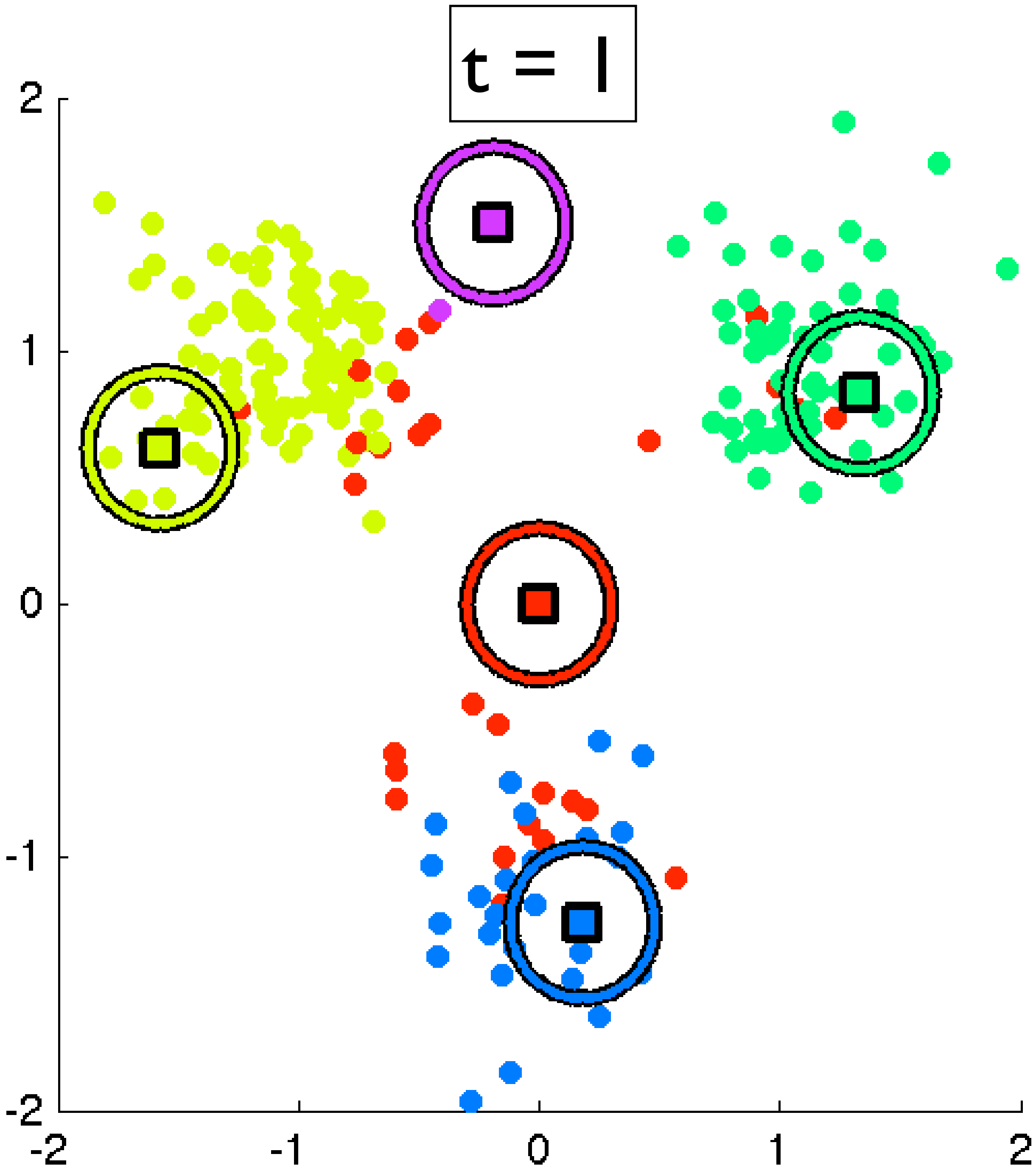


- Assign all points to one cluster
- For t = 1, ..., T
    - ◇ For n = 1, ..., N

      $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

    - ◇ Sample cluster parameters

EPPF: Calculating posterior

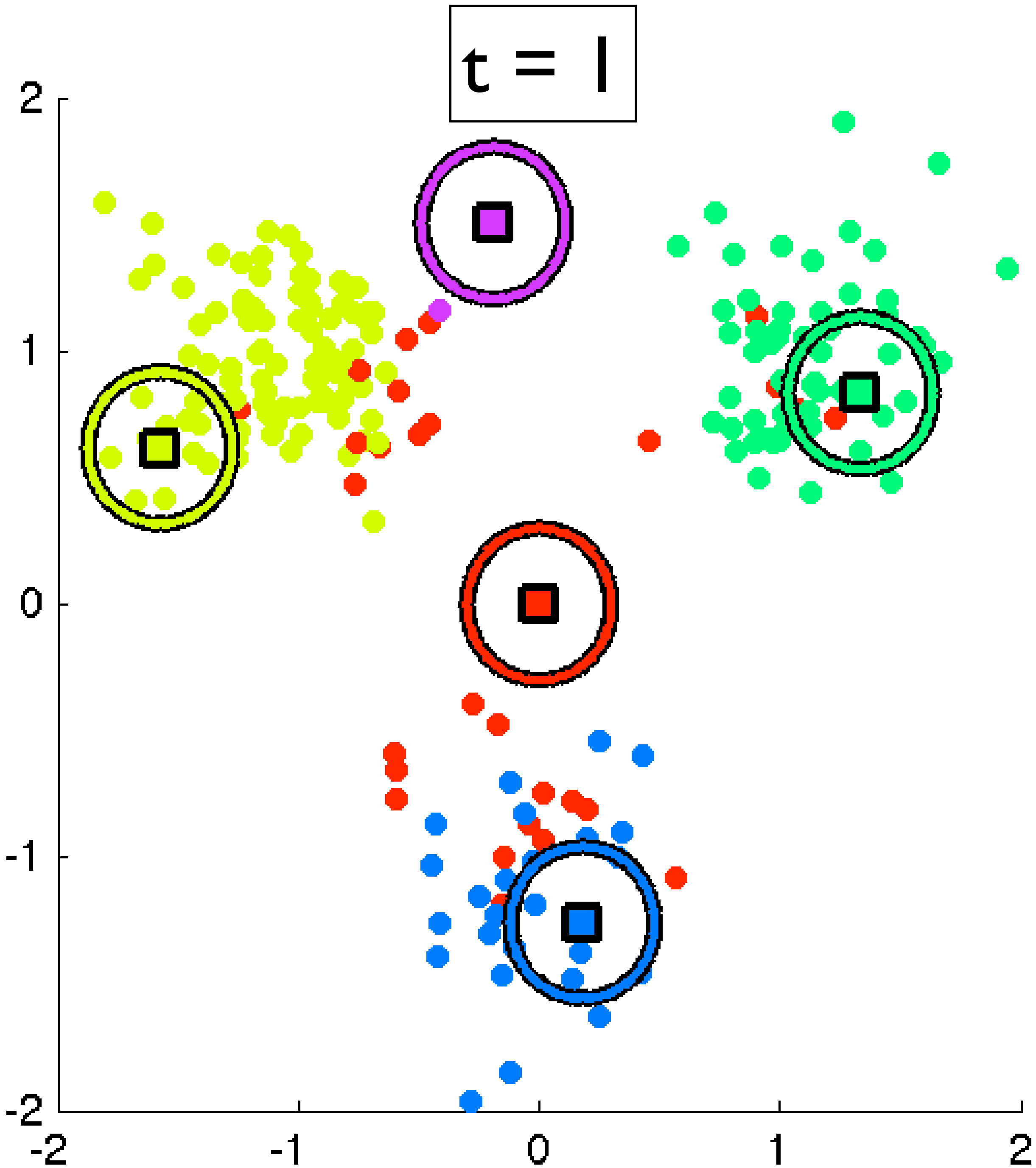


- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

$$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ Sample cluster parameters

EPPF: Calculating posterior

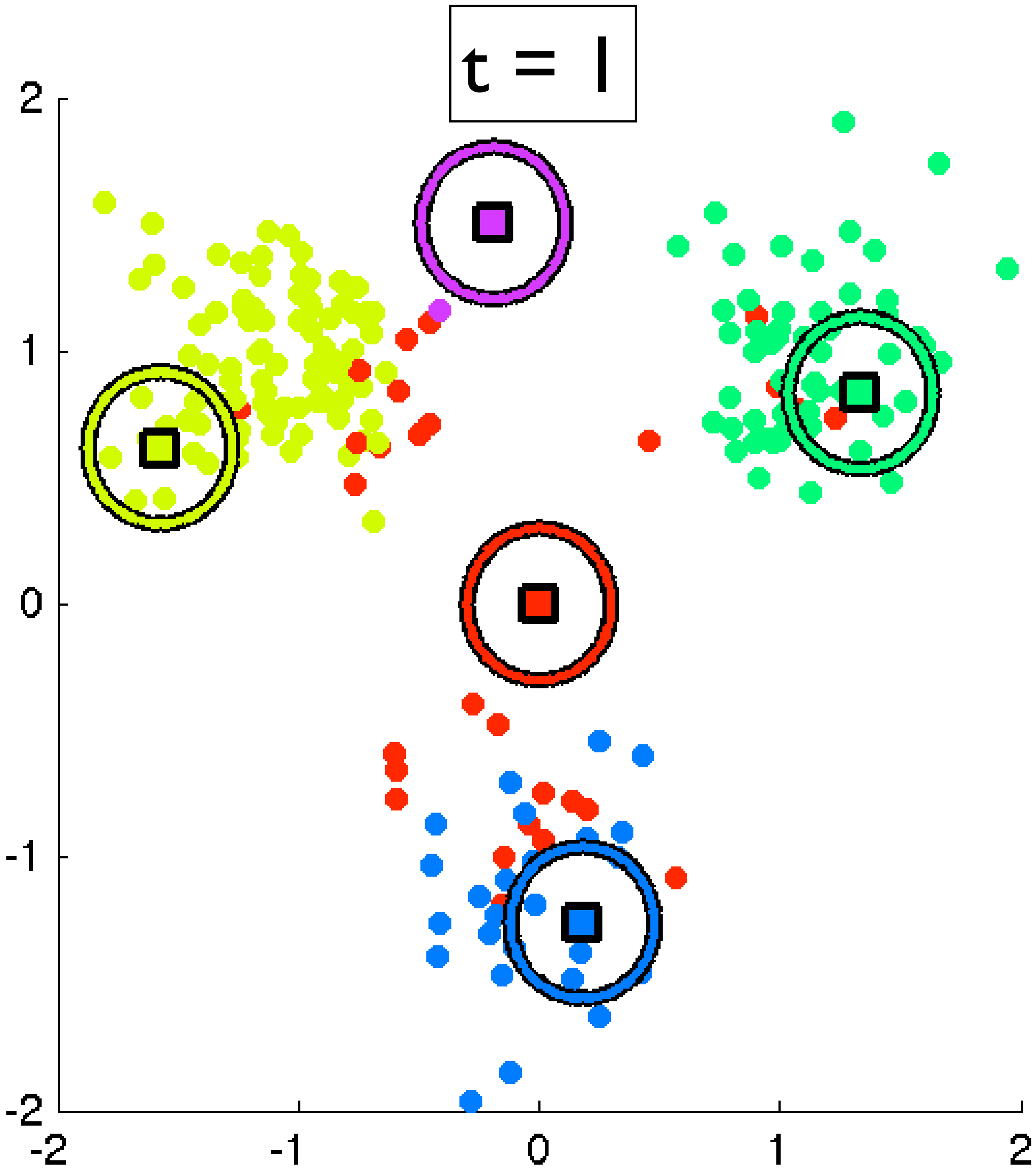


- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ Sample cluster parameters

EPPF: Calculating posterior



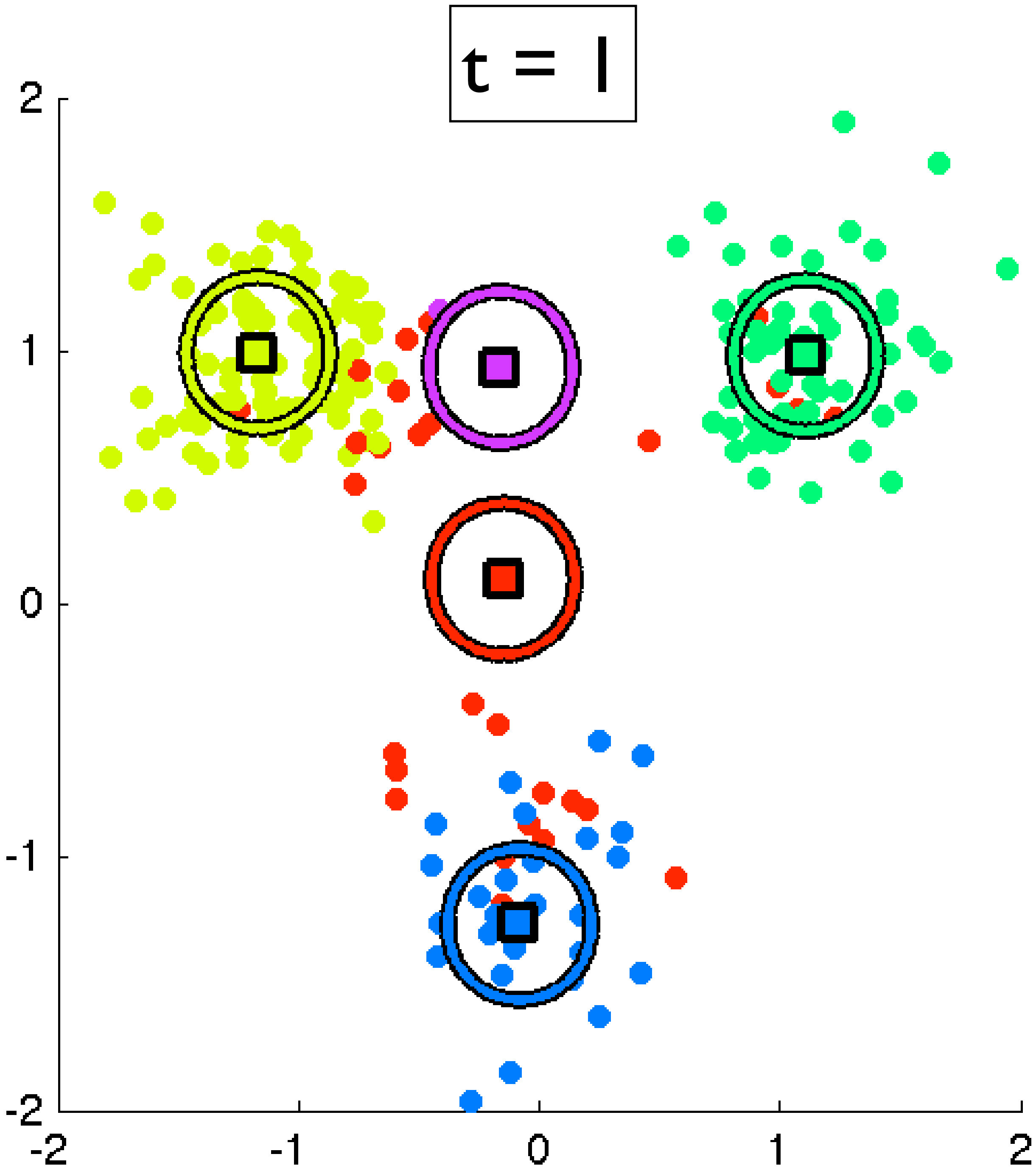
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

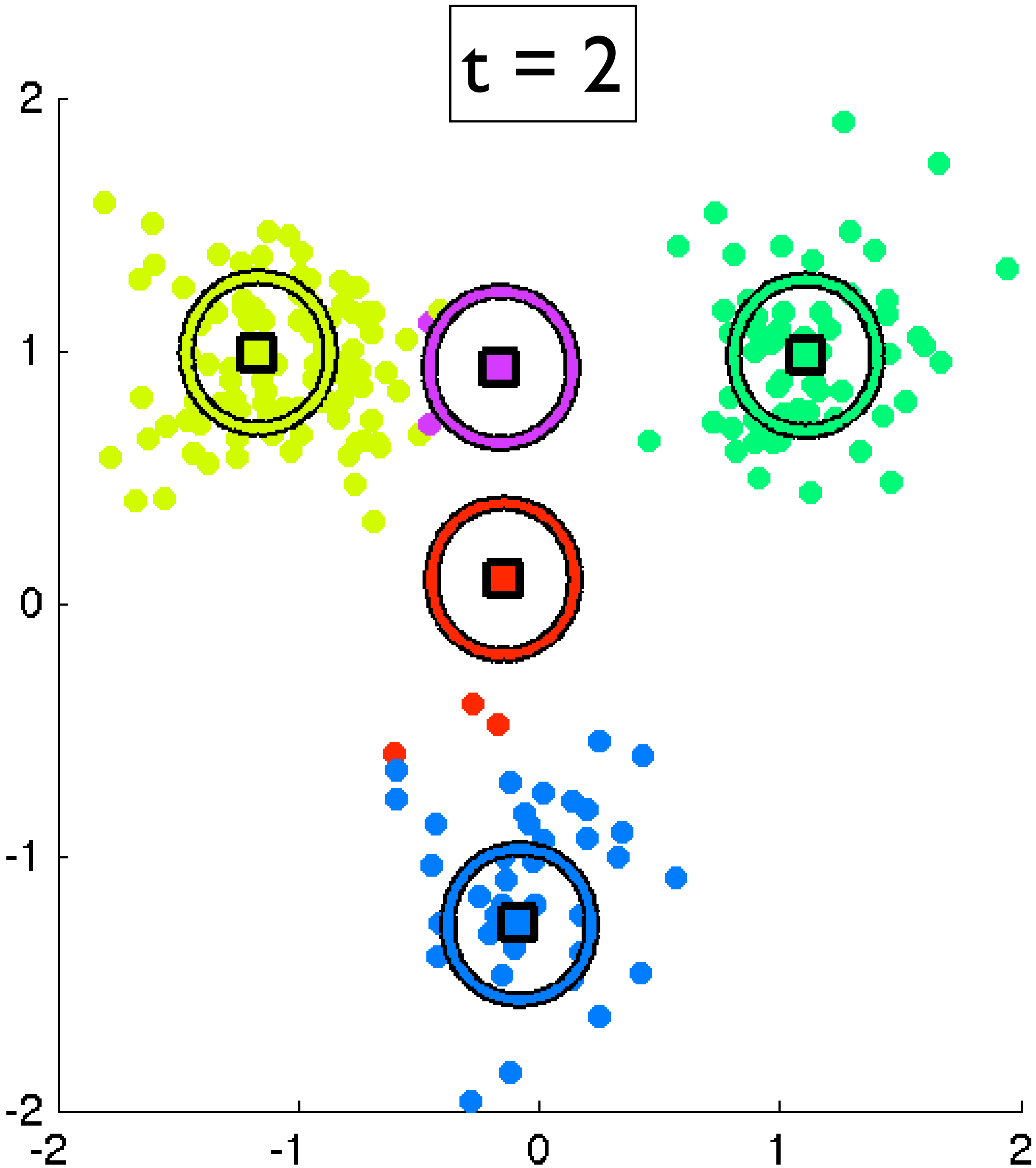
    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N
    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$
  - ◇ For k = 1, ..., K
    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



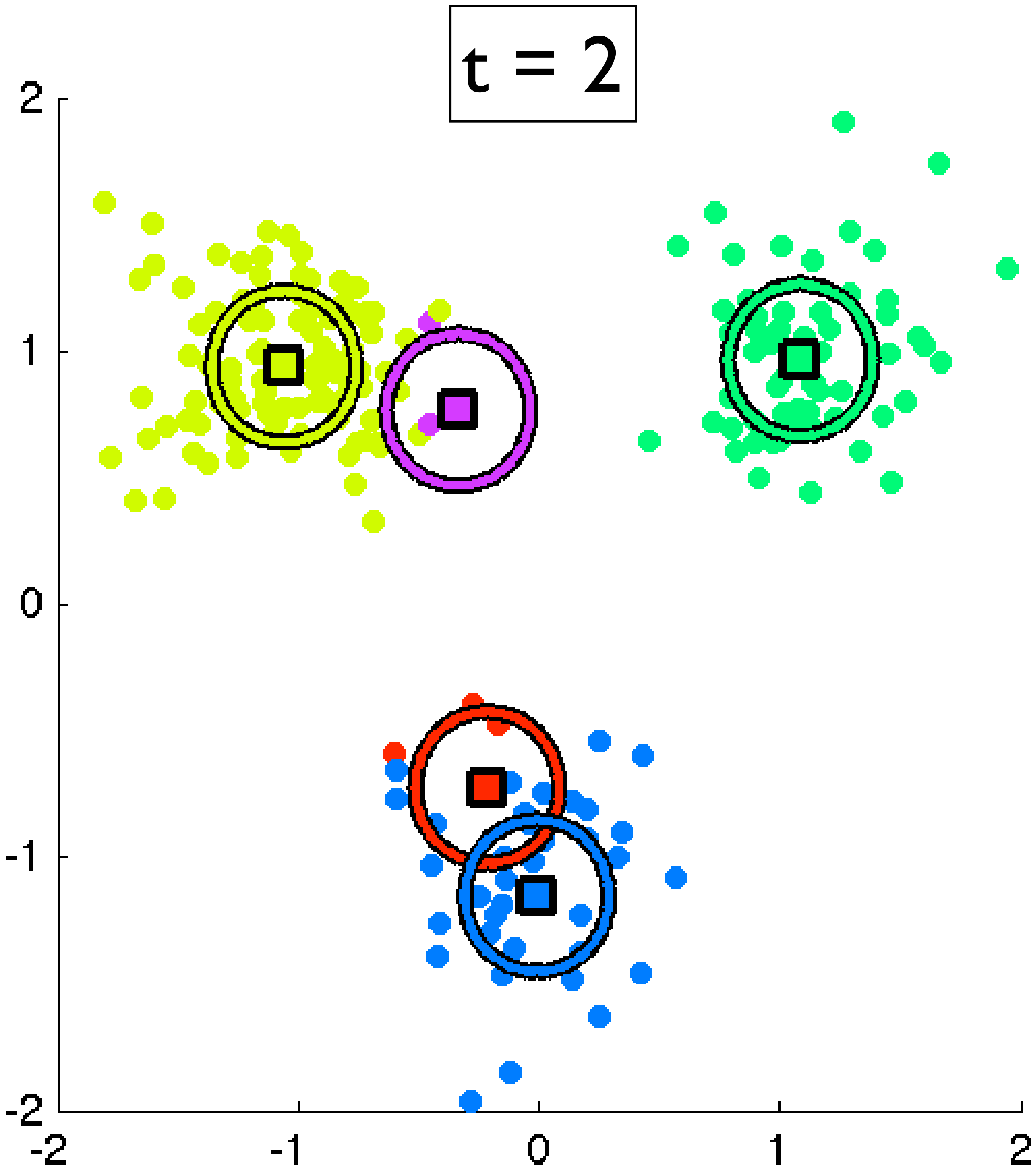
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

$$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

$$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



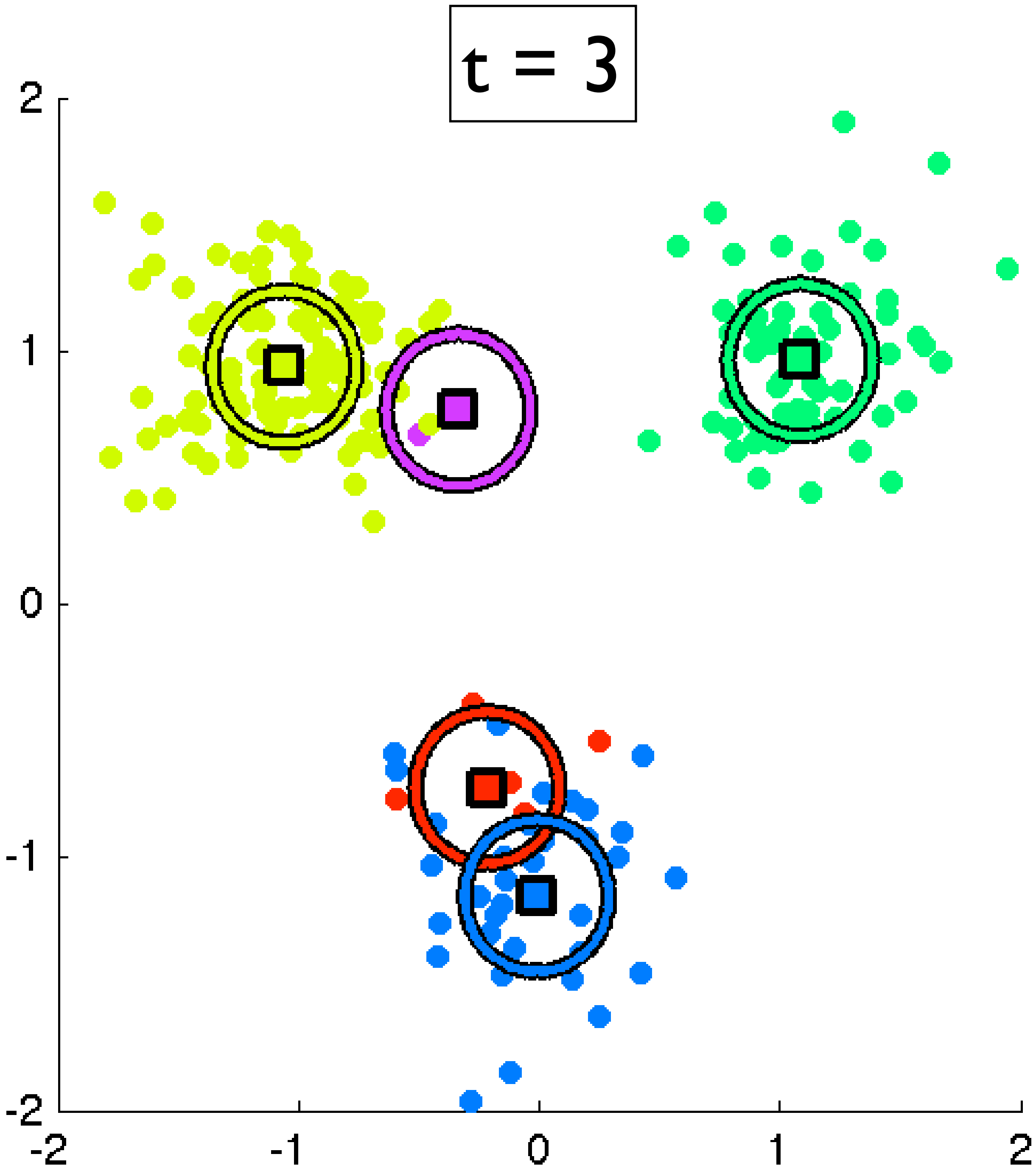
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior

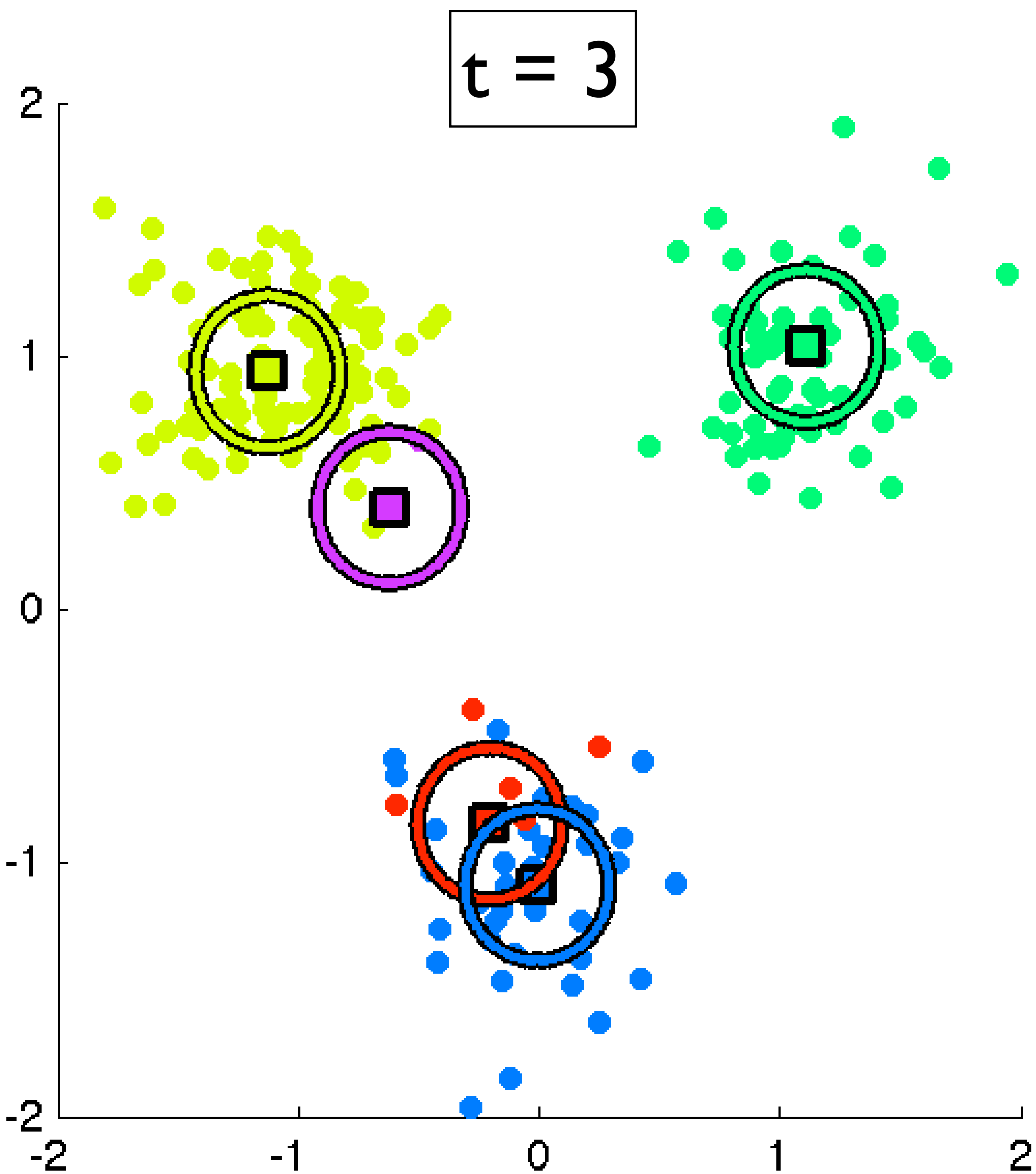


- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$
  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



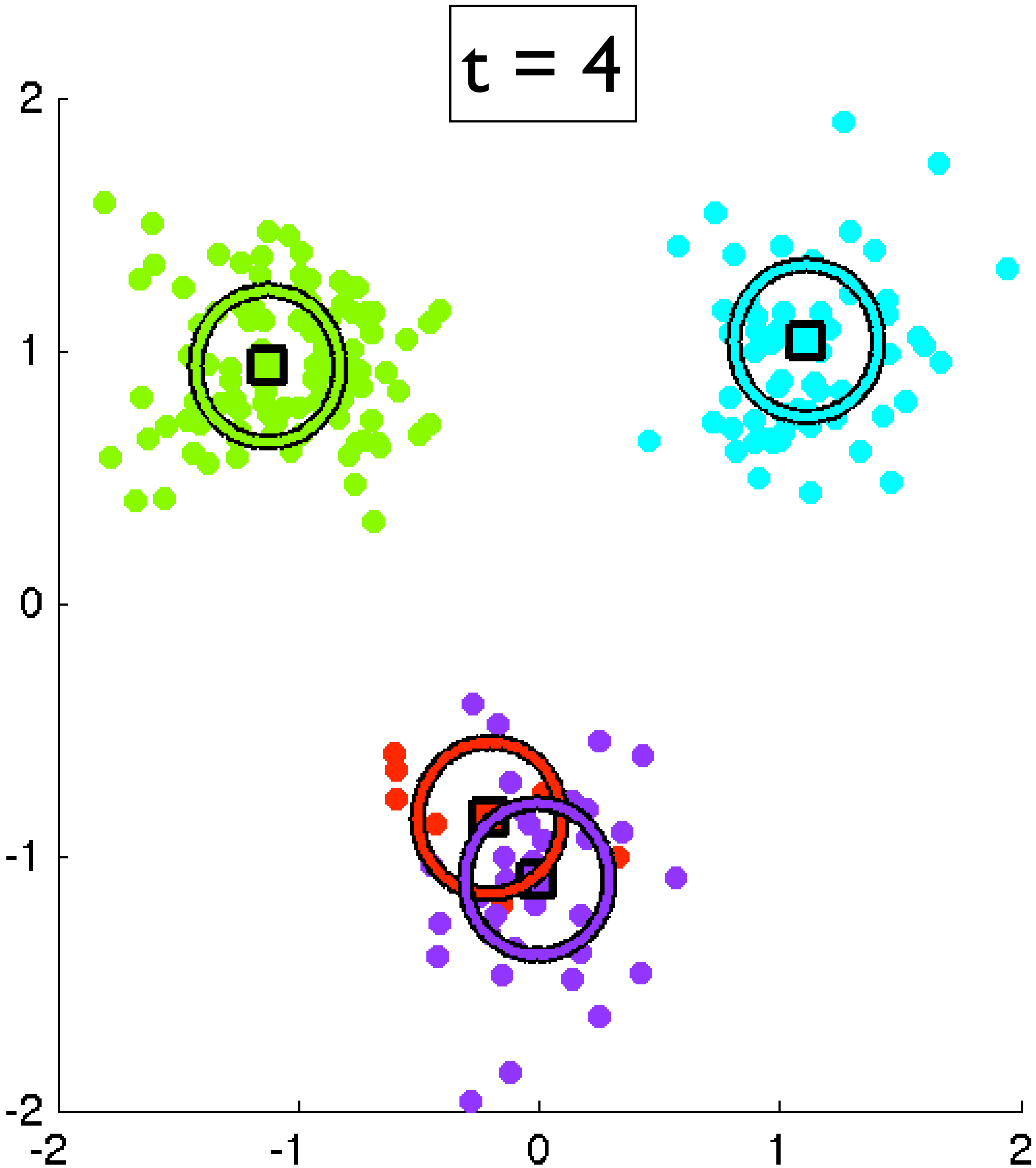
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



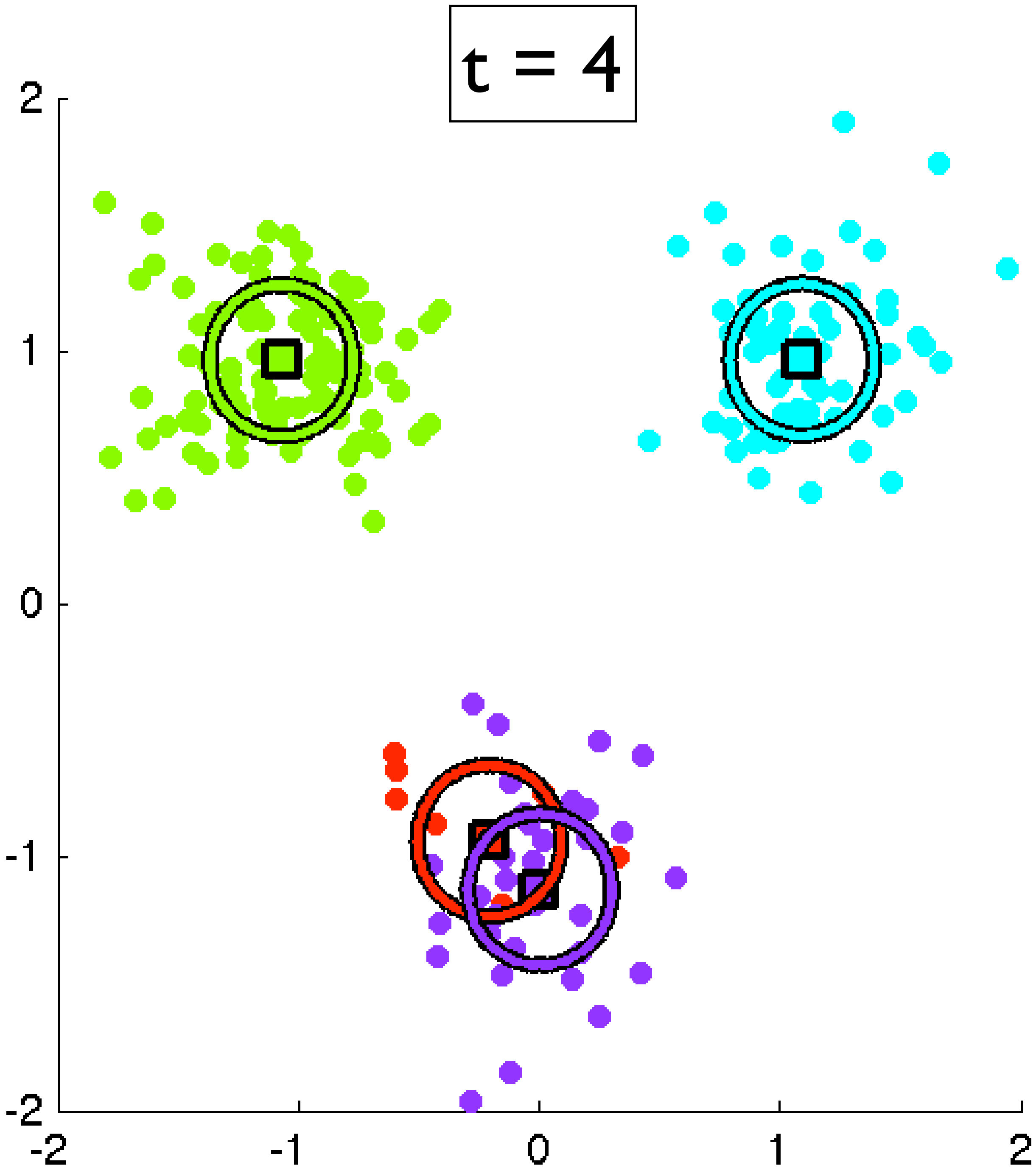
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



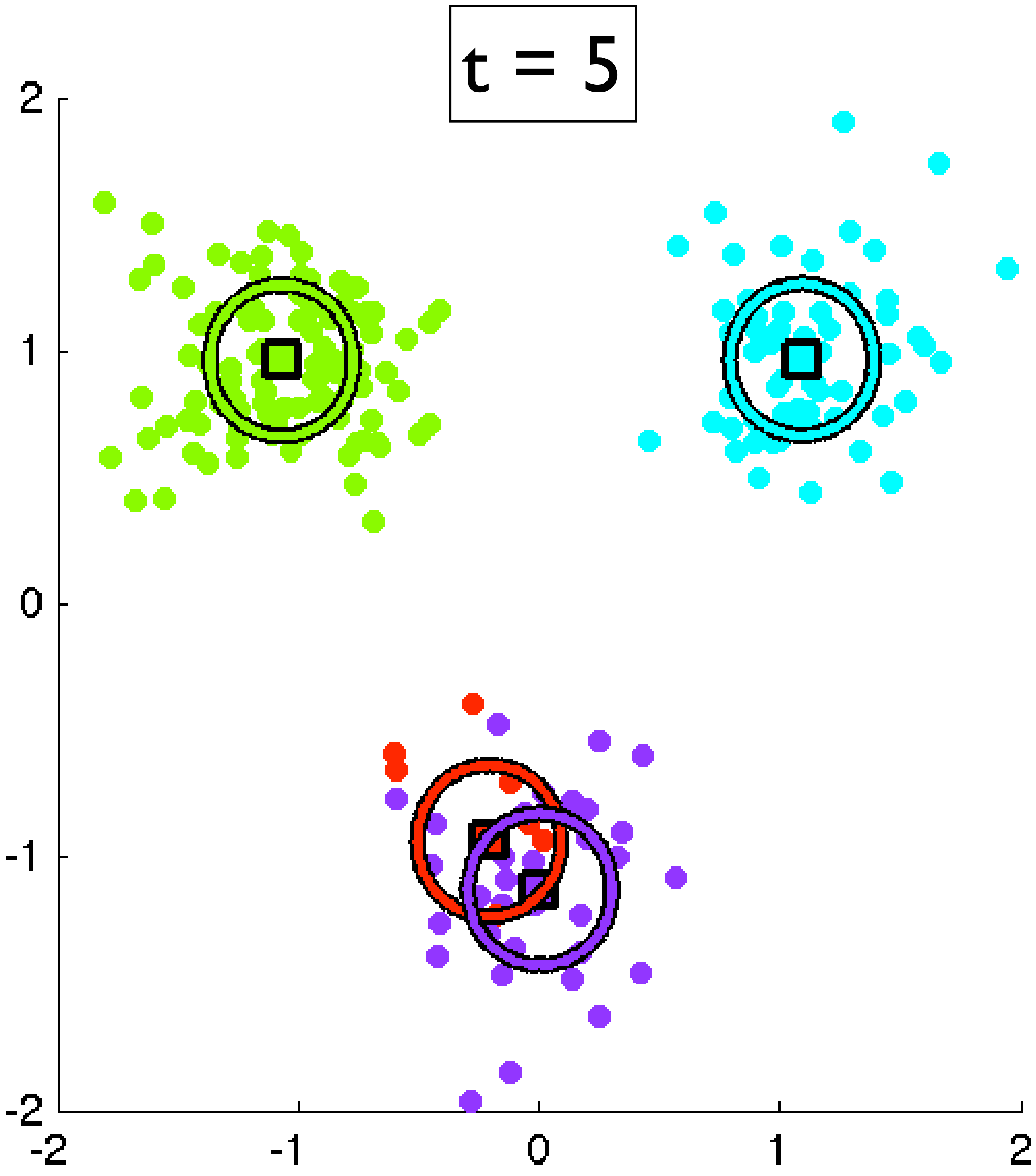
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



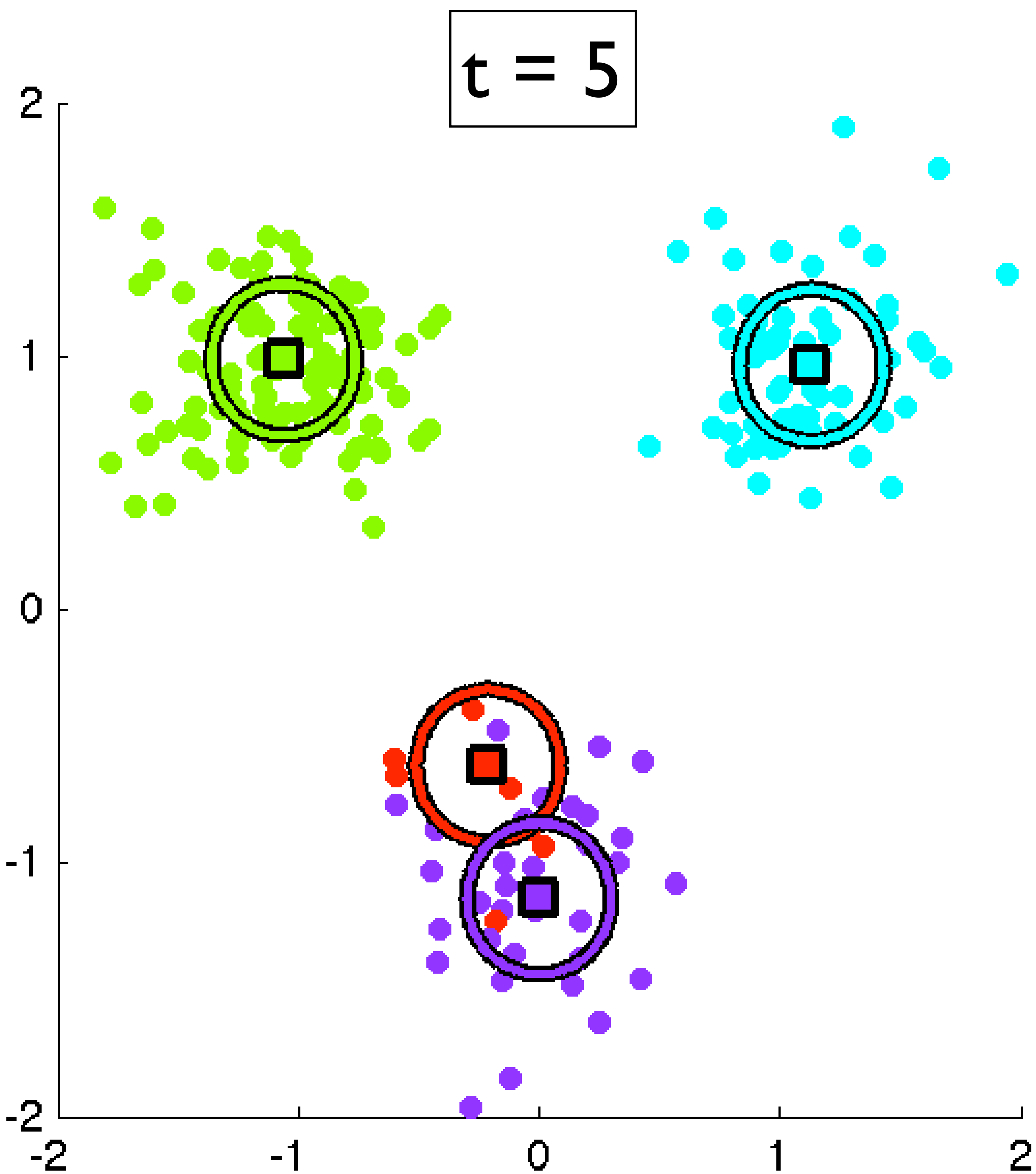
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior

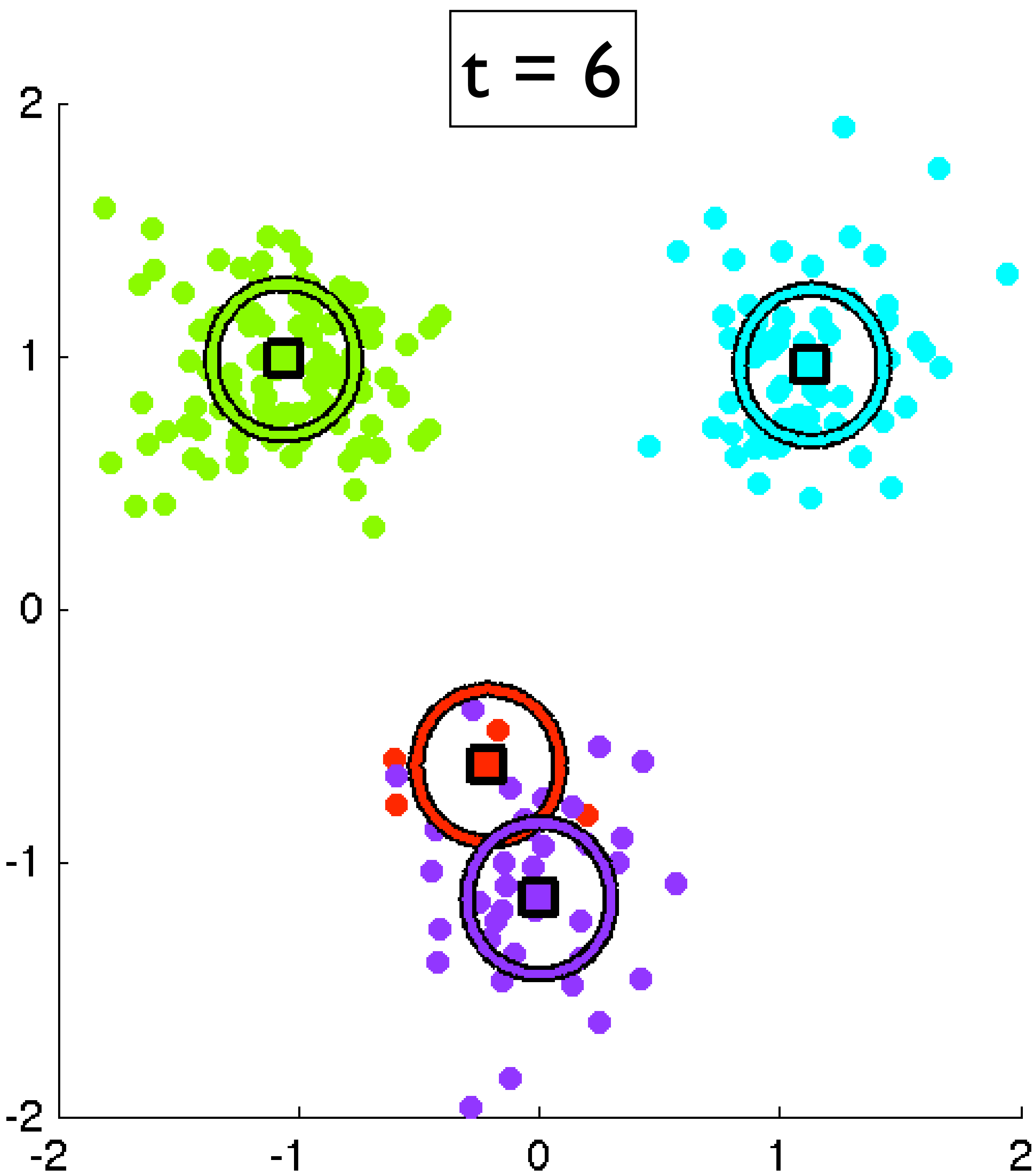


- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$
  - ◇ For k = 1, ..., K

    $\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$

EPPF: Calculating posterior



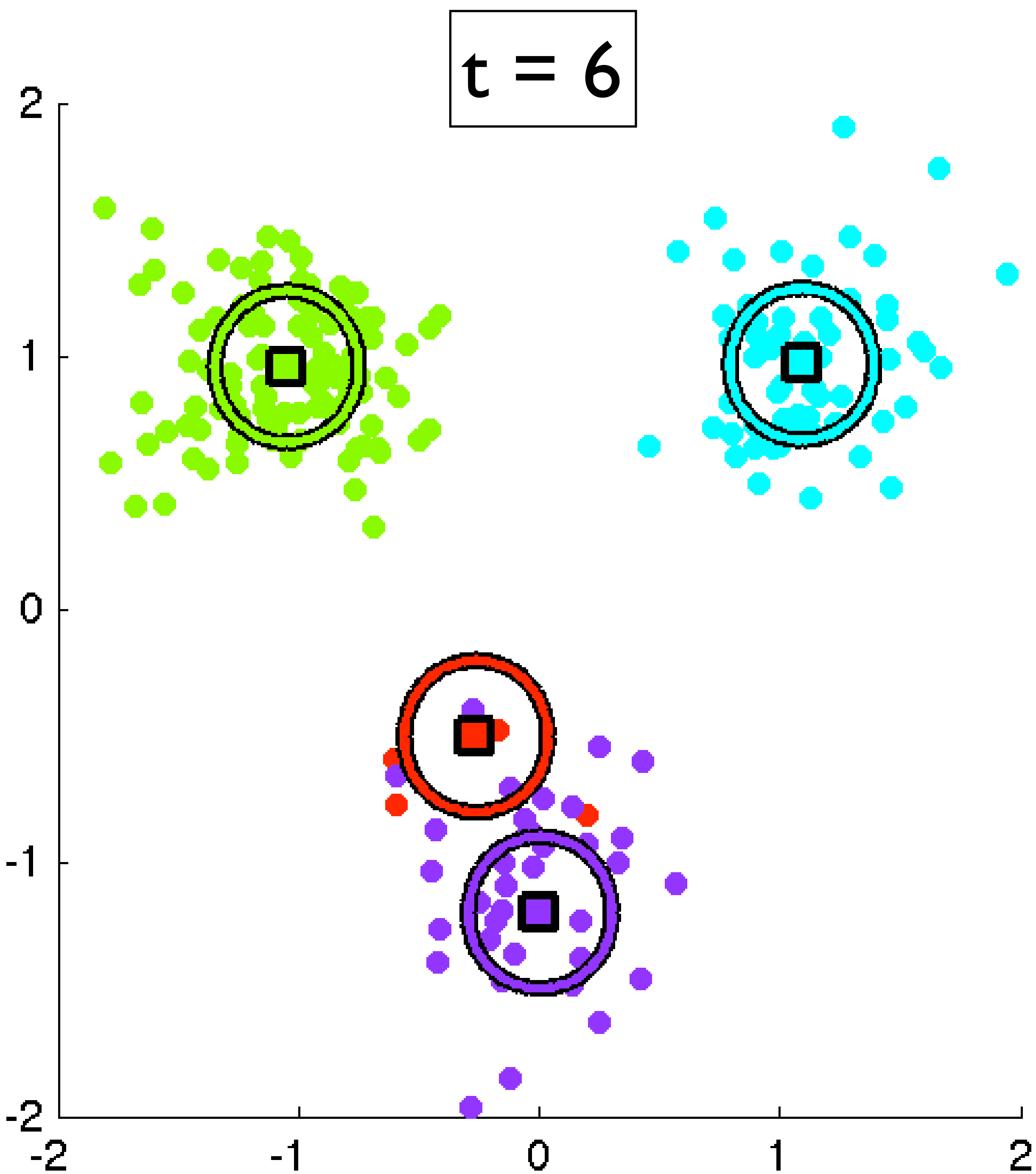
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

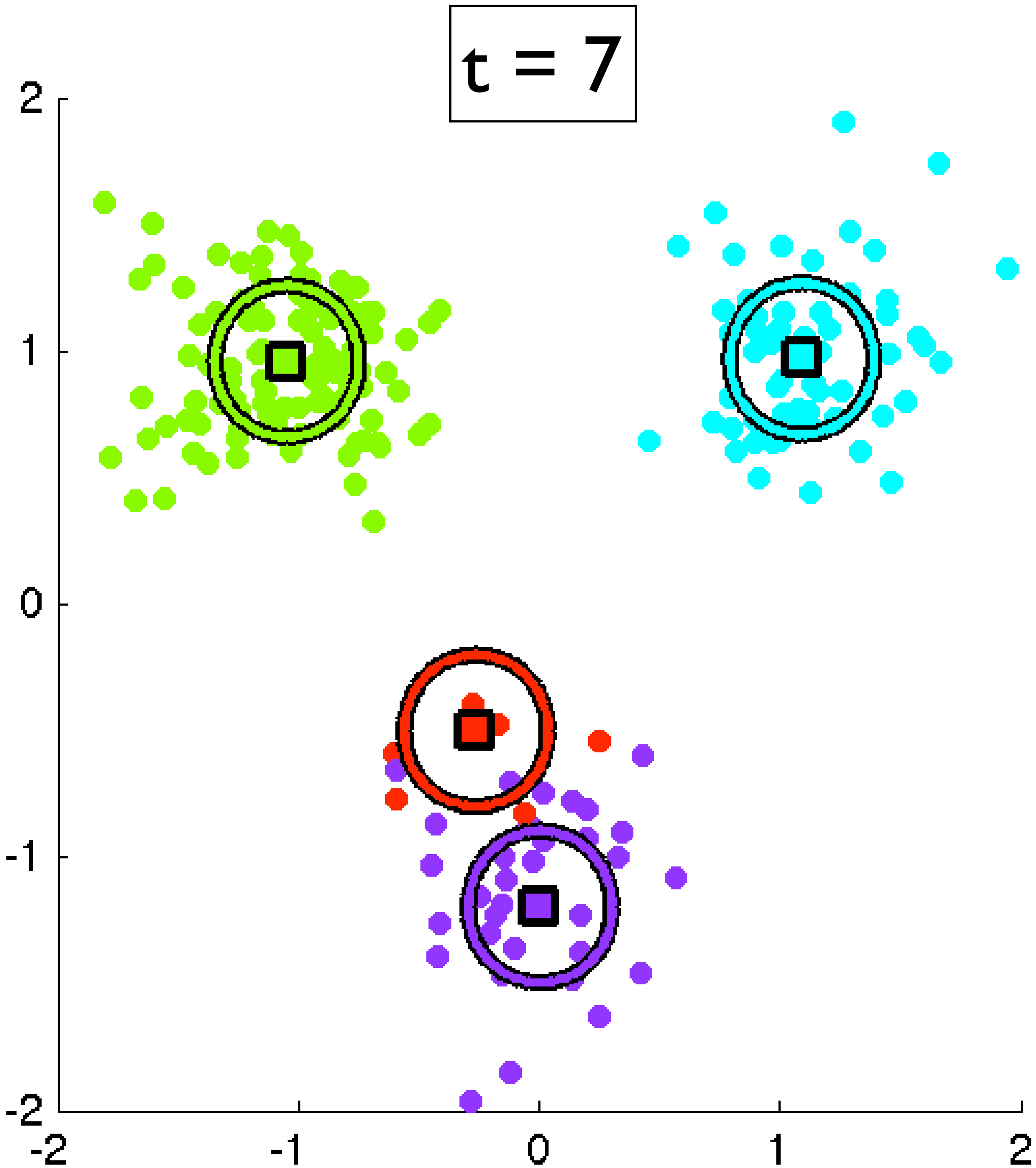
    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N
    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$
  - ◇ For k = 1, ..., K
    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



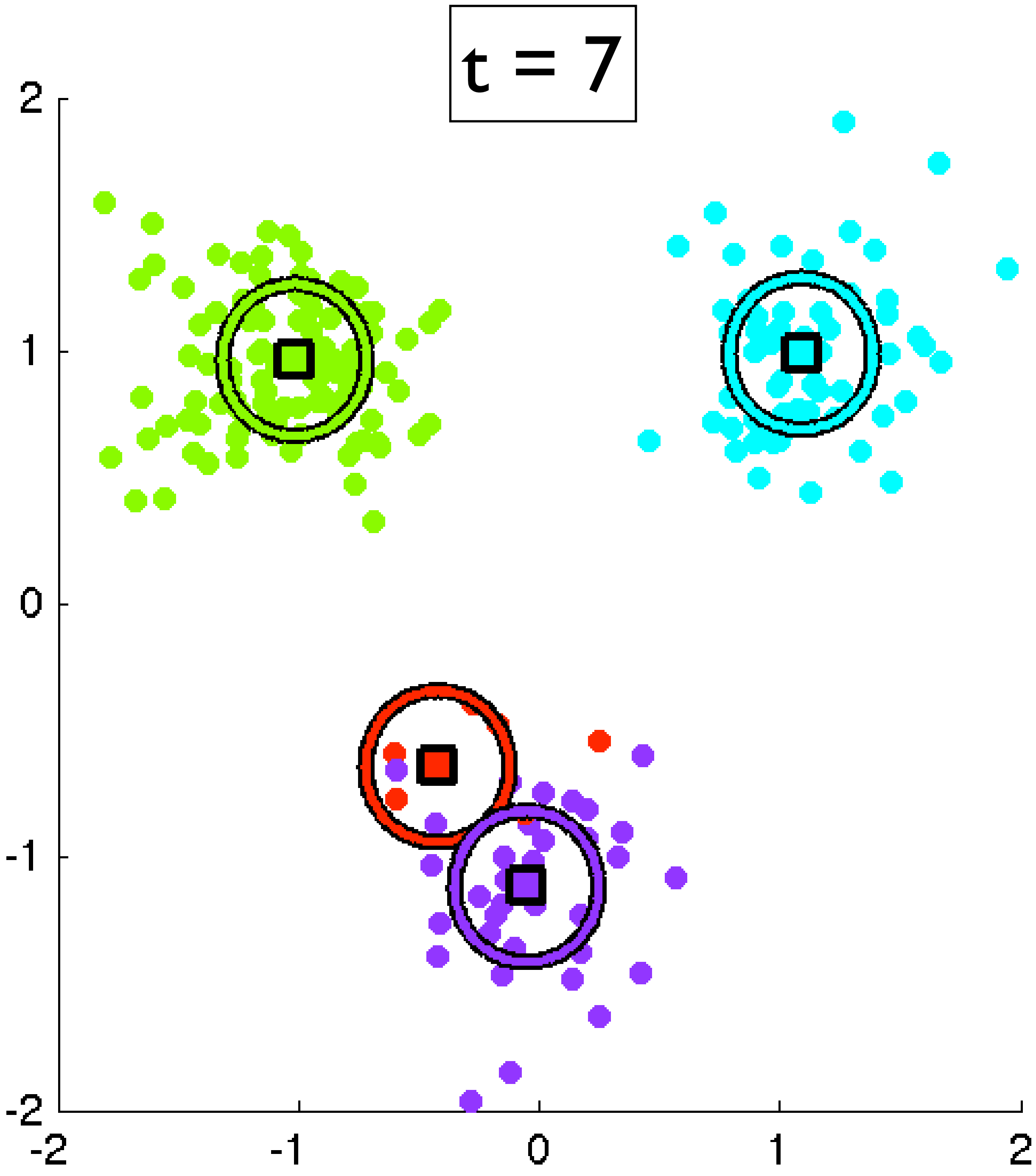
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior

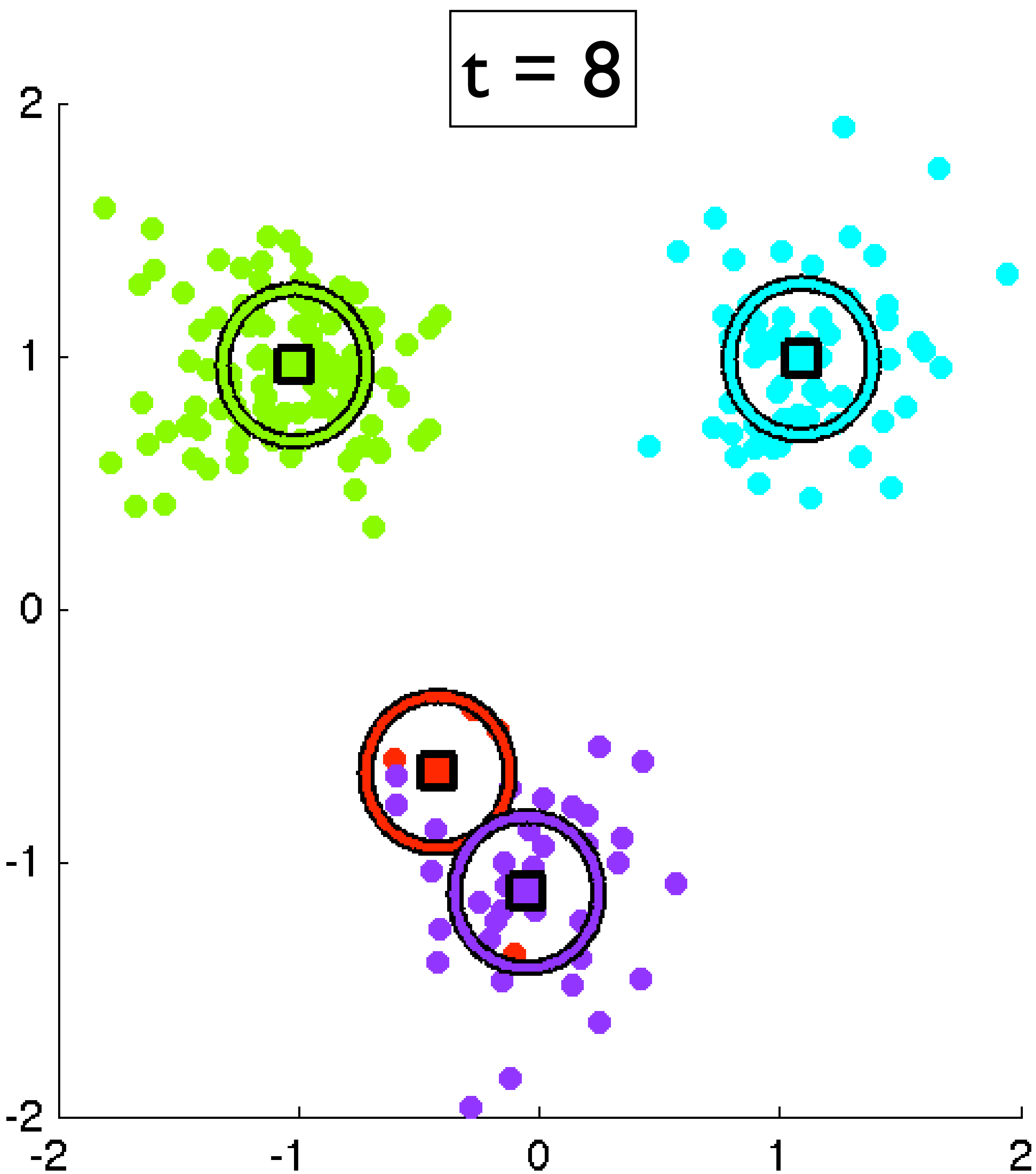


- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $Z_n \sim \mathbb{P}(Z_n|X, \mu, Z_{-n})$
  - ◇ For k = 1, ..., K

    $\mu_k \sim \mathbb{P}(\mu_k|X, Z, \mu_{-k})$

EPPF: Calculating posterior



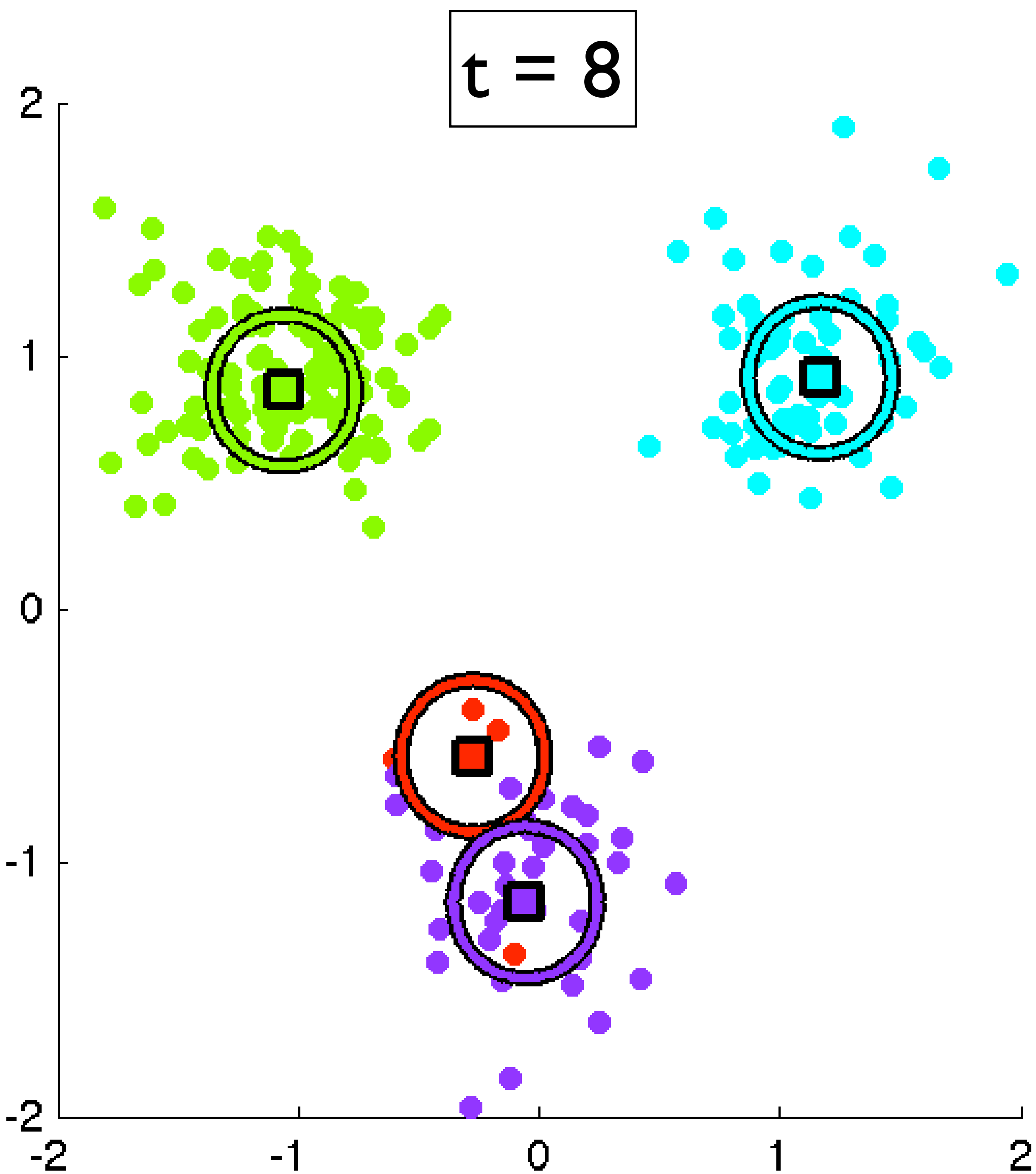
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



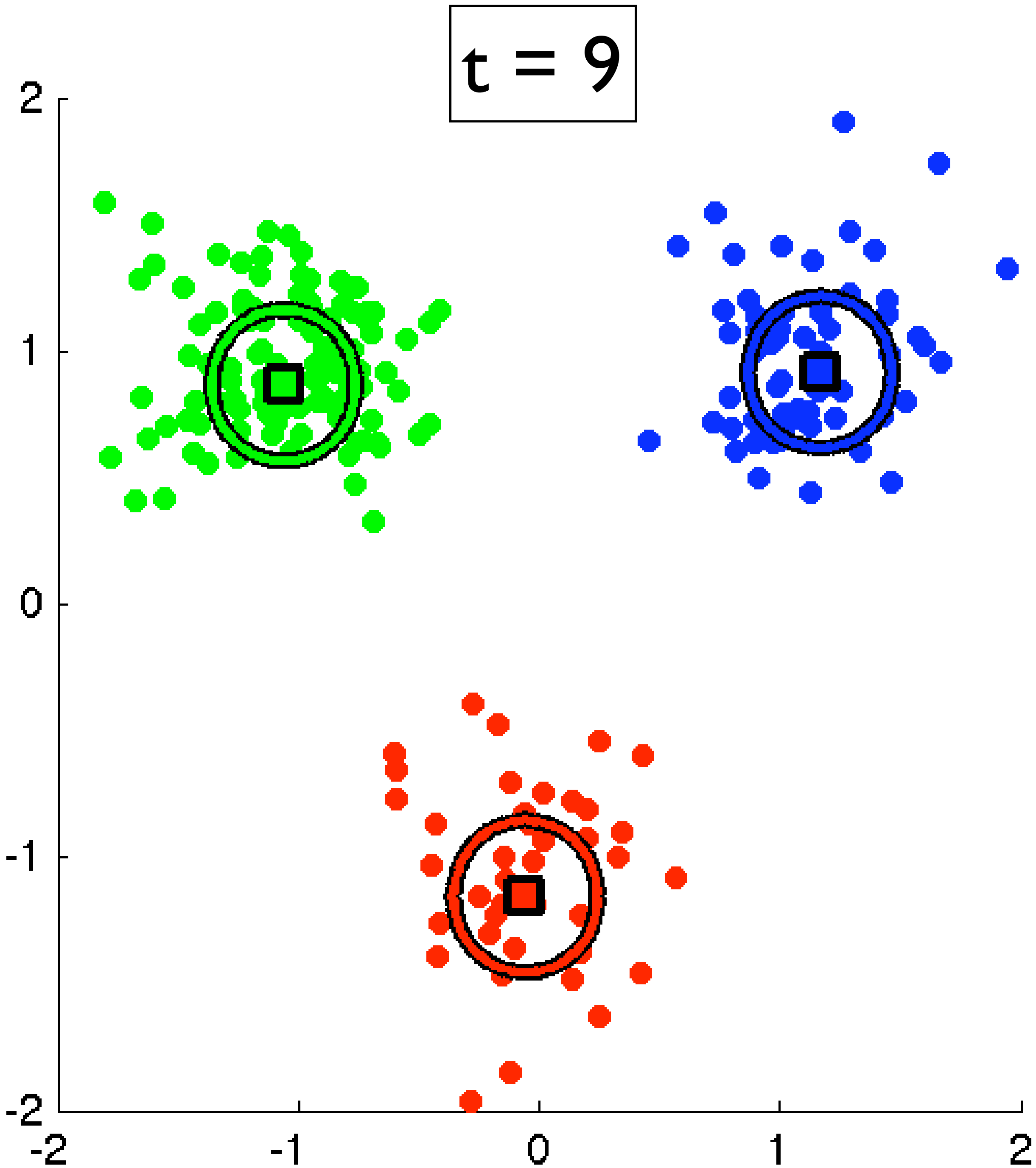
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



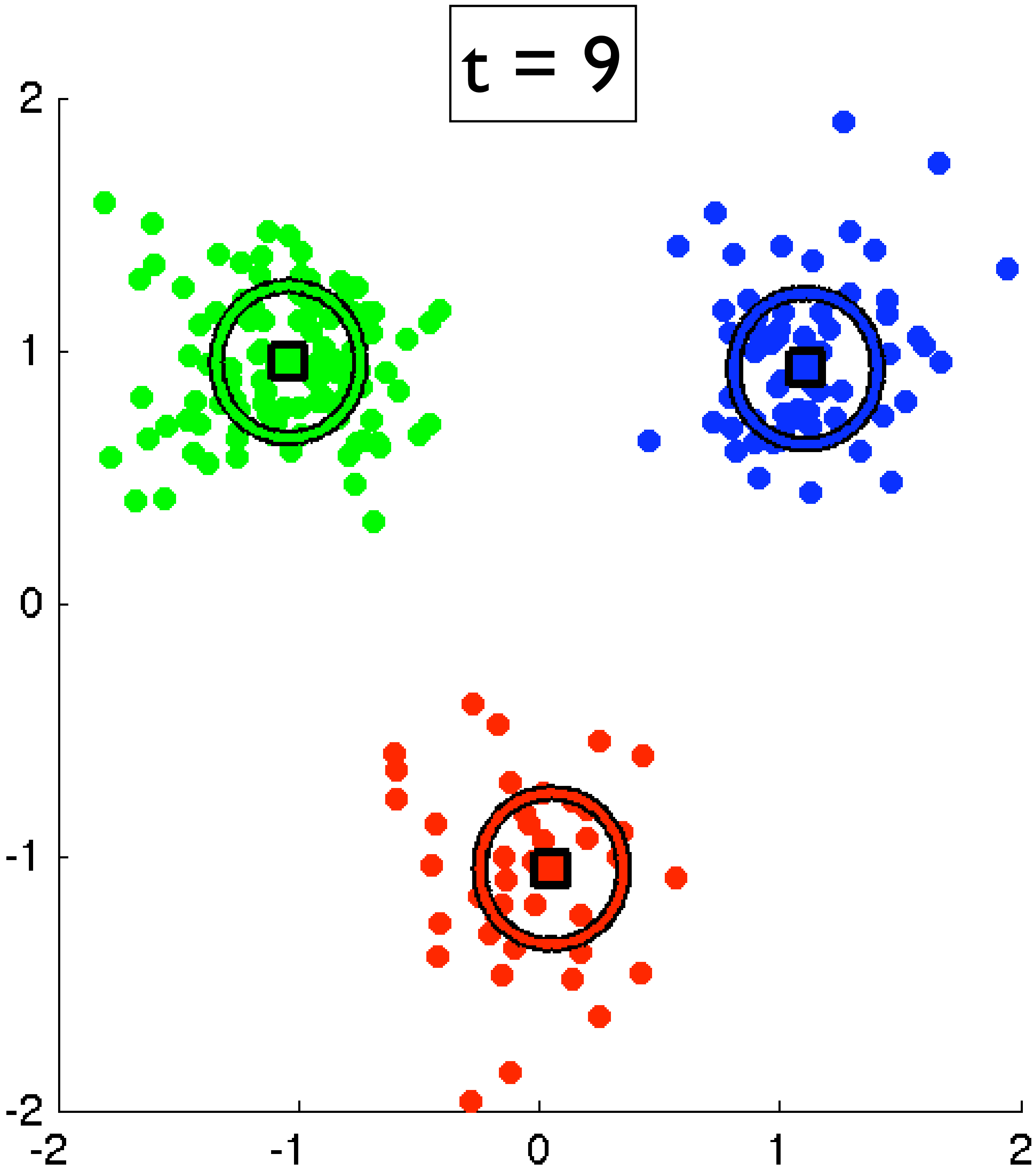
- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N

    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

  - ◇ For k = 1, ..., K

    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior



- Assign all points to one cluster
- For t = 1, ..., T
  - ◇ For n = 1, ..., N
    $$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$
  - ◇ For k = 1, ..., K
    $$\mu_k \sim \mathbb{P}(\mu_k | X, Z, \mu_{-k})$$

EPPF: Calculating posterior

Gibbs sampling: potential issues

EPPF: Calculating posterior

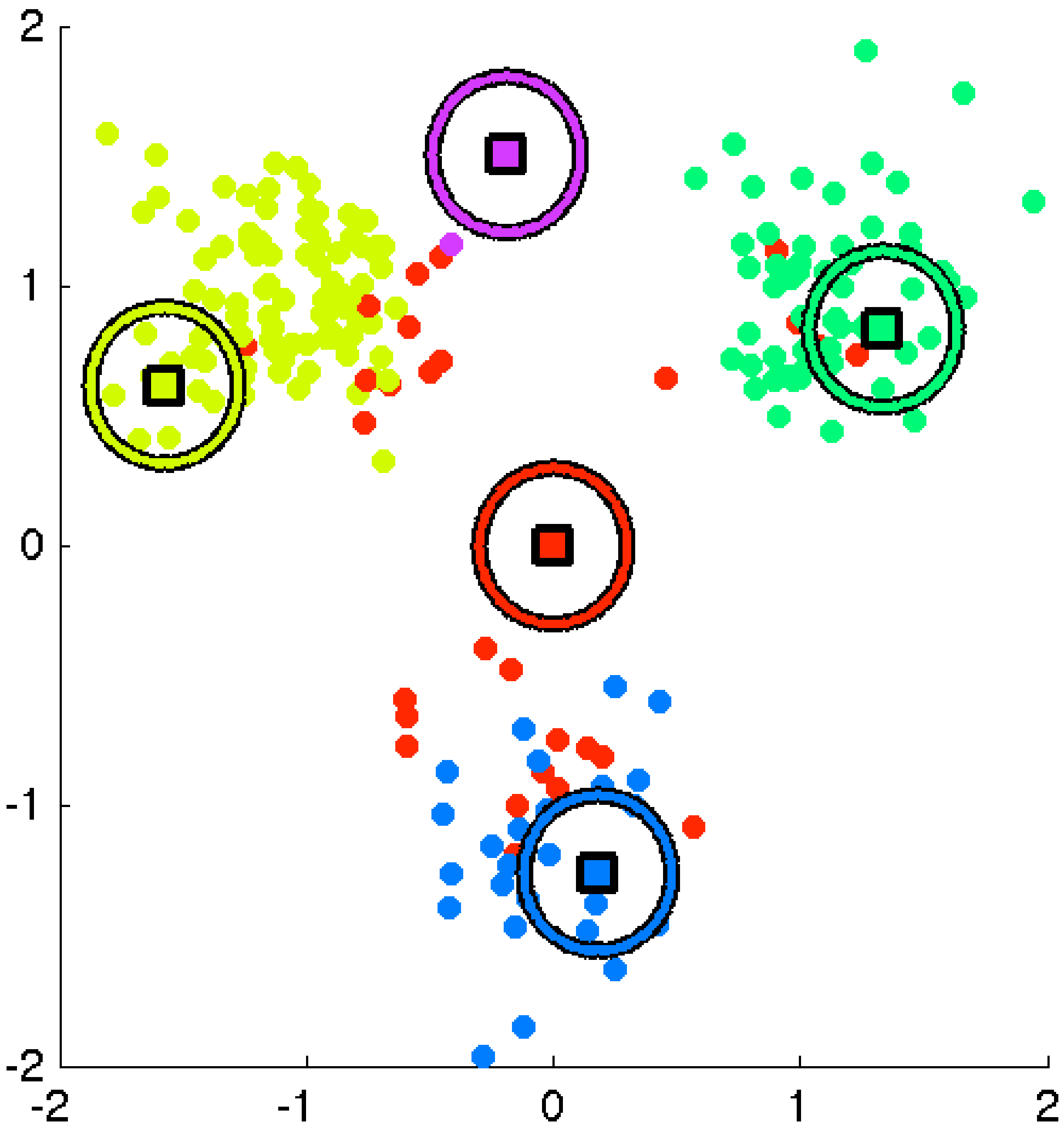
Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter

EPPF: Calculating posterior

Gibbs sampling: potential issues

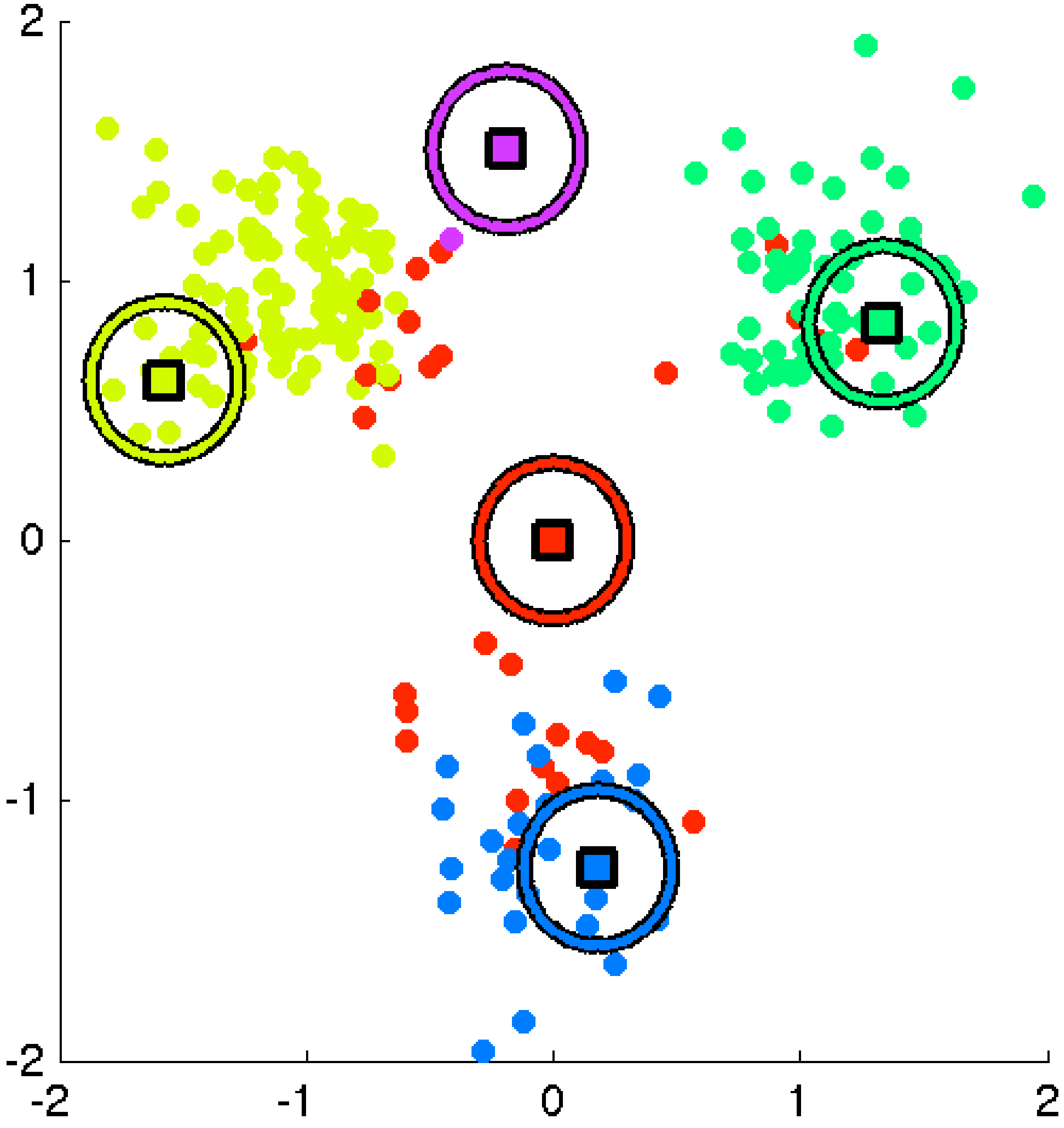
- Bad mixing from dependence on cluster parameter



EPPF: Calculating posterior

Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter



Instead try:

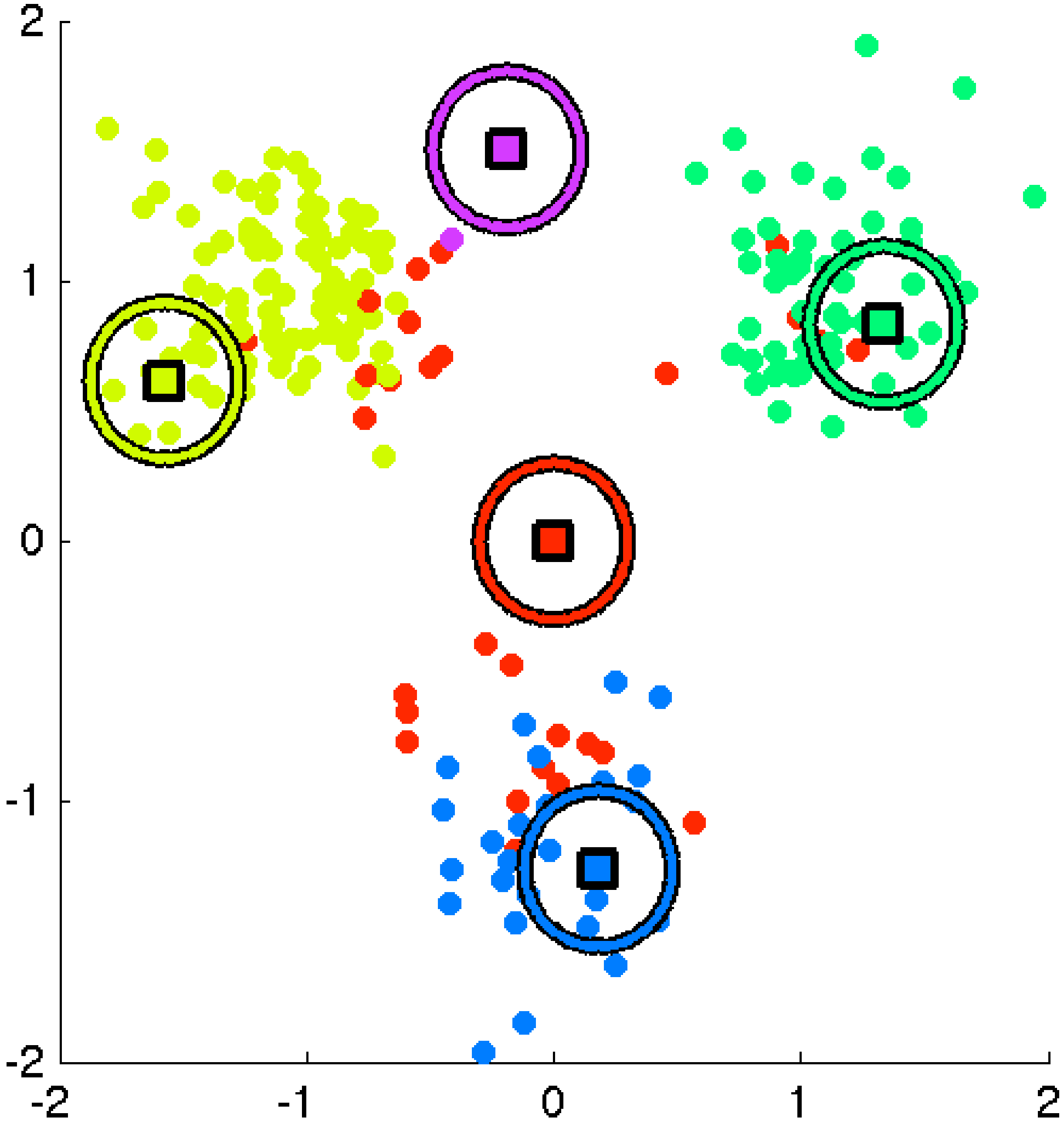
collapsed sampler

[Neal 1992; MacEachern 1994; Neal 2000]

EPPF: Calculating posterior

Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter



Instead try:

collapsed sampler

- Instead of $\mathbb{P}(Z, \mu | X)$ learn $\mathbb{P}(Z | X)$

[Neal 1992; MacEachern 1994; Neal 2000]

EPPF: Calculating posterior

Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter
- Bad mixing since each indicator depends on rest

EPPF: Calculating posterior

Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter
- Bad mixing since each indicator depends on rest

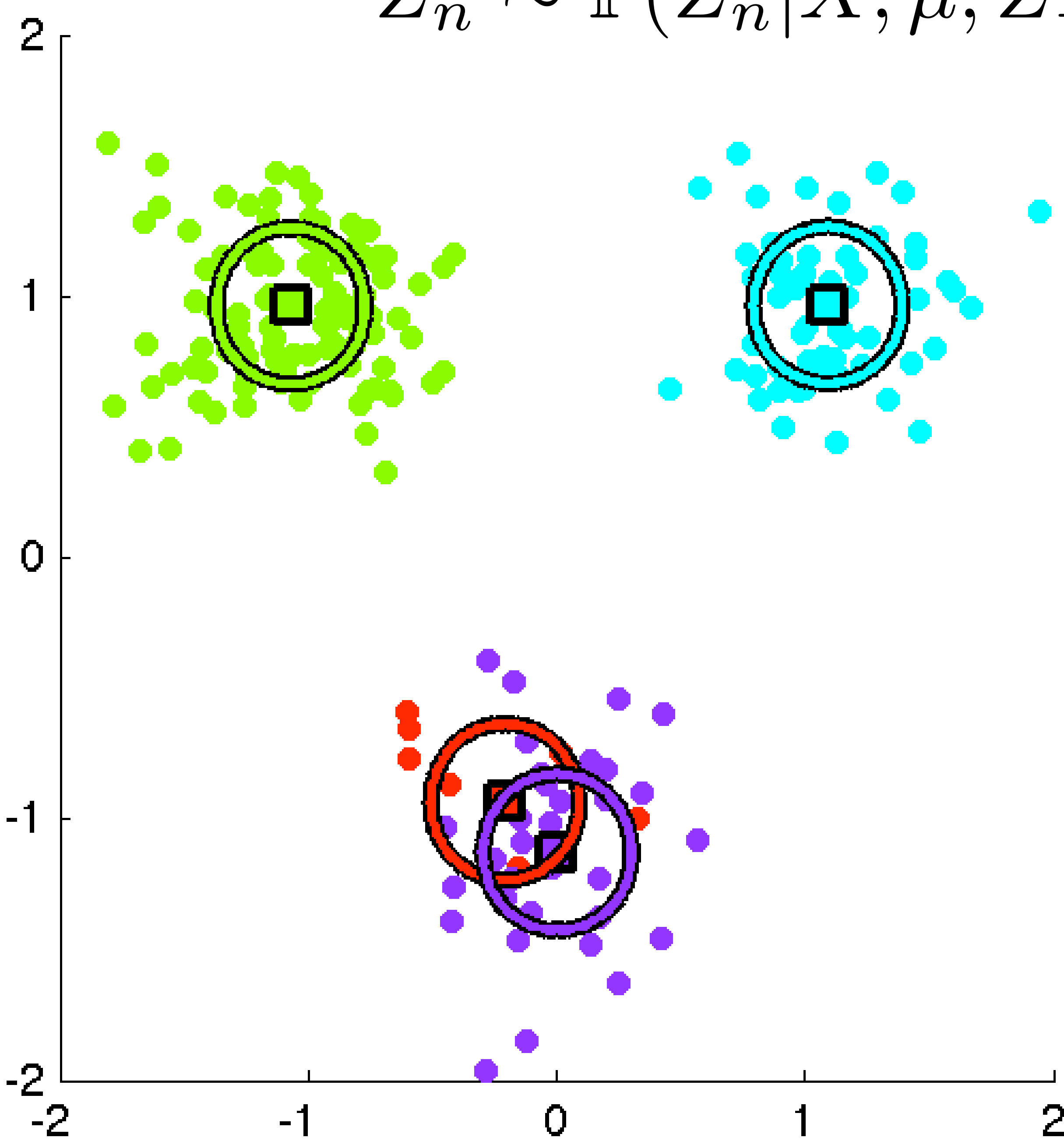
$$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

EPPF: Calculating posterior

Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter
- Bad mixing since each indicator depends on rest

$$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$

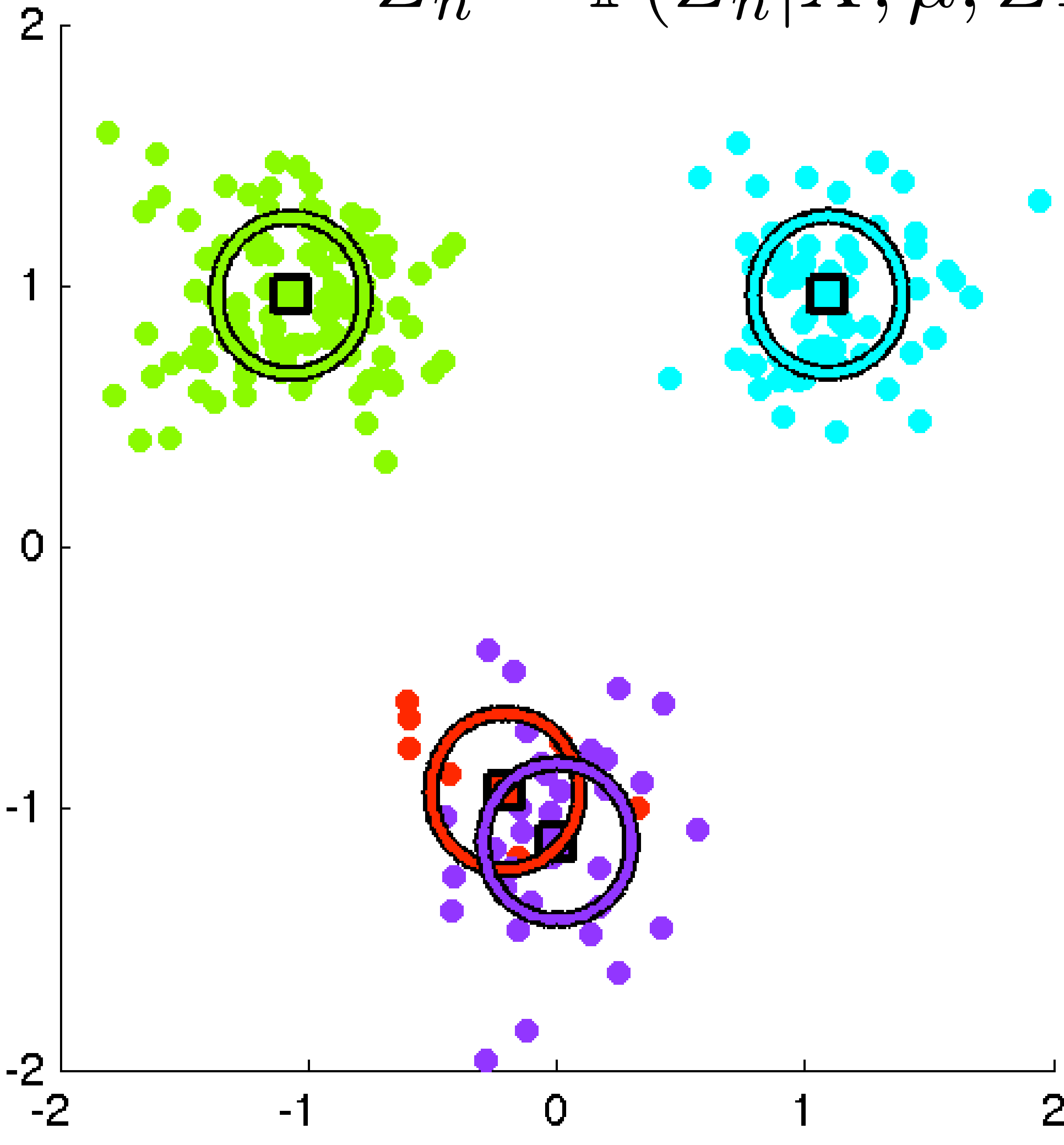


EPPF: Calculating posterior

Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter
- Bad mixing since each indicator depends on rest

$$Z_n \sim \mathbb{P}(Z_n | X, \mu, Z_{-n})$$



Instead try:

split-merge sampler

[Jain, Neal 2000]

EPPF: Calculating posterior

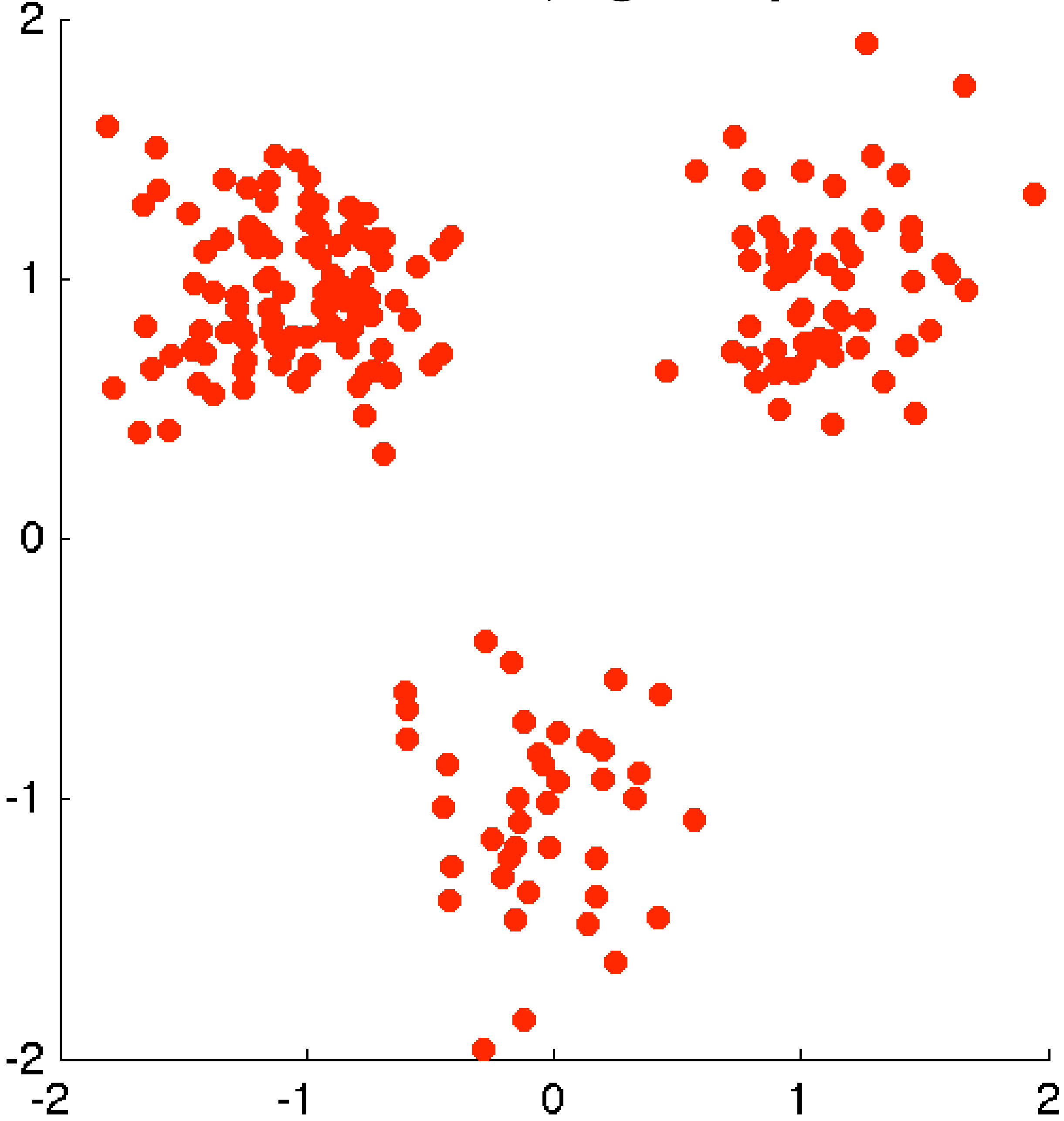
Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter
- Bad mixing since each indicator depends on rest
- Non-conjugate prior

EPPF: Calculating posterior

Gibbs sampling: potential issues

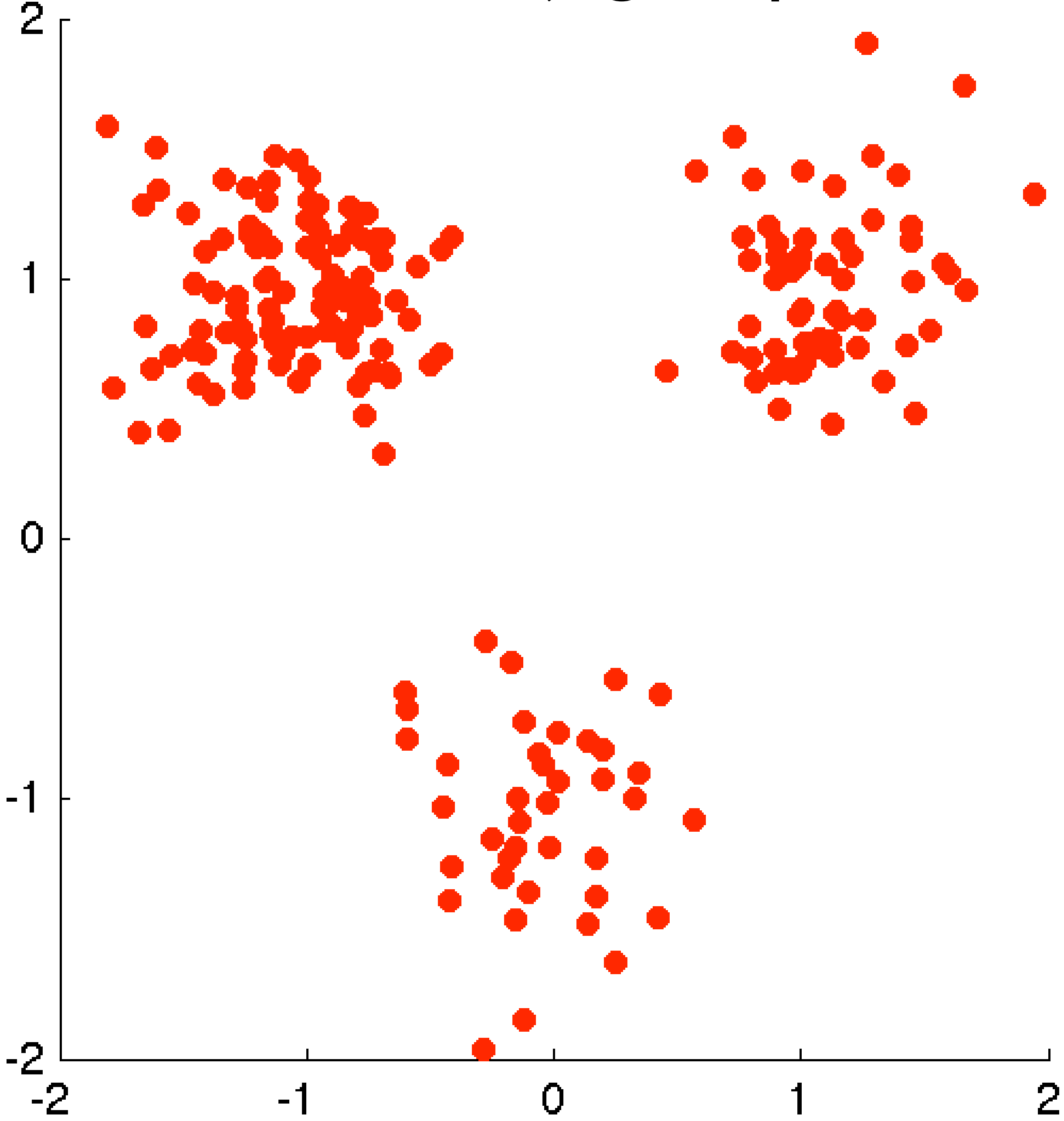
- Bad mixing from dependence on cluster parameter
- Bad mixing since each indicator depends on rest
- Non-conjugate prior



EPPF: Calculating posterior

Gibbs sampling: potential issues

- Bad mixing from dependence on cluster parameter
- Bad mixing since each indicator depends on rest
- Non-conjugate prior



Instead try:

Metropolis Hastings, auxiliary variables, etc

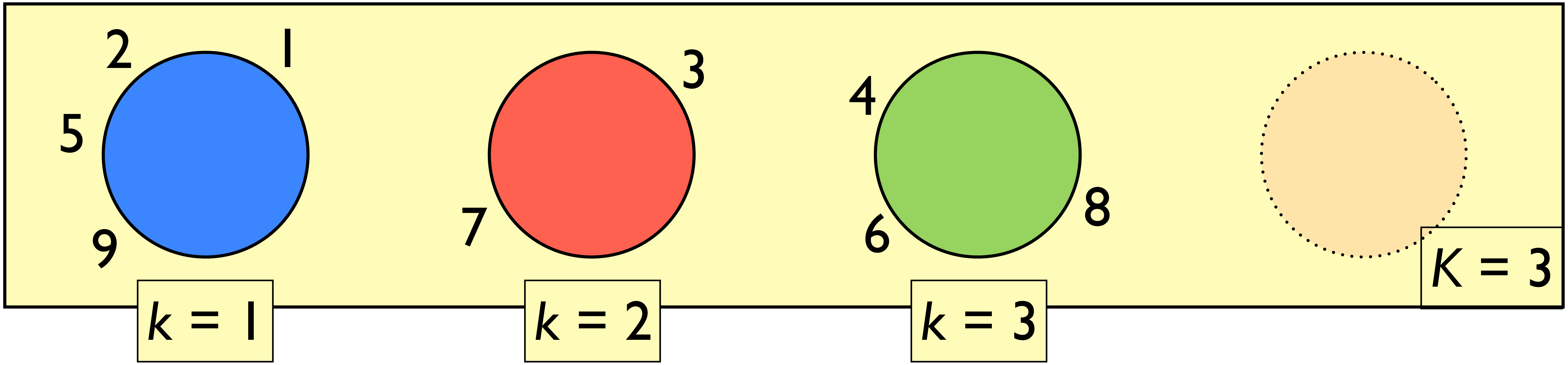
[Neal 2000]

Cluster labels

- For previous Gibbs sampler, choose by computational convenience

Cluster labels

Order of appearance



Cluster labels

Order of appearance

k

n

	1	2	3	4
1				
2				
3				
4				
5				
6				
7				
8				
9				

[Pitman 2006]

Cluster labels

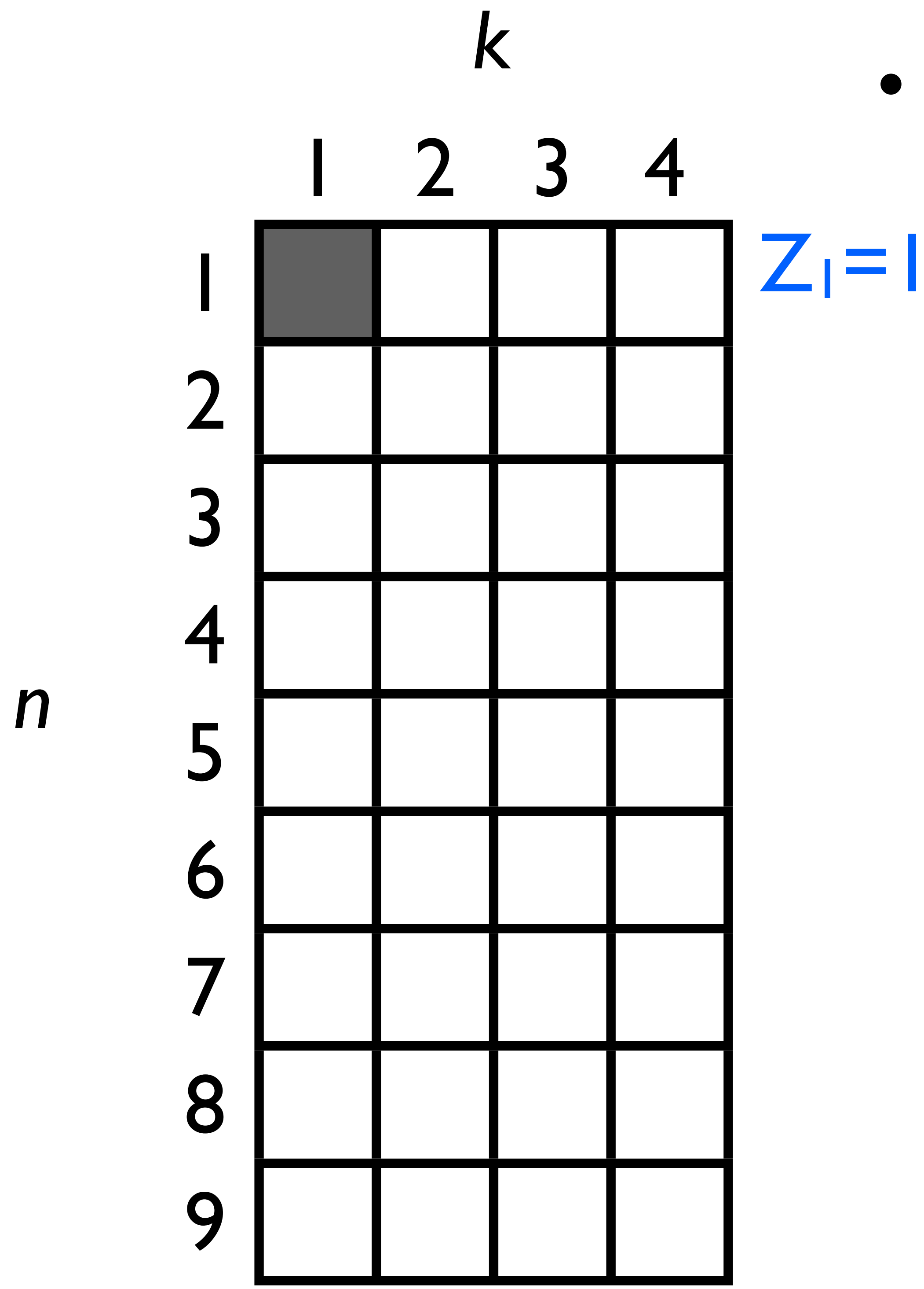
Order of appearance

n \ *k*	1	2	3	4
1				
2				
3				
4				
5				
6				
7				
8				
9				

- Recursively: *n*th person sits
    - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
    - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

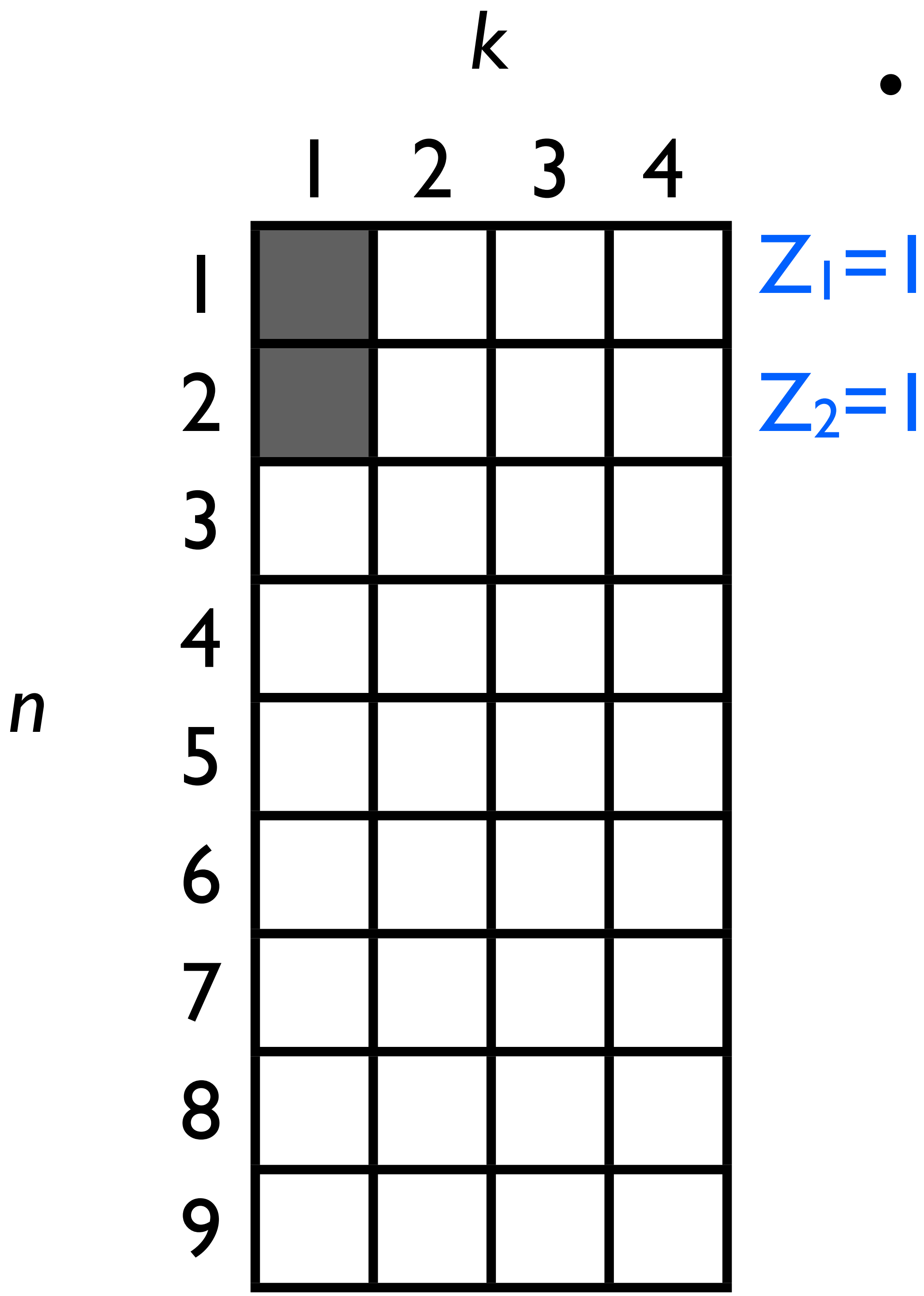
Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

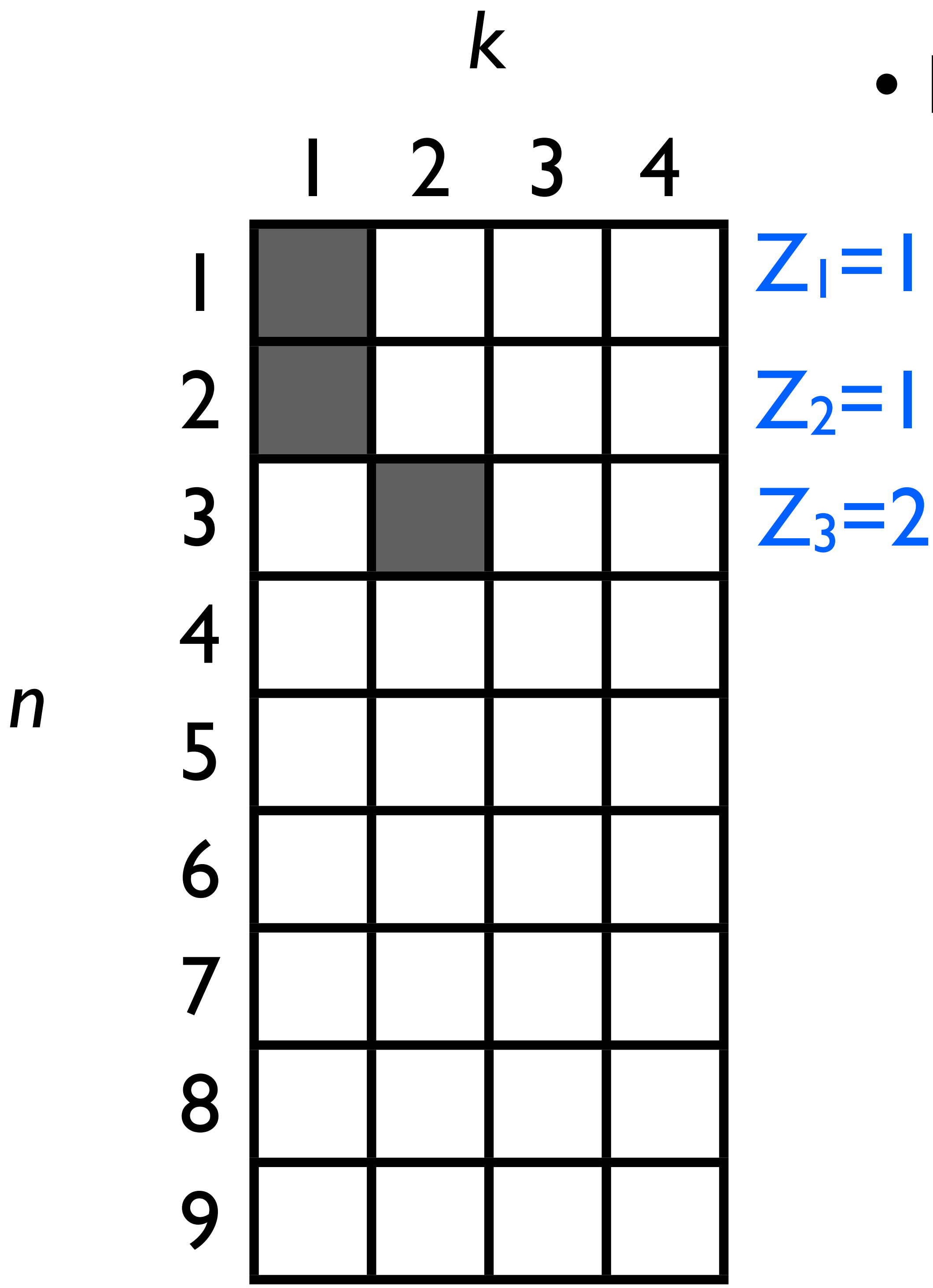
Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

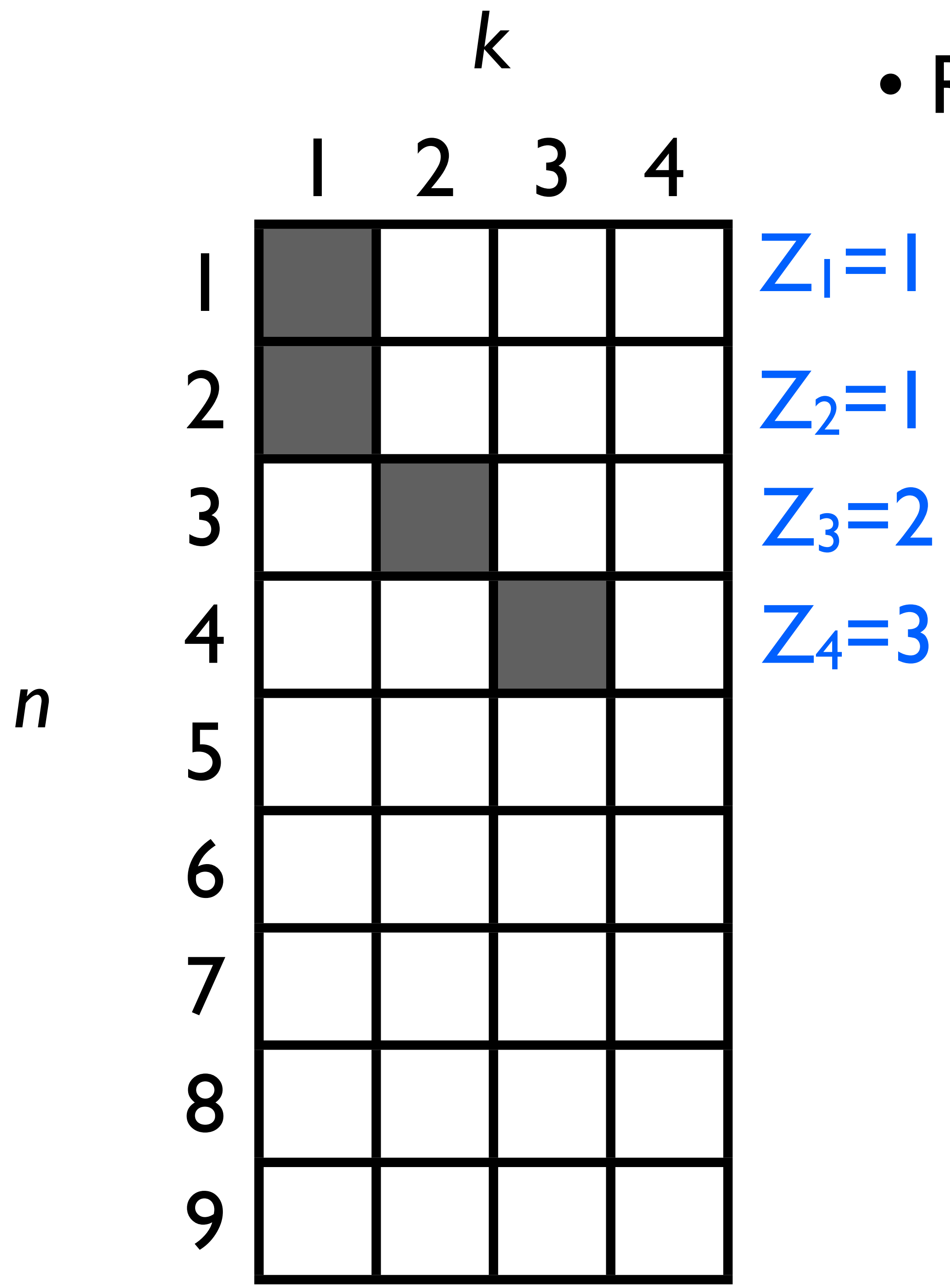
Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\ \text{people there})$
  - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

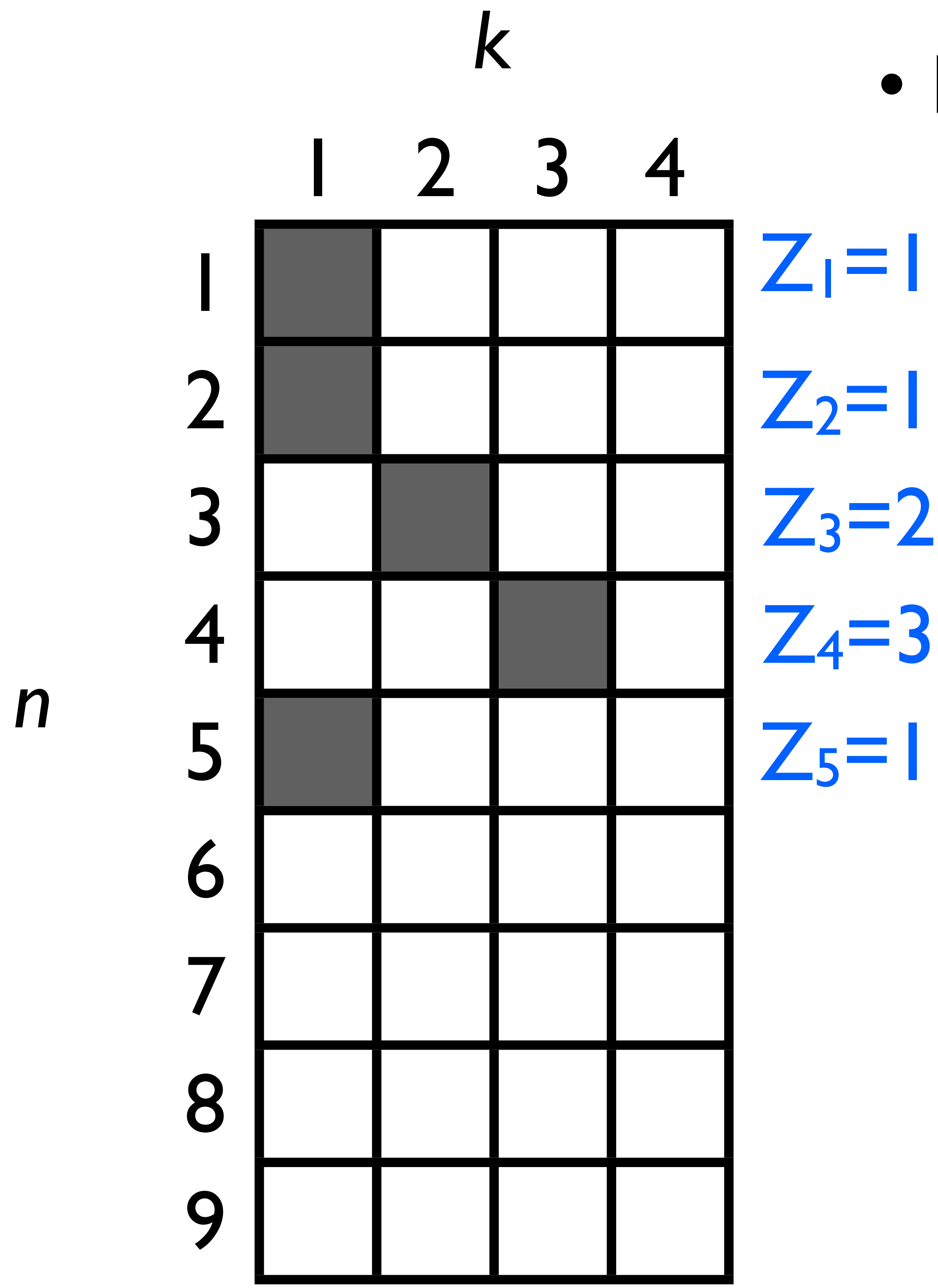
Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

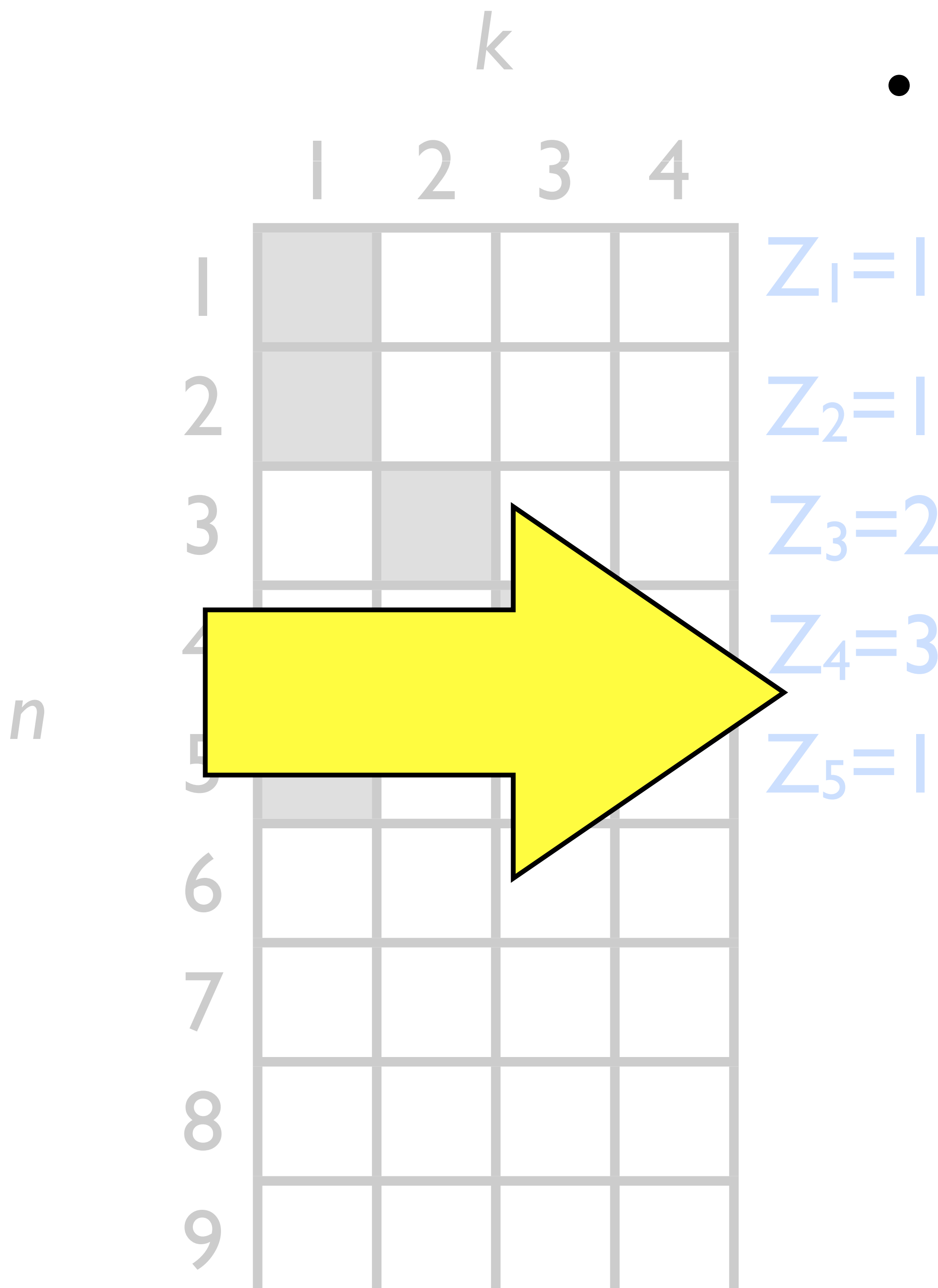
Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

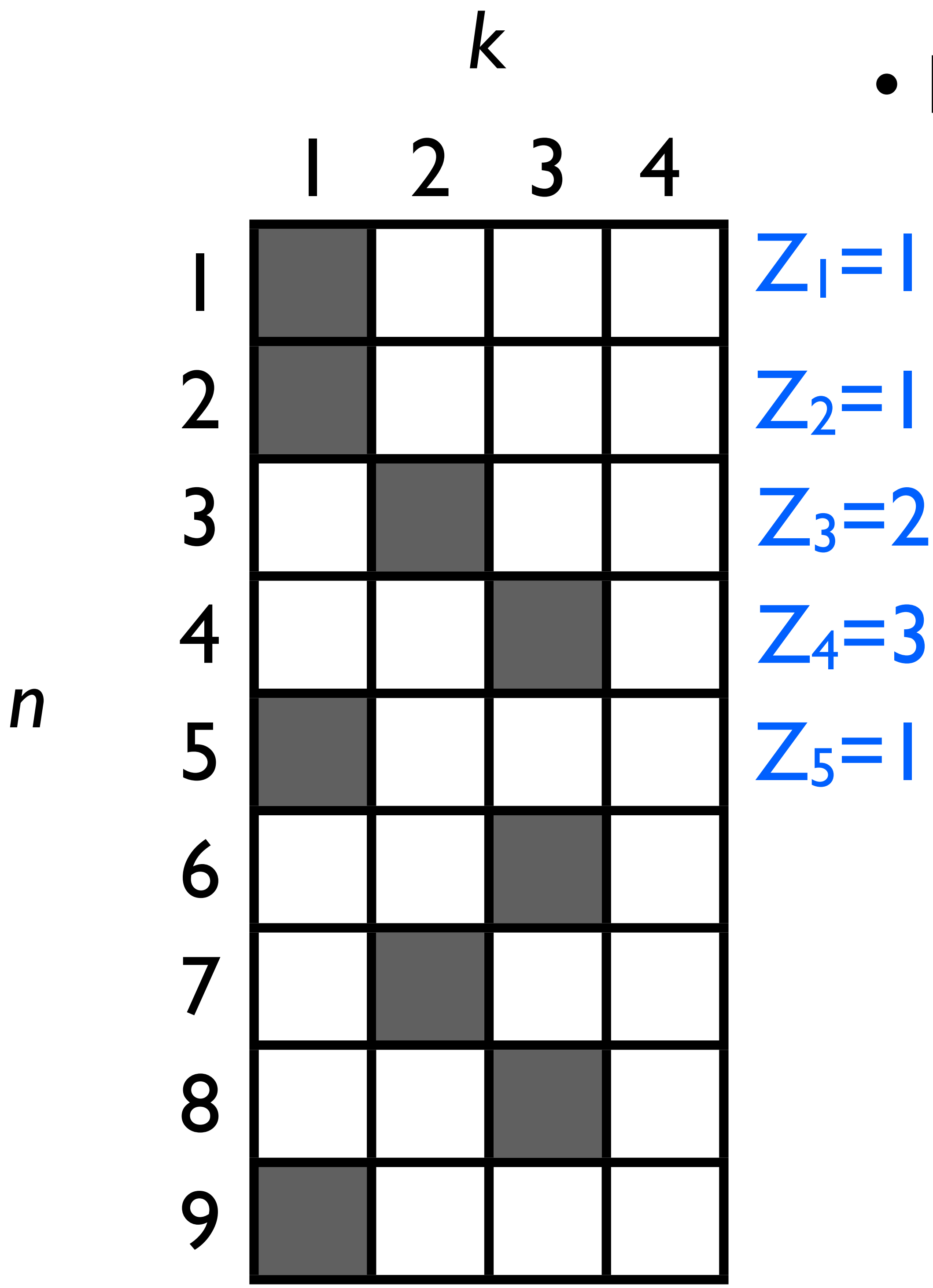
Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\text{\# people there})$
  - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

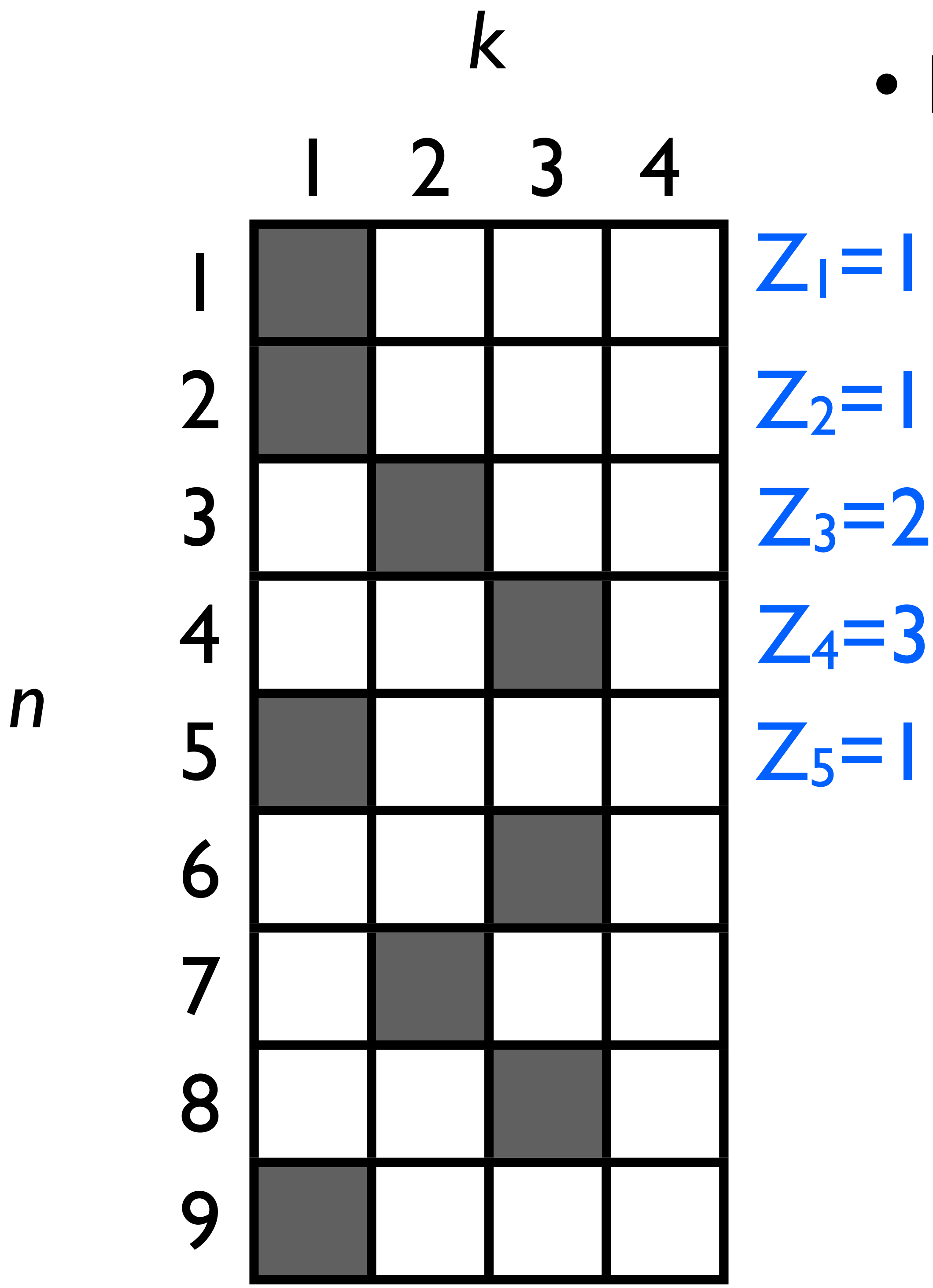
Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$

Cluster labels

Order of appearance

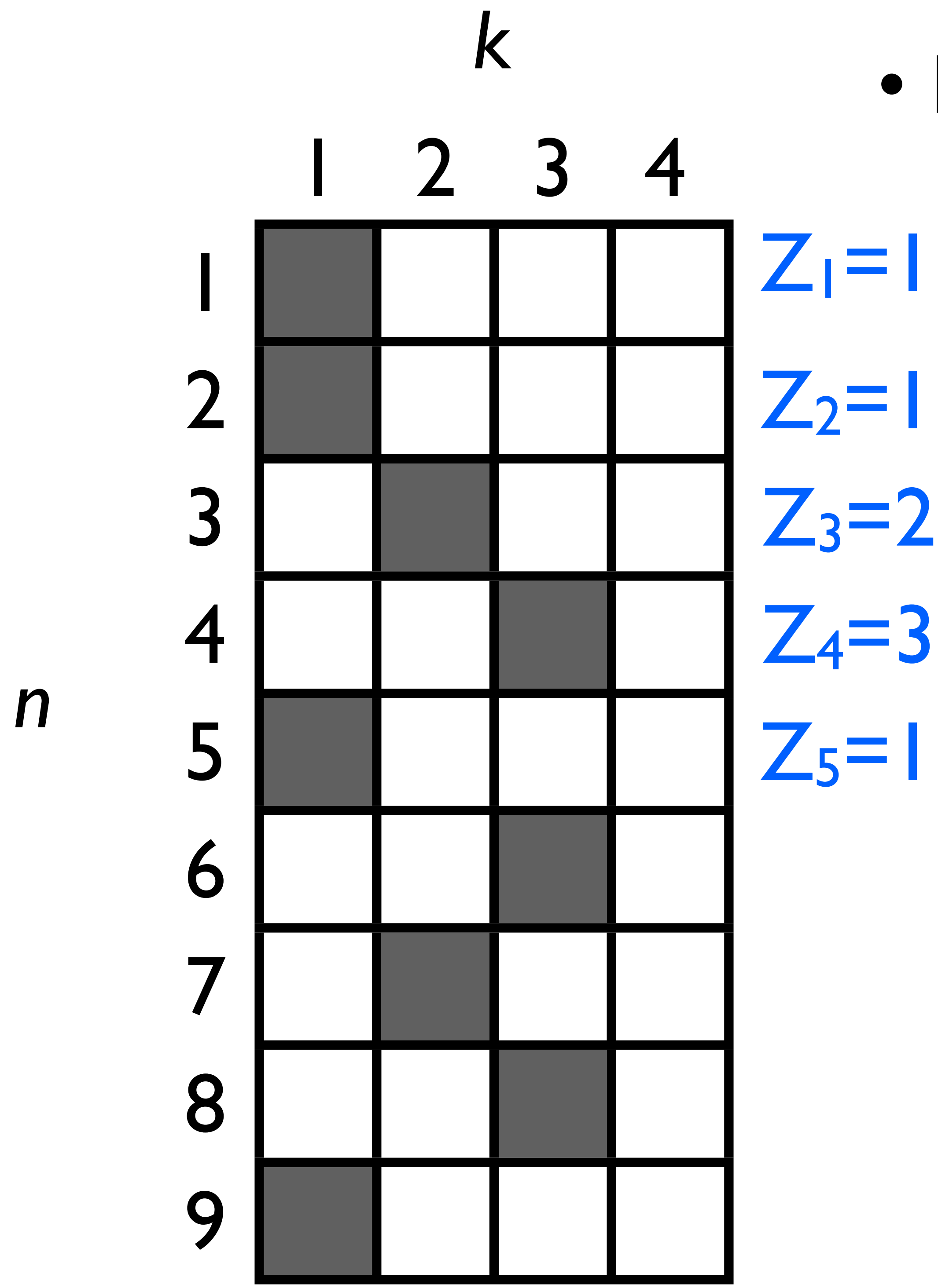


- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$

- The clustering is exchangeable

Cluster labels

Order of appearance

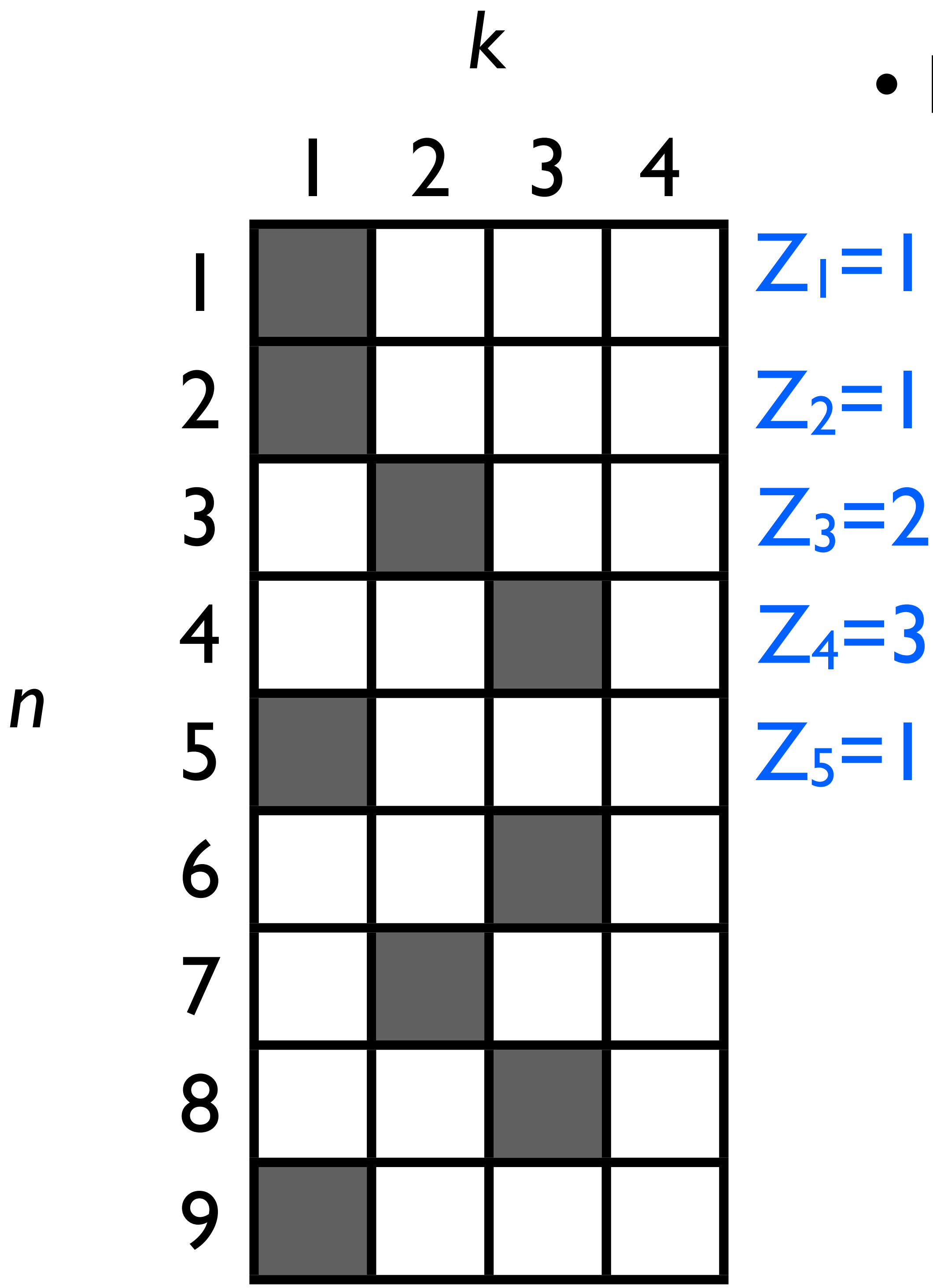


- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

- The clustering is exchangeable
- Z_n here NOT exchangeable

Cluster labels

Order of appearance



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$

- The clustering is exchangeable
- Z_n here NOT exchangeable
- A matrix is a clustering and an integer labeling

Outline

I. Clusters

- Overview
- Distribution
  - ◇ Clusters (Example: Chinese restaurant process)
  - ◇ Data given clusters (Example: Gaussian mixture)
  - ◇ Posterior
- Proportions
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model
  - ◇ Posterior
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model
  - ◇ Posterior
- Random probability measure

II. Features

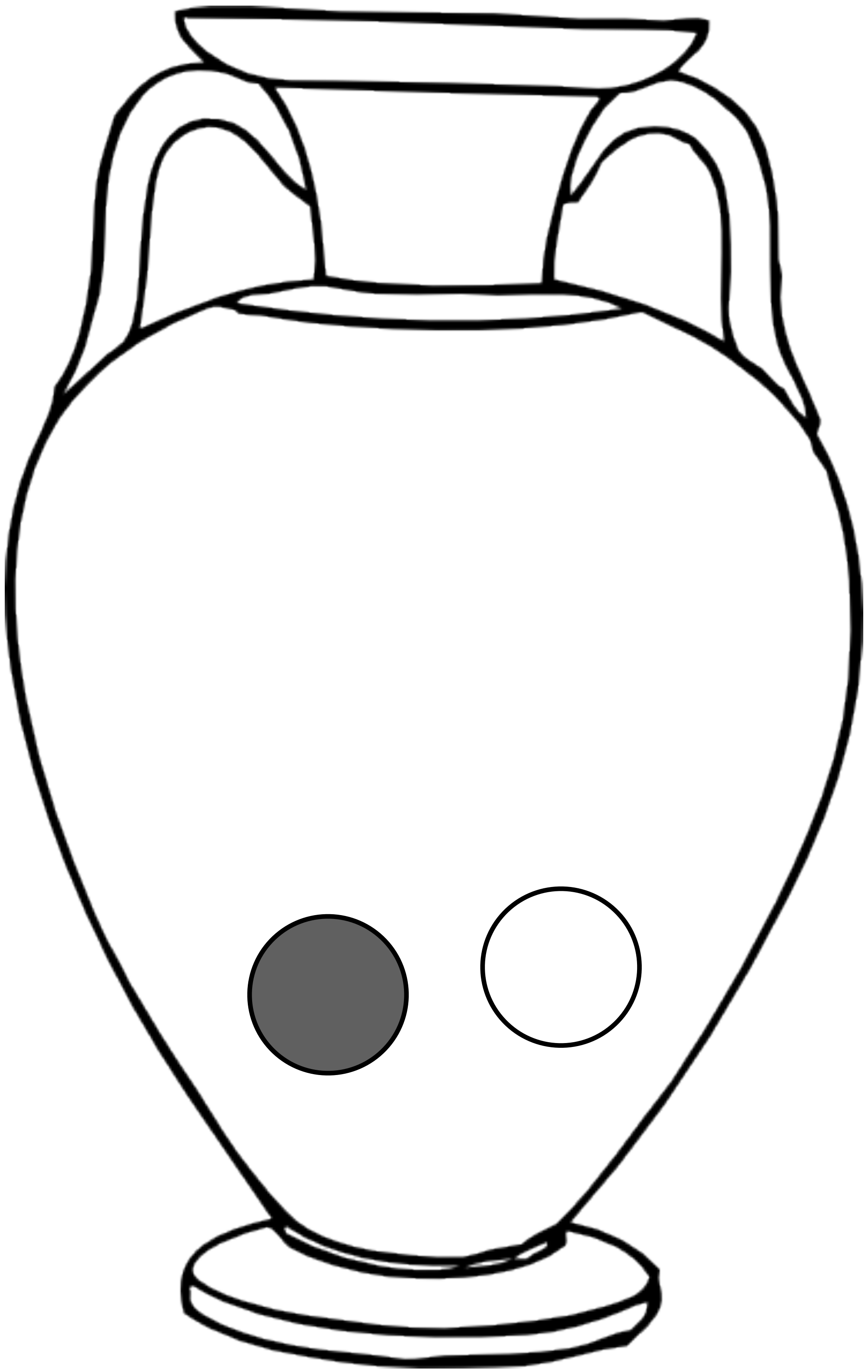
Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model (Example: CRP stick-breaking)
  - ◇ Posterior
- Random probability measure

II. Features

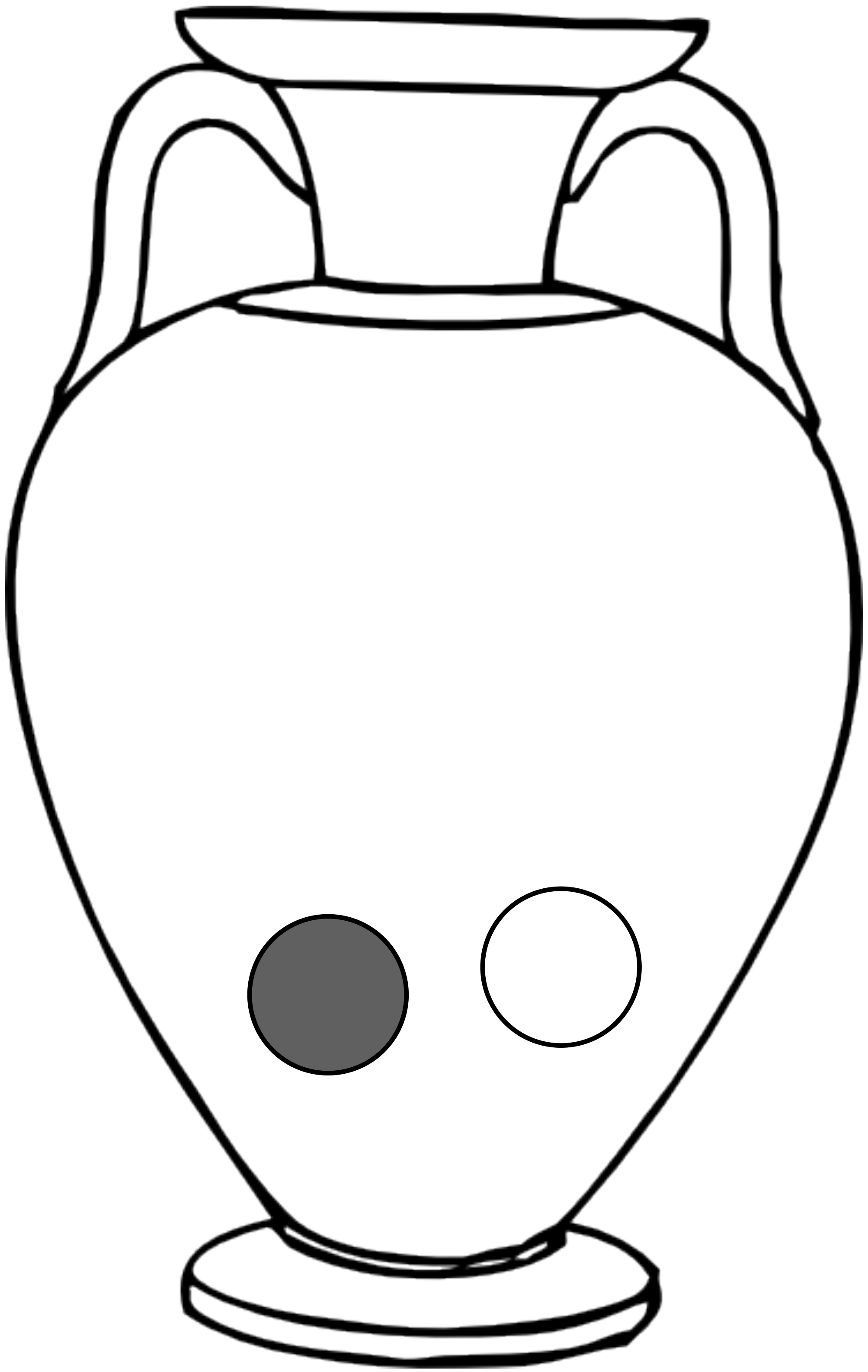
Aside: Polya Urn



- G_0 initial gray balls
- W_0 initial white balls

[Polya 1930; Freedman 1965]

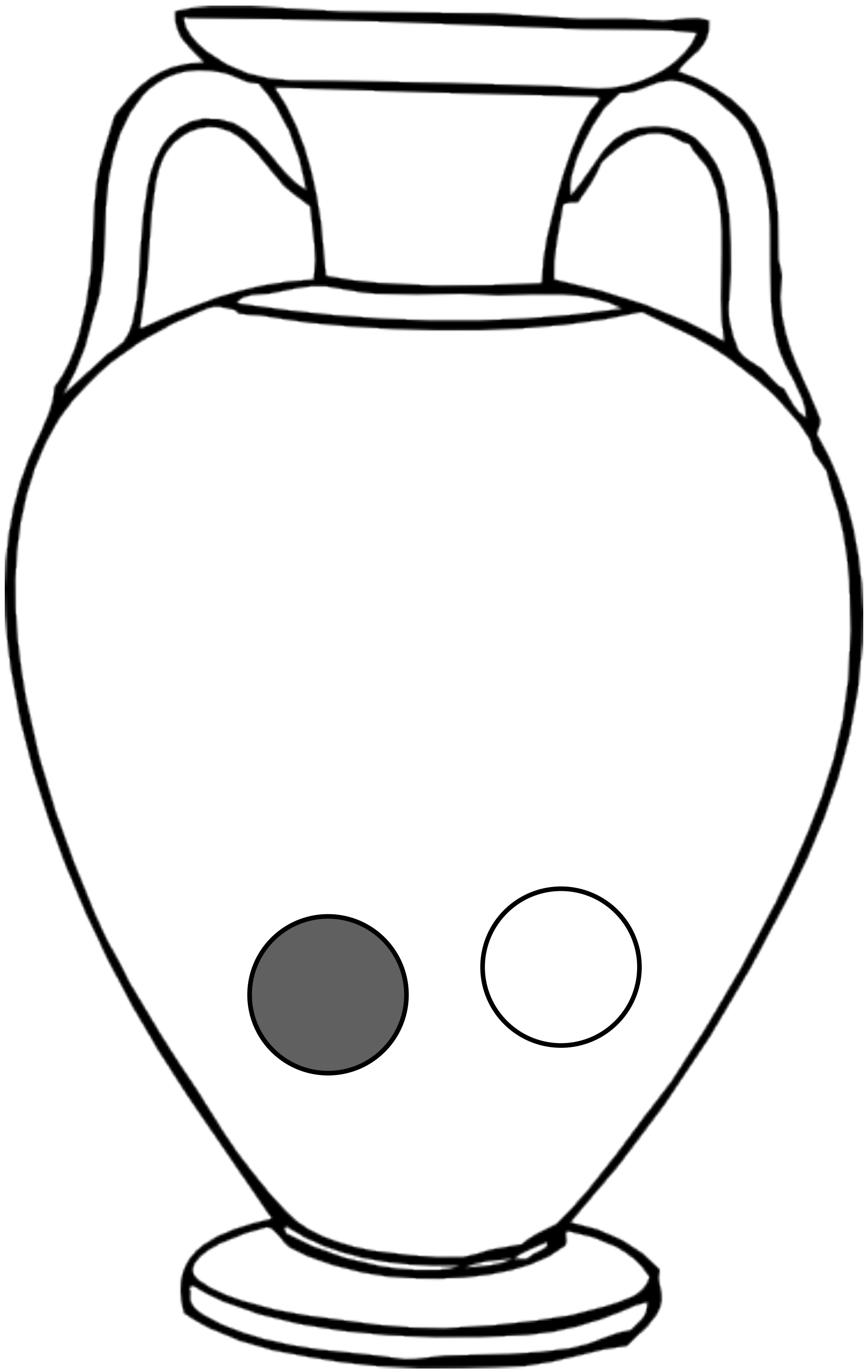
Aside: Polya Urn



- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...

[Polya 1930; Freedman 1965]

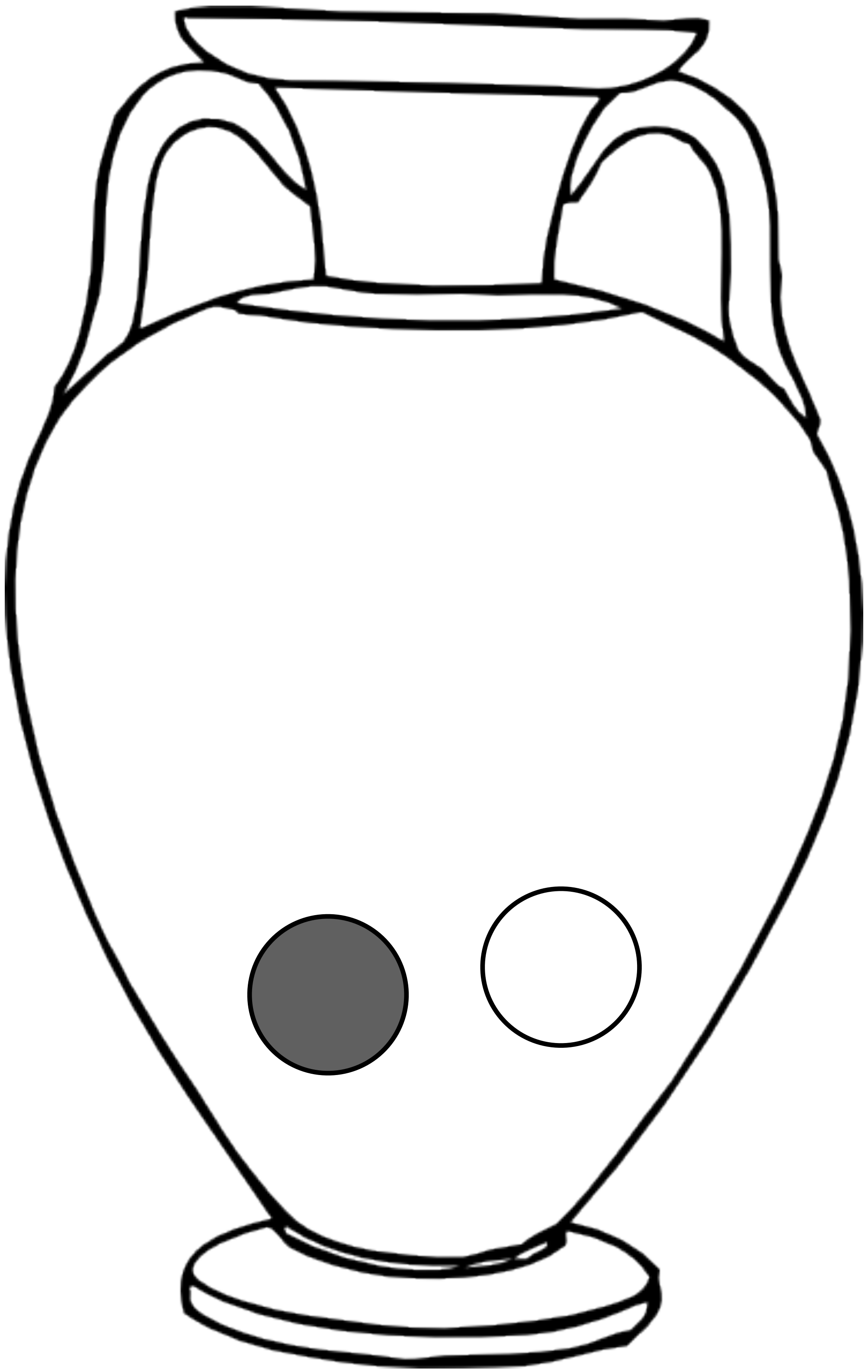
Aside: Polya Urn



- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
    - ◇ Draw a ball uniformly from the urn

[Polya 1930; Freedman 1965]

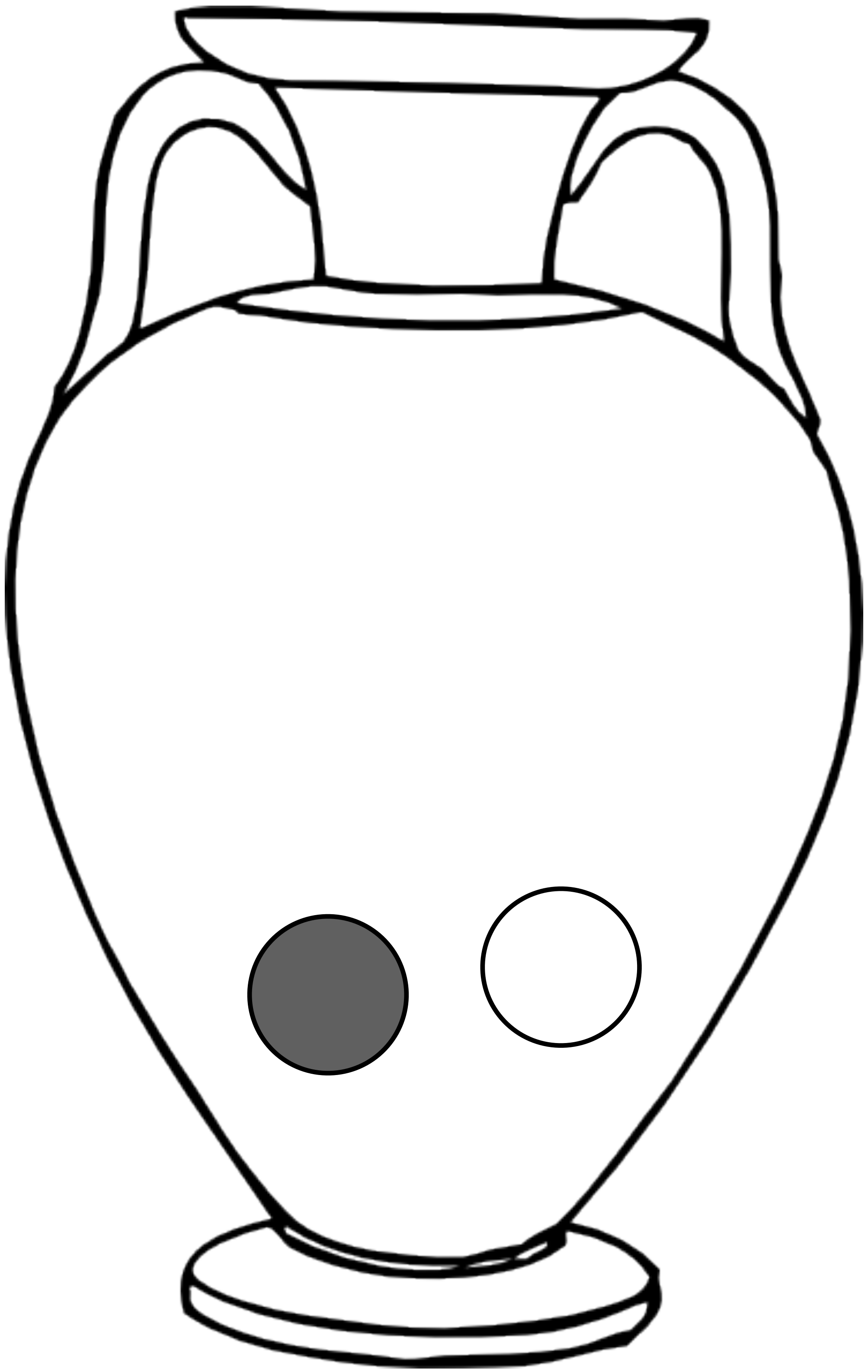
Aside: Polya Urn



- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color

[Polya 1930; Freedman 1965]

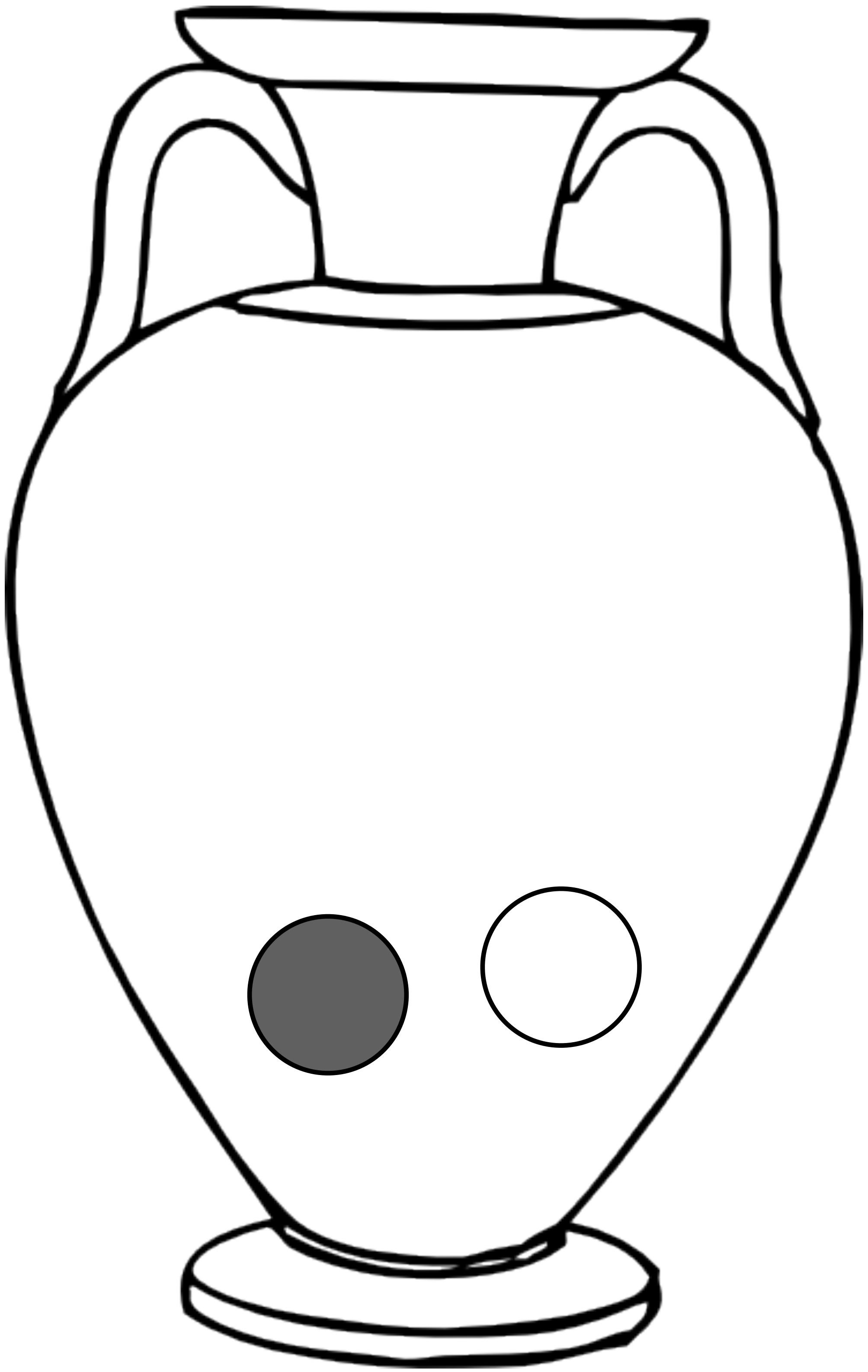
Aside: Polya Urn



- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color

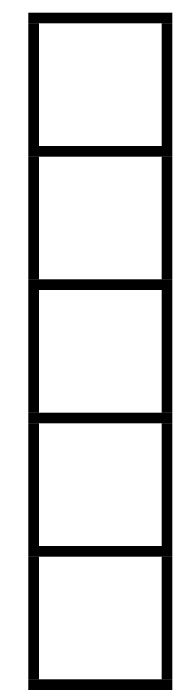
- Example: $G_0 = 1$, $W_0 = 1$

Aside: Polya Urn



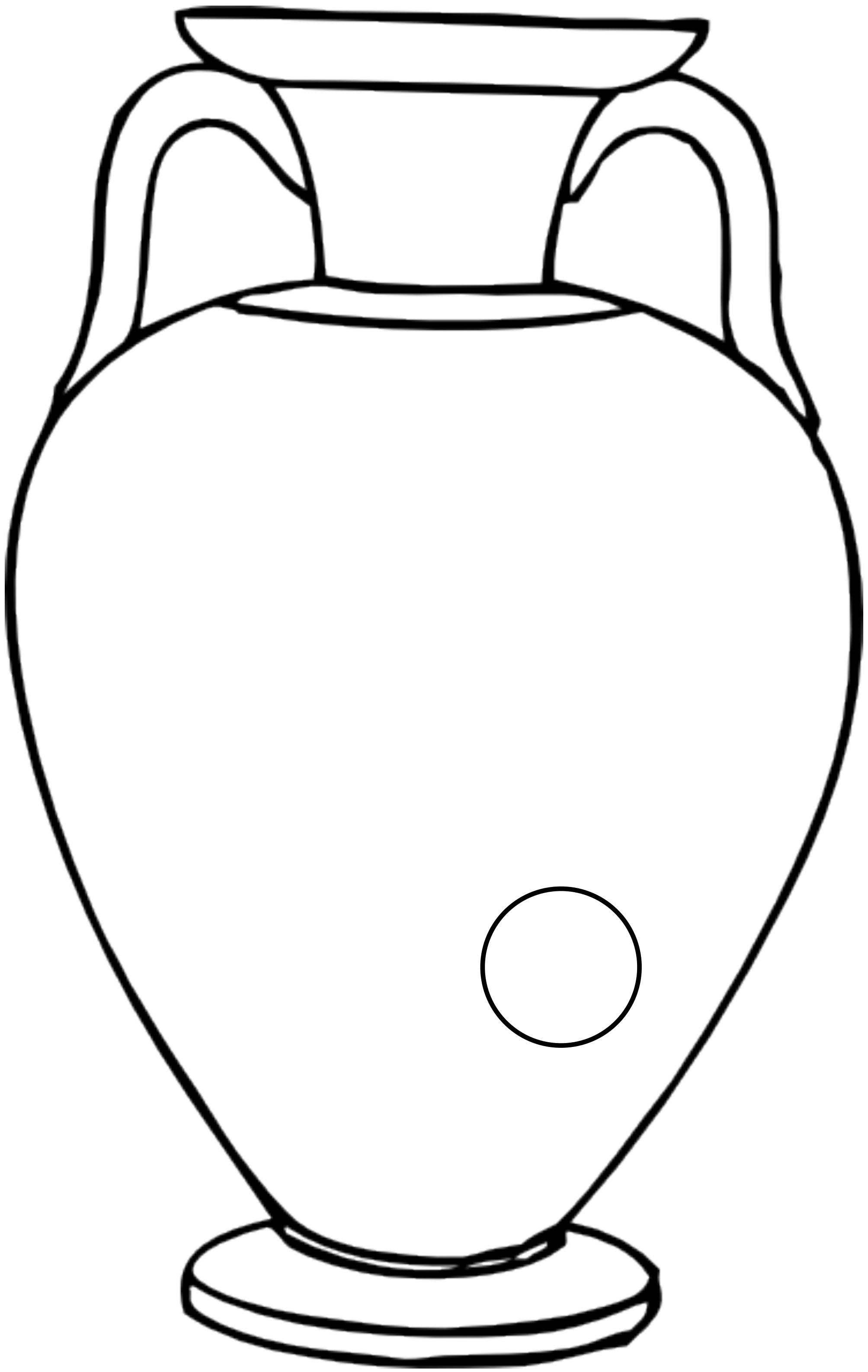
- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color

n

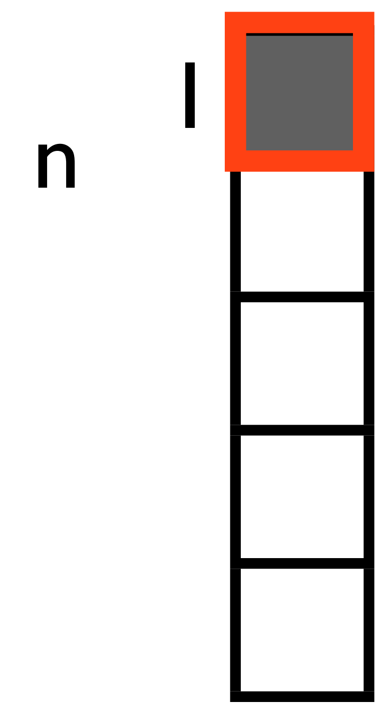


- Example: $G_0 = 1$, $W_0 = 1$

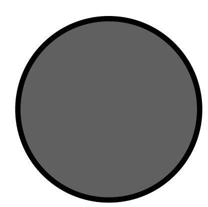
Aside: Polya Urn



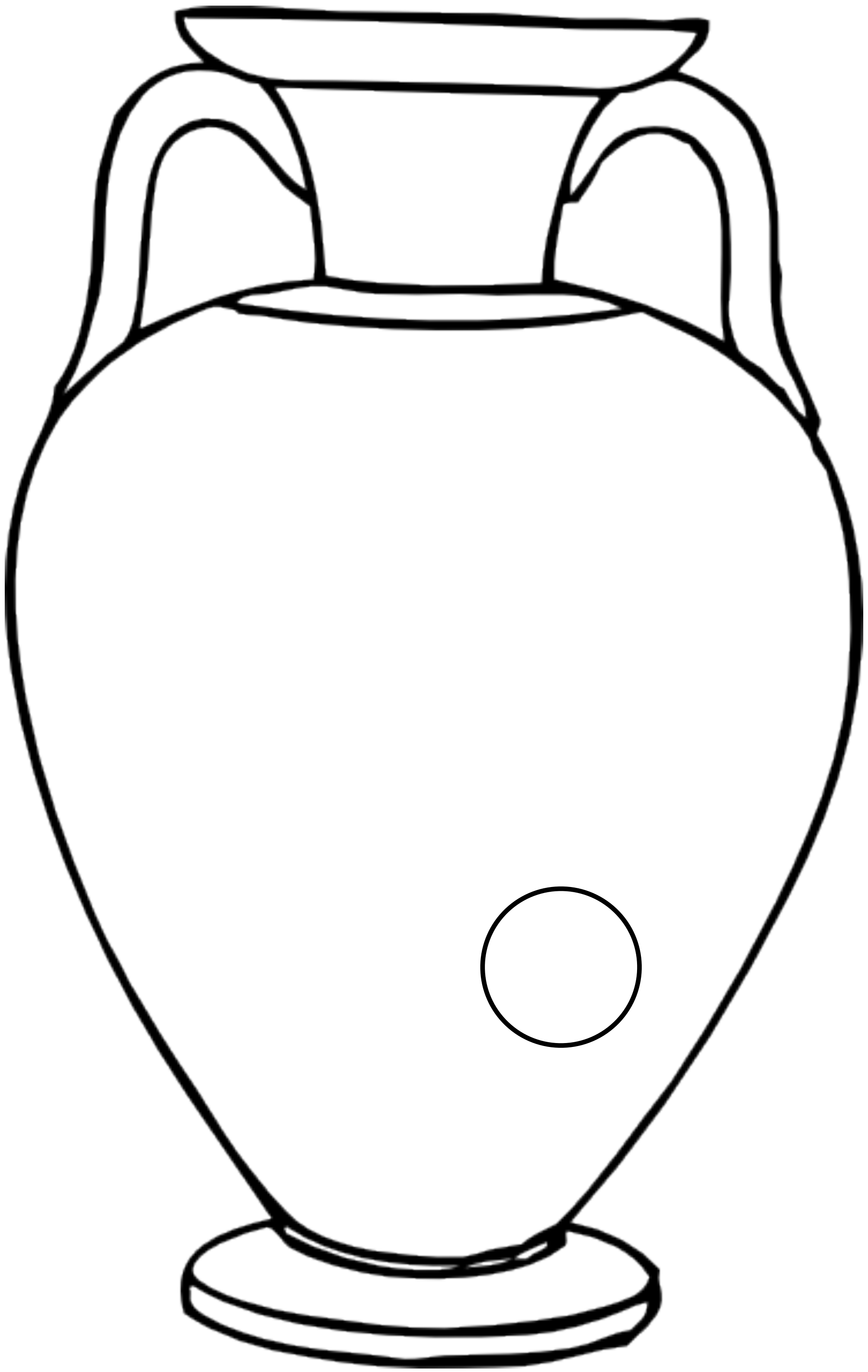
- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
    - ◇ Draw a ball uniformly from the urn
    - ◇ Put it back with another ball of the same color



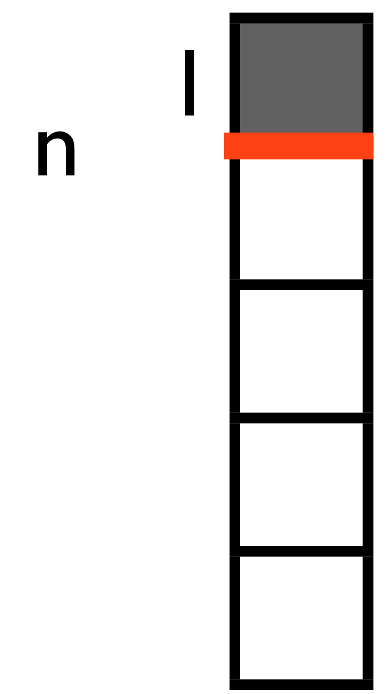
- Example: G_0 = 1, W_0 = 1



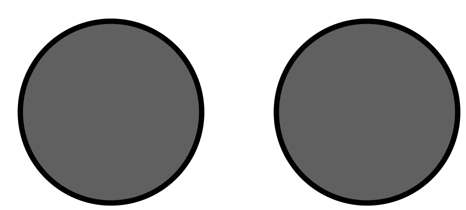
Aside: Polya Urn



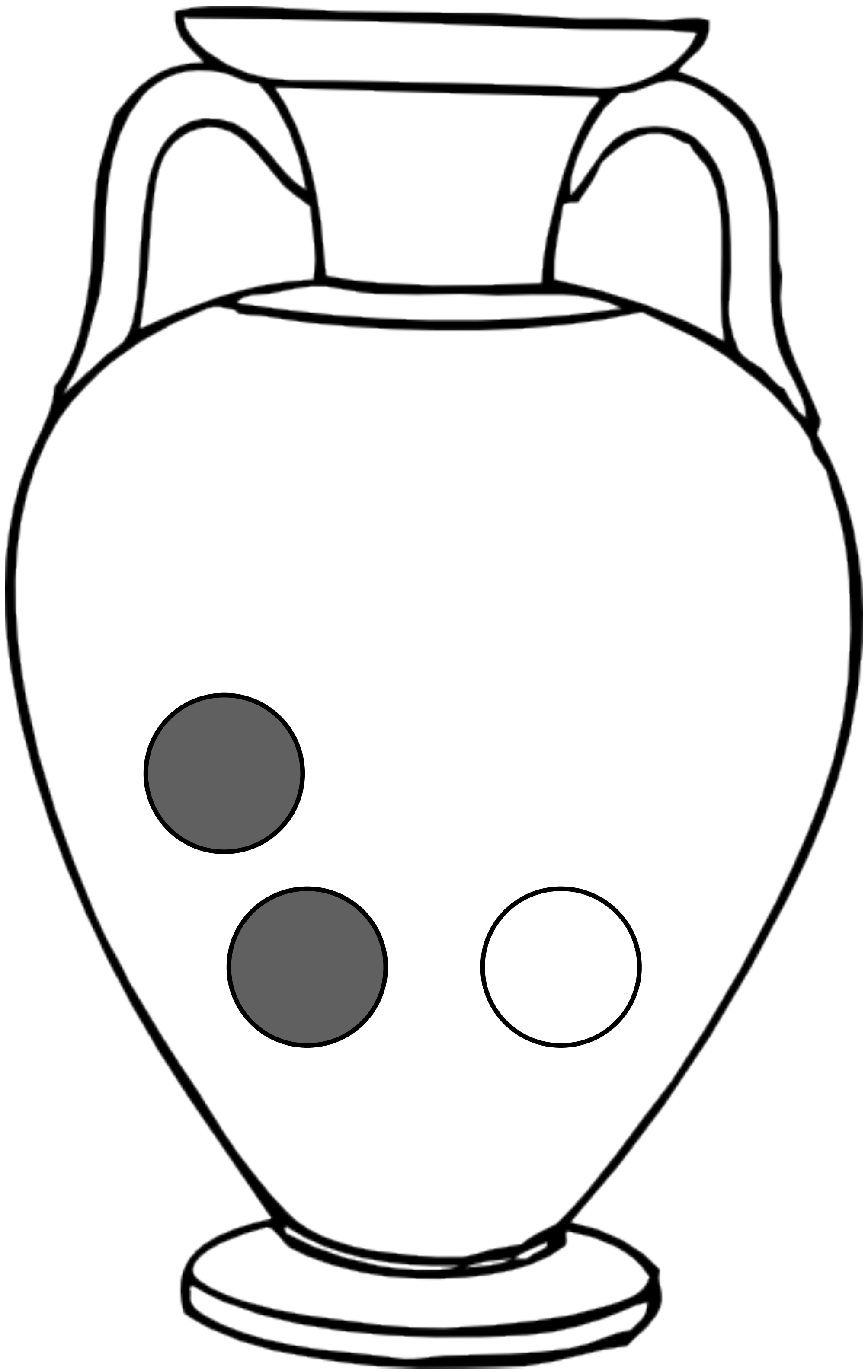
- G_0 initial gray balls
- W_0 initial white balls
- $n = 1, 2, \ldots$
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color



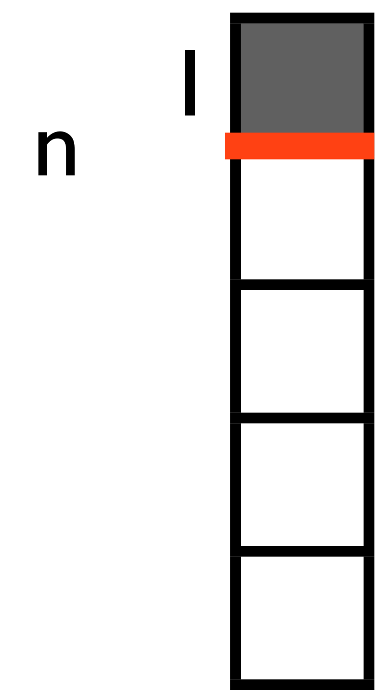
- Example: $G_0 = 1$, $W_0 = 1$



Aside: Polya Urn

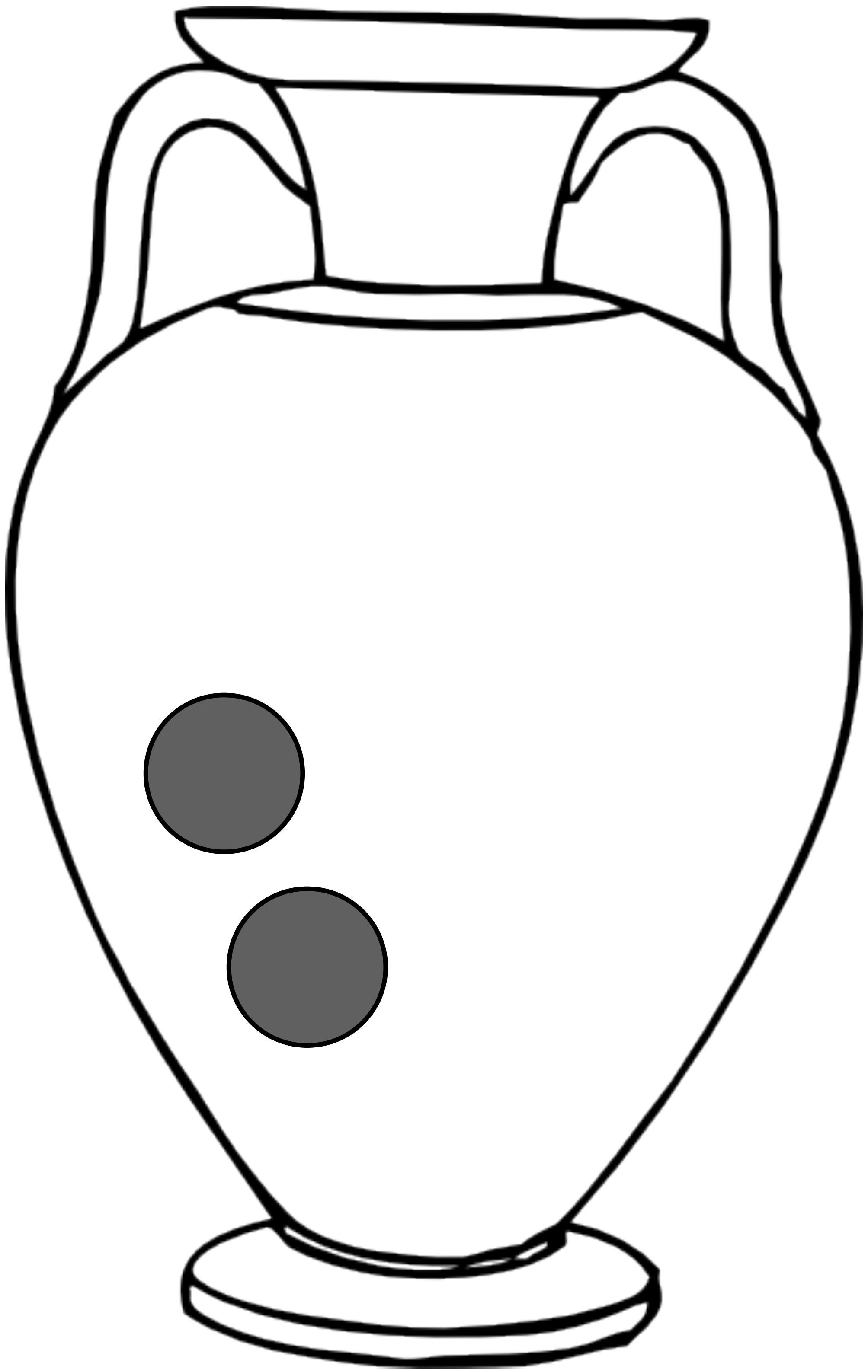


- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color

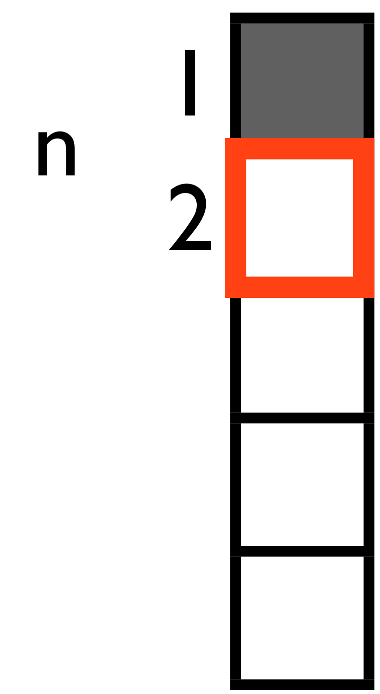


- Example: $G_1 = 2$, $W_1 = 1$

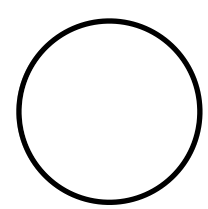
Aside: Polya Urn



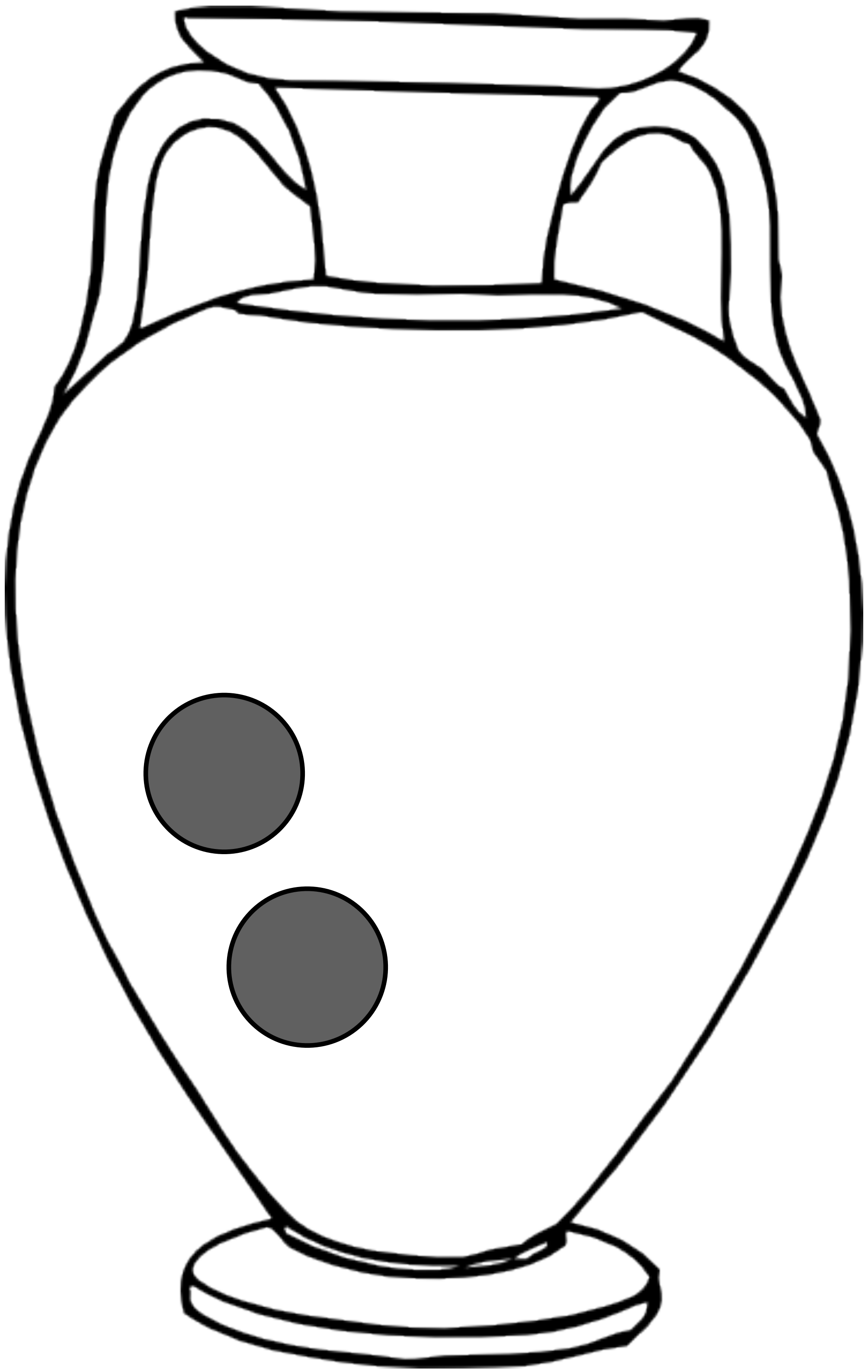
- G_0 initial gray balls
- W_0 initial white balls
- $n = 1, 2, \ldots$
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color



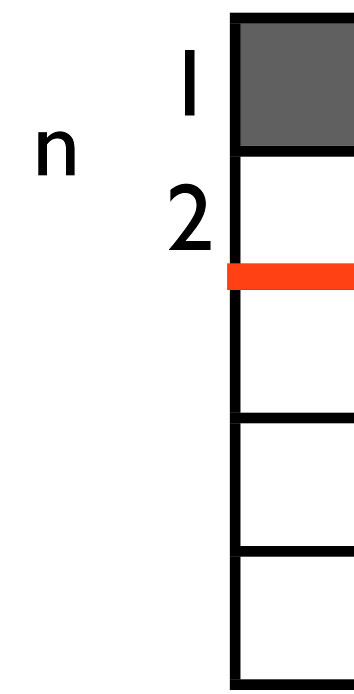
- Example: $G_1 = 2$, $W_1 = 1$



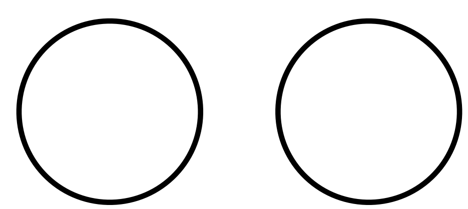
Aside: Polya Urn



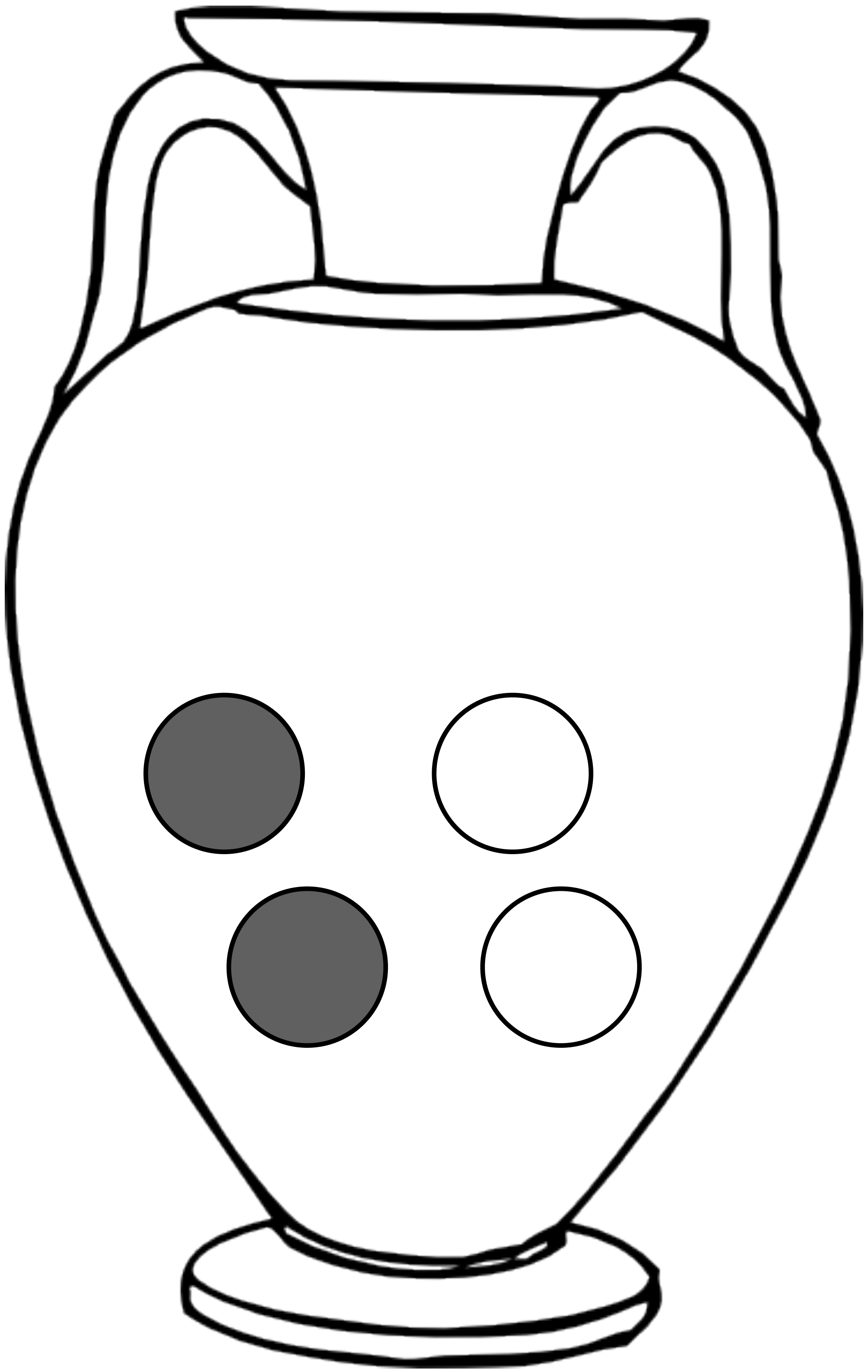
- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color



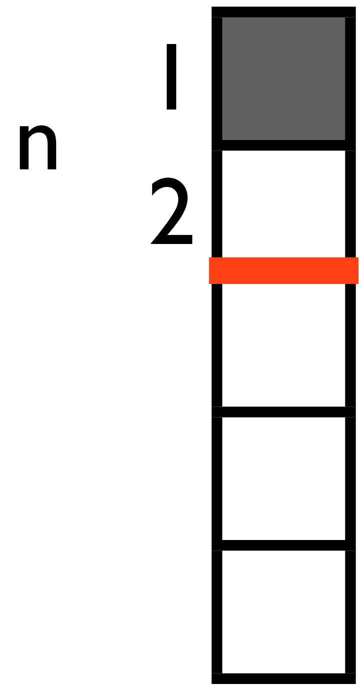
- Example: G_1 = 2, W_1 = 1



Aside: Polya Urn

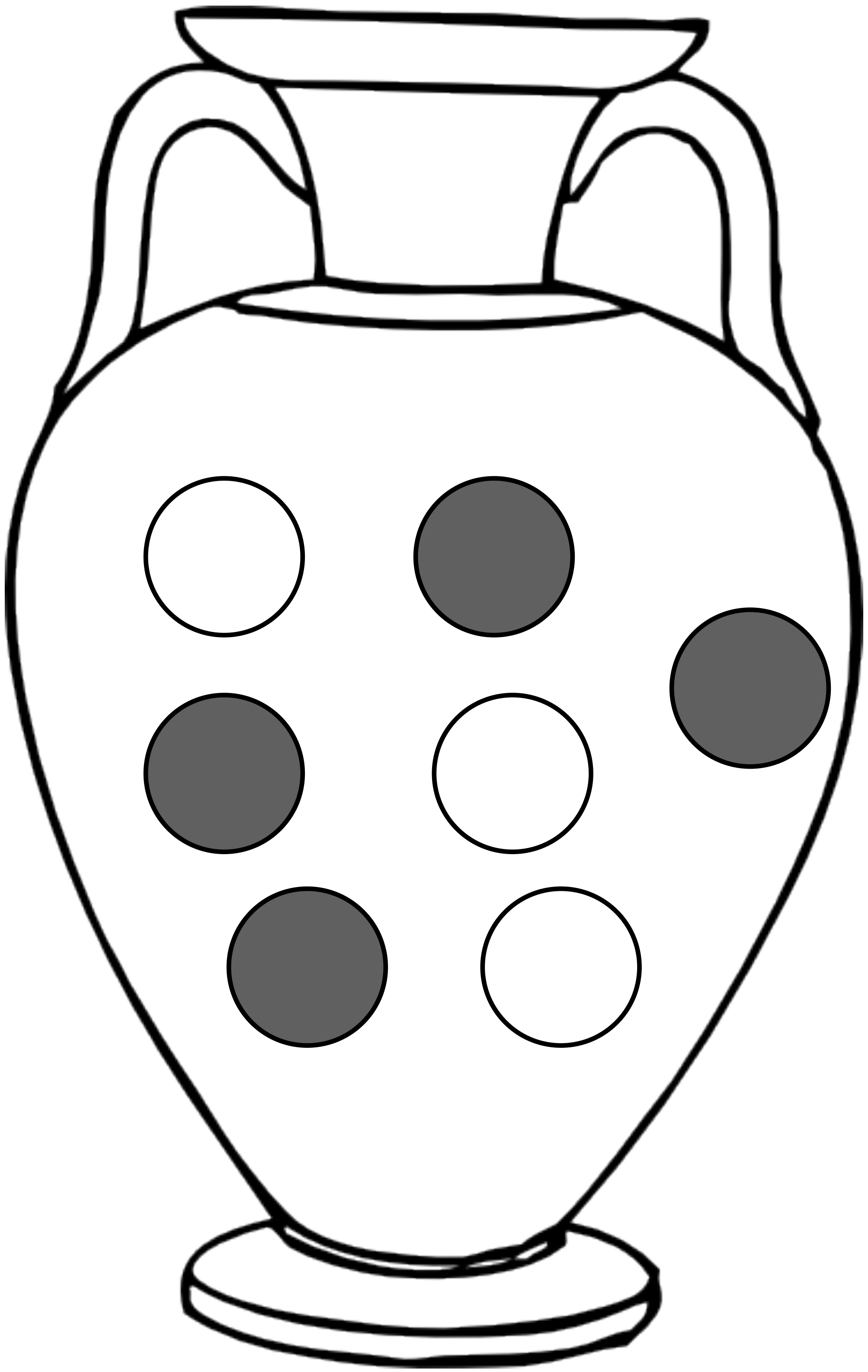


- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color



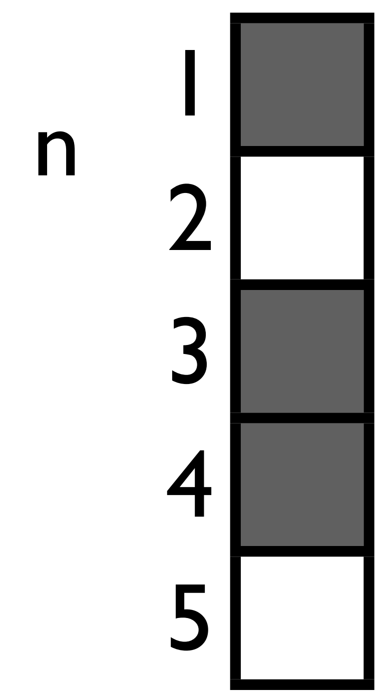
- Example: $G_2 = 2$, $W_2 = 2$

Aside: Polya Urn

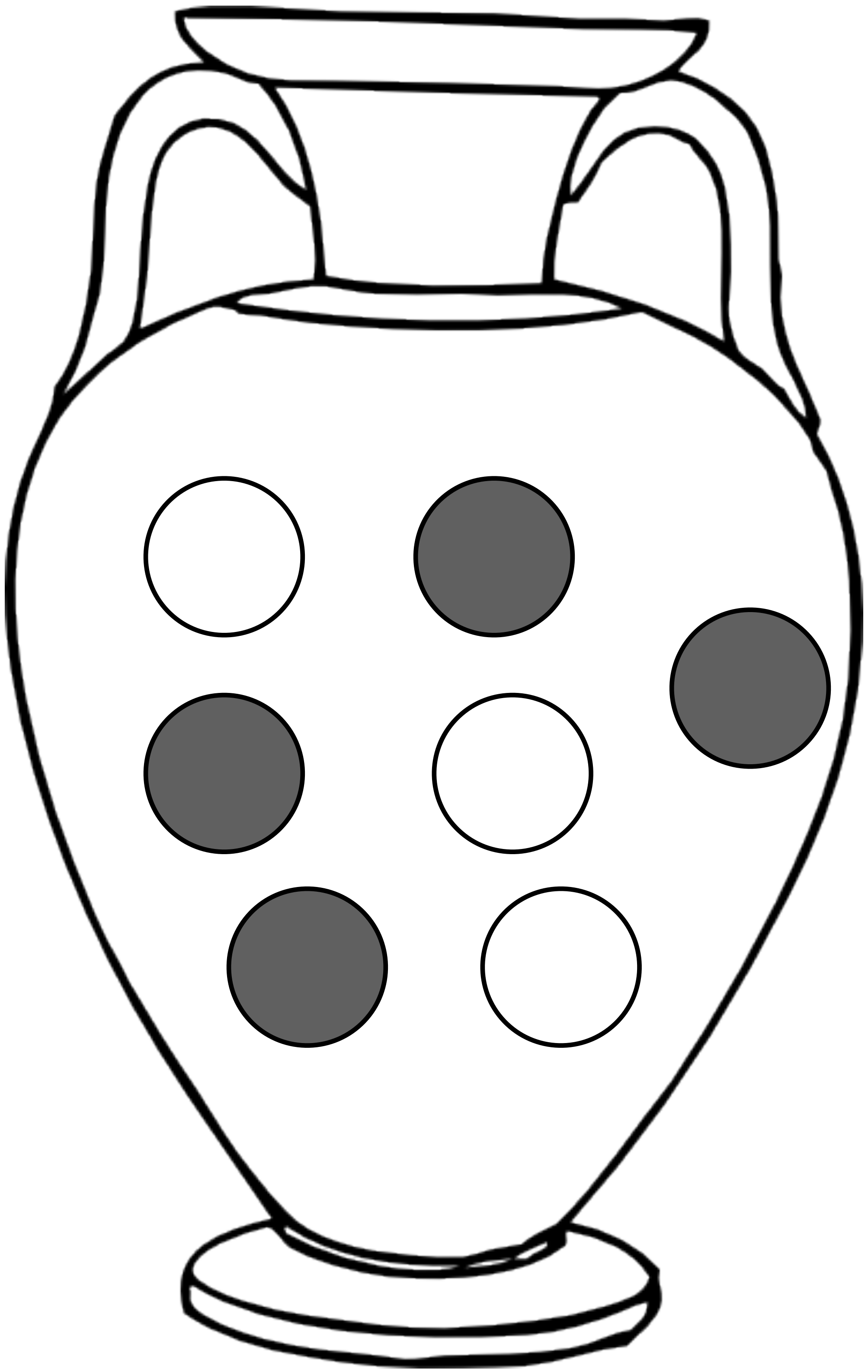


- G_0 initial gray balls
- W_0 initial white balls
- $n = 1, 2, \ldots$
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color

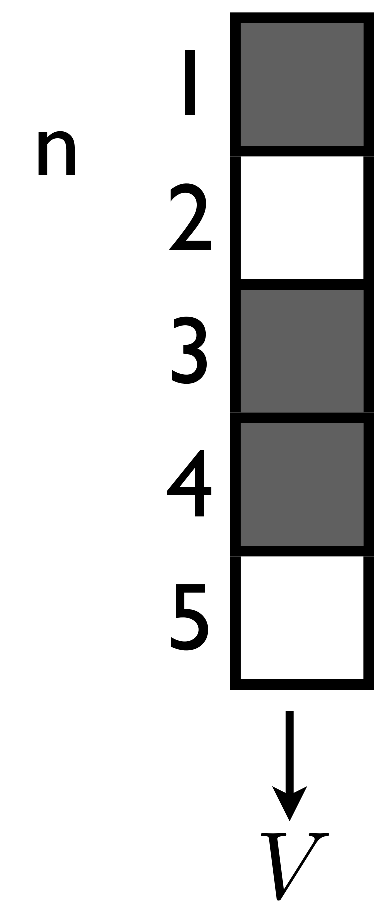
- Example: $G_5 = 4$, $W_5 = 3$



Aside: Polya Urn

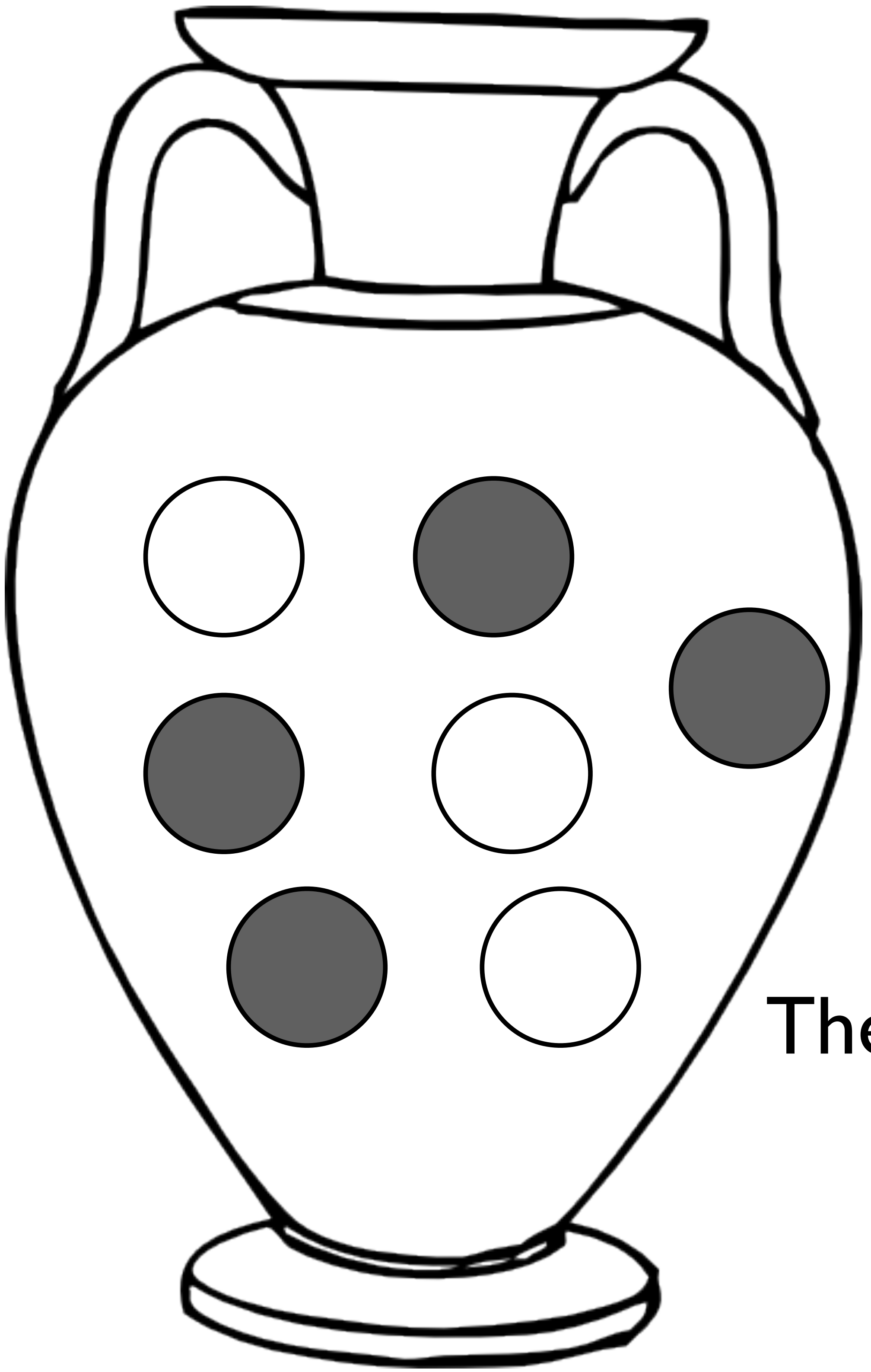


- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color

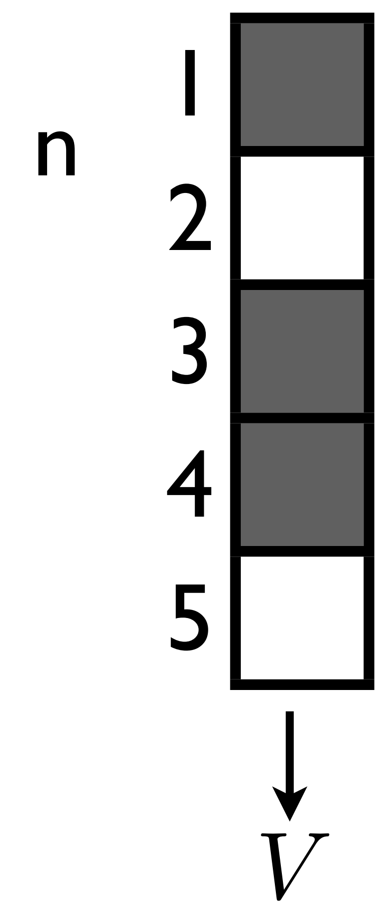


[Polya 1930; Freedman 1965]

Aside: Polya Urn



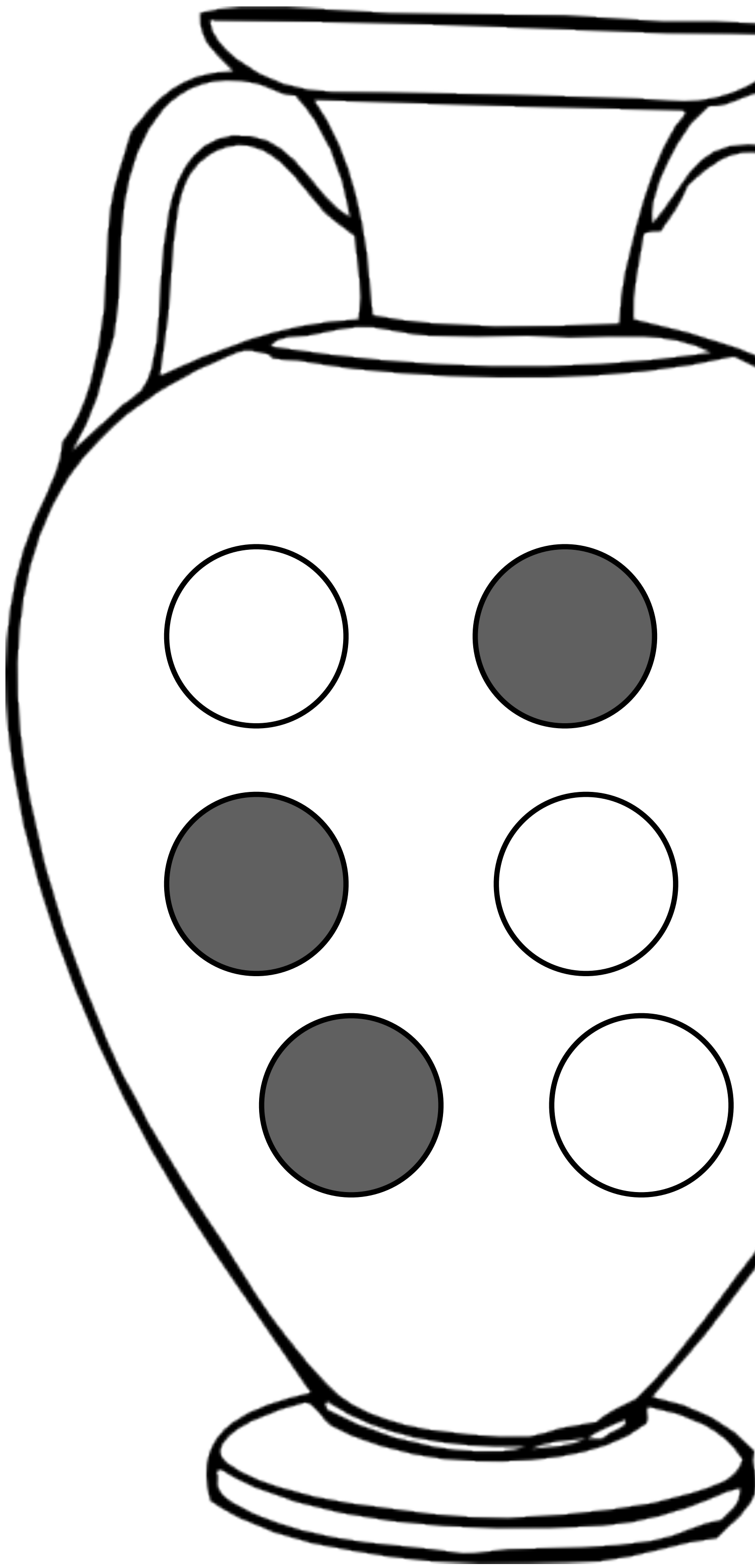
- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color



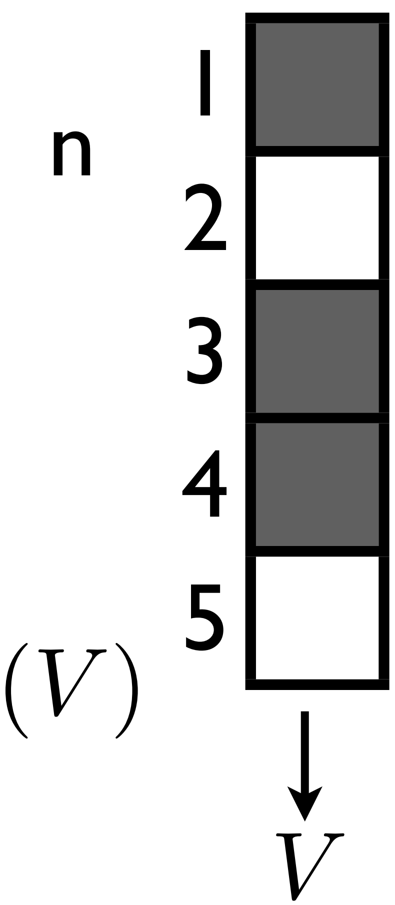
Then $\exists V \sim \mathrm{Beta}(G_0, W_0)$

[Polya 1930; Freedman 1965]

Aside: Polya Urn



- G_0 initial gray balls
- W_0 initial white balls
- n = 1, 2, ...
  - ◇ Draw a ball uniformly from the urn
  - ◇ Put it back with another ball of the same color



Then $\exists V \sim \text{Beta}(G_0, W_0)$

s.t. $G_{n+1} - G_n \overset{iid}{\sim} \text{Bernoulli}(V)$

[Polya 1930; Freedman 1965]

CRP as Polya urns

CRP as Polya urns

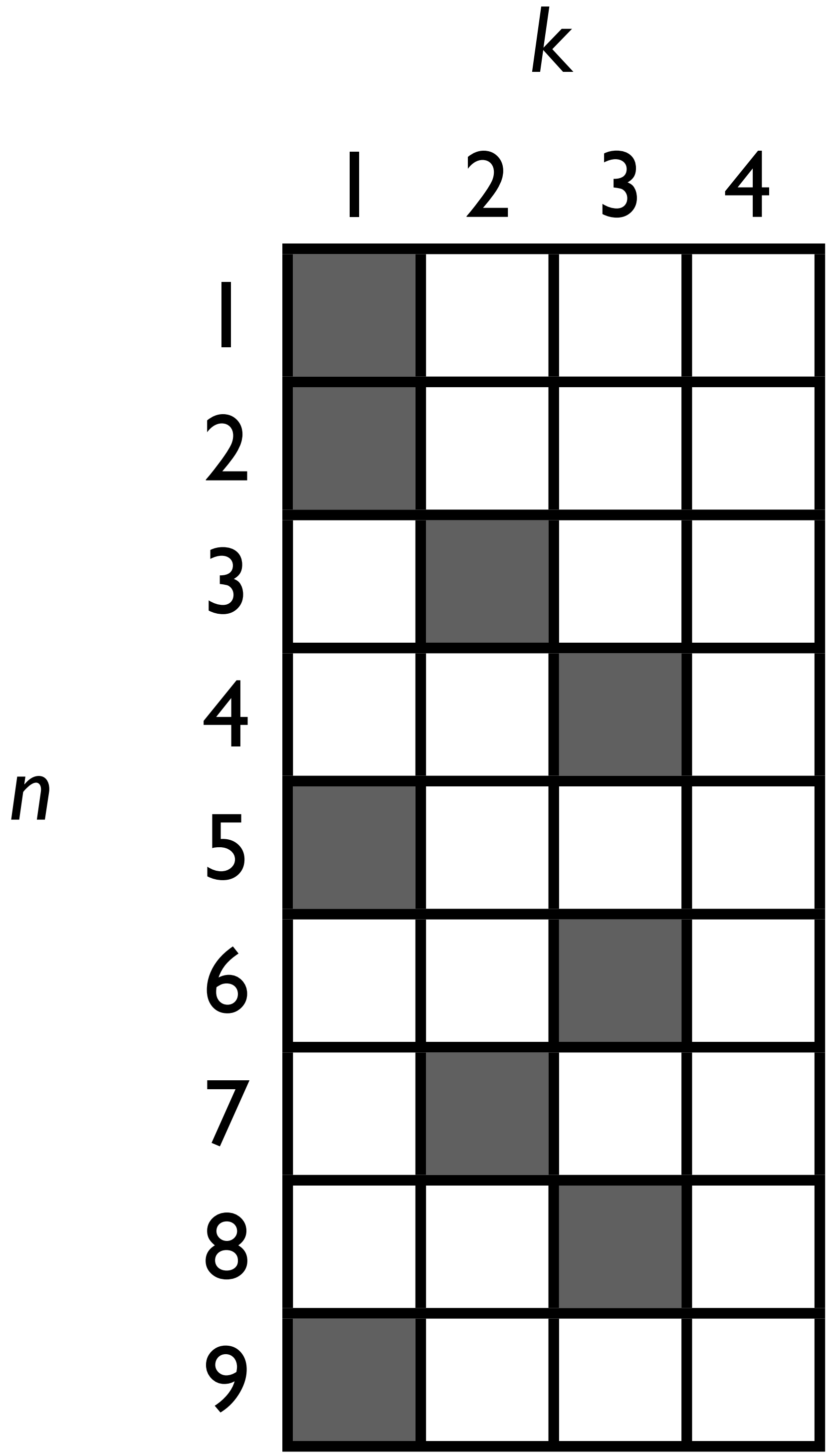
- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability

    $$\propto (\#\ \text{people there})$$

  - at new table *K+1* with probability

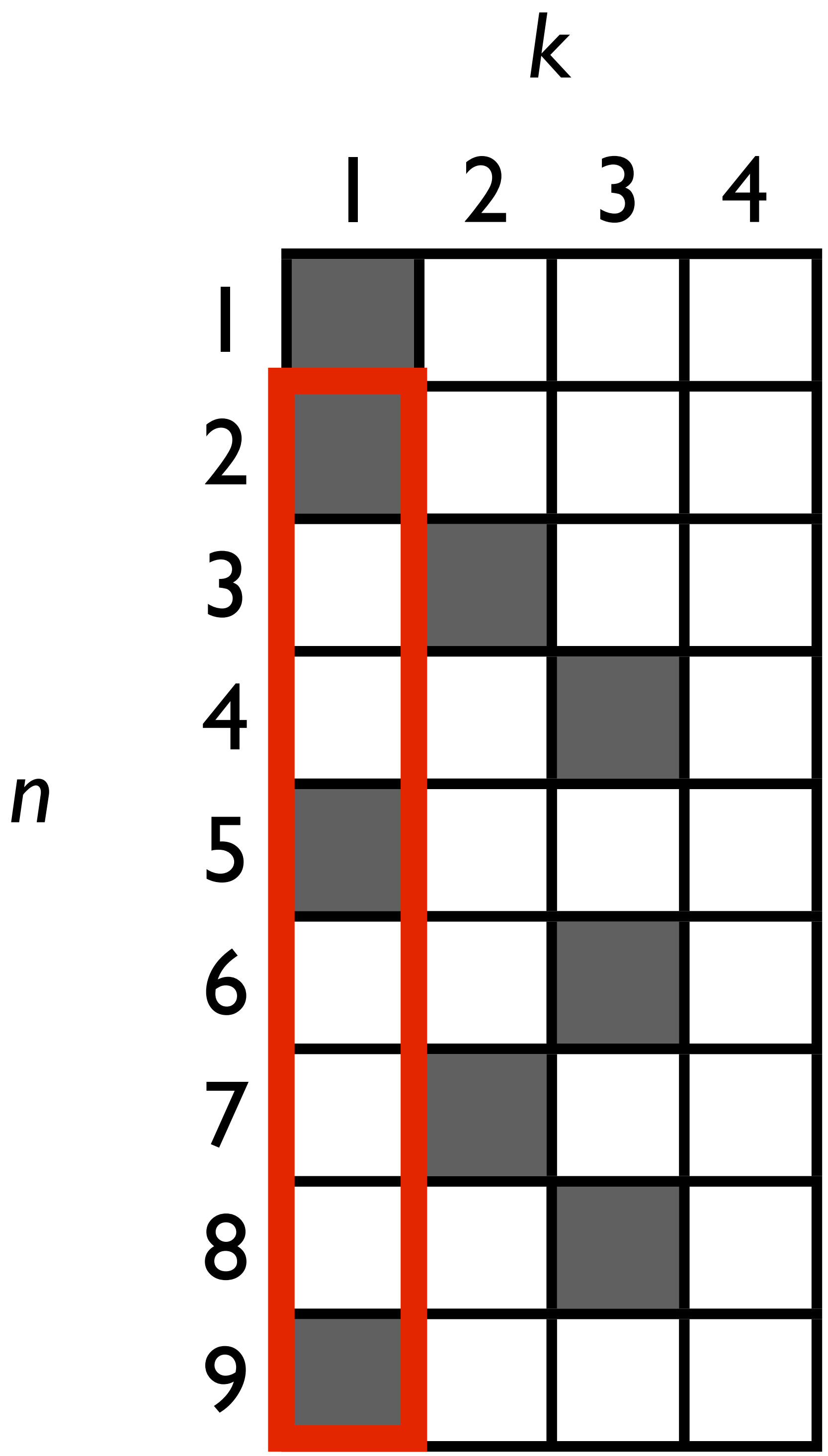
    $$\propto \theta$$

CRP as Polya urns



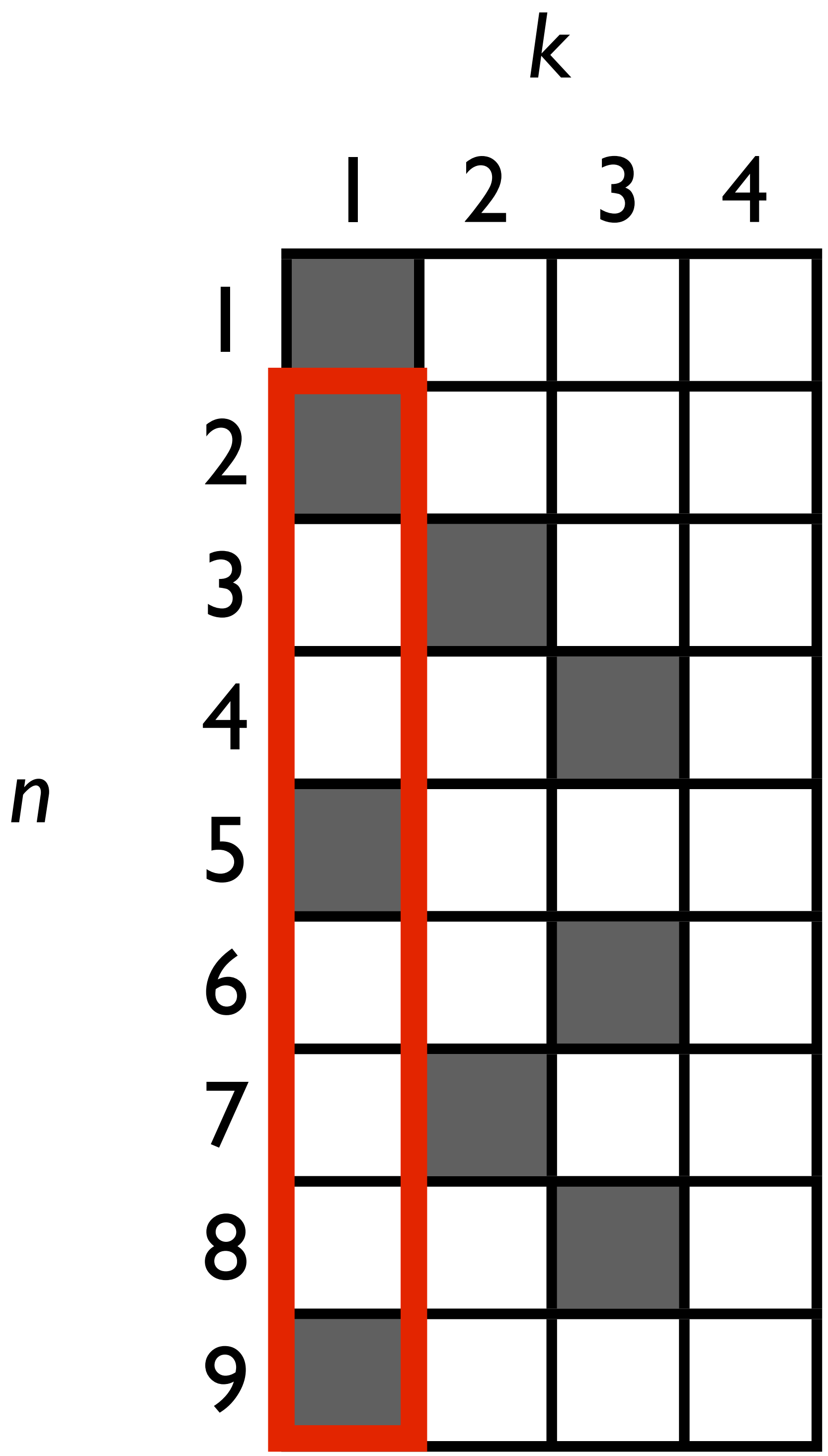
- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

CRP as Polya urns



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\text{\# people there})$
  - at new table *K*+*1* with probability $\propto \theta$

CRP as Polya urns

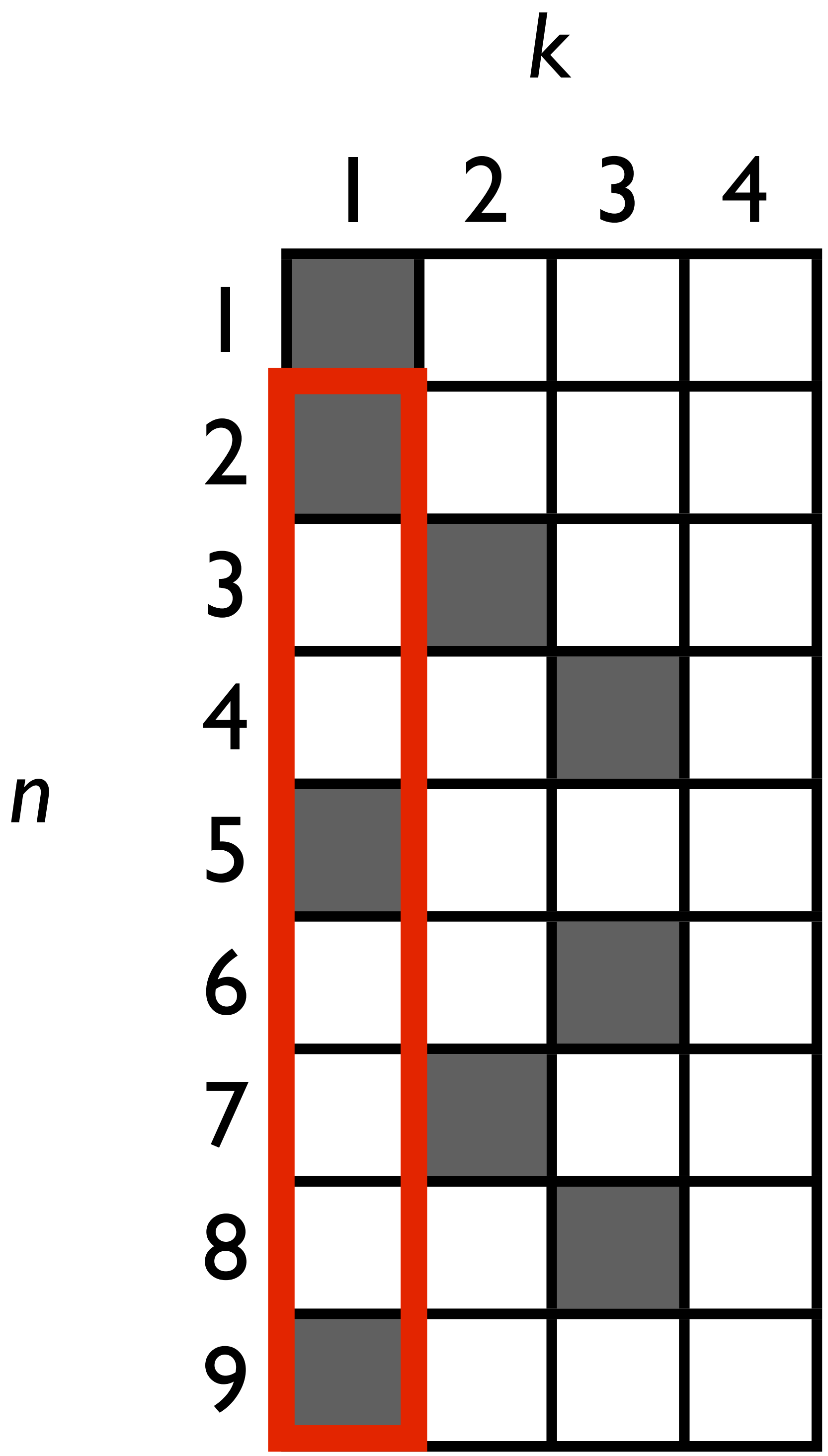


- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto$ (# people there)
  - at new table *K*+*1* with probability $\propto \theta$

- First cluster: Polya urn with

$$G_{1,0} = 1, W_{1,0} = \theta$$

CRP as Polya urns

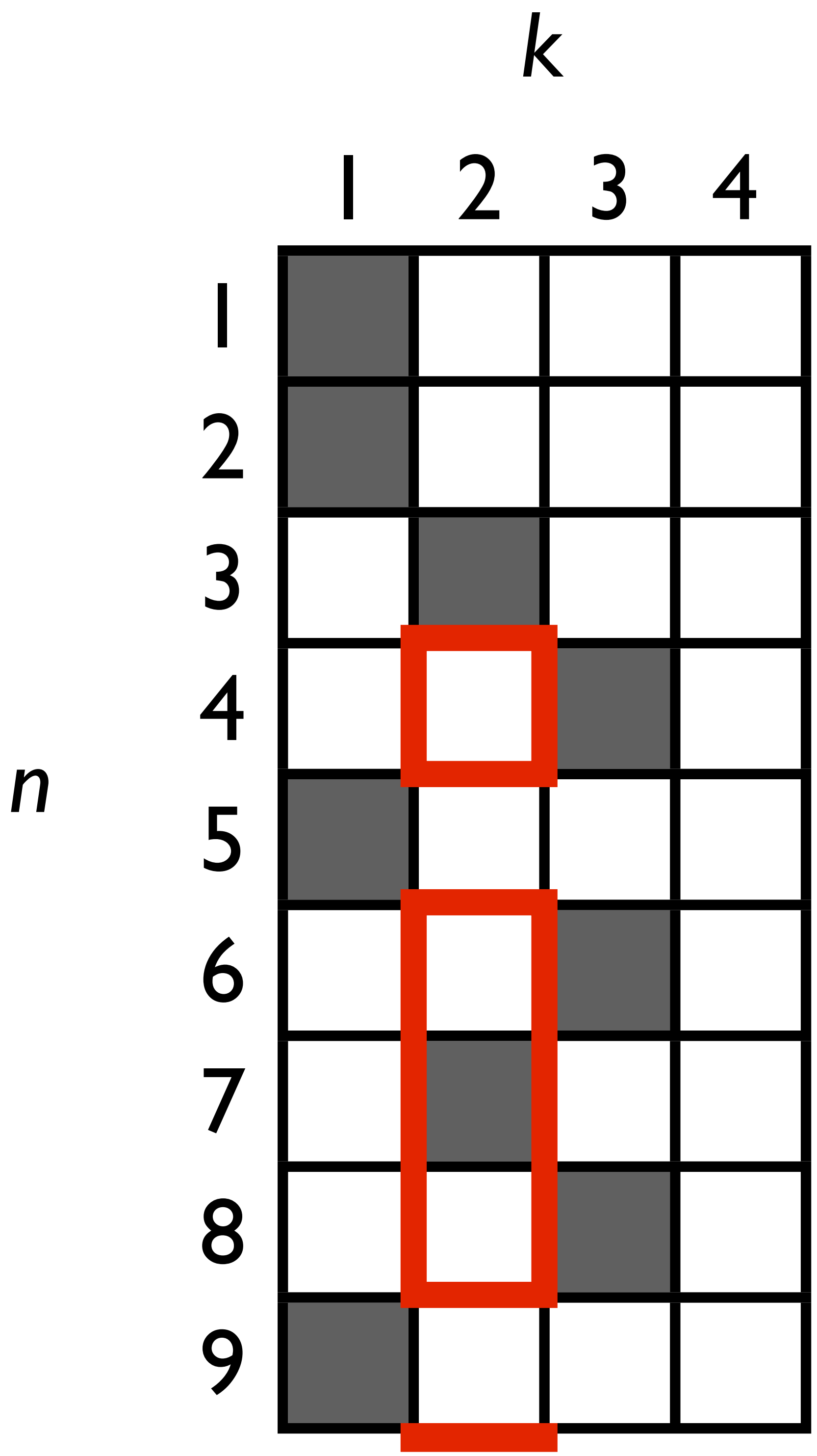


- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$

- First cluster: Polya urn with

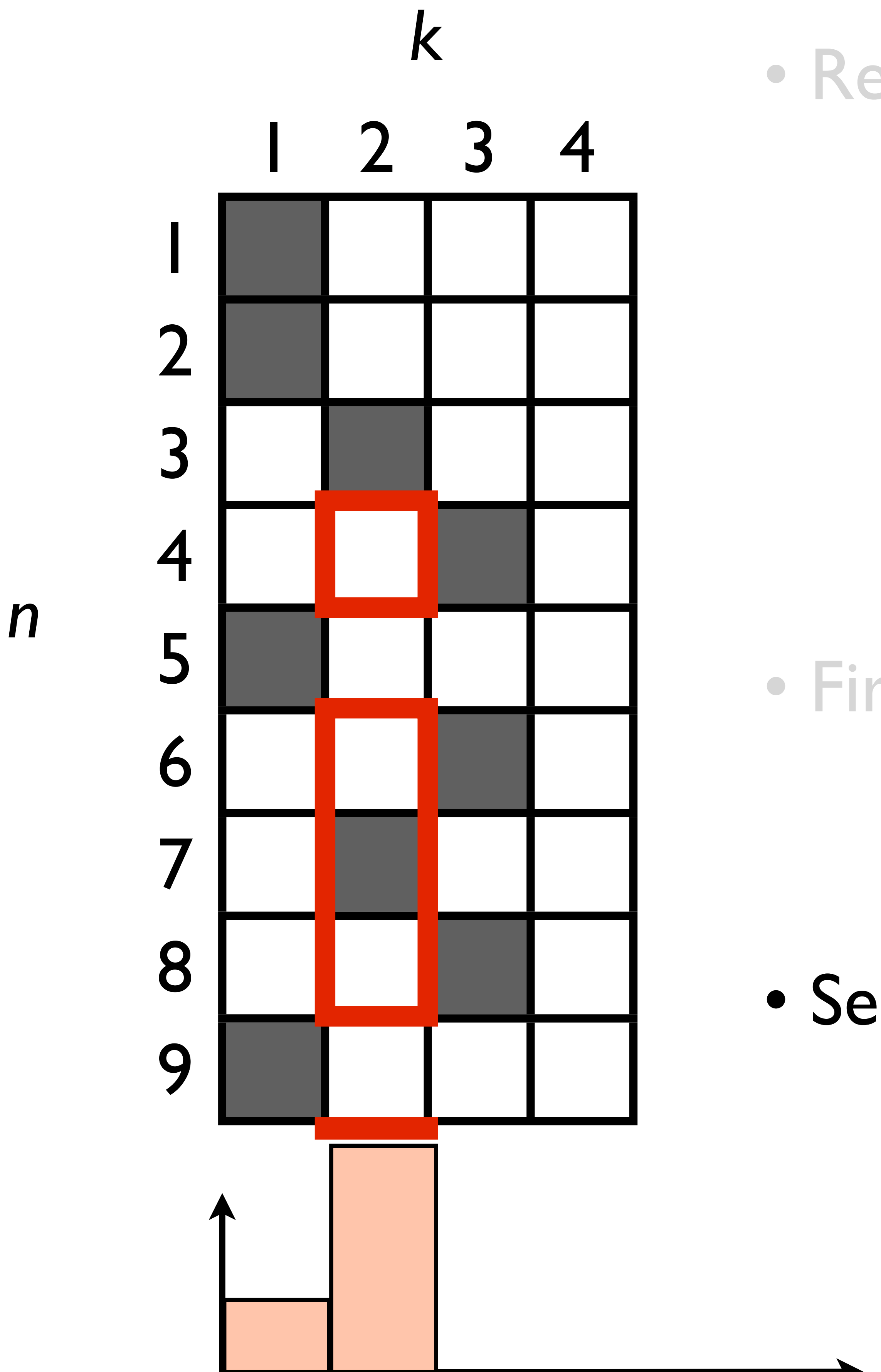
$$G_{1,0} = 1, W_{1,0} = \theta$$
$$V_1 \sim \text{Beta}(1, \theta)$$

CRP as Polya urns



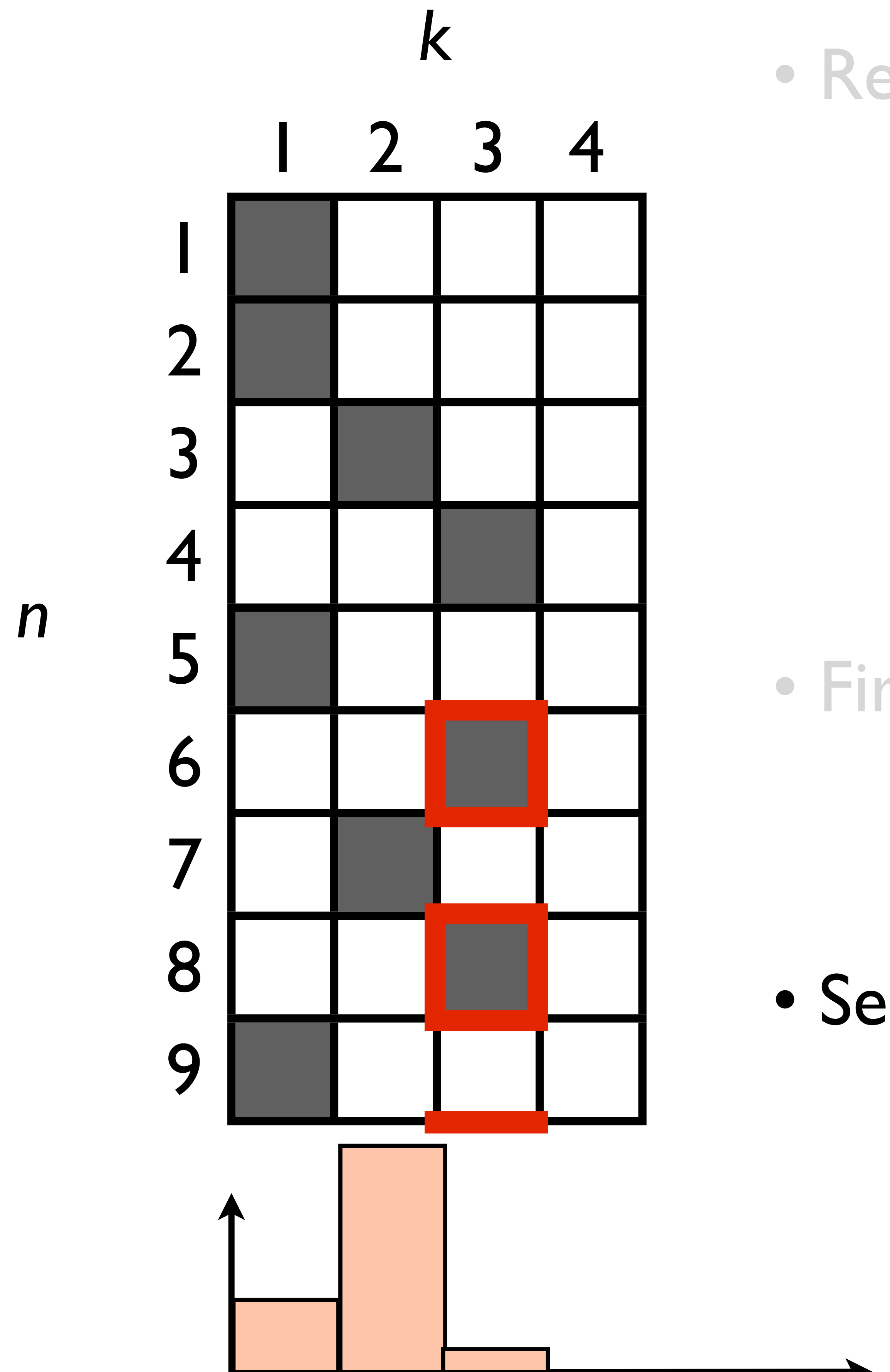
- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$
- First cluster: Polya urn with
  $$G_{1,0} = 1, W_{1,0} = \theta$$
  $$V_1 \sim \text{Beta}(1, \theta)$$
- Second cluster if not in first: Polya urn
  $$G_{2,0} = 1, W_{2,0} = \theta$$

CRP as Polya urns



- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K*+*1* with probability $\propto \theta$
- First cluster: Polya urn with
  $$G_{1,0} = 1, W_{1,0} = \theta$$
  $$V_1 \sim \text{Beta}(1, \theta)$$
- Second cluster if not in first: Polya urn
  $$G_{2,0} = 1, W_{2,0} = \theta$$
  $$V_2 \sim \text{Beta}(1, \theta)$$

CRP as Polya urns



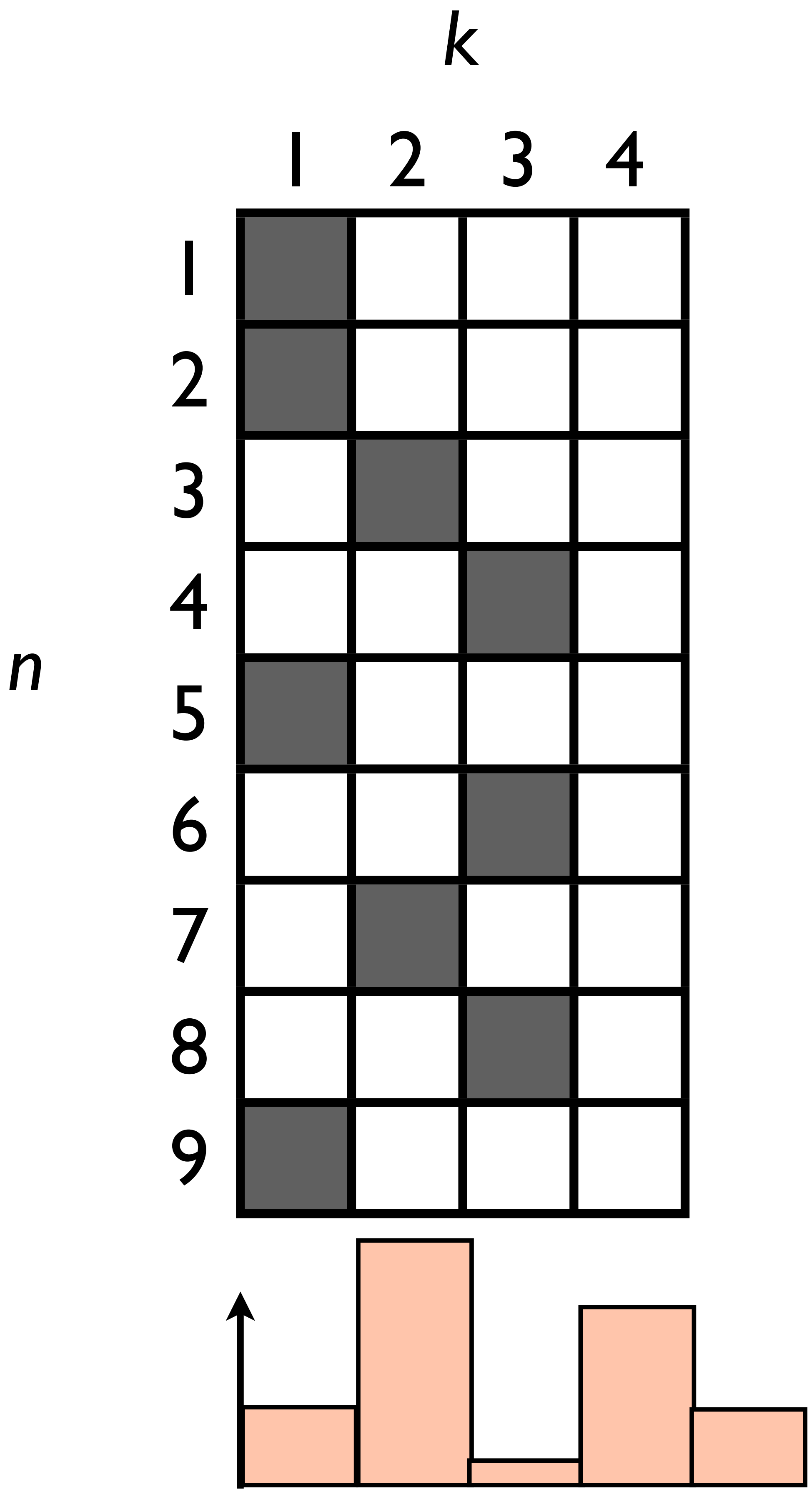
- Recursively: *n*th person sits
  - at table *k* (of *K*) with probability $\propto (\#\text{ people there})$
  - at new table *K+1* with probability $\propto \theta$
- First cluster: Polya urn with

$$G_{1,0} = 1, W_{1,0} = \theta$$
$$V_1 \sim \text{Beta}(1, \theta)$$

- Second cluster if not in first: Polya urn

$$G_{2,0} = 1, W_{2,0} = \theta$$
$$V_2 \sim \text{Beta}(1, \theta)$$

CRP as Polya urns

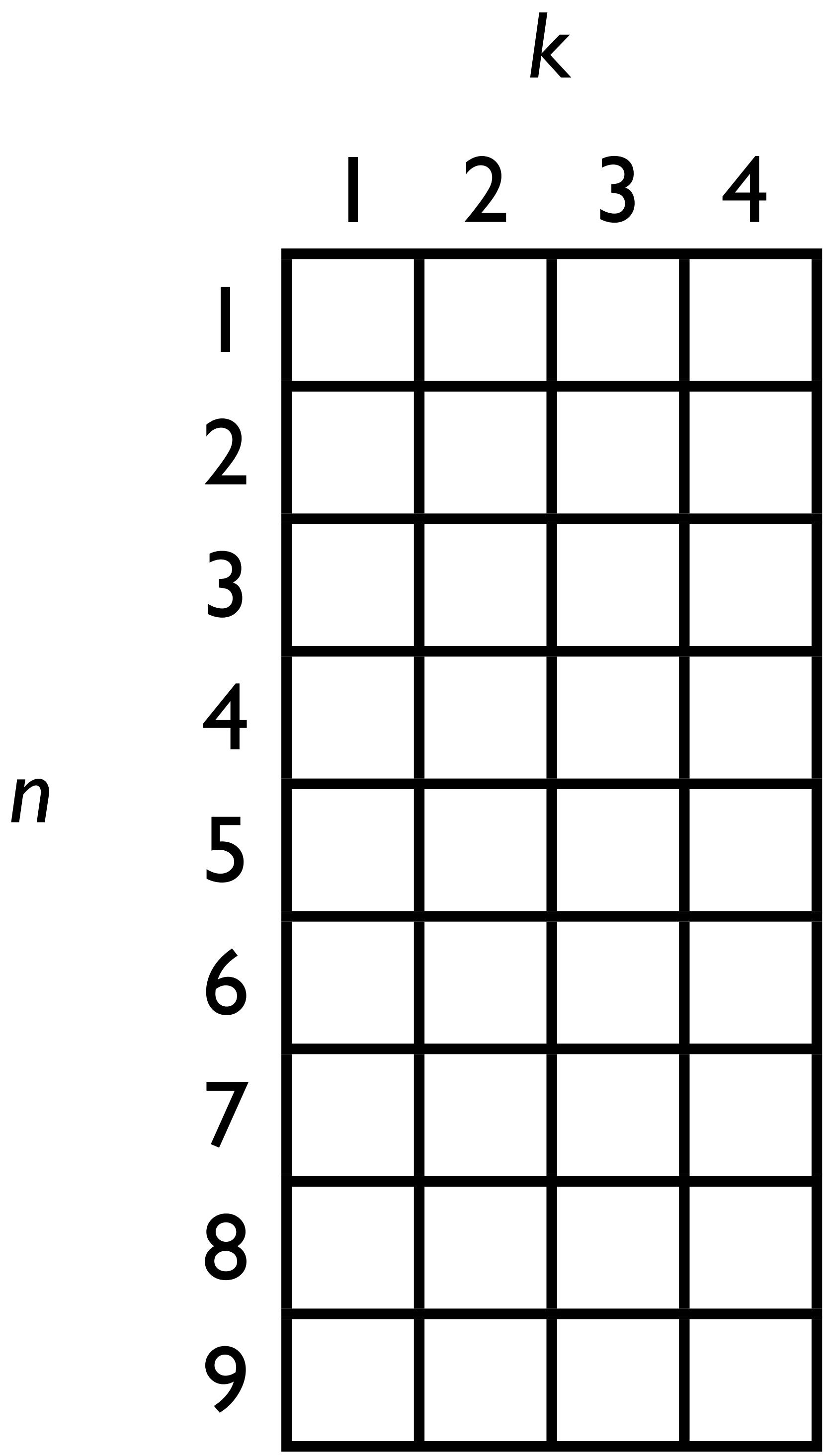


CRP as Polya urns

Another way to generate the CRP:

n \ *k*	1	2	3	4
1				
2				
3				
4				
5				
6				
7				
8				
9				

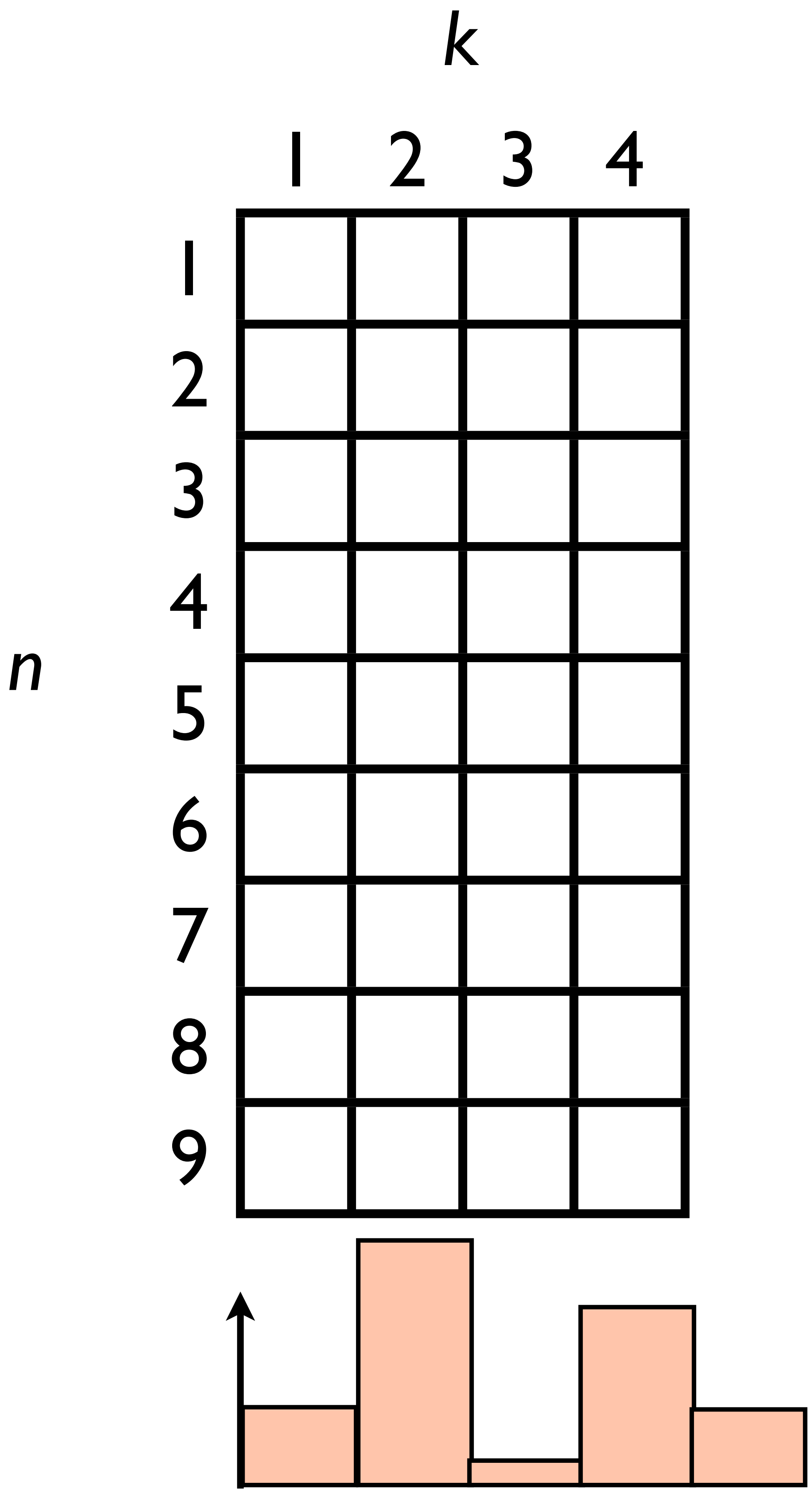
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables

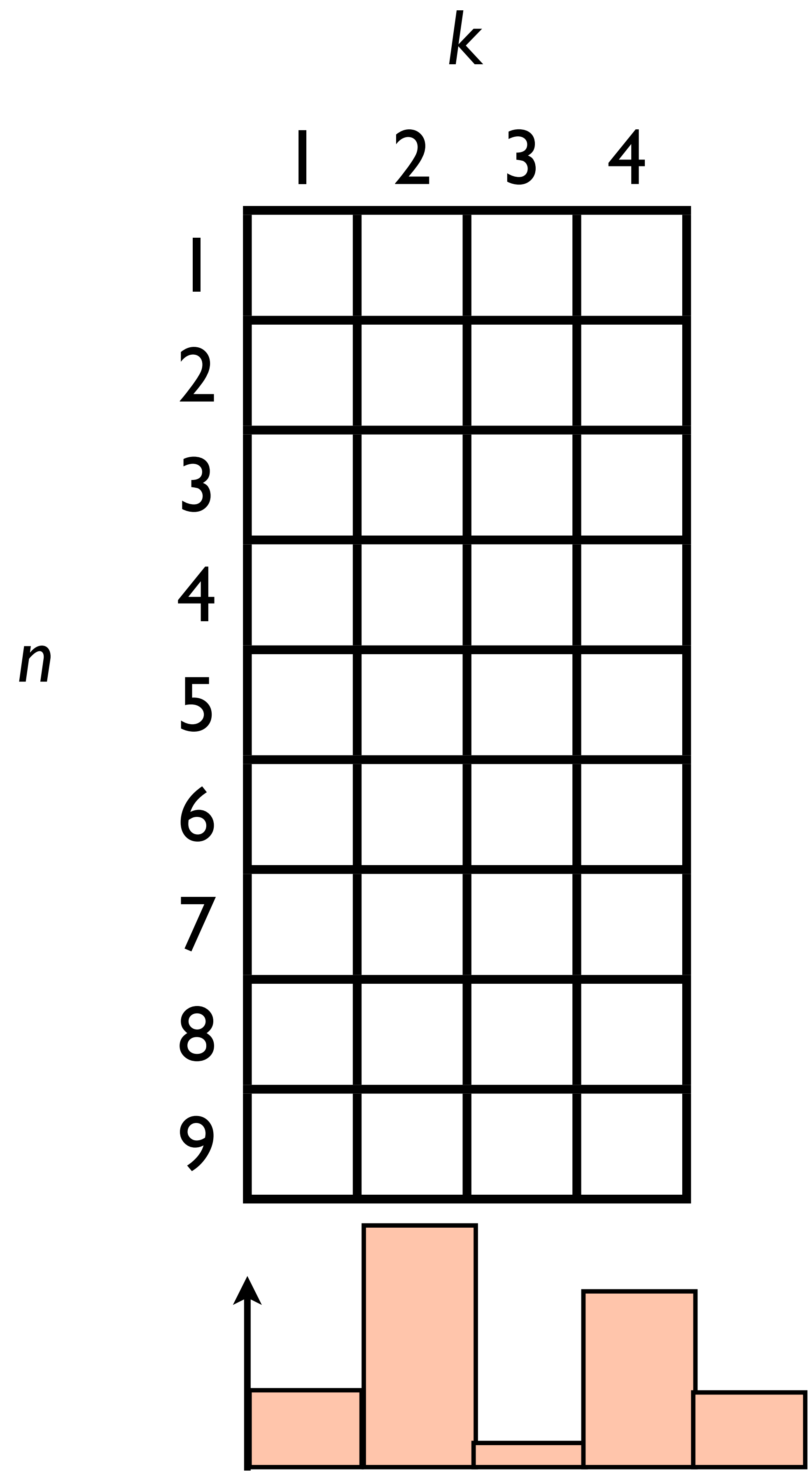
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables

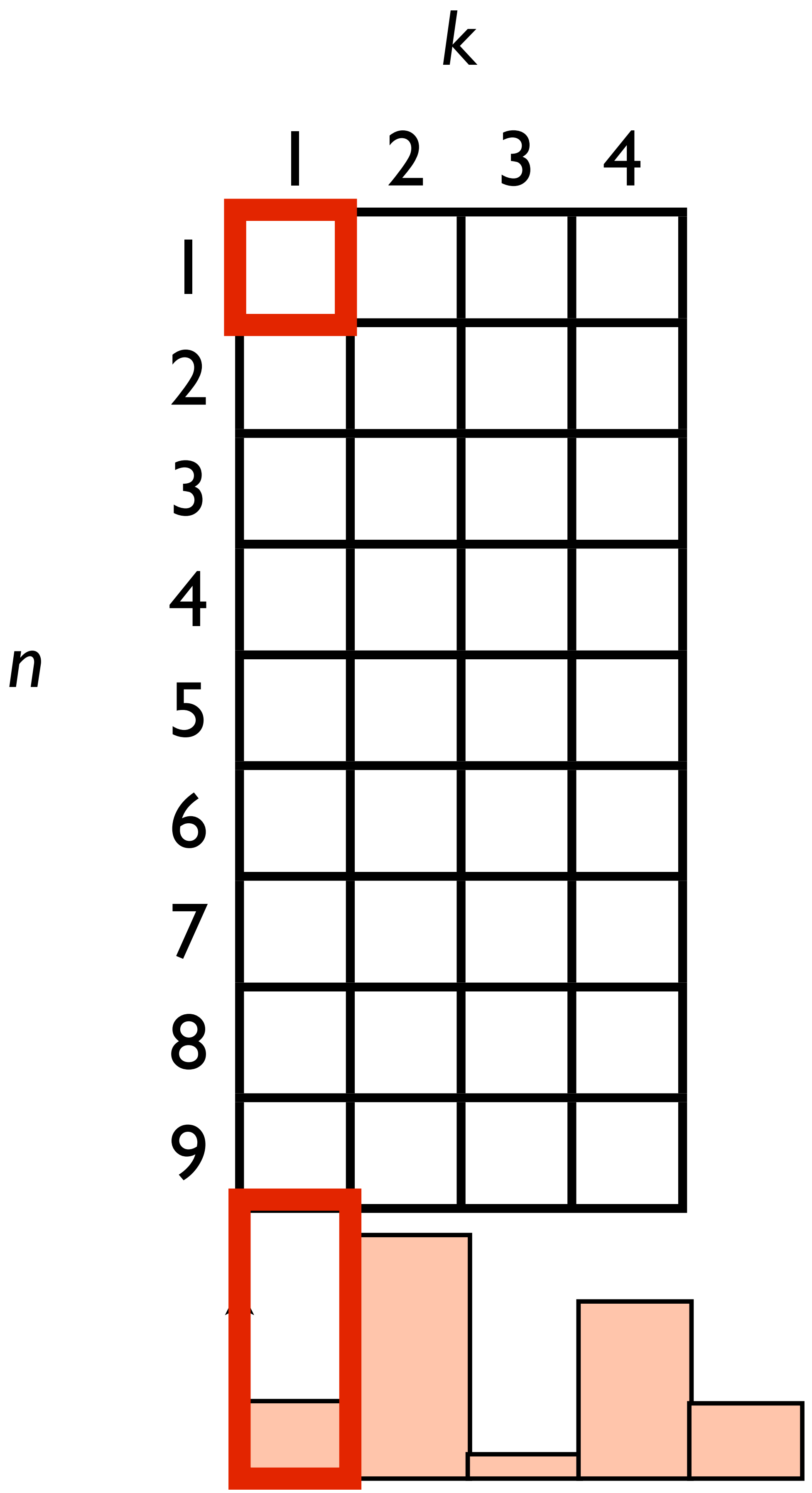
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables
- For each n, Bernoulli coin flips until success

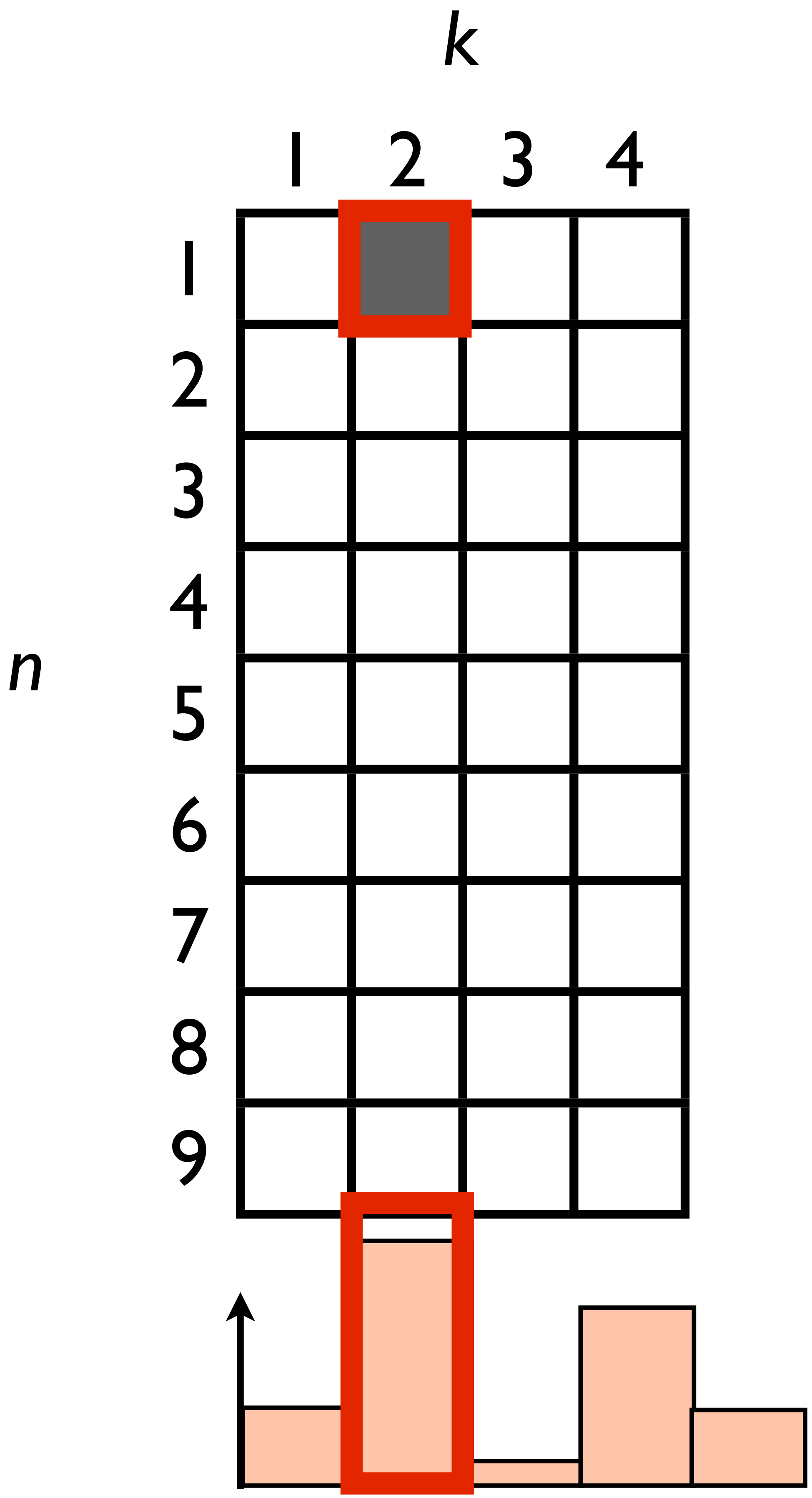
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables
- For each n, Bernoulli coin flips until success

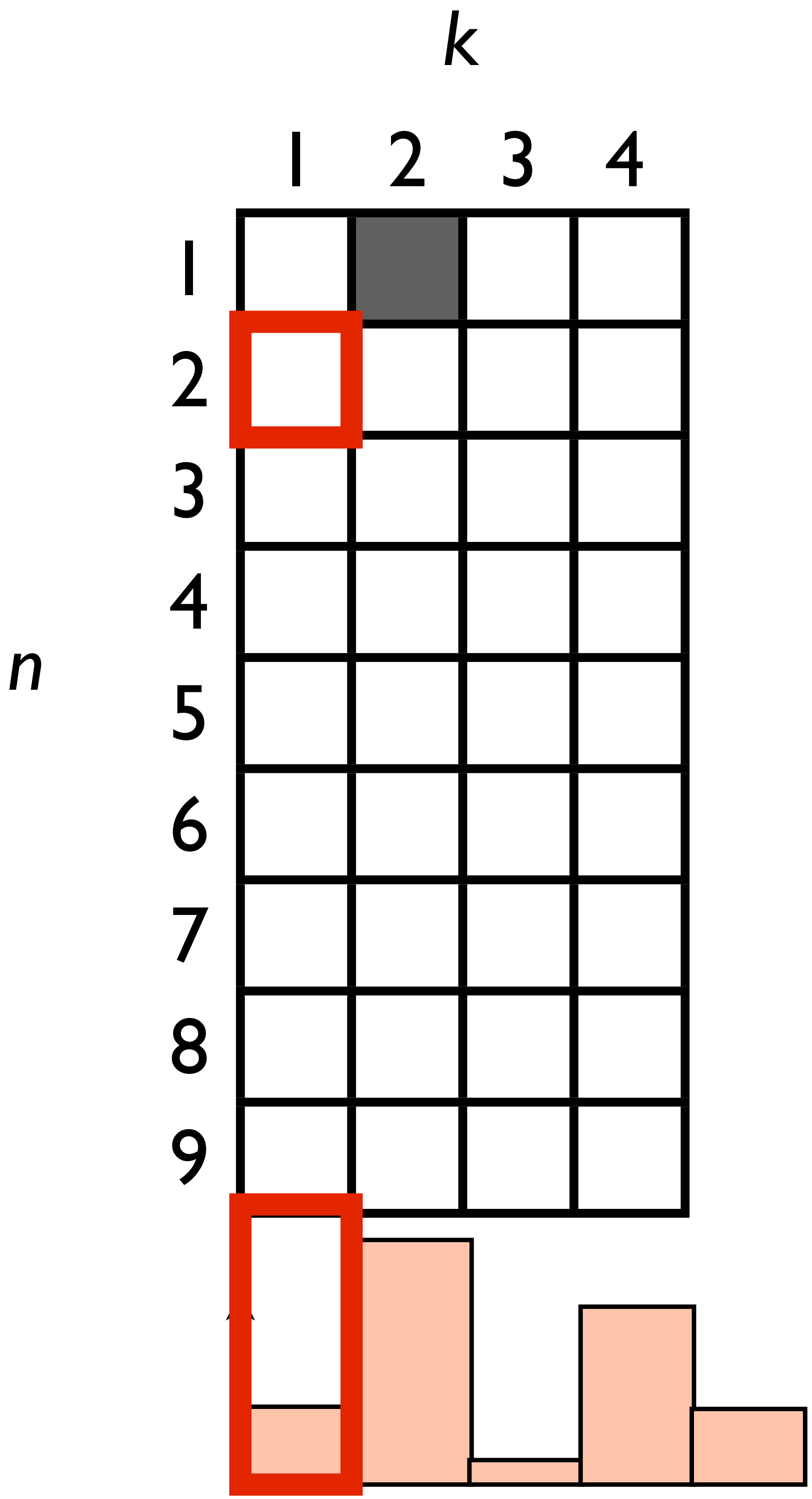
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables
- For each n, Bernoulli coin flips until success

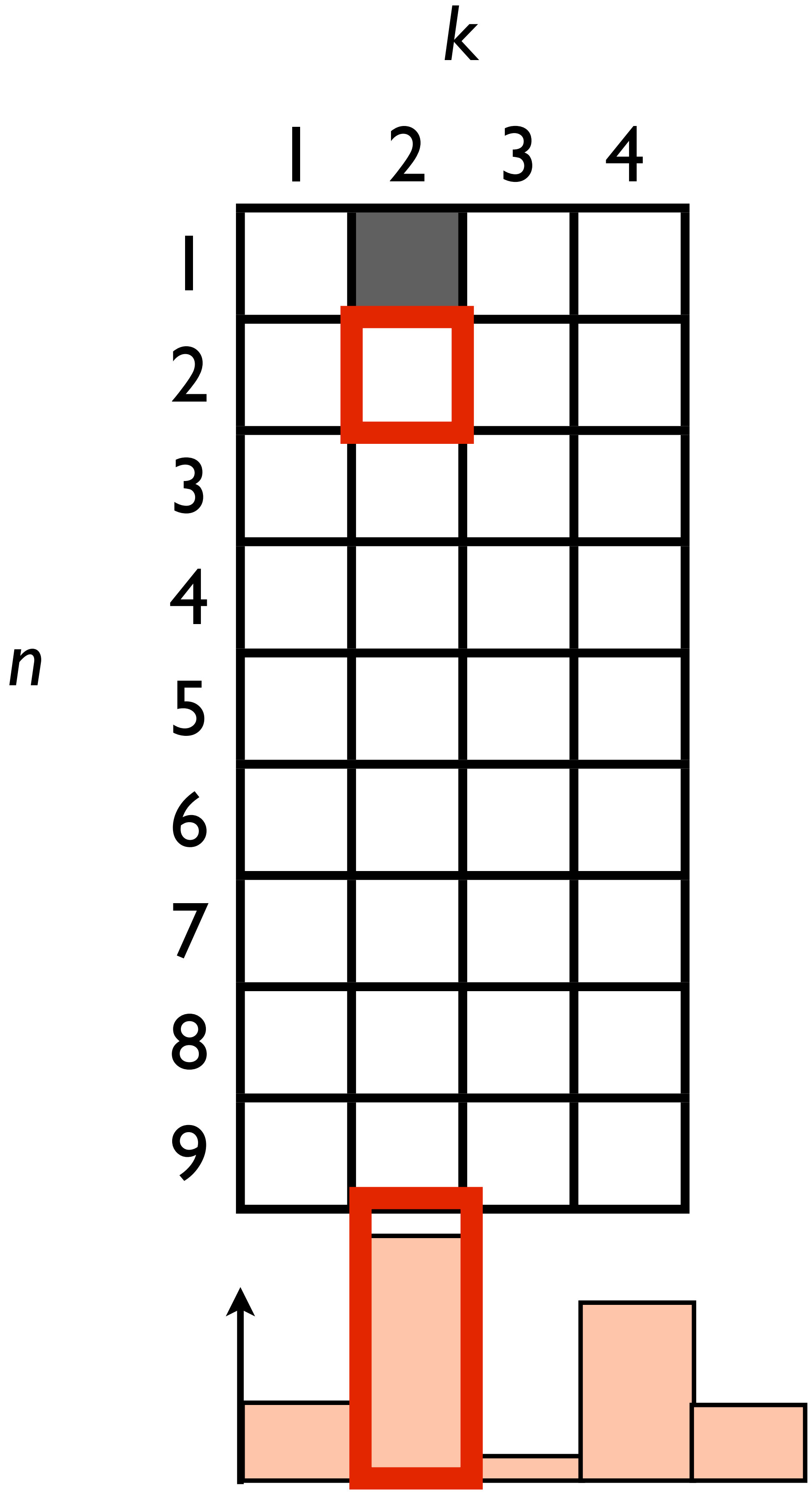
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables
- For each n, Bernoulli coin flips until success

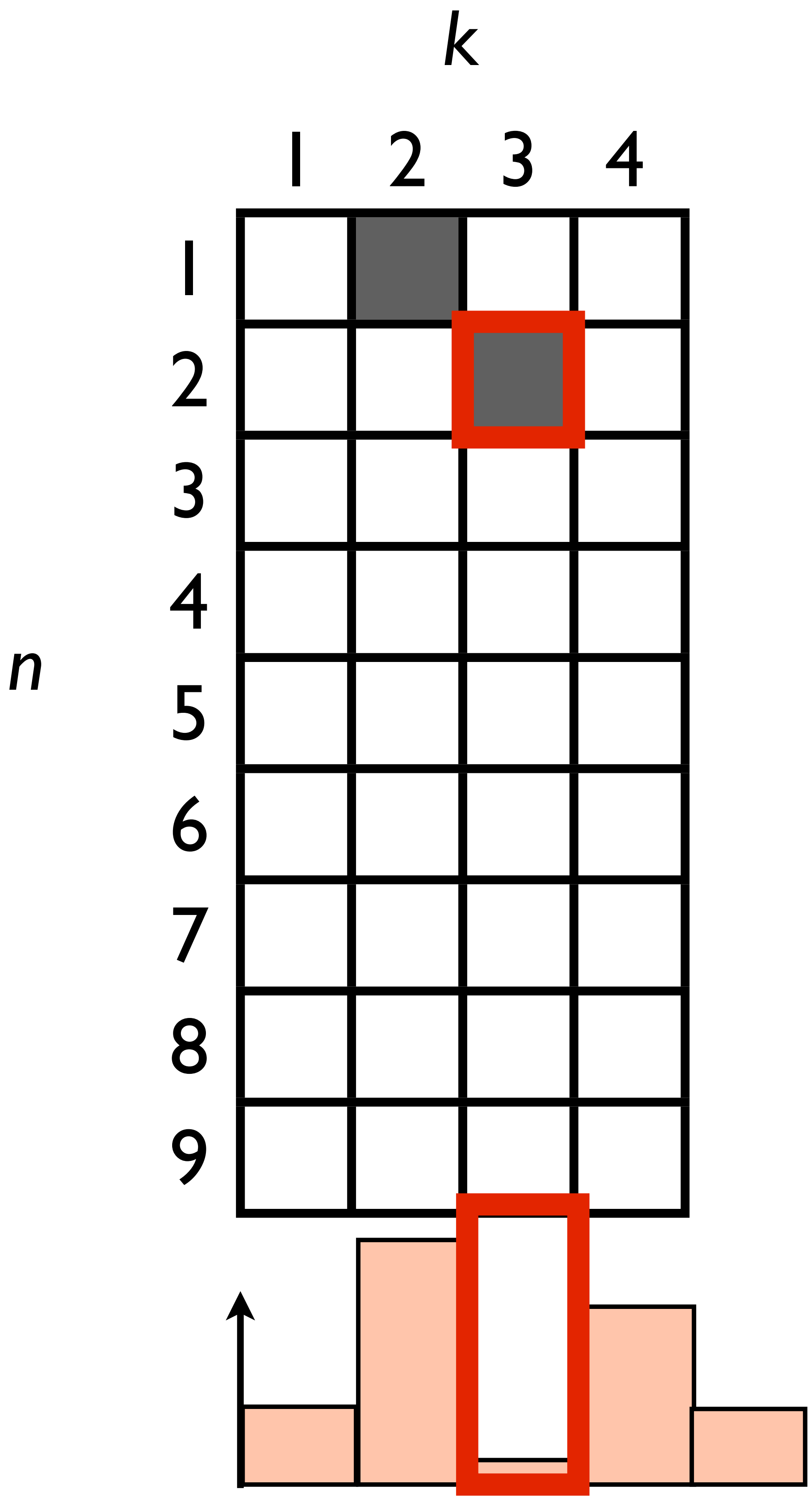
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables
- For each n, Bernoulli coin flips until success

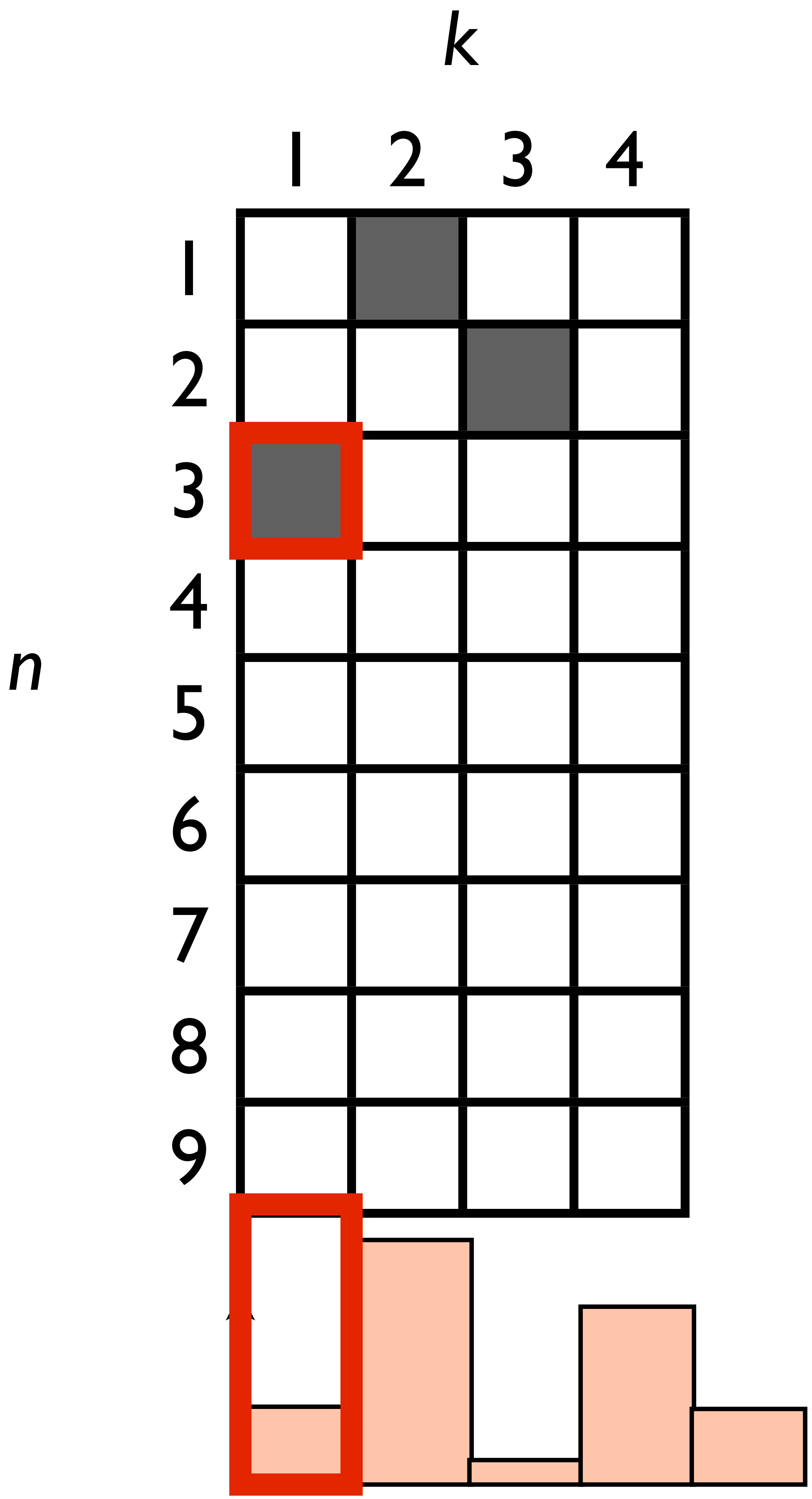
CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables
- For each n, Bernoulli coin flips until success

CRP as Polya urns



Another way to generate the CRP:

- Draw random beta variables
- For each n, Bernoulli coin flips until success

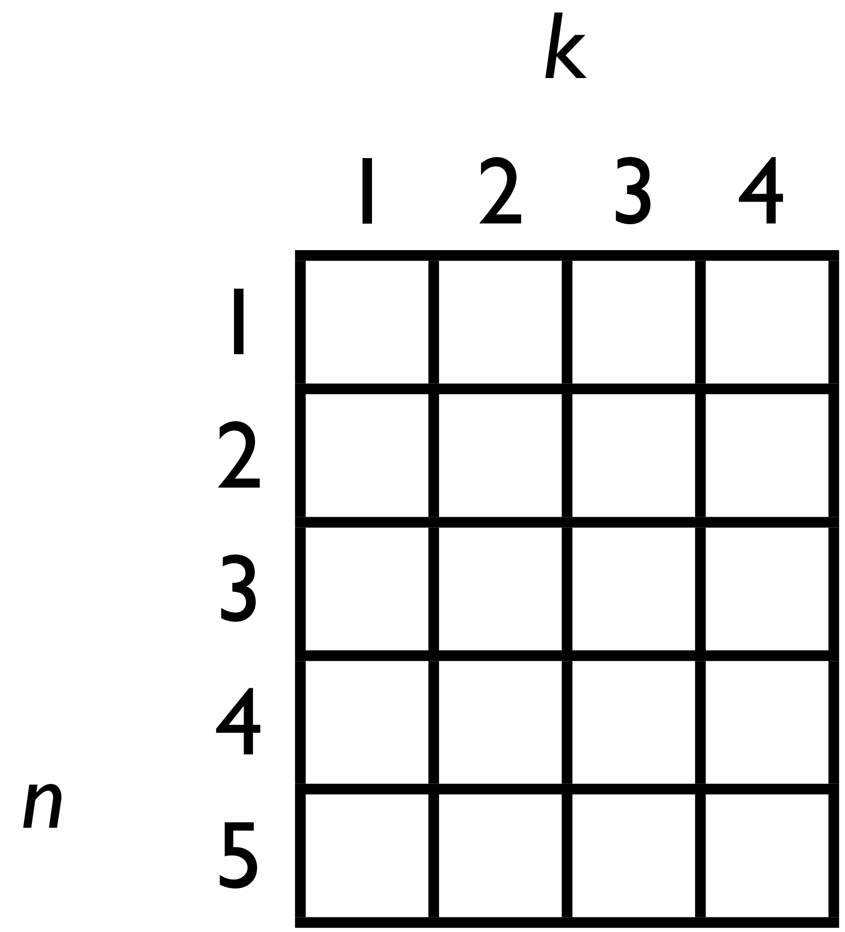
CRP as Polya urns

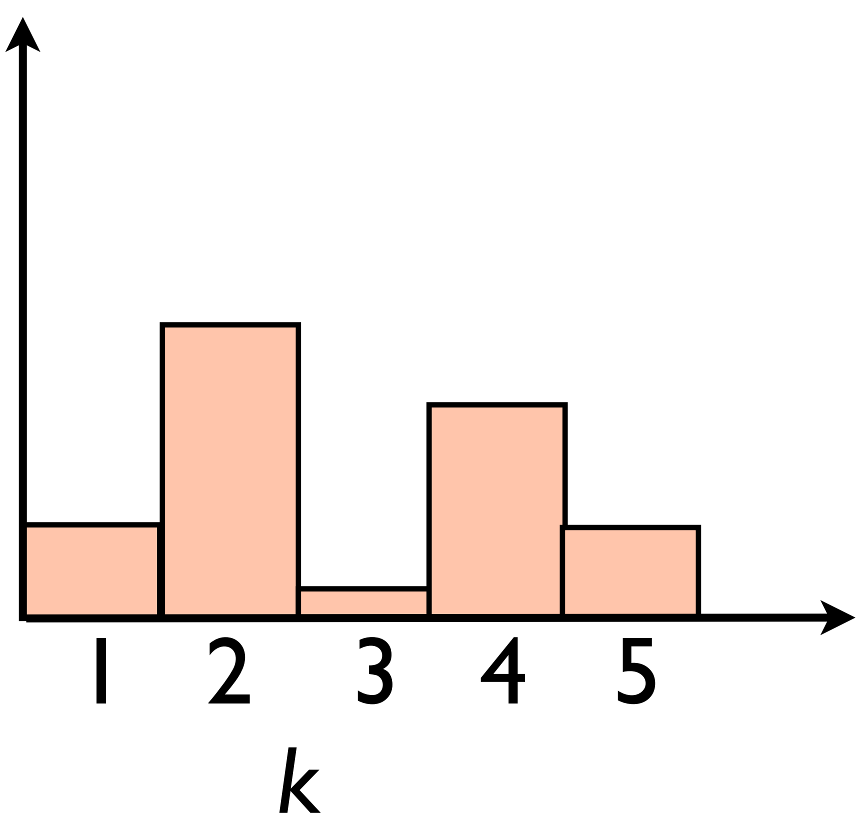
k

n	1	2	3	4
1				
2				
3				
4				
5				

Yet another way to generate the CRP:

CRP as Polya urns

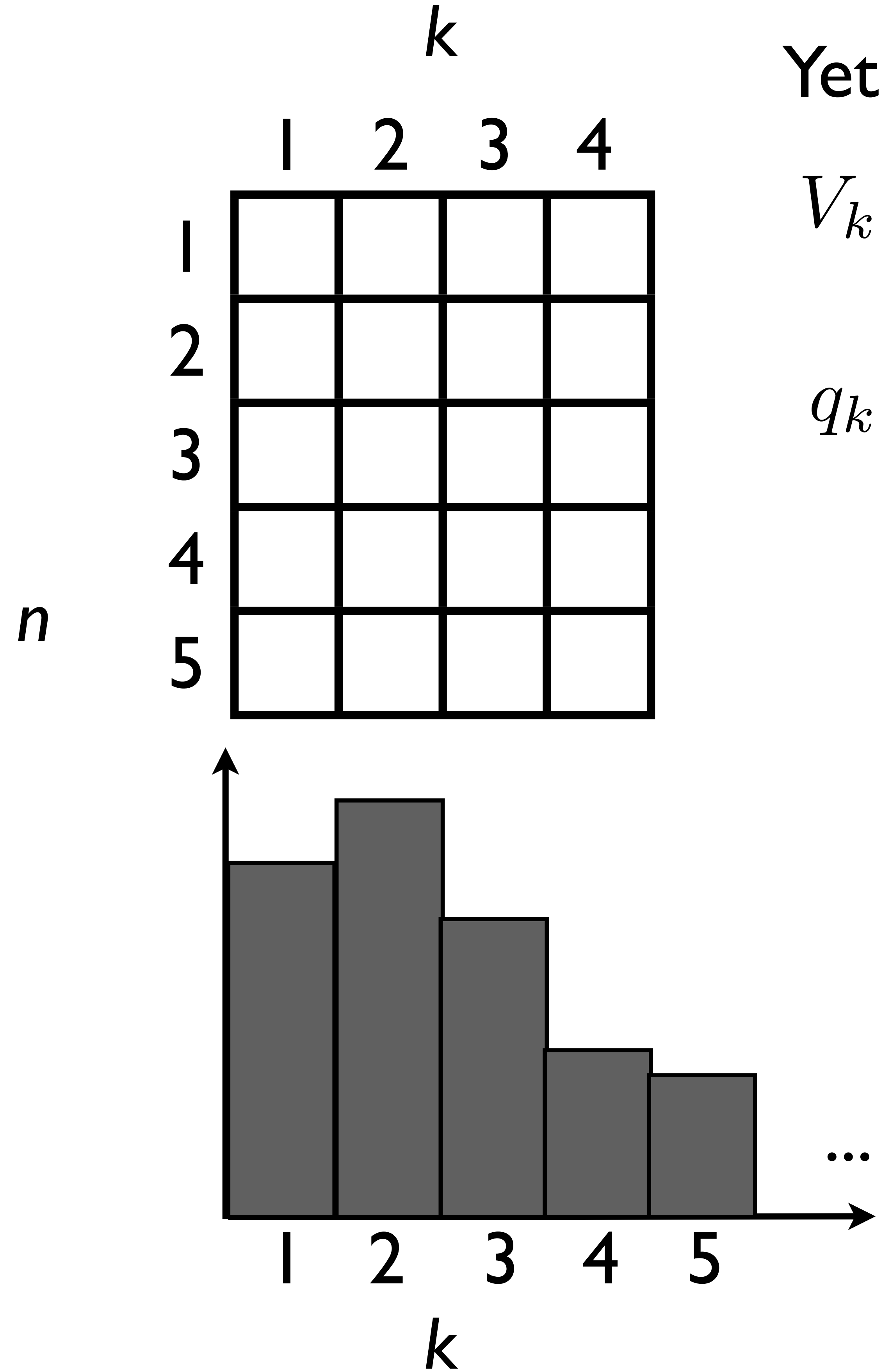




Yet another way to generate the CRP:

$$V_k \overset{iid}{\sim} \mathrm{Beta}(1, \theta), \quad k = 1, 2, \ldots$$

CRP as Polya urns

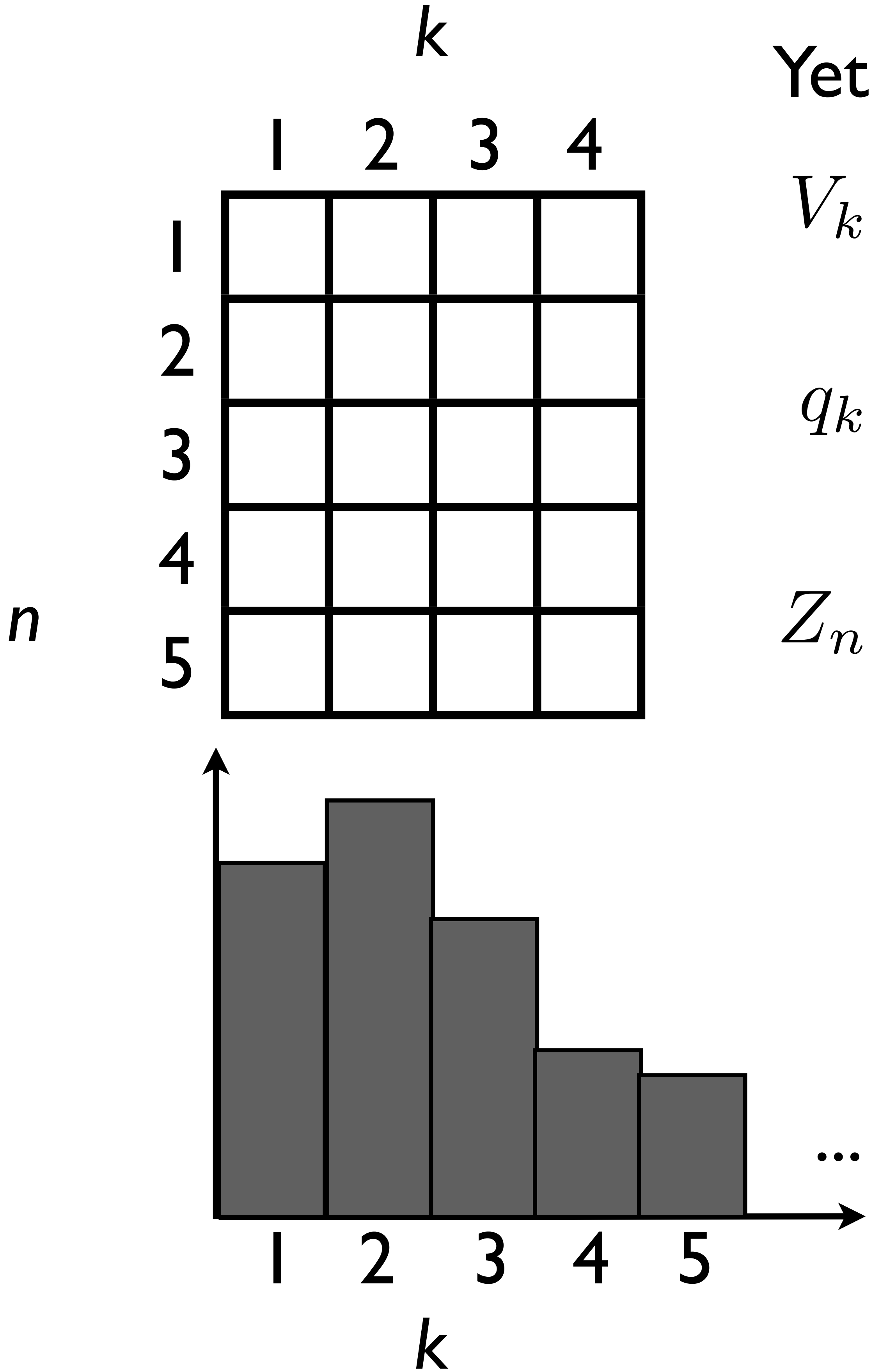


Yet another way to generate the CRP:

$$V_k \stackrel{iid}{\sim} \mathrm{Beta}(1, \theta), \quad k = 1, 2, \ldots$$

$$q_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \quad k = 1, 2, \ldots$$

CRP as Polya urns



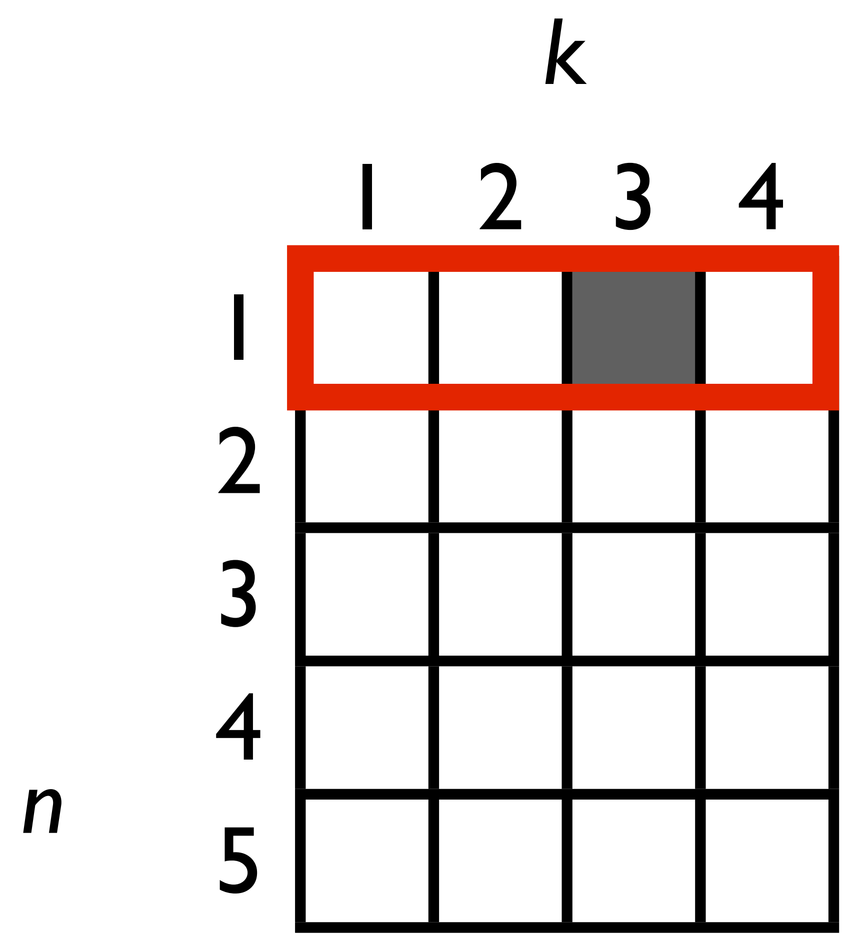
Yet another way to generate the CRP:

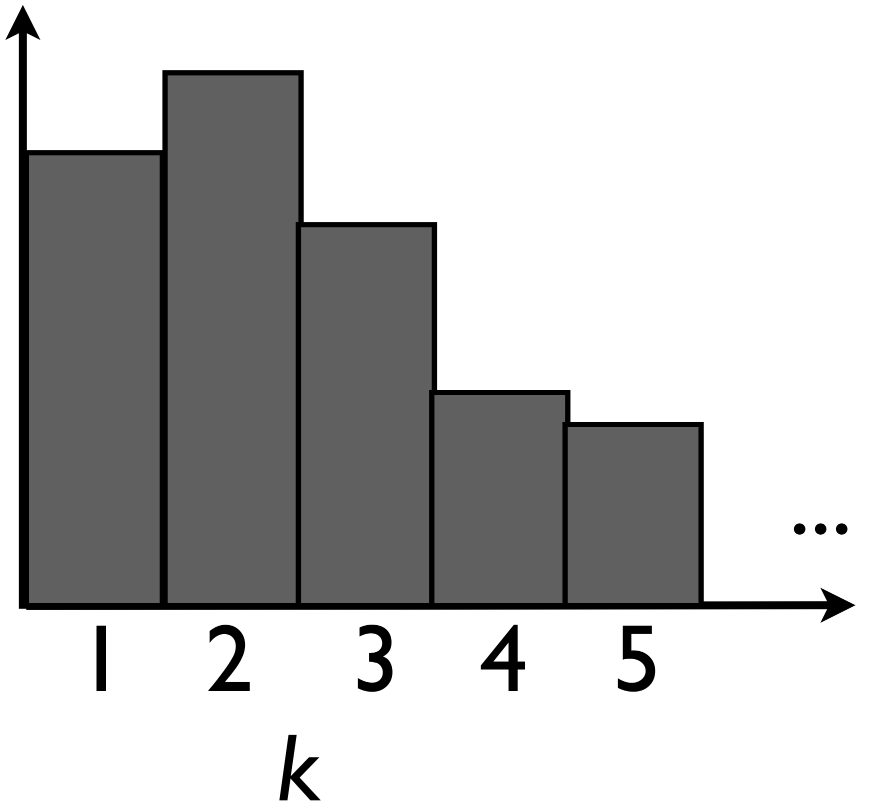
$$V_k \stackrel{iid}{\sim} \mathrm{Beta}(1, \theta), \quad k = 1, 2, \ldots$$

$$q_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \quad k = 1, 2, \ldots$$

$$Z_n \stackrel{iid}{\sim} q, \quad n = 1, 2, \ldots$$

CRP as Polya urns





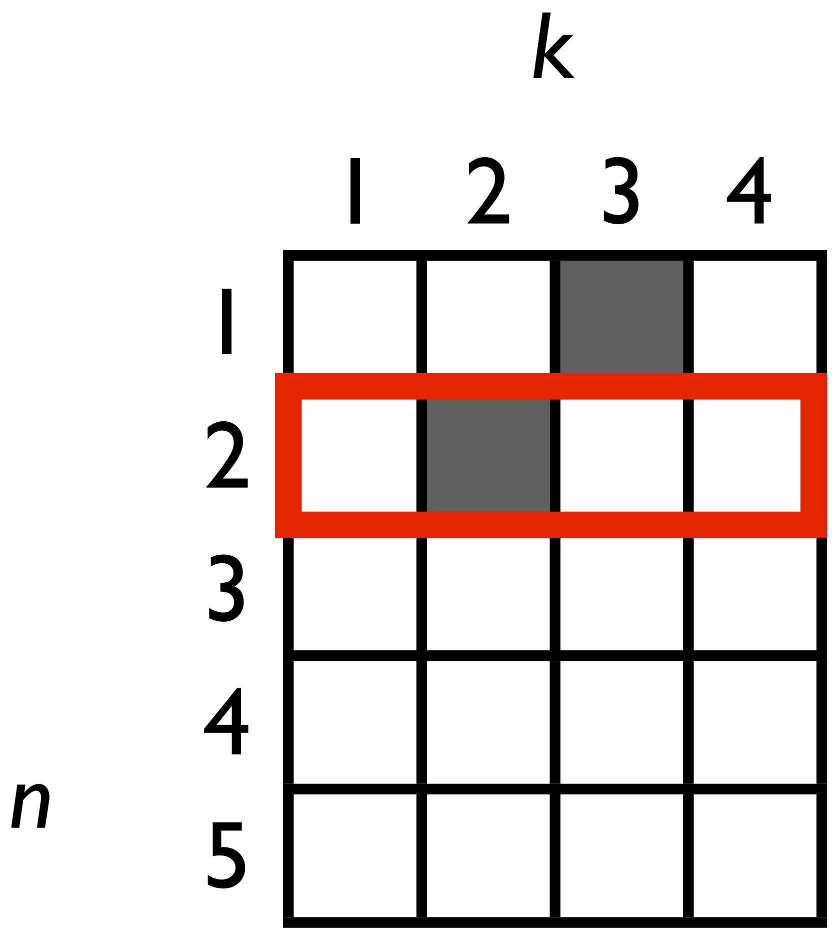
Yet another way to generate the CRP:

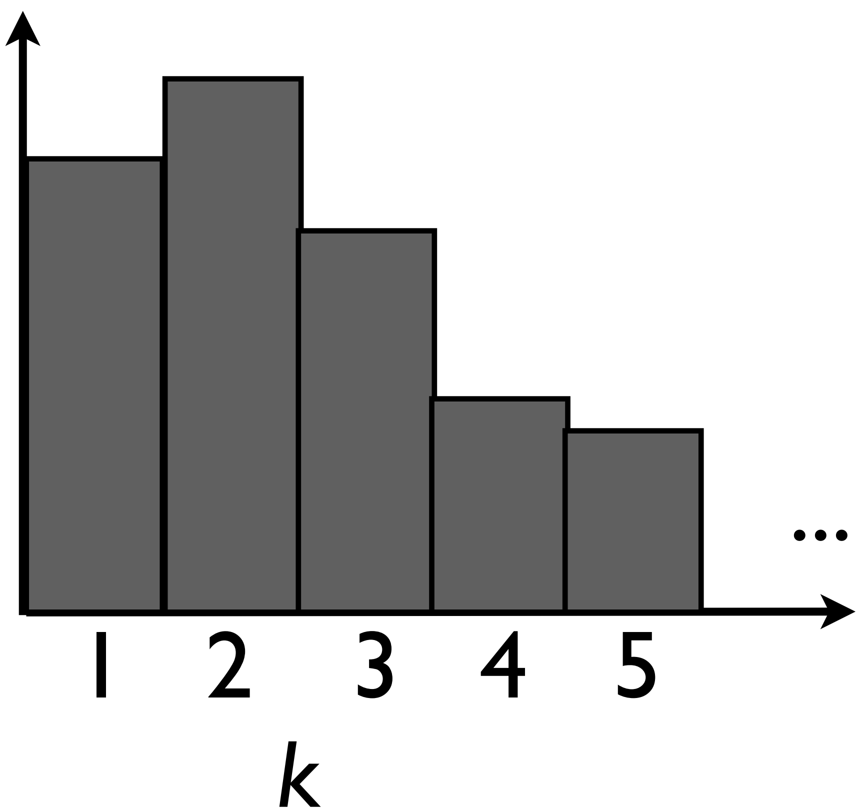
$$V_k \stackrel{iid}{\sim} \mathrm{Beta}(1, \theta), \quad k = 1, 2, \ldots$$

$$q_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \quad k = 1, 2, \ldots$$

$$Z_n \stackrel{iid}{\sim} q, \quad n = 1, 2, \ldots$$

CRP as Polya urns





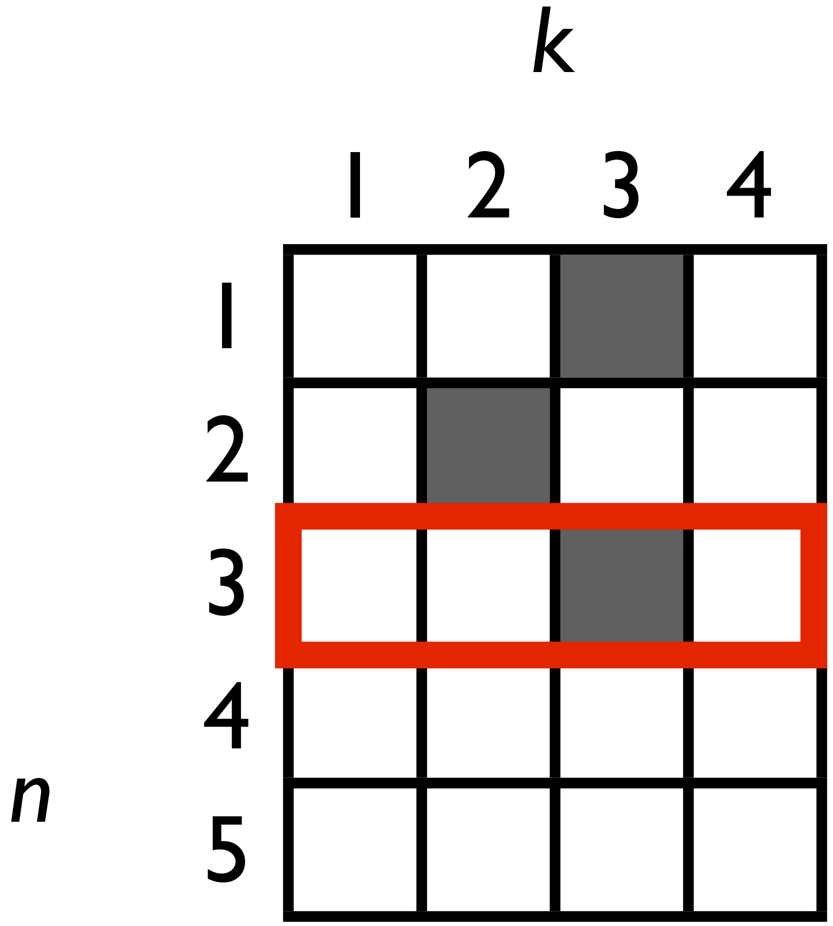
Yet another way to generate the CRP:

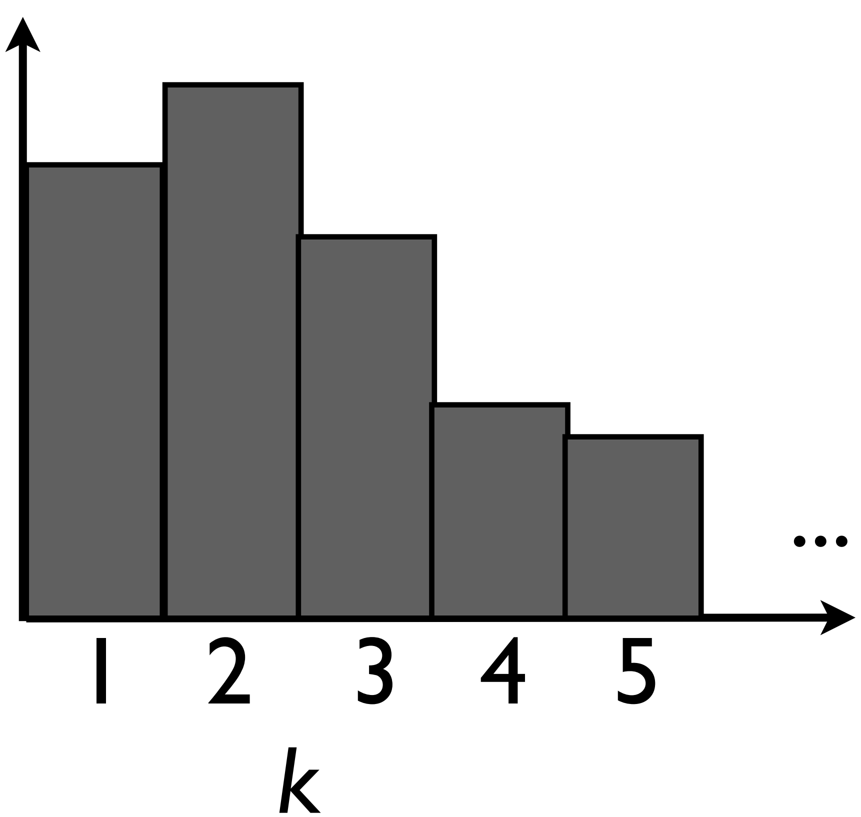
$$V_k \overset{iid}{\sim} \text{Beta}(1, \theta), \quad k = 1, 2, \ldots$$

$$q_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \quad k = 1, 2, \ldots$$

$$Z_n \overset{iid}{\sim} q, \quad n = 1, 2, \ldots$$

CRP as Polya urns





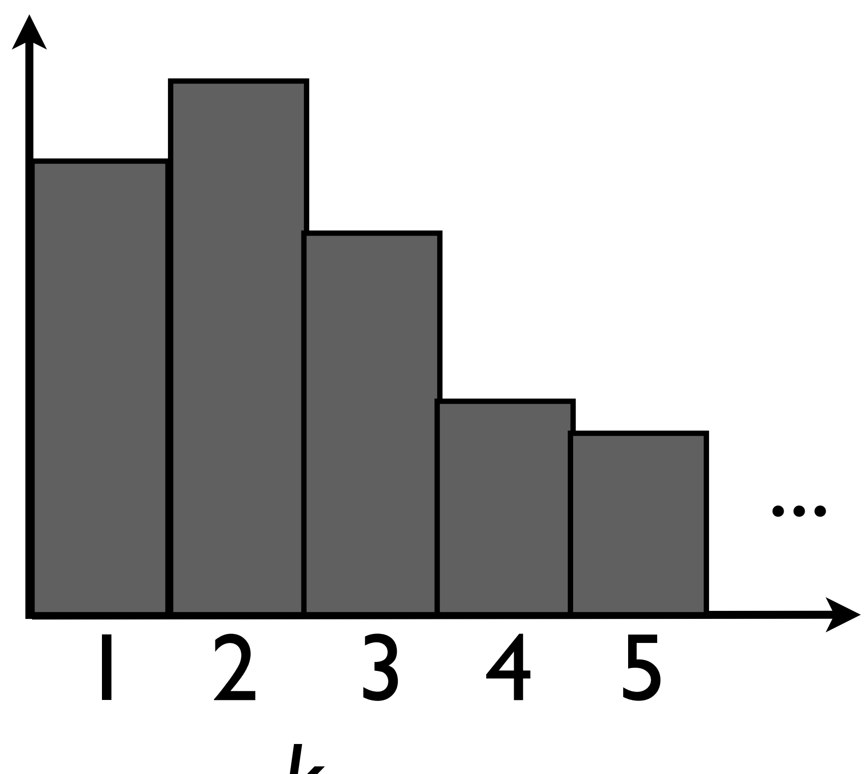
Yet another way to generate the CRP:

$$V_k \stackrel{iid}{\sim} \mathrm{Beta}(1, \theta), \quad k = 1, 2, \ldots$$

$$q_k = V_k \prod_{j=1}^{k-1} (1 - V_j), \quad k = 1, 2, \ldots$$

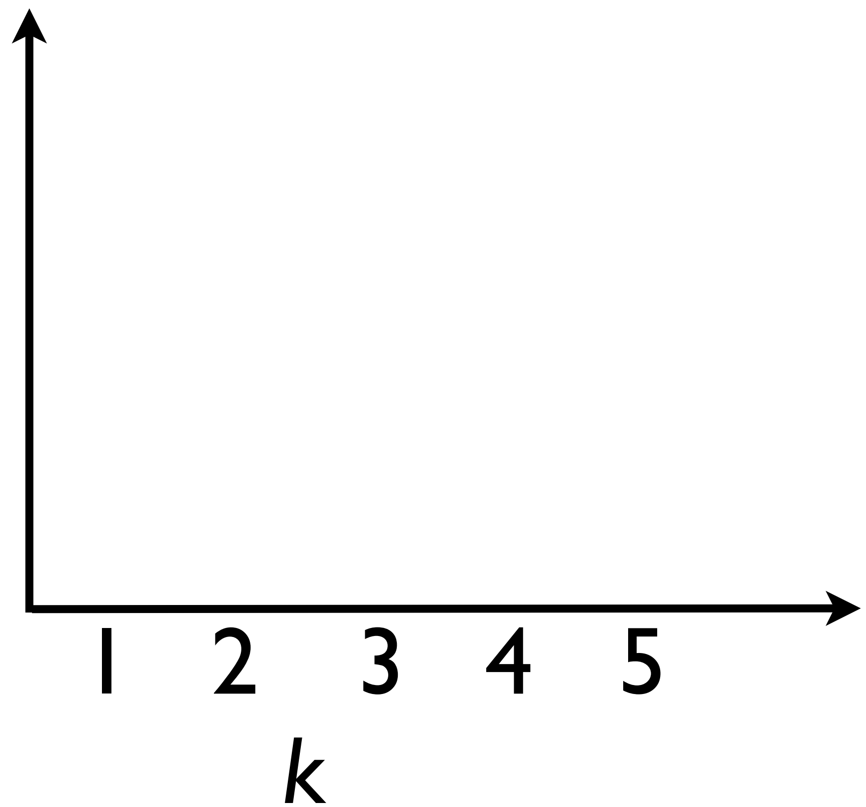
$$Z_n \stackrel{iid}{\sim} q, \quad n = 1, 2, \ldots$$

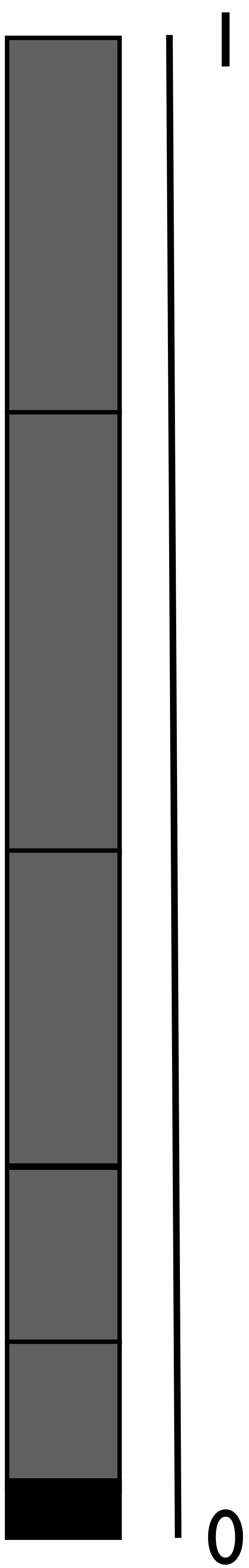
Stick-breaking



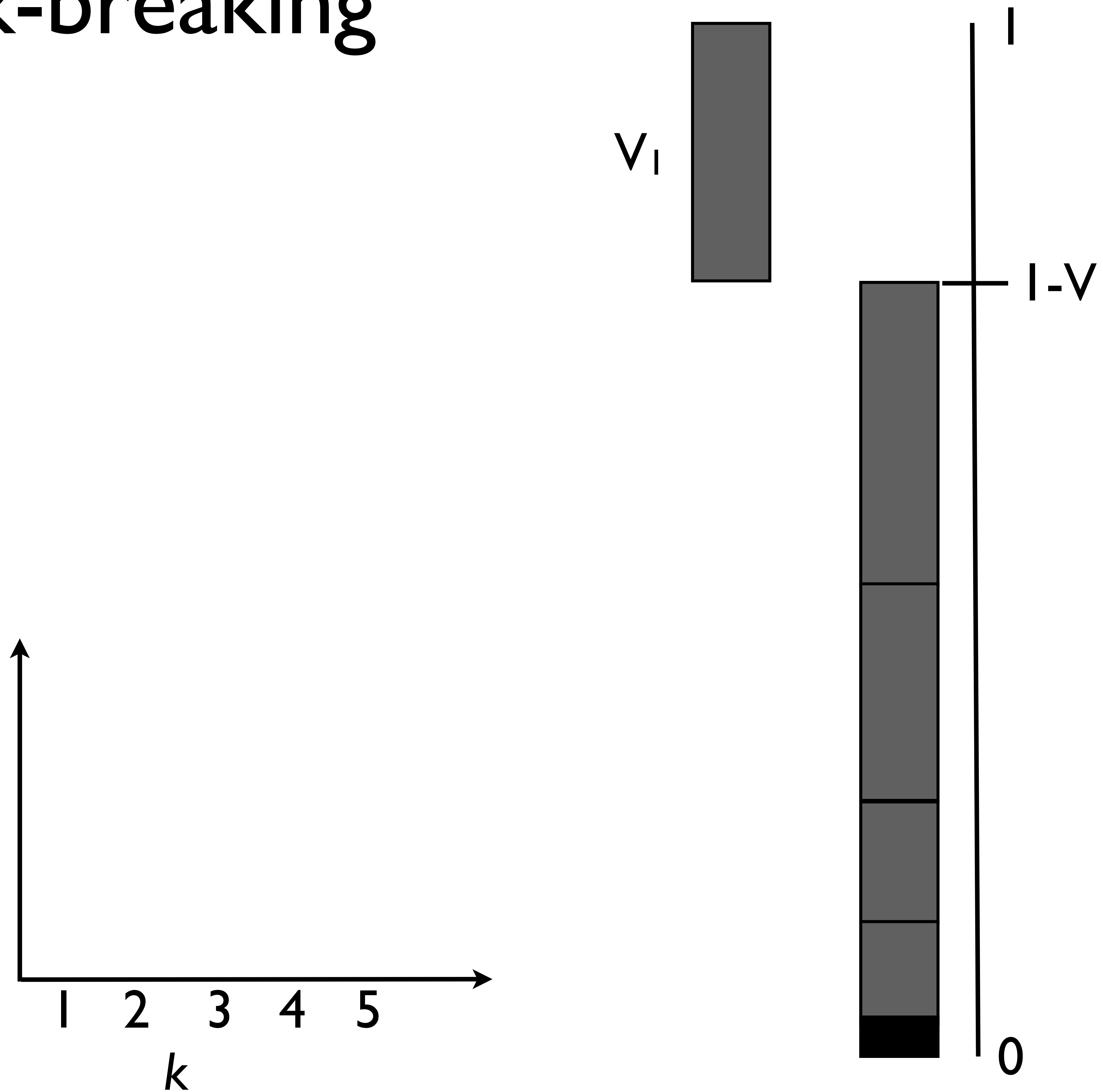
[McCloskey 1965; Patil and Taillie 1977; Sethuraman 1984; Ishwaran, James 2001]

Stick-breaking

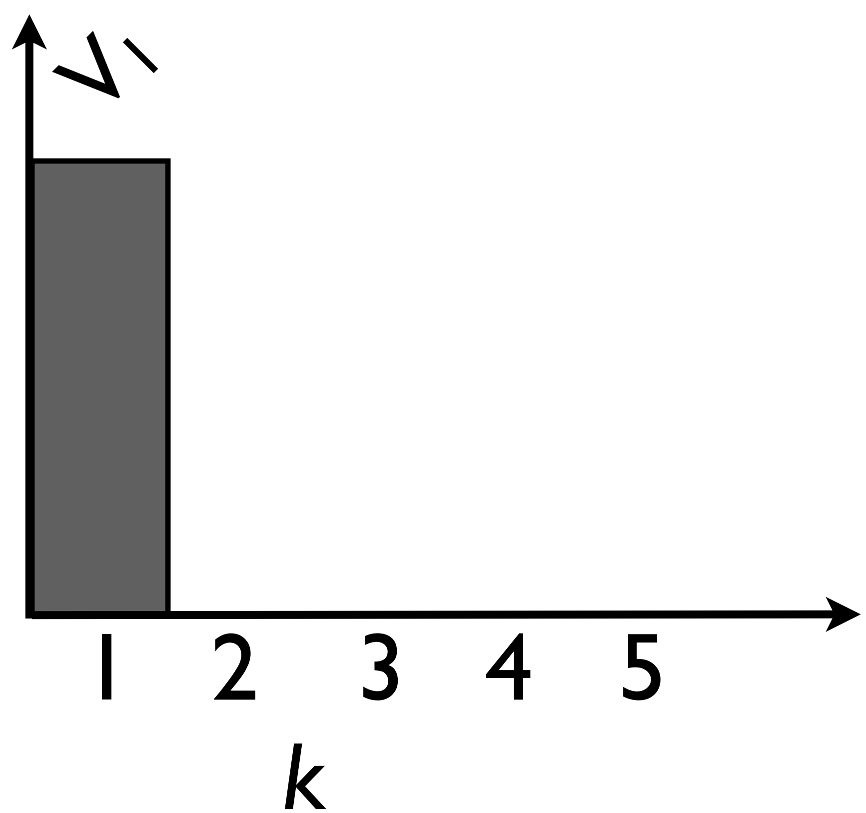


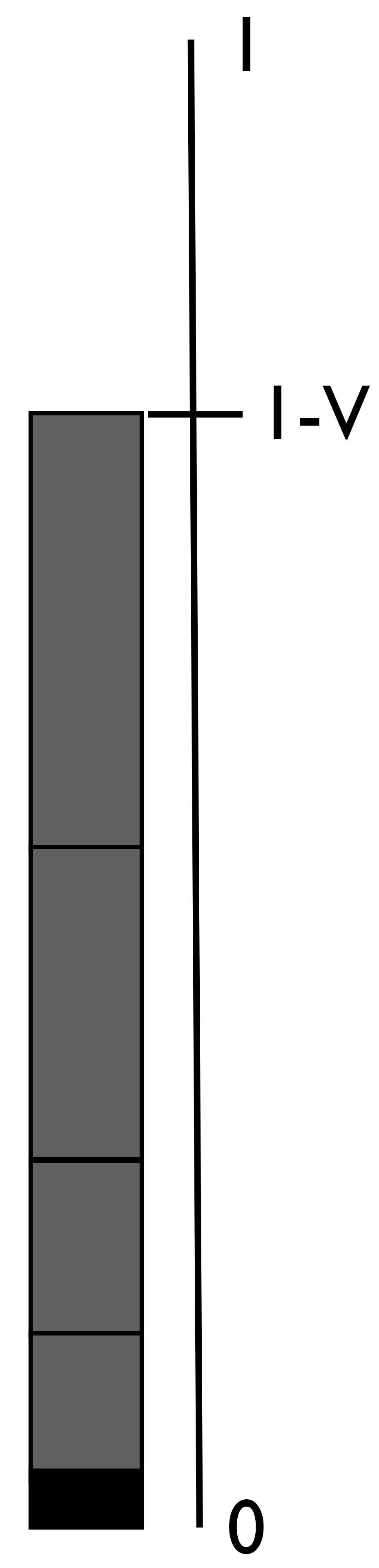


Stick-breaking

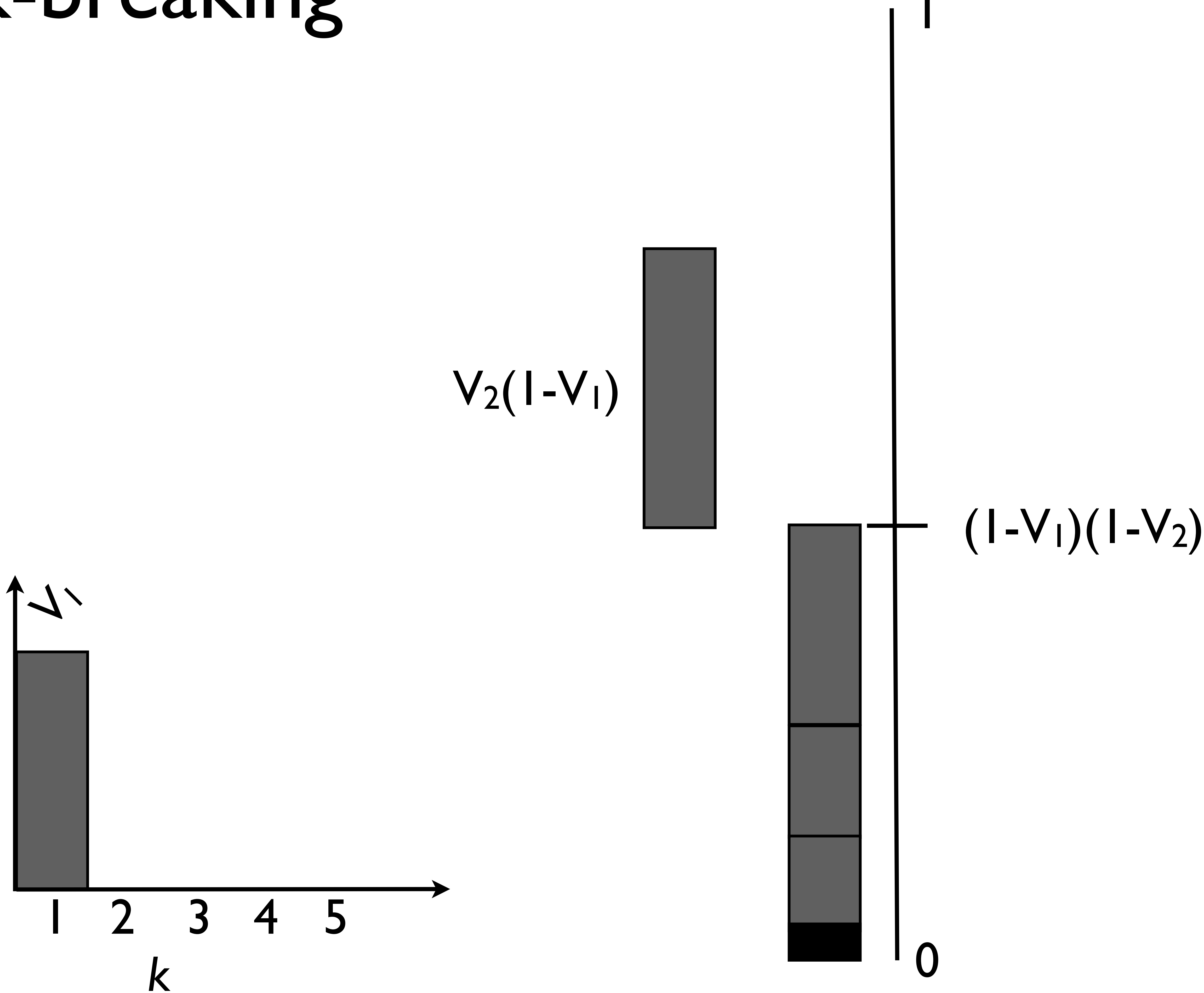


Stick-breaking

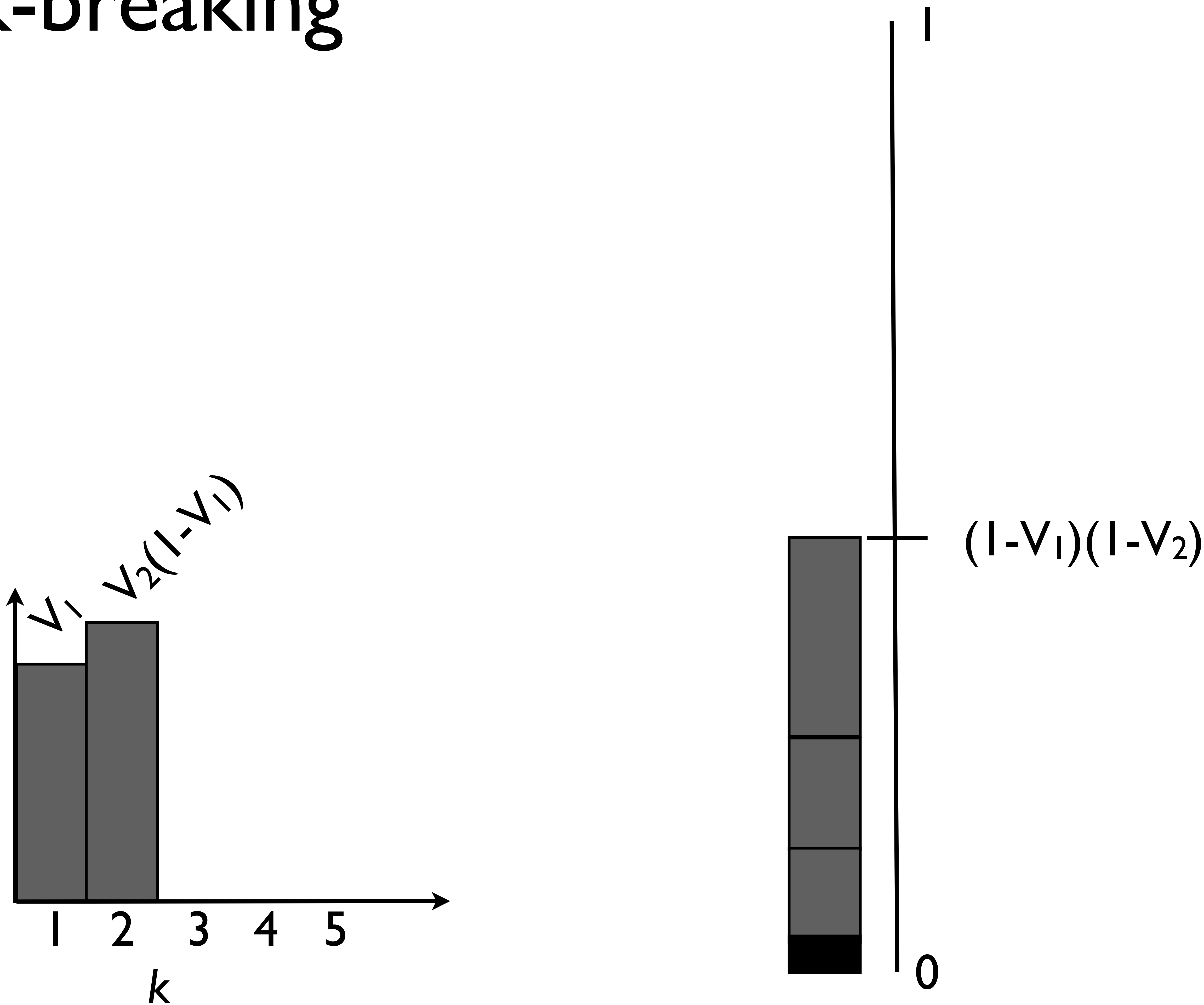




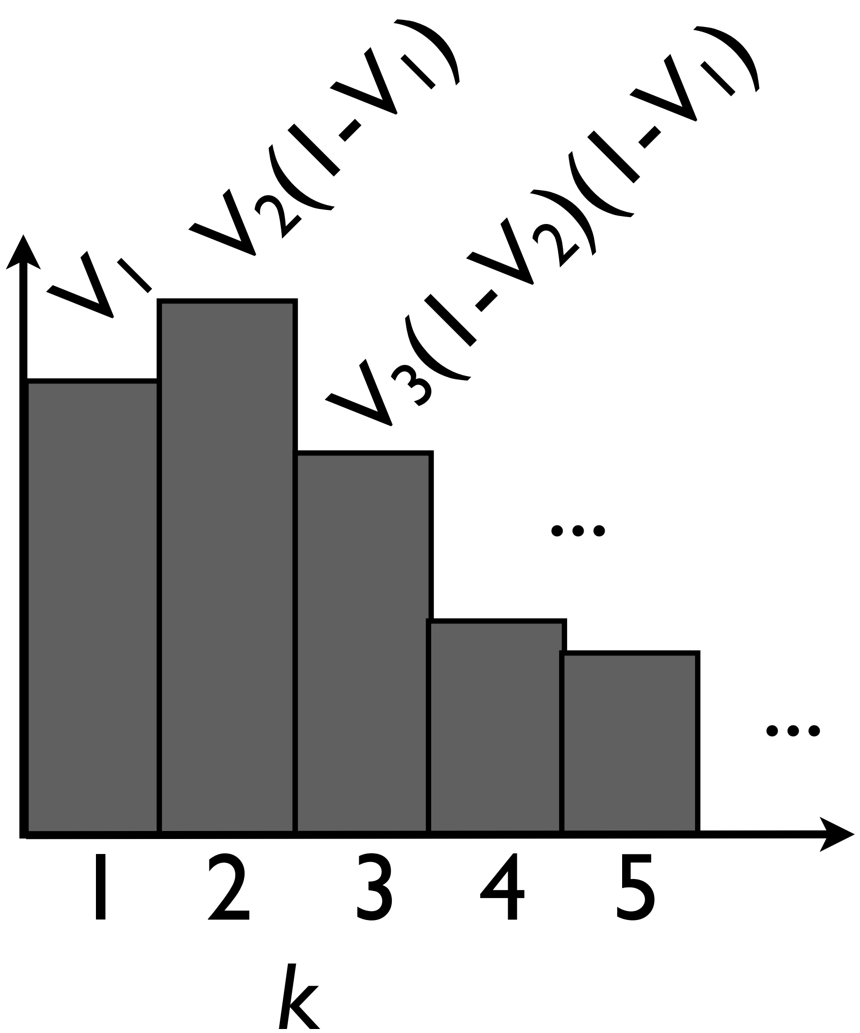
Stick-breaking

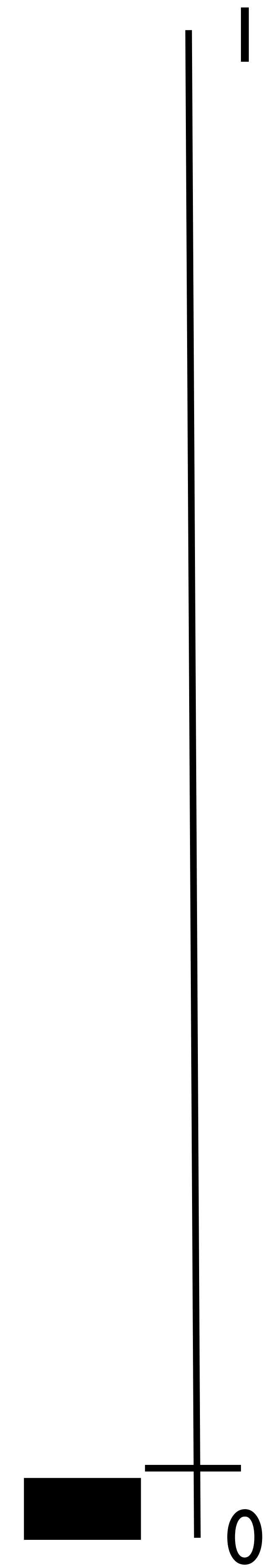


Stick-breaking

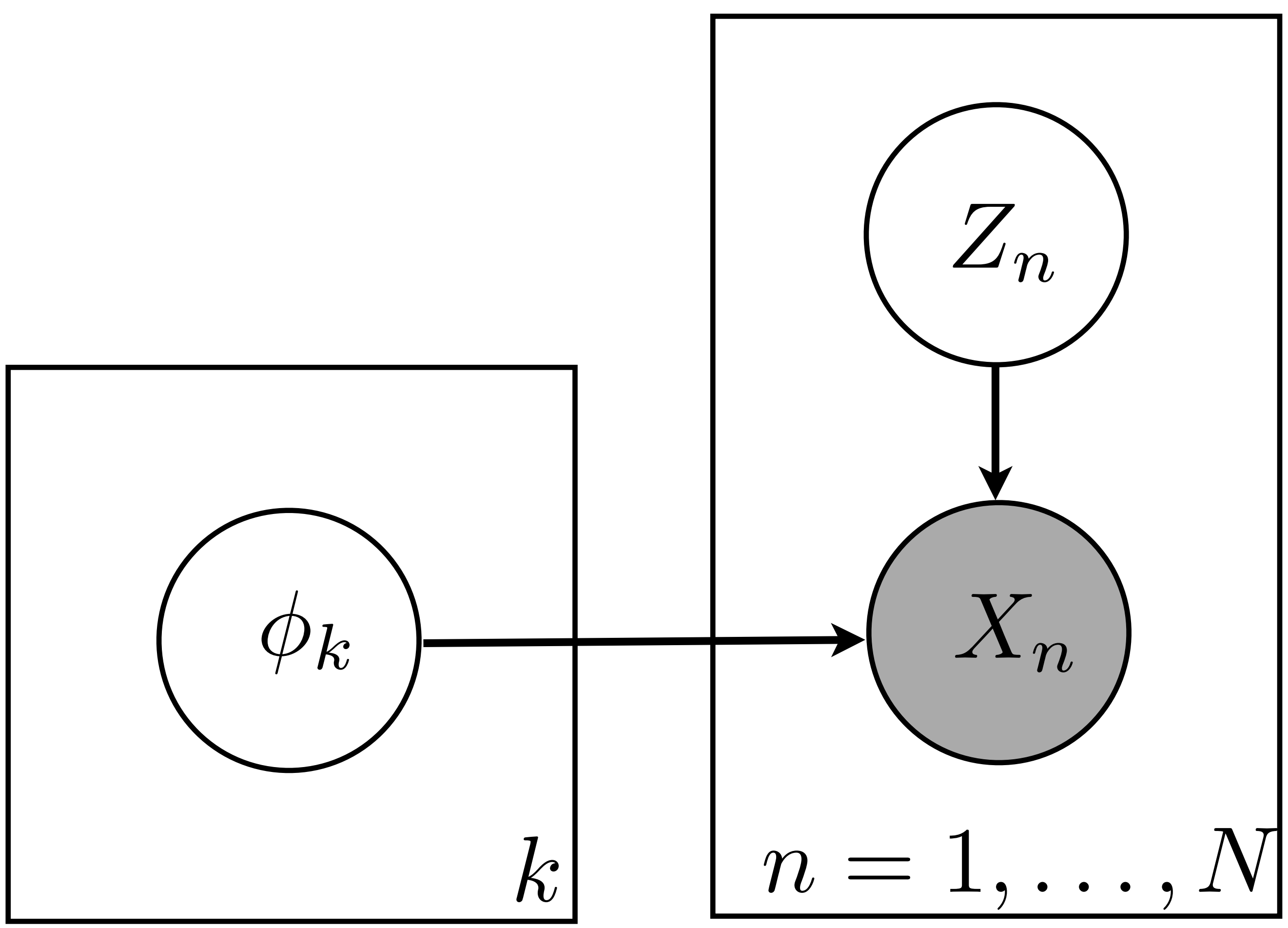


Stick-breaking





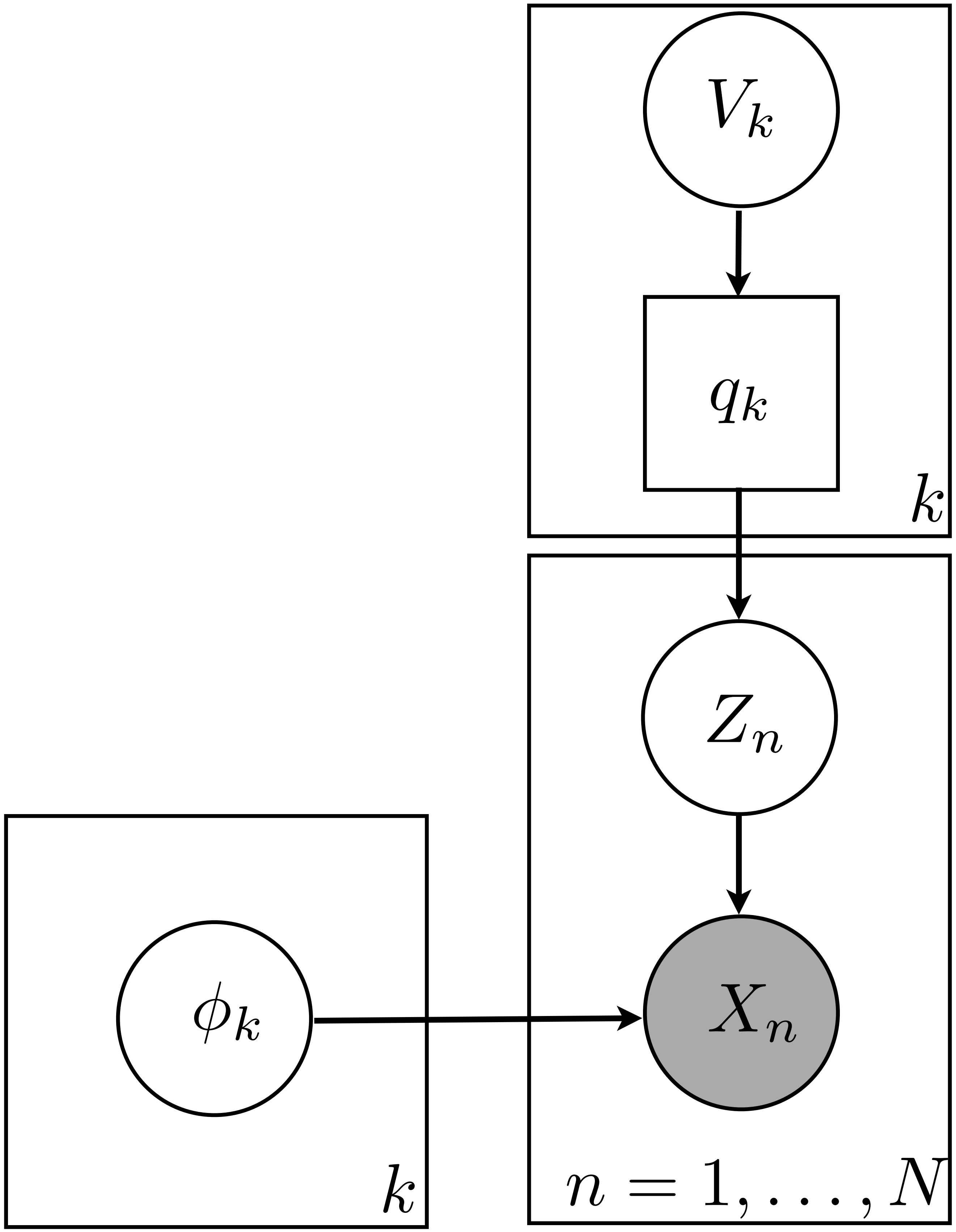
Stick-breaking: part of full gen model



$$\phi_k \overset{iid}{\sim} H$$

$$X_n \overset{indep}{\sim} F(\phi_{Z_n})$$

Stick-breaking: part of full gen model



$$V_k \overset{iid}{\sim} \text{Beta}(1, \theta)$$

$$q_k = V_k \prod_{j=1}^{k-1} (1 - V_j)$$

$$Z_n \overset{iid}{\sim} q$$

$$\phi_k \overset{iid}{\sim} H$$

$$X_n \overset{indep}{\sim} F(\phi_{Z_n})$$

Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model (Example: CRP stick-breaking)
  - ◇ Posterior
- Random probability measure

II. Features

Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model (Example: CRP stick-breaking)
  - ◇ Posterior
- Random probability measure

II. Features

Stick-breaking: calculating posterior

Why use stick-breaking?

- More general models
- May want to infer the stick lengths

Stick-breaking: calculating posterior

MCMC

Stick-breaking: calculating posterior

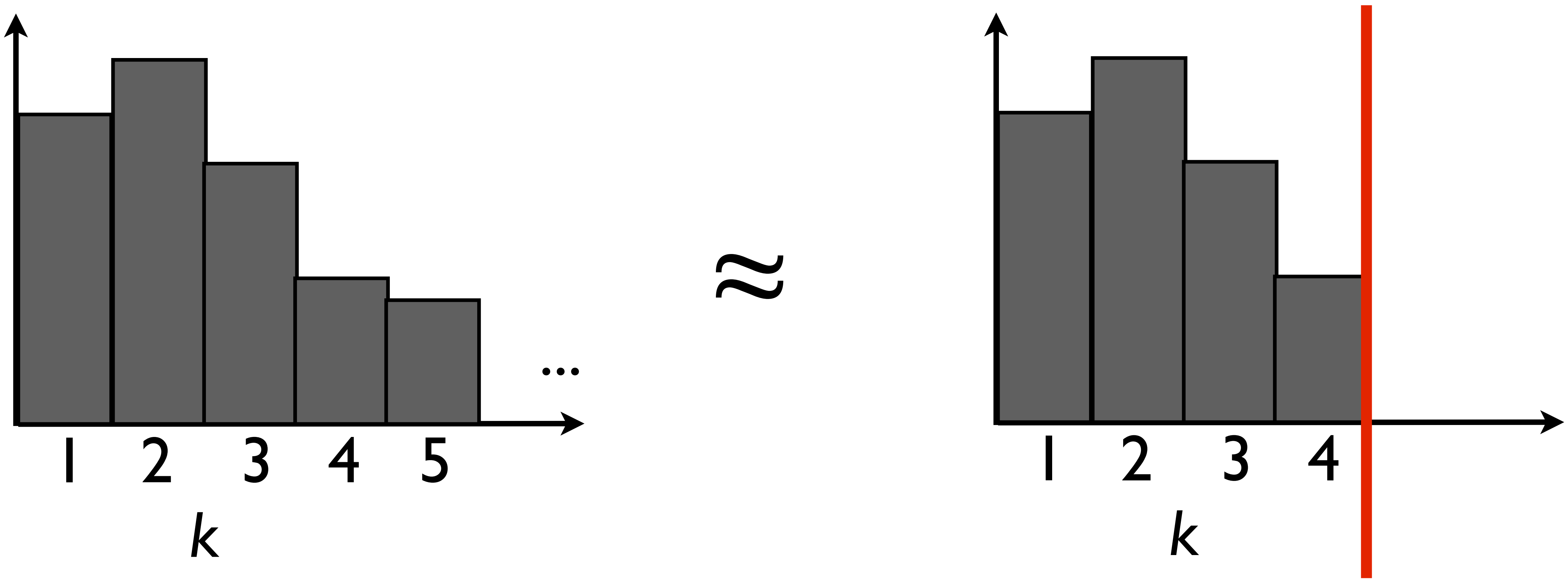
MCMC

- Finite approximation

Stick-breaking: calculating posterior

MCMC

- Finite approximation



[Ishwaran, Zarepour 2000]

Stick-breaking: calculating posterior

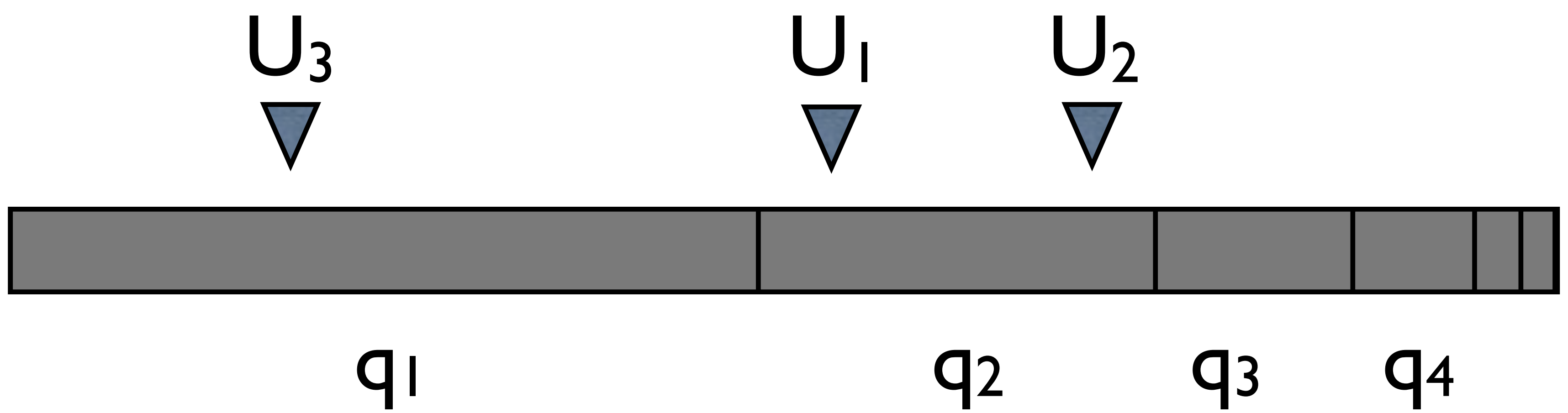
MCMC

- Finite approximation
- Retrospective sampling

Stick-breaking: calculating posterior

MCMC

- Finite approximation
- Retrospective sampling



[Papaspiliopoulos, Roberts 2008]

Stick-breaking: calculating posterior

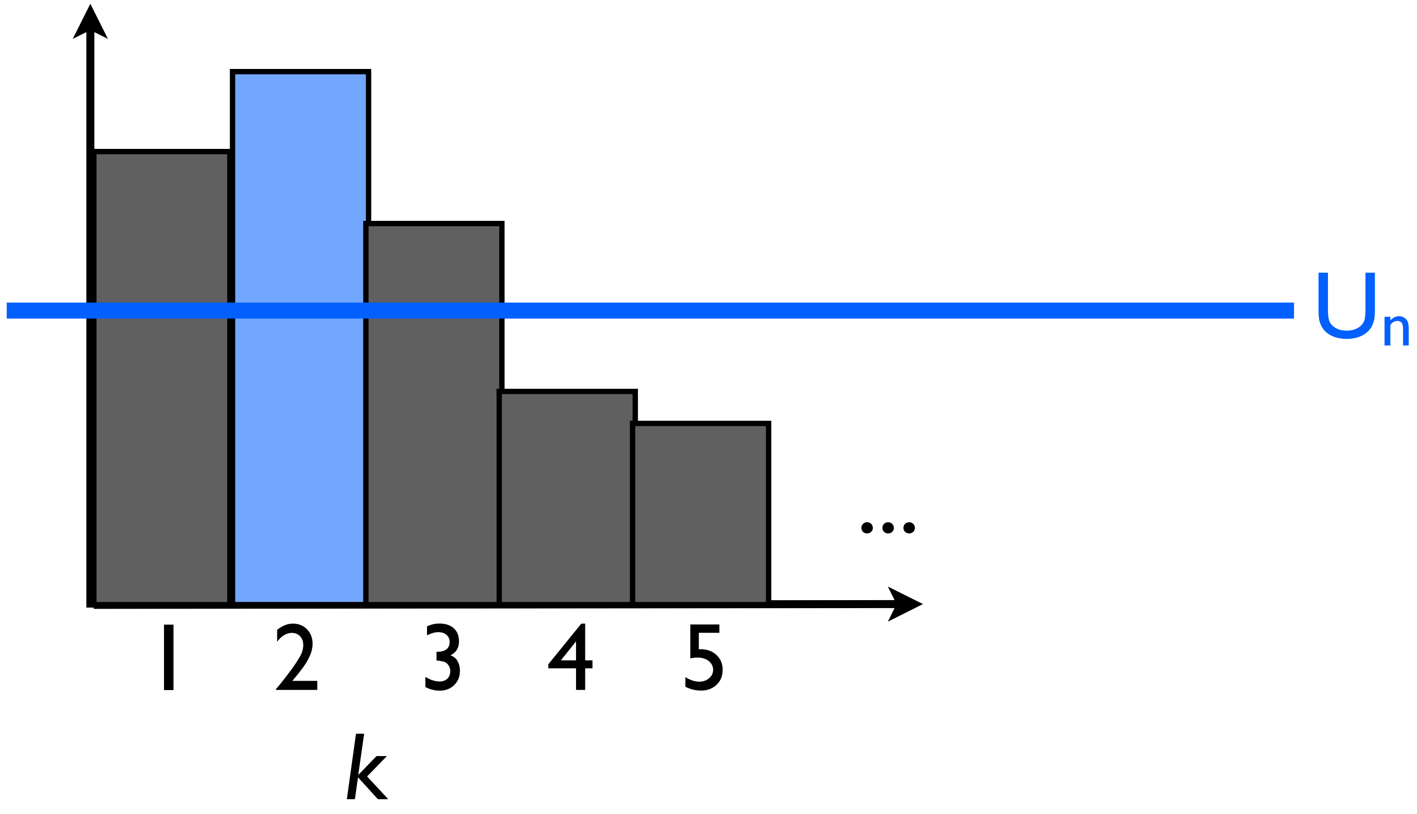
MCMC

- Finite approximation
- Retrospective sampling
- Slice sampling

Stick-breaking: calculating posterior

MCMC

- Finite approximation
- Retrospective sampling
- Slice sampling



[Walker 2007]

Stick-breaking: calculating posterior

MCMC

- Finite approximation
- Retrospective sampling
- Slice sampling

Variational methods

- Mean field

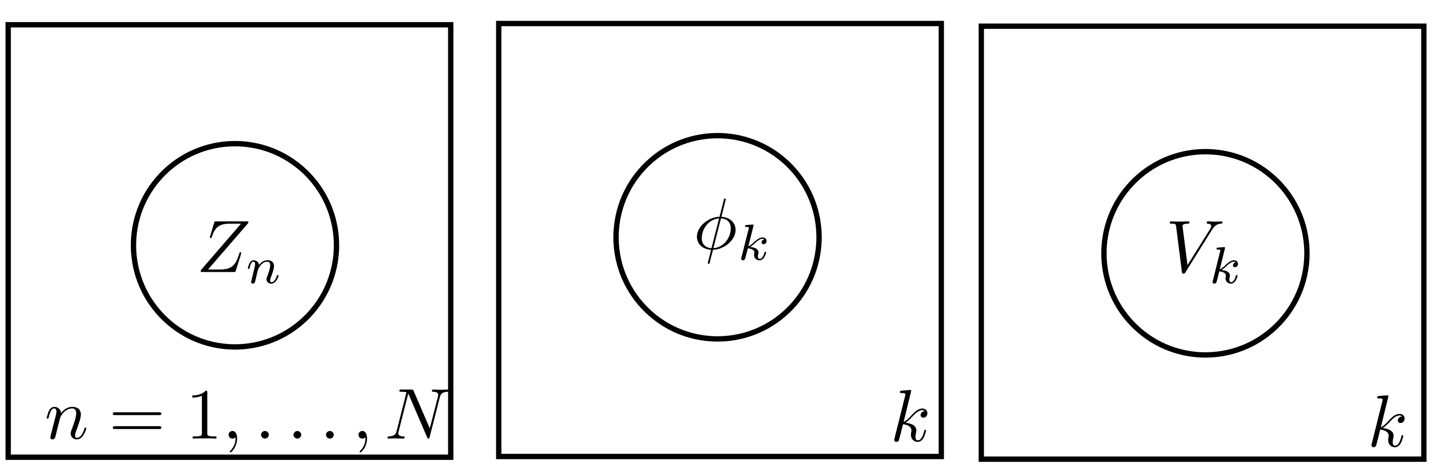
Stick-breaking: calculating posterior

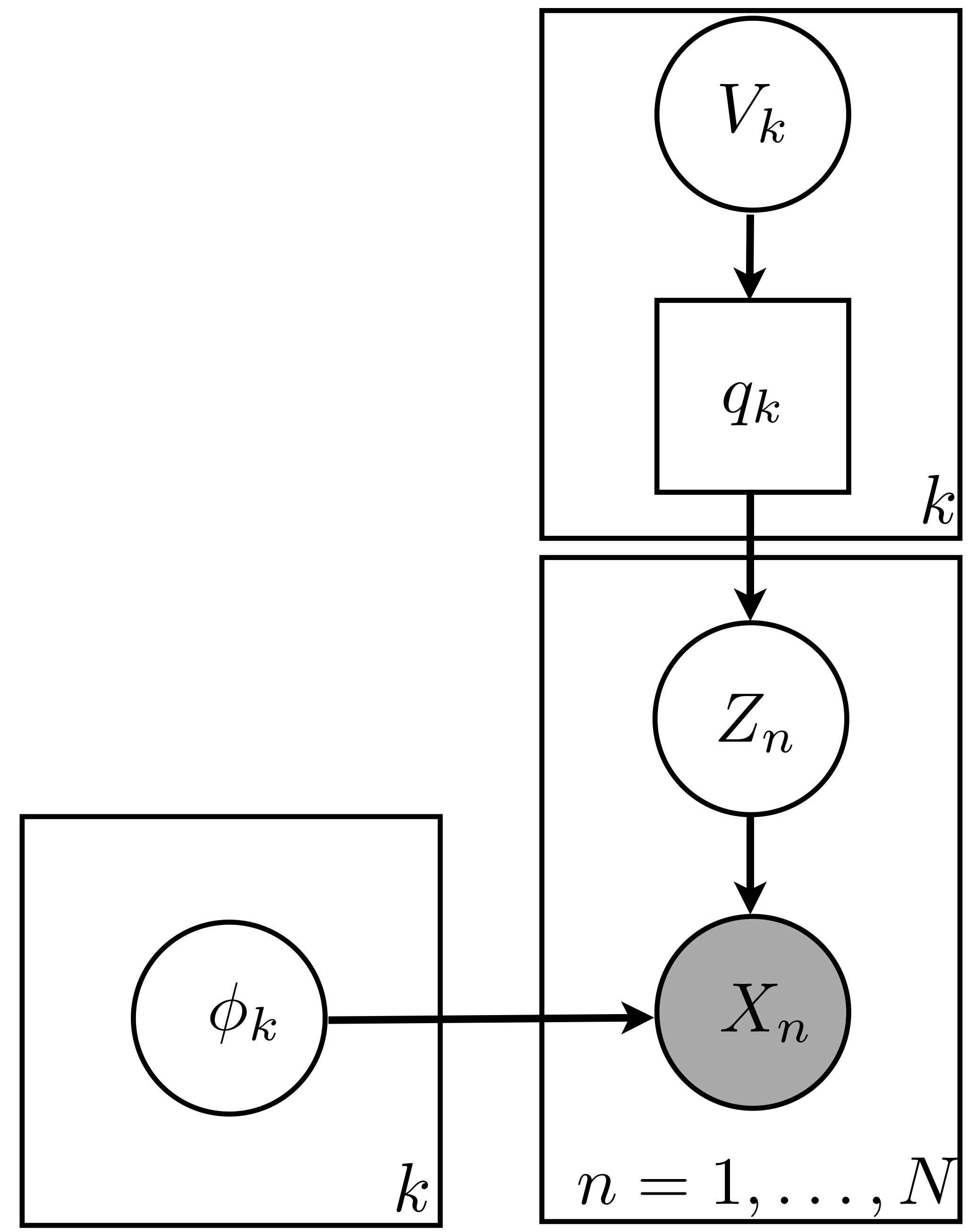
MCMC

- Finite approximation
- Retrospective sampling
- Slice sampling

Variational methods

- Mean field





[Blei, Jordan 2004]

Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model (Example: CRP stick-breaking)
  - ◇ Posterior
- Random probability measure

II. Features

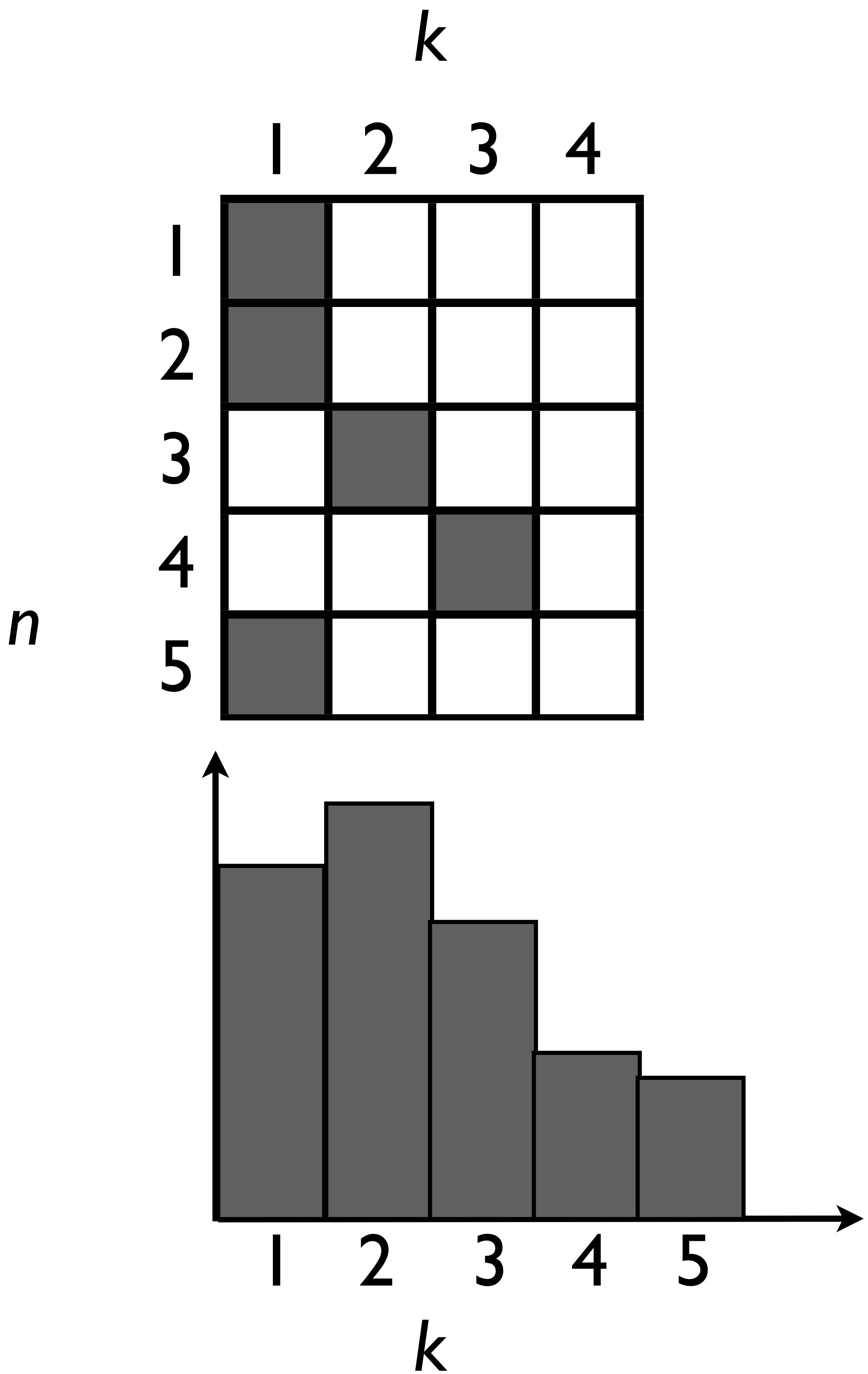
Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model
  - ◇ Posterior
- Random probability measure

II. Features

Stick-breaking: extensions



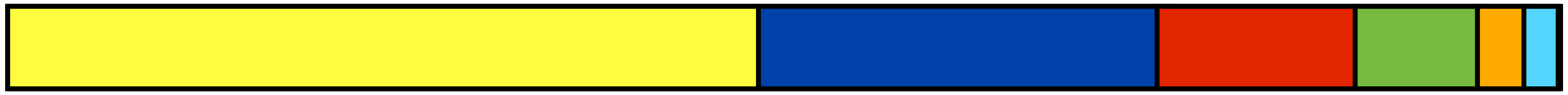
Connections

Exchangeable clustering	Chinese restaurant
EPPF	CRP
?	CRP stick-breaking

Kingman paintbox

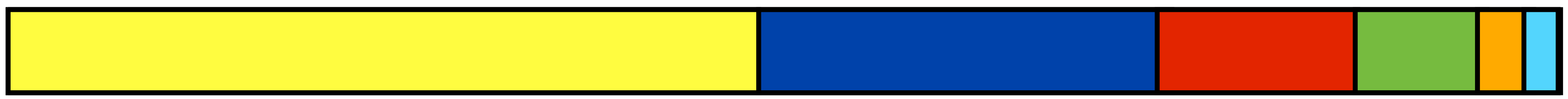
[Kingman 1978]

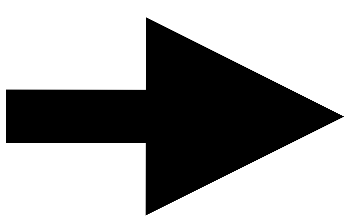
Kingman paintbox

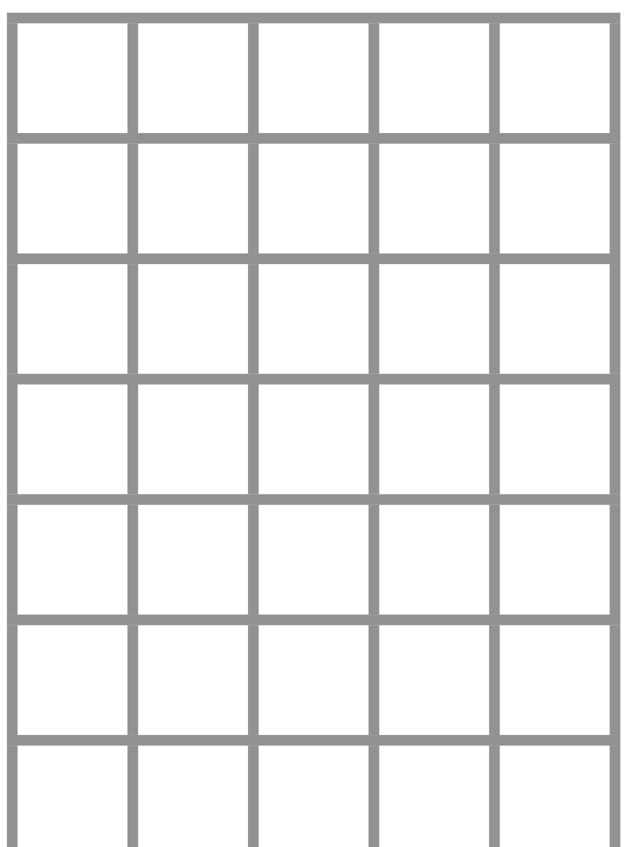


[Kingman 1978]

Kingman paintbox

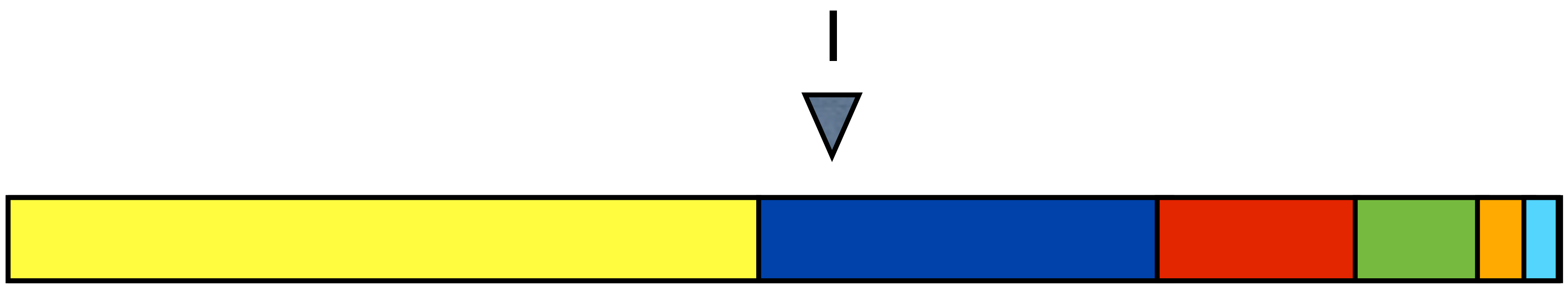


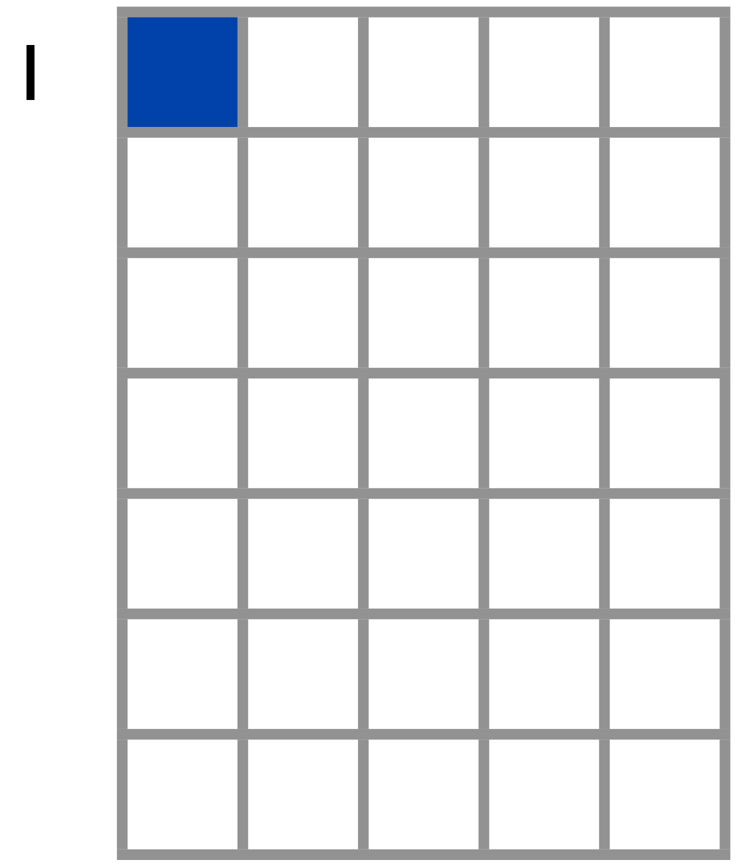




[Kingman 1978]

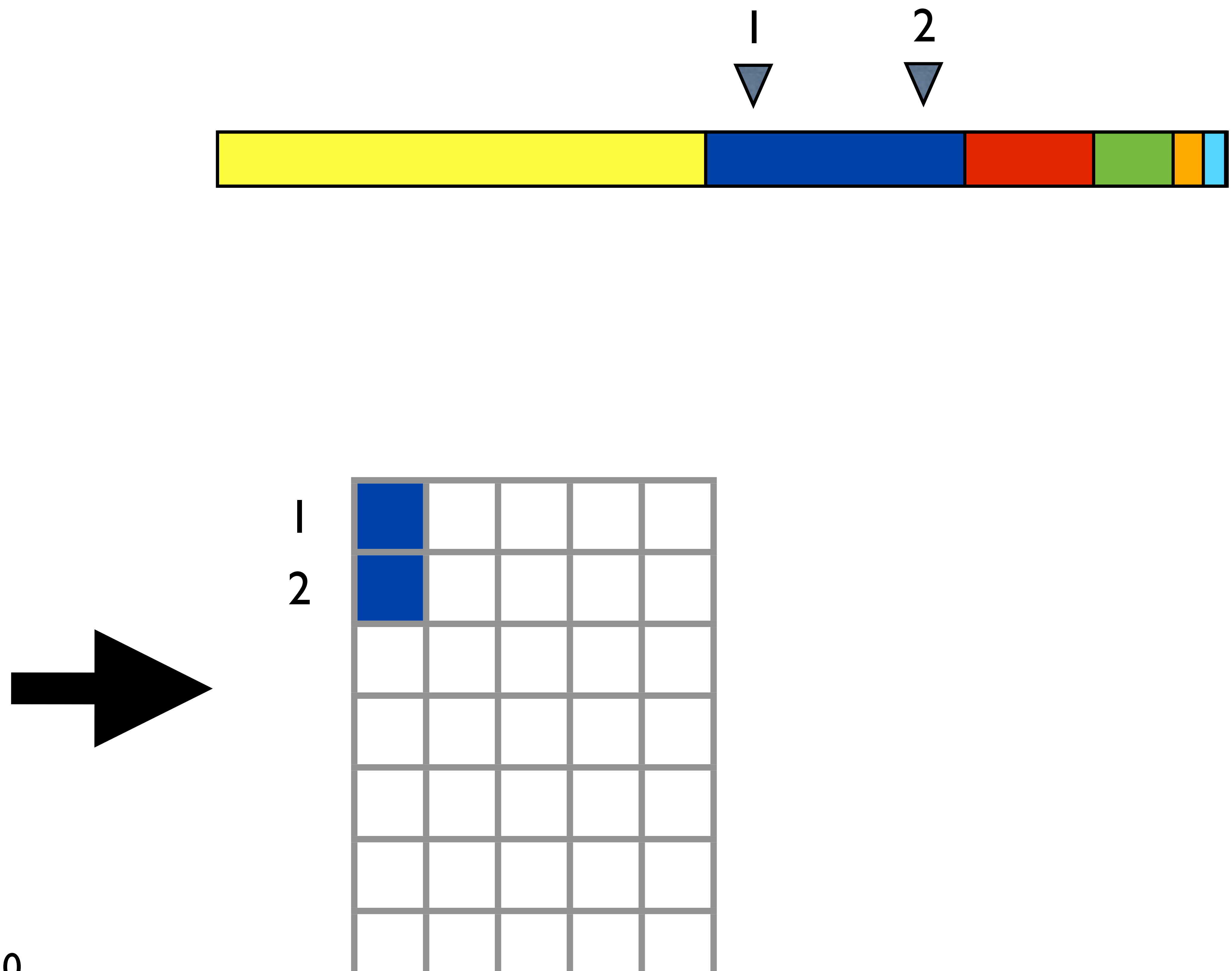
Kingman paintbox





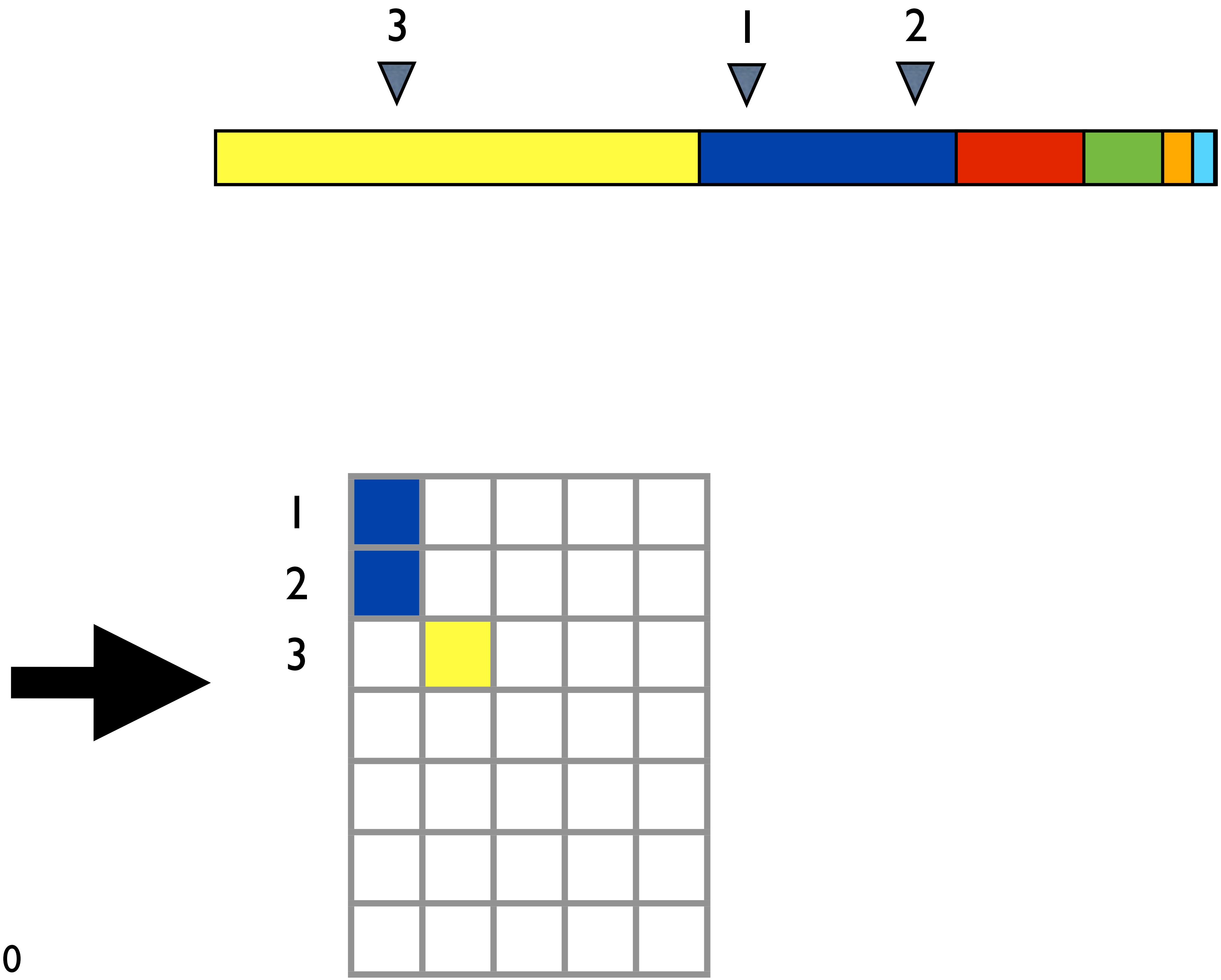
[Kingman 1978]

Kingman paintbox



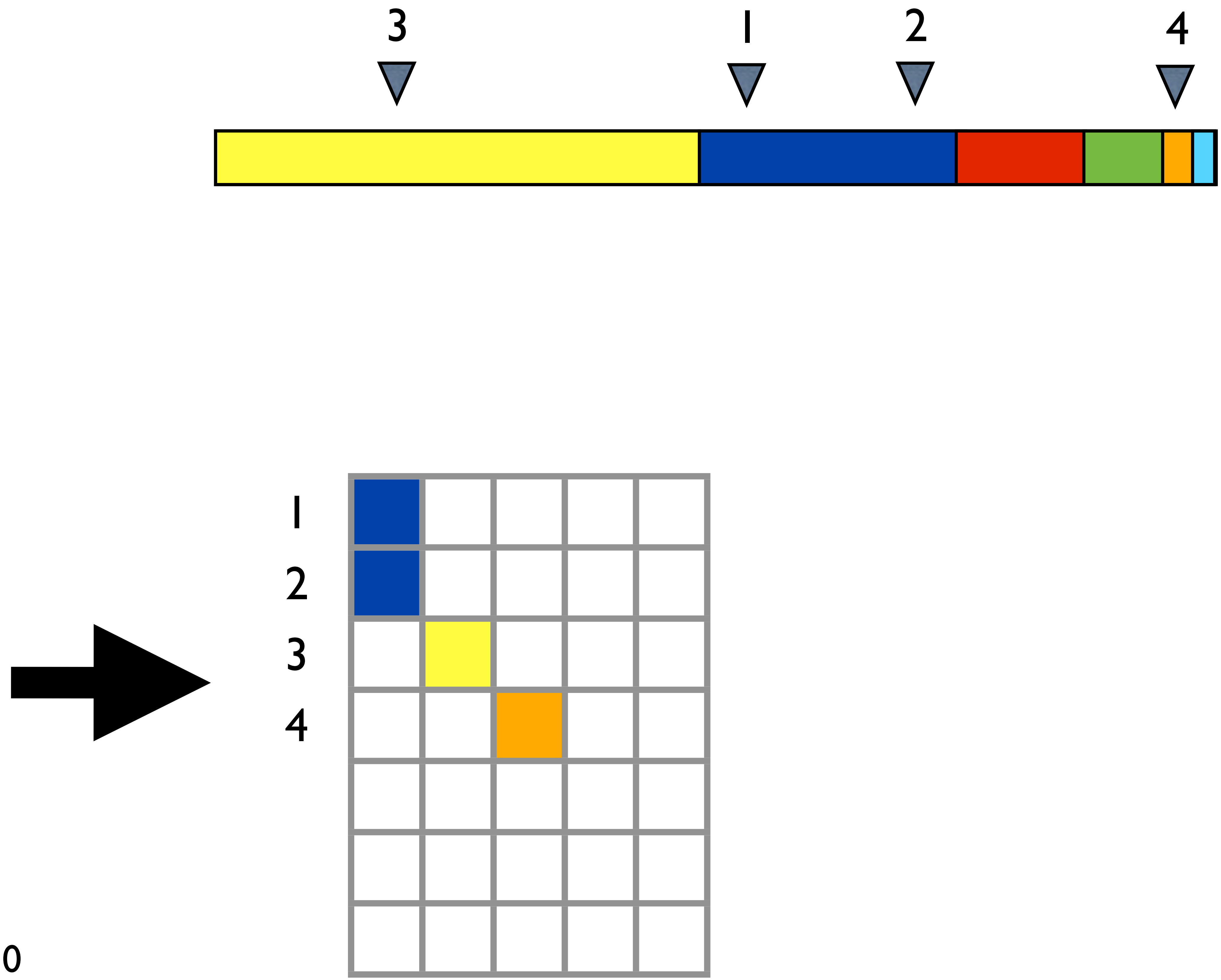
[Kingman 1978]

Kingman paintbox



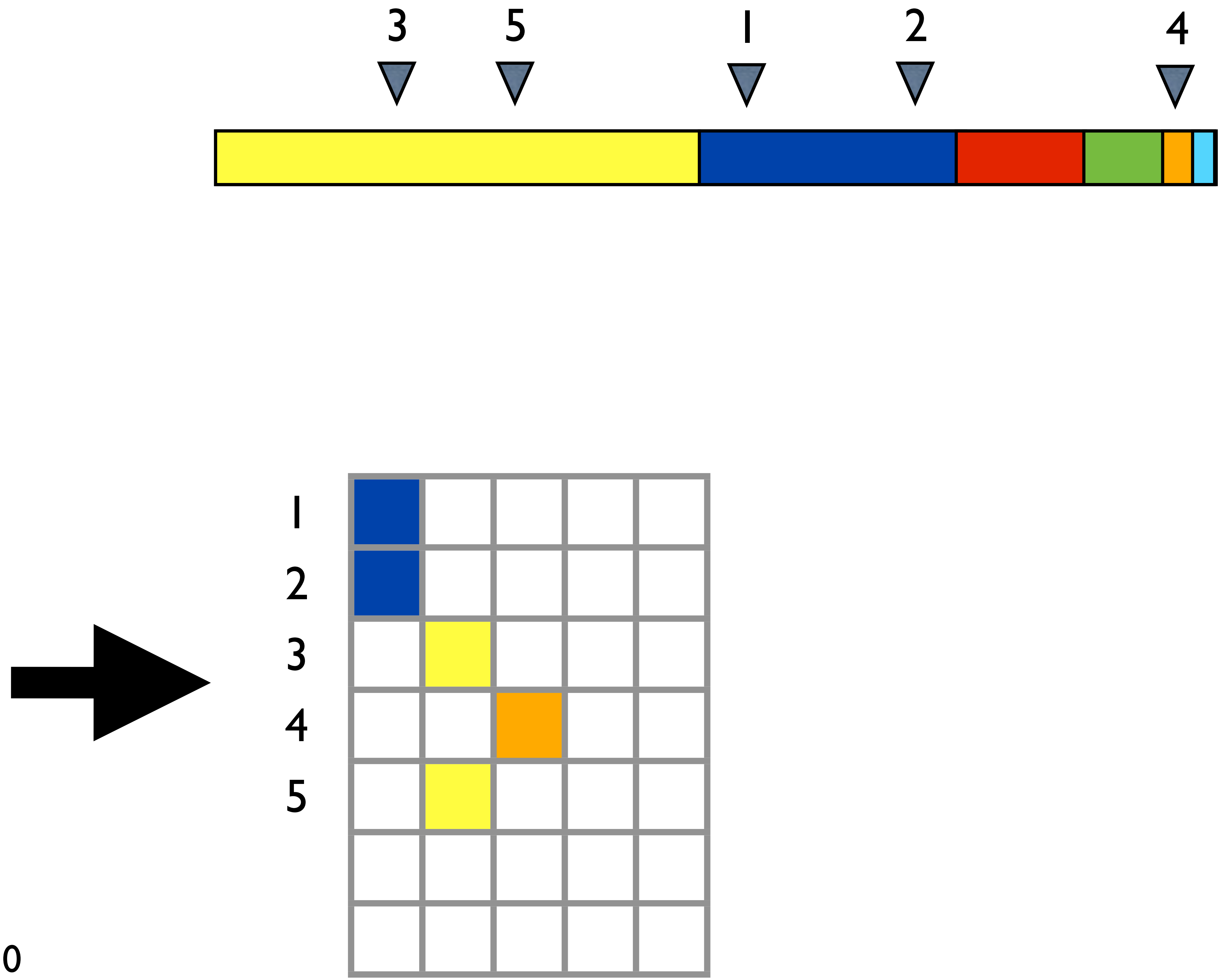
[Kingman 1978]

Kingman paintbox



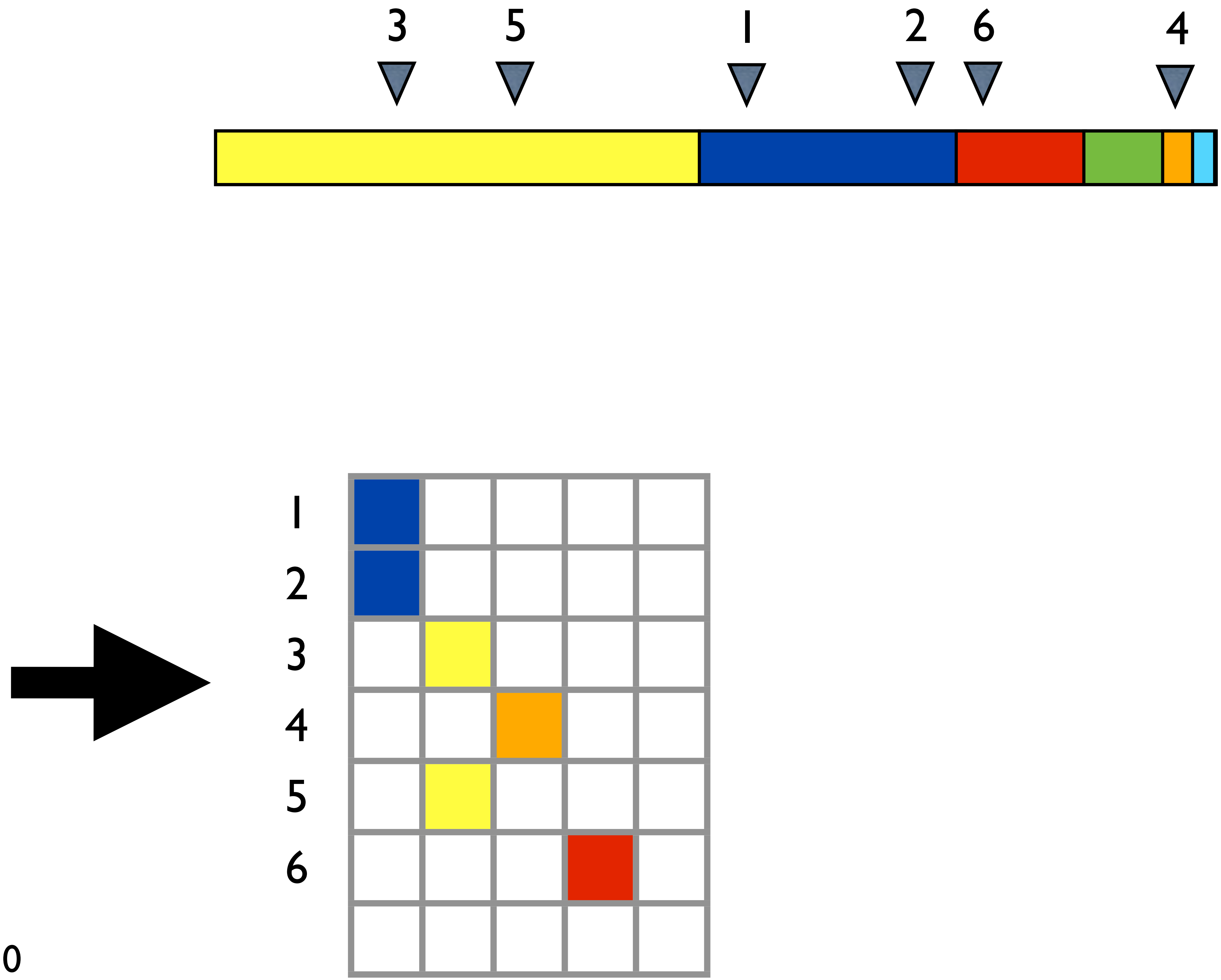
[Kingman 1978]

Kingman paintbox



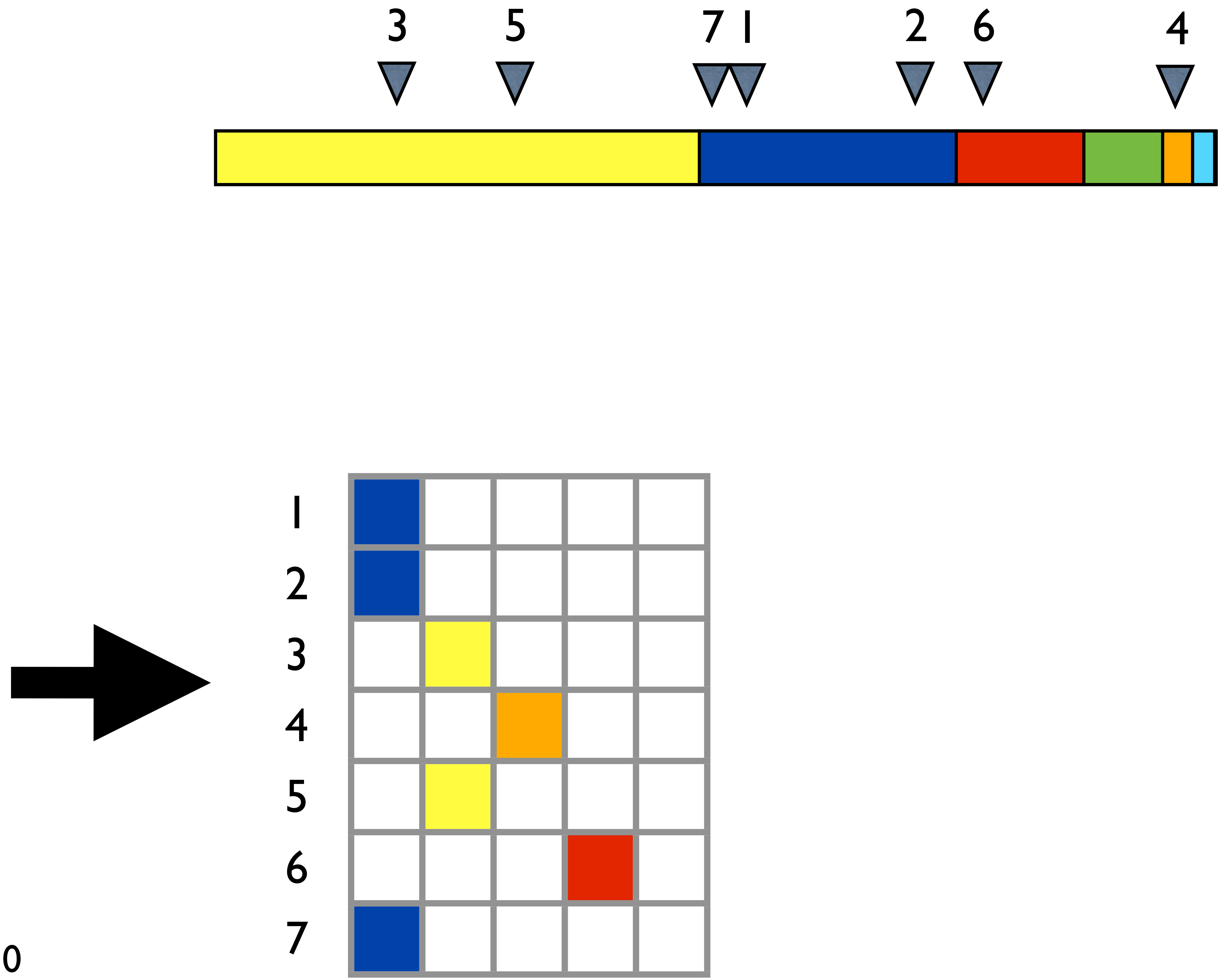
[Kingman 1978]

Kingman paintbox



[Kingman 1978]

Kingman paintbox



[Kingman 1978]

Connections

Exchangeable clustering	Chinese restaurant
EPPF	CRP
Kingman paintbox	CRP stick-breaking

Outline

I. Clusters

- Overview
- Distribution
- Proportions
  - ◇ Generative model
  - ◇ Posterior
- Random probability measure

II. Features

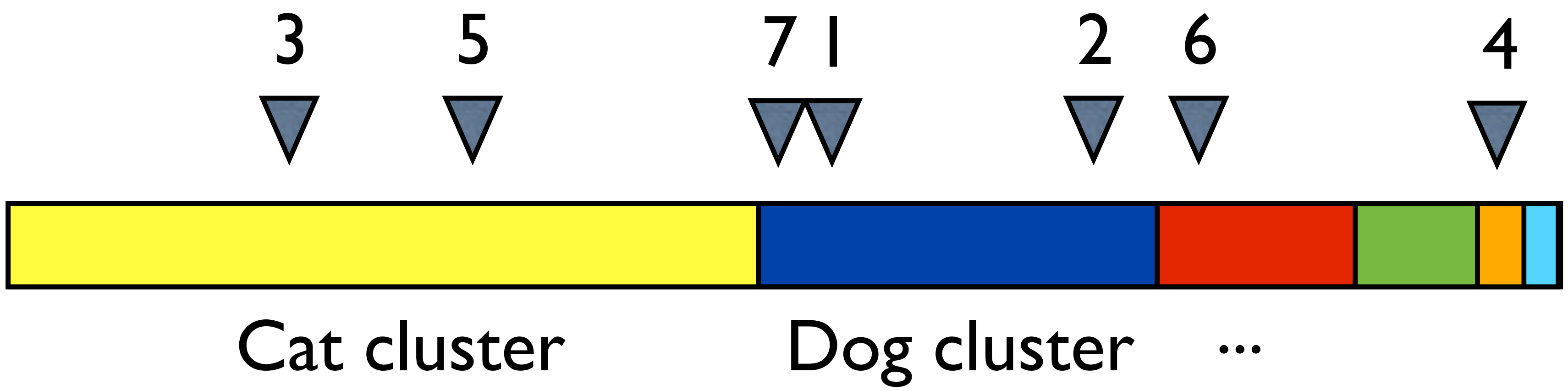
Outline

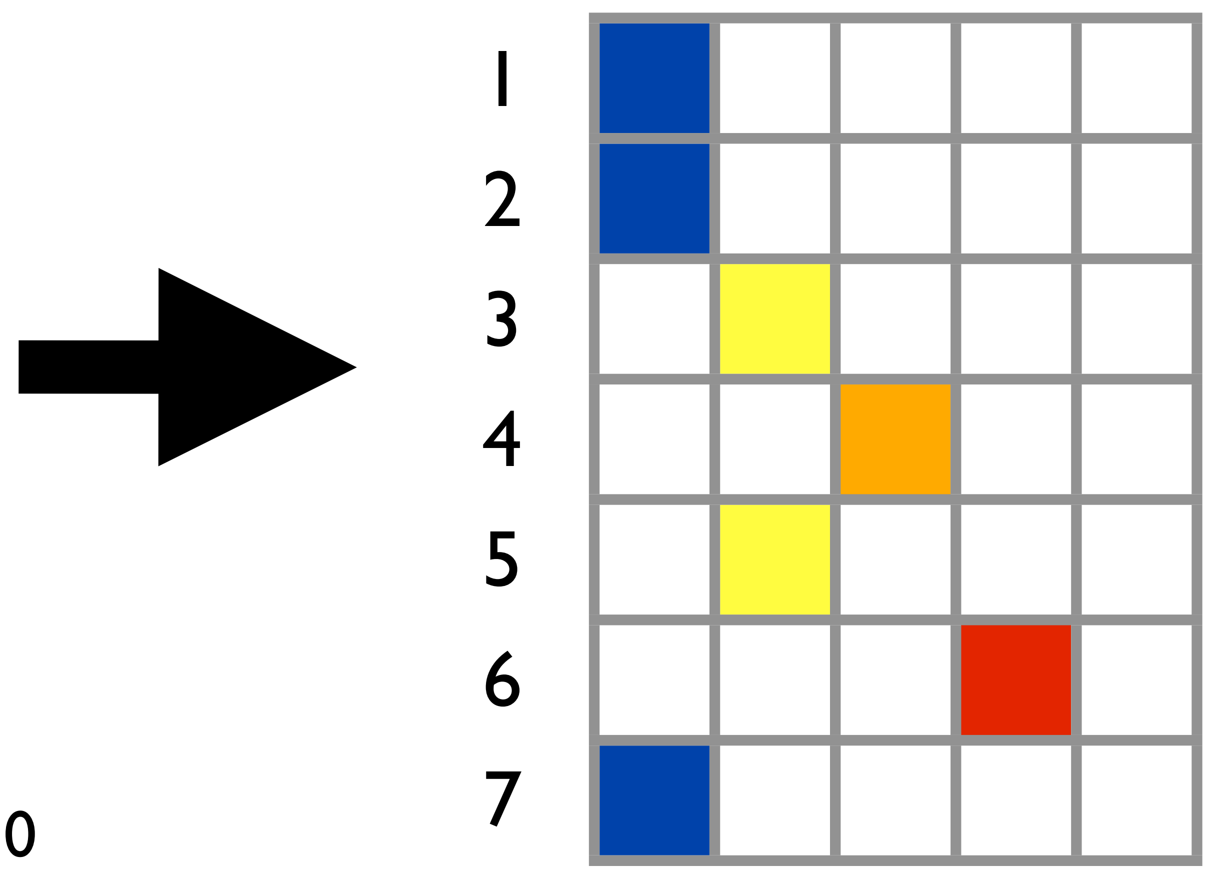
I. Clusters

- Overview
- Distribution
- Proportions
- Random probability measure

II. Features

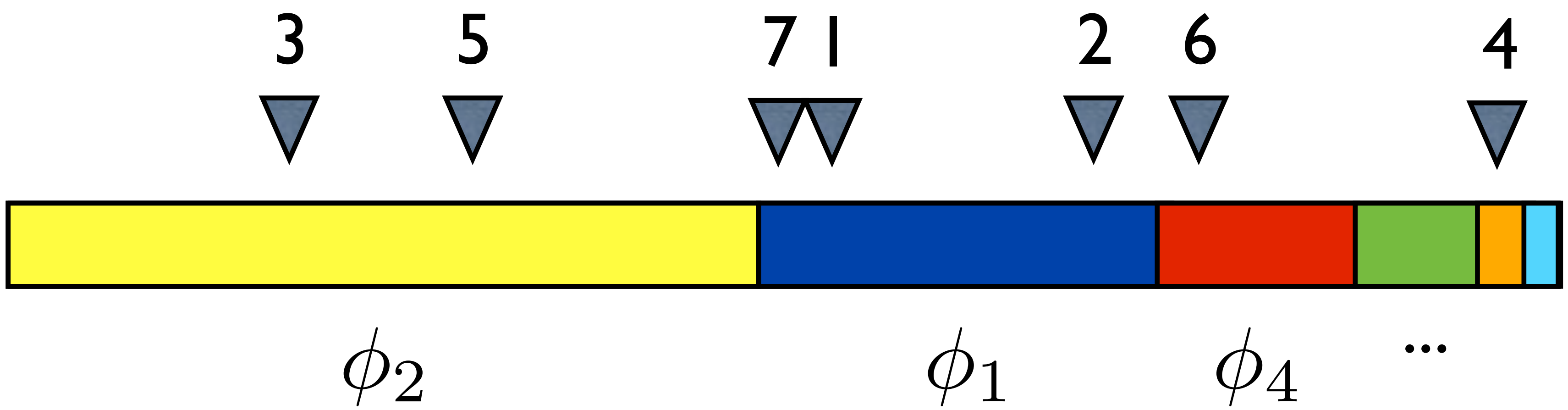
Kingman paintbox

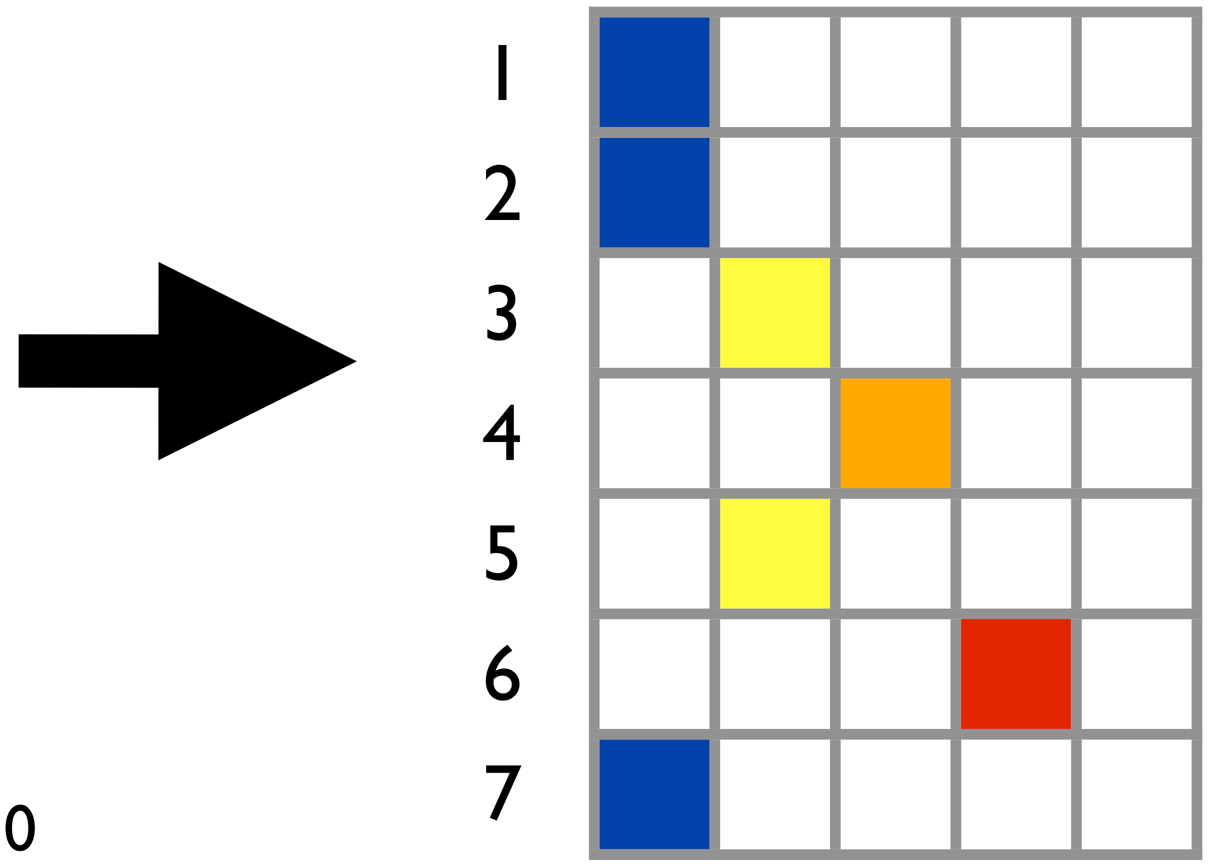




[Kingman 1978]

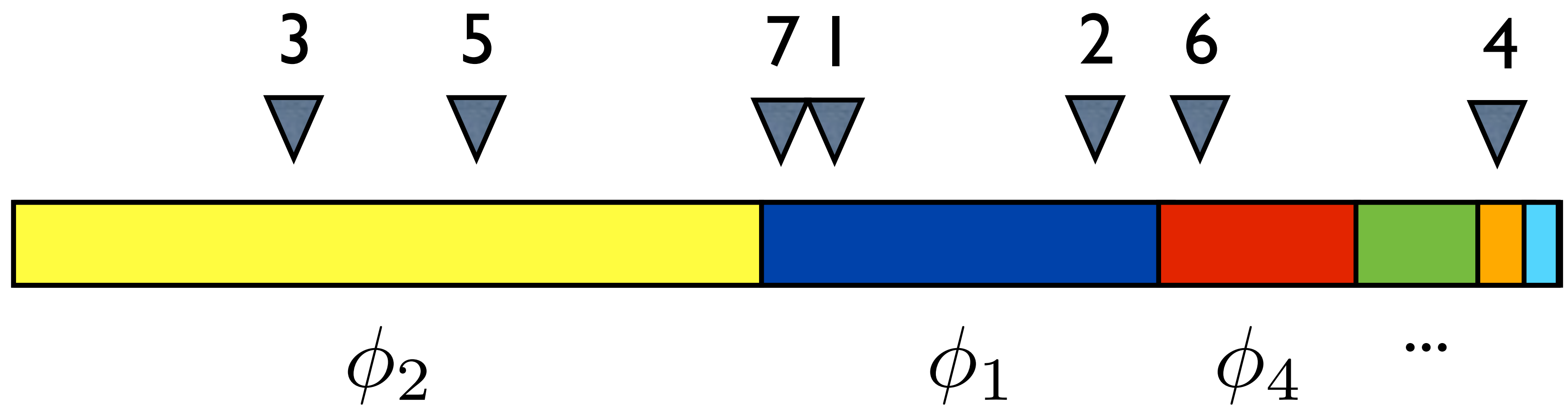
Kingman paintbox

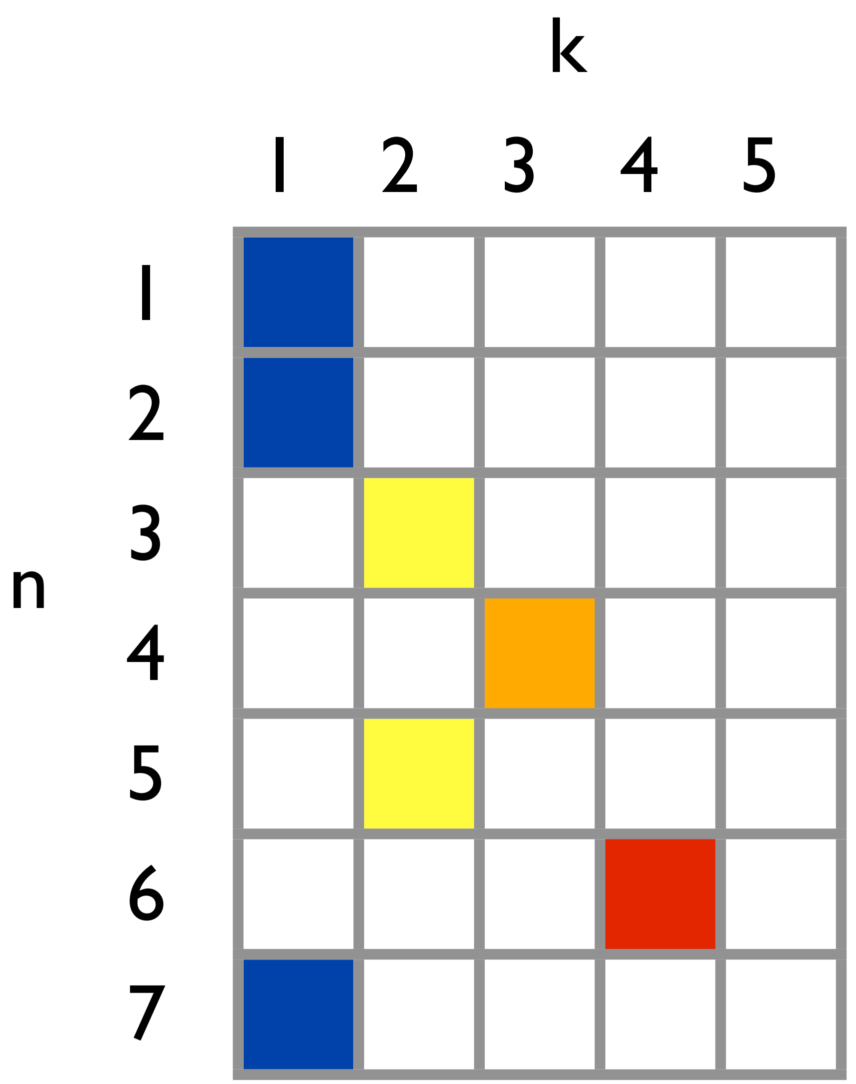




[Kingman 1978]

Cluster labels





Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

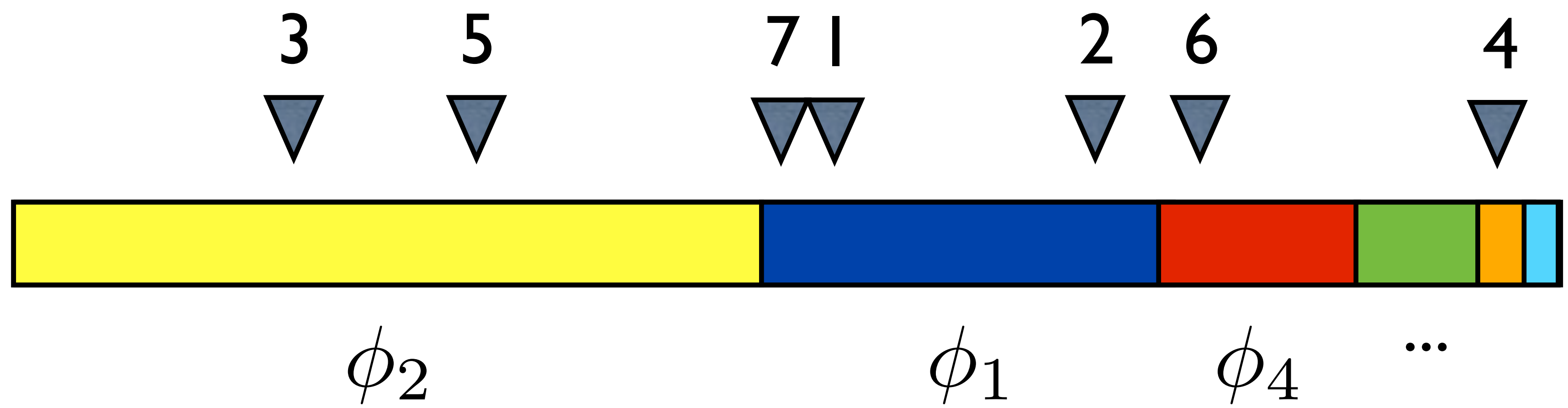
$Z_4 = 3$

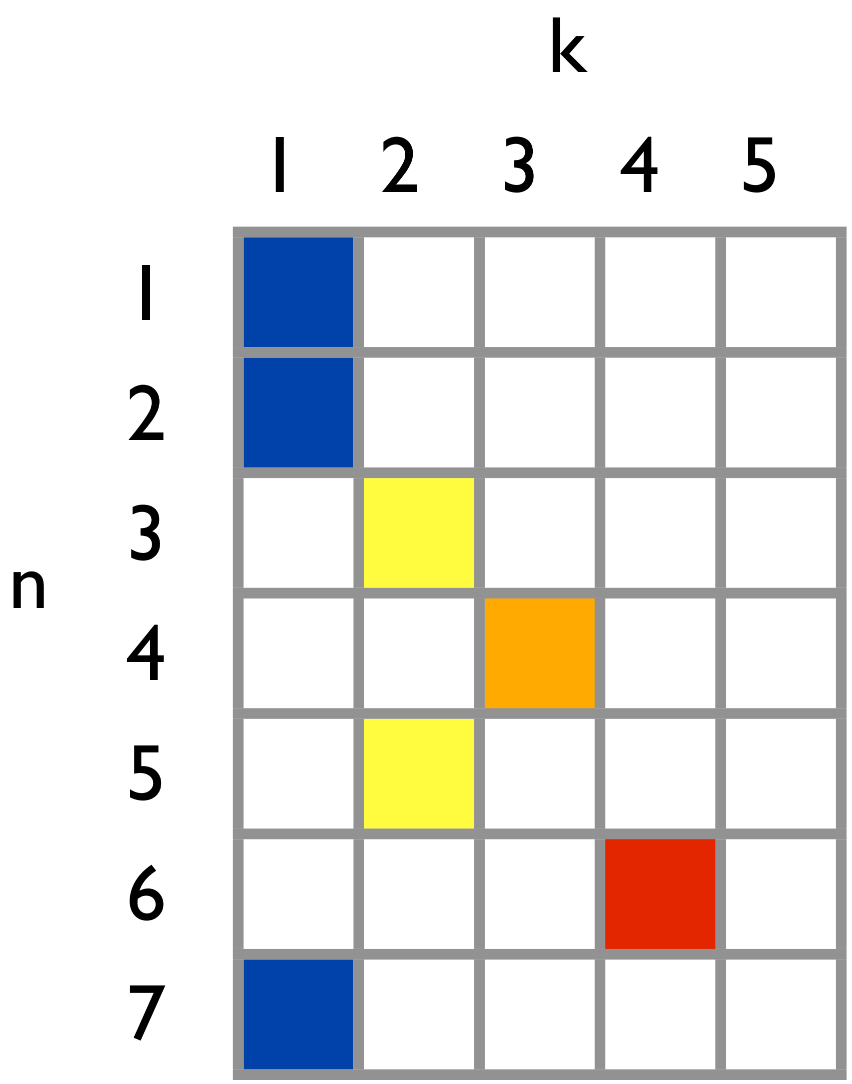
$Z_5 = 2$

$Z_6 = 4$

$Z_7 = 1$

Cluster labels





Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

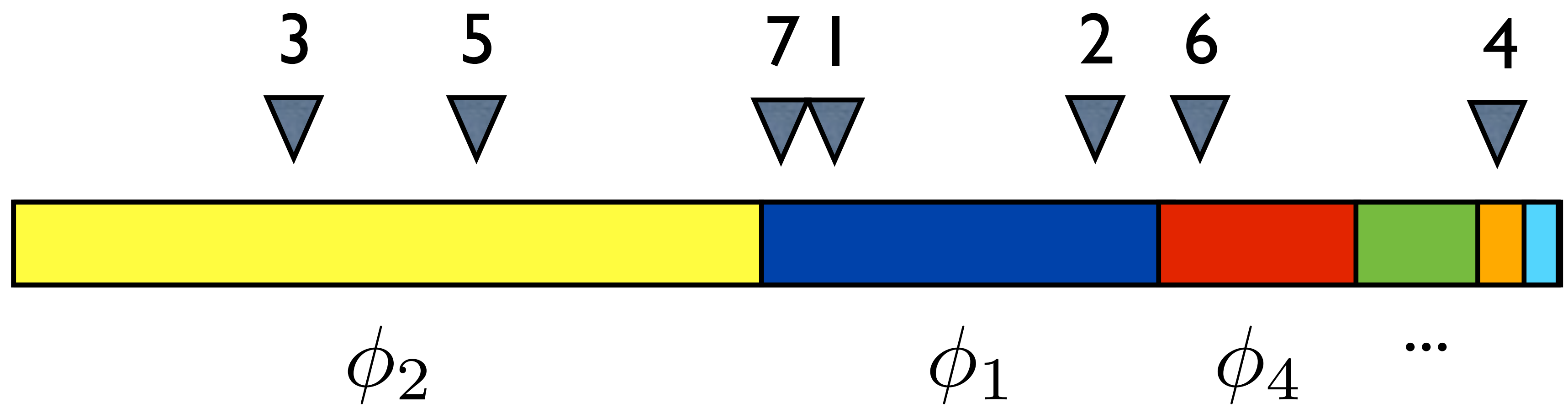
$Z_5 = 2$

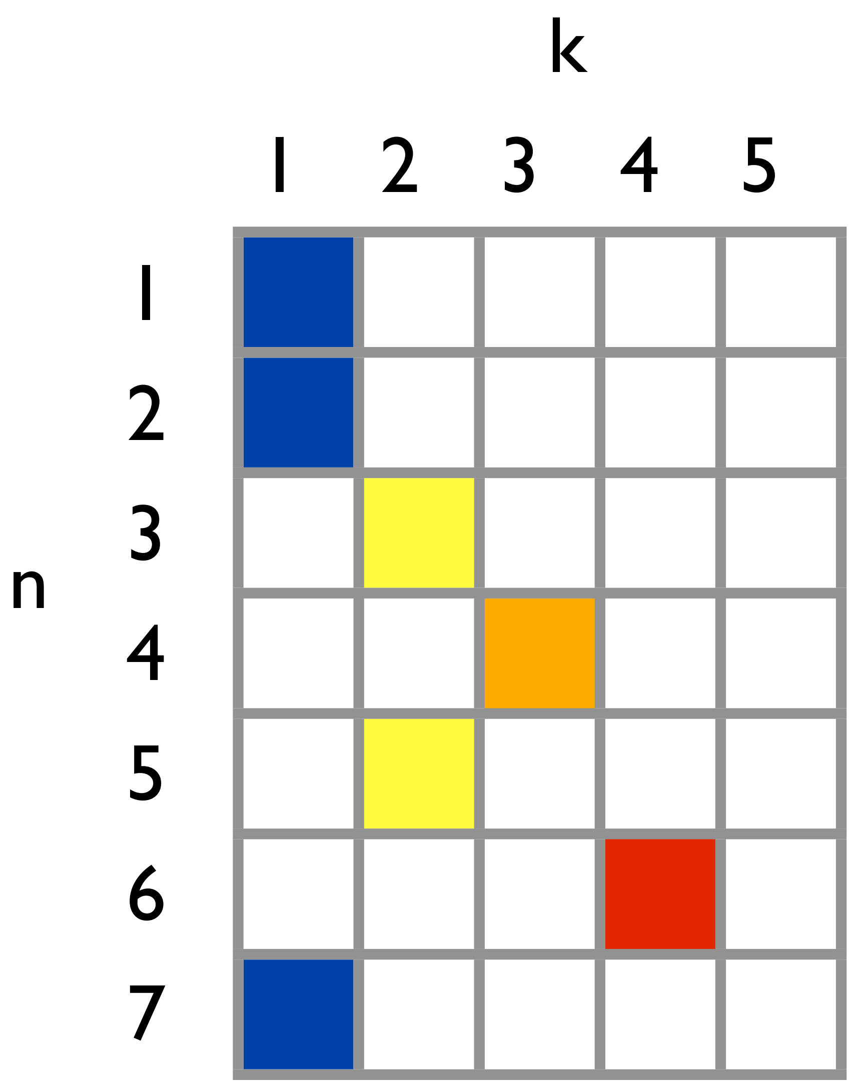
$Z_6 = 4$

$Z_7 = 1$

Random labels

Cluster labels





Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$

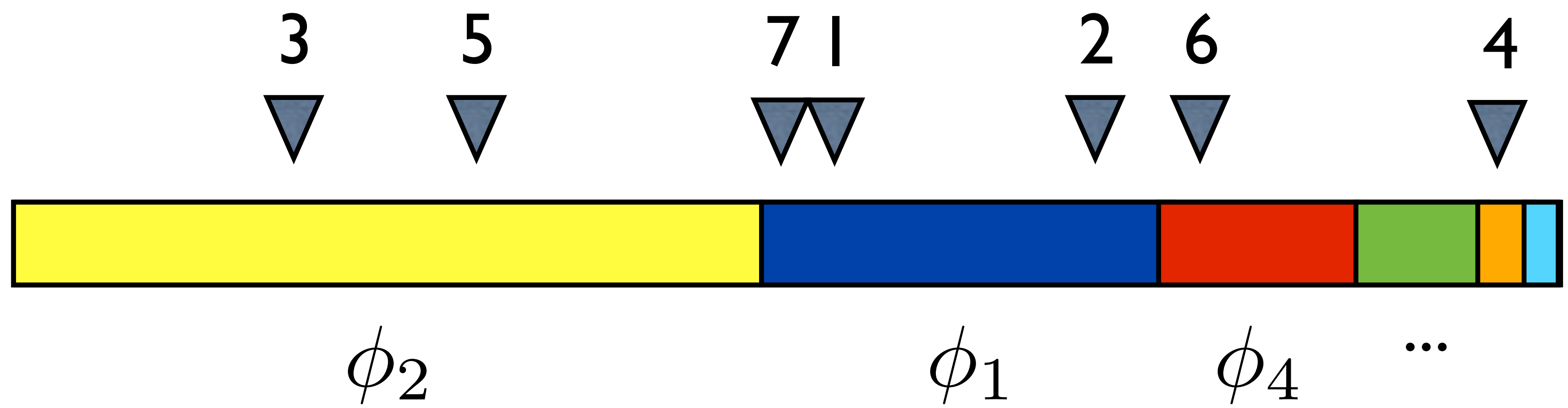
$Z_6 = 4$

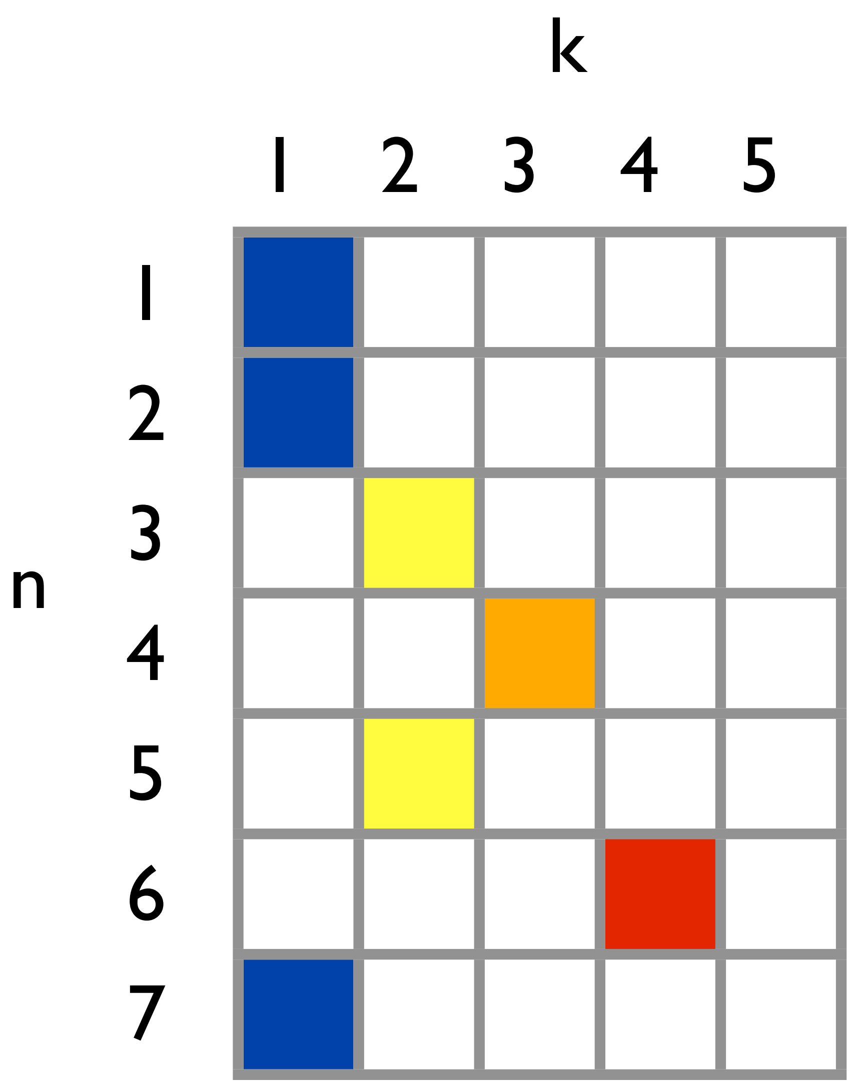
$Z_7 = 1$

Random labels

$Z'_1 = \phi_1$

Cluster labels





Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$

$Z_6 = 4$

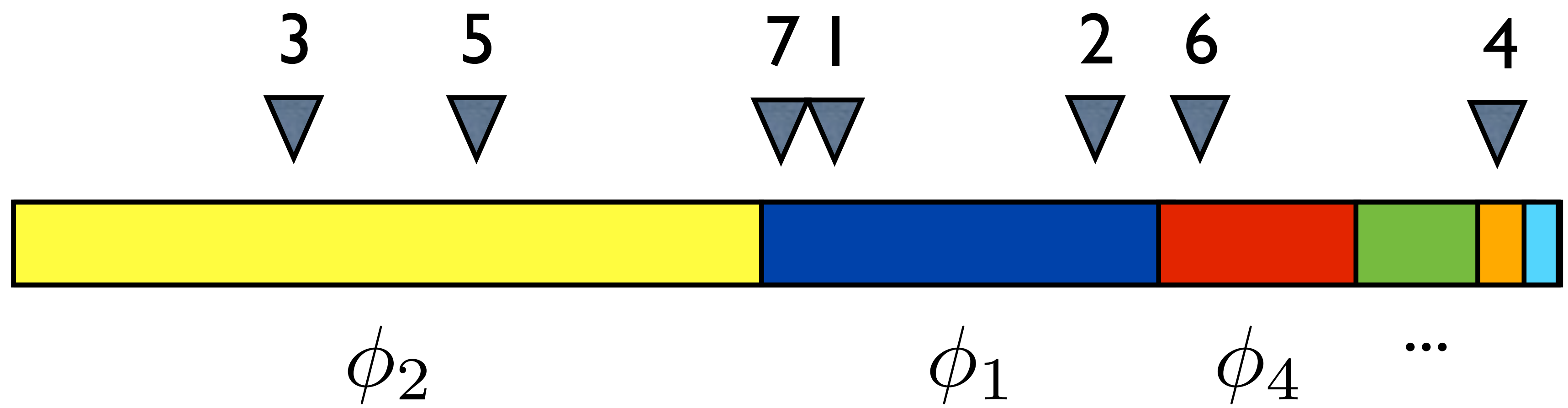
$Z_7 = 1$

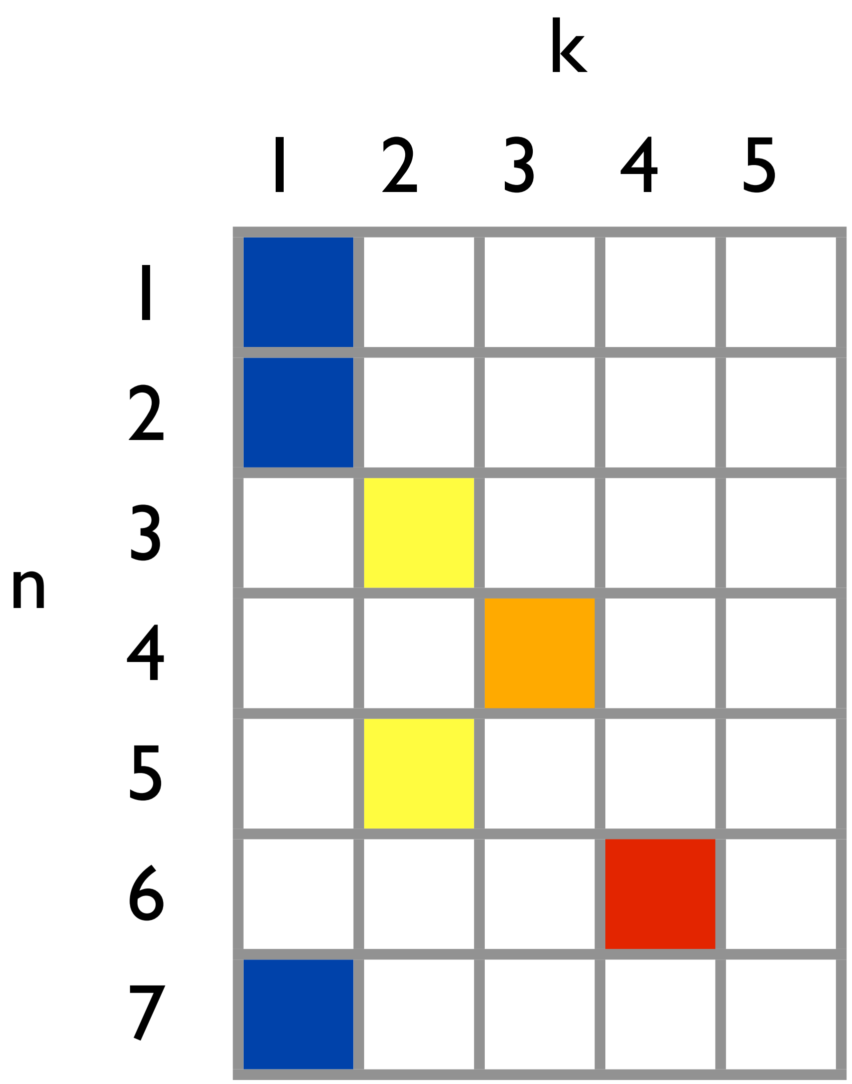
Random labels

$Z'_1 = \phi_1$

$Z'_2 = \phi_1$

Cluster labels





Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$

$Z_6 = 4$

$Z_7 = 1$

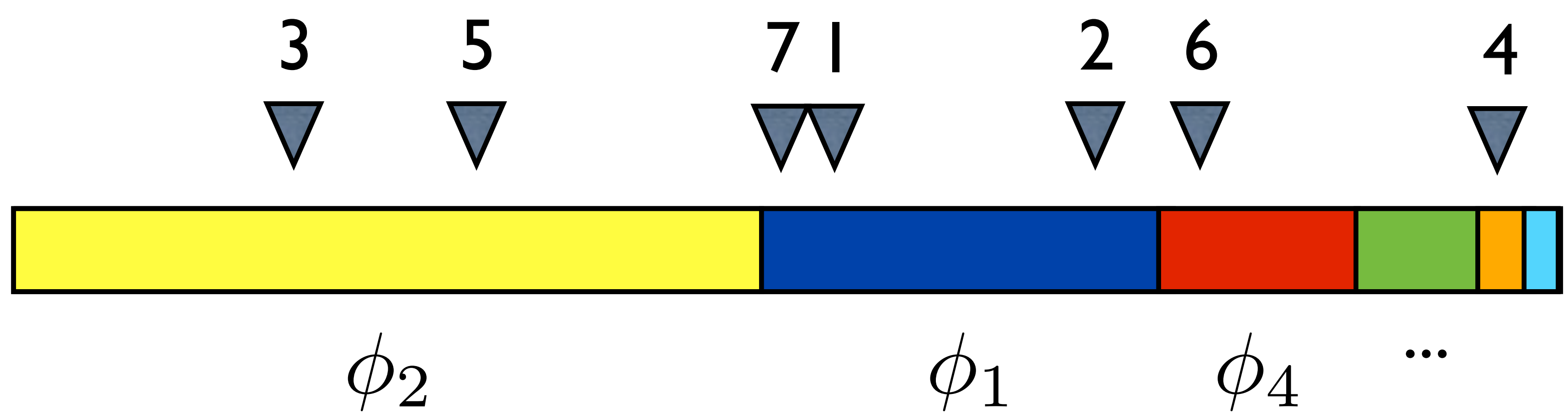
Random labels

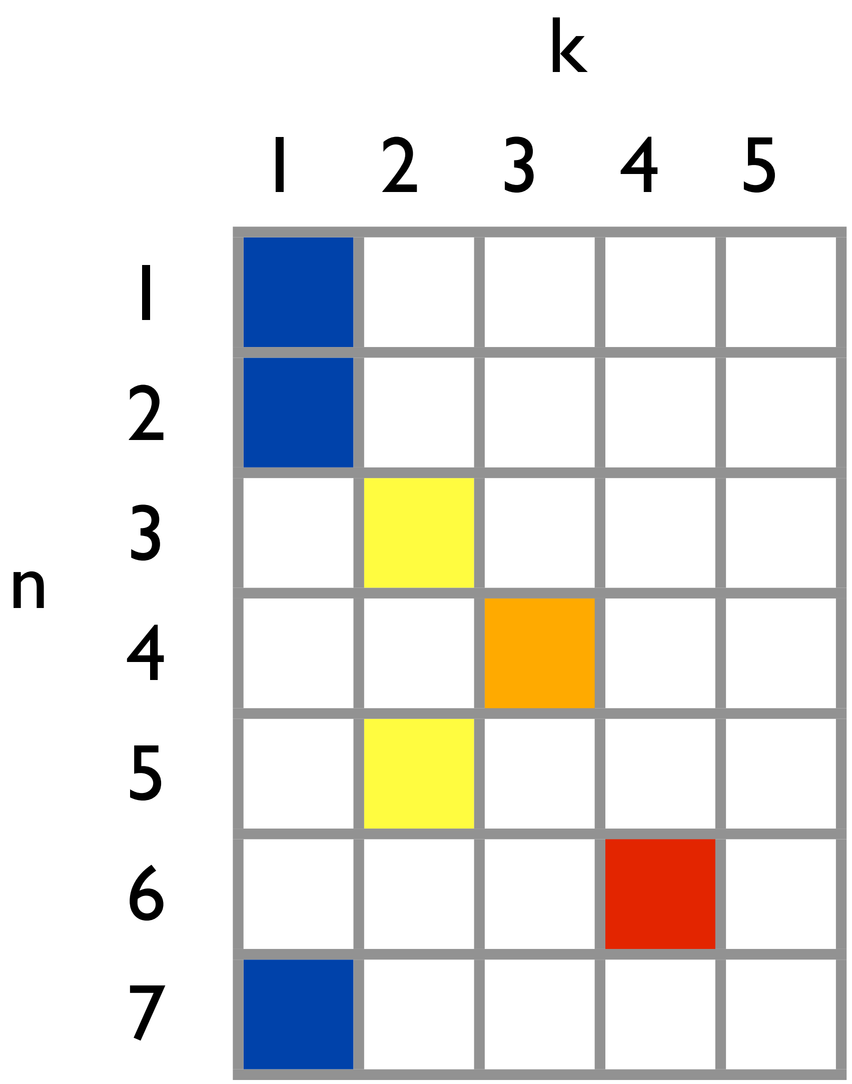
$Z'_1 = \phi_1$

$Z'_2 = \phi_1$

$Z'_3 = \phi_2$

Cluster labels





Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$

$Z_6 = 4$

$Z_7 = 1$

Random labels

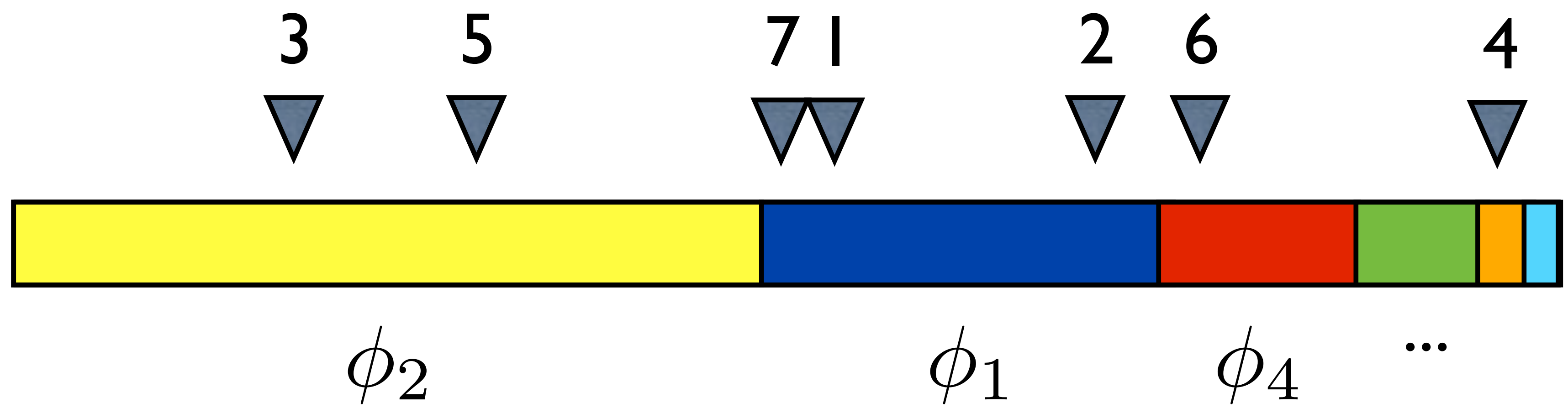
$Z'_1 = \phi_1$

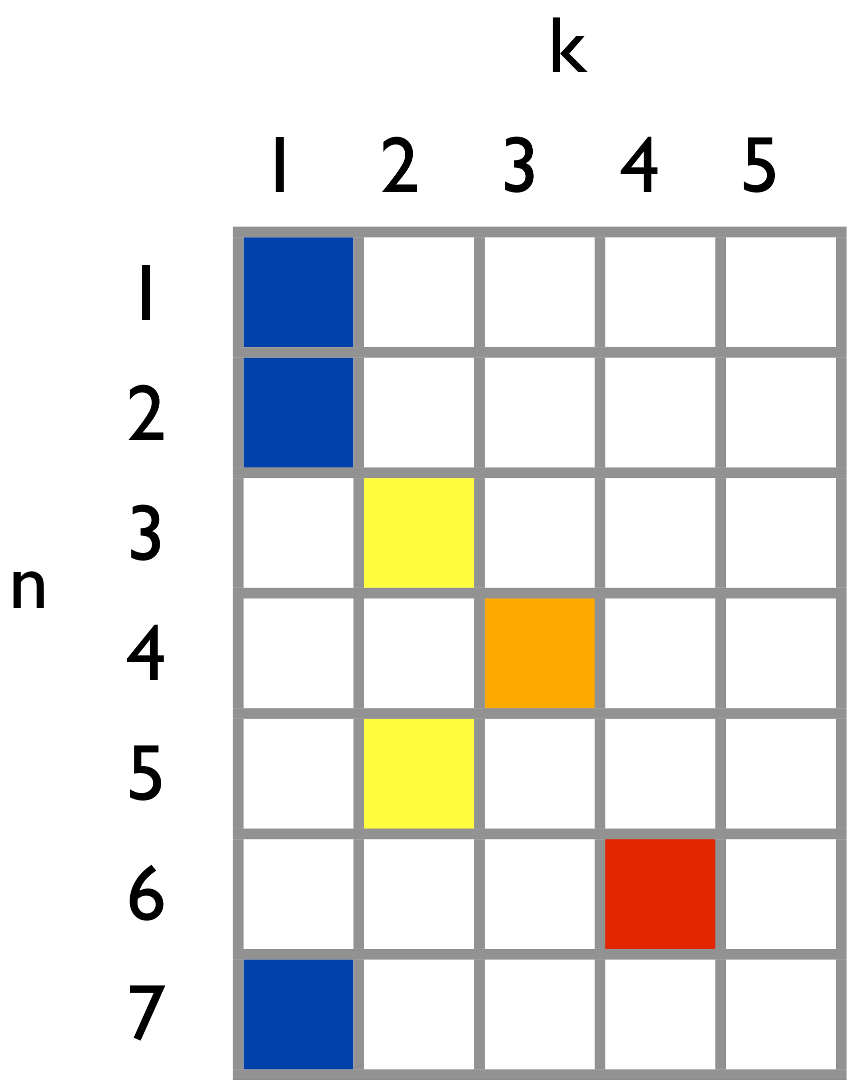
$Z'_2 = \phi_1$

$Z'_3 = \phi_2$

$Z'_4 = \phi_3$

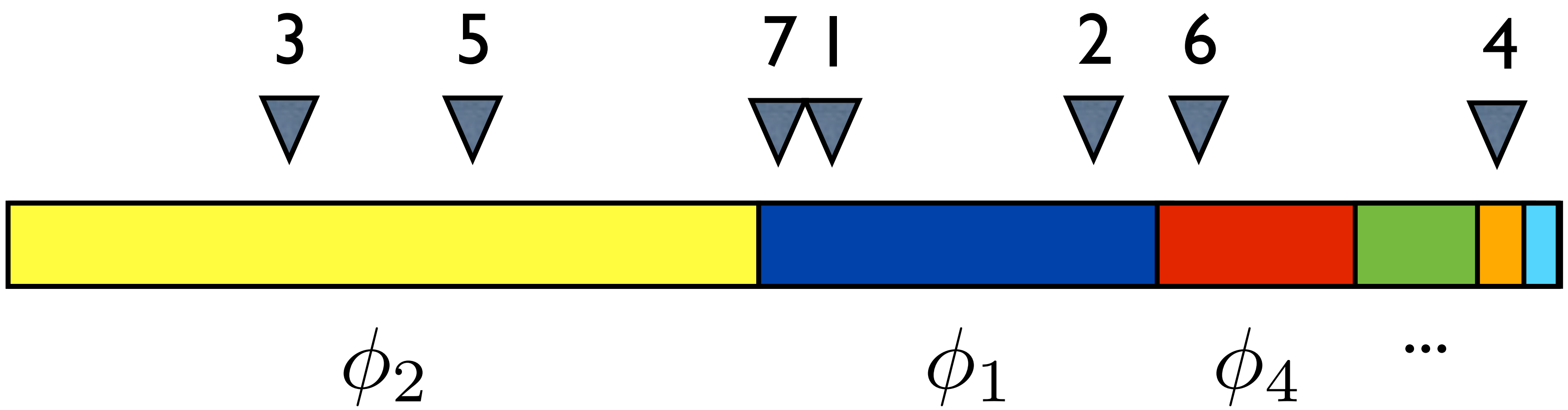
Cluster labels

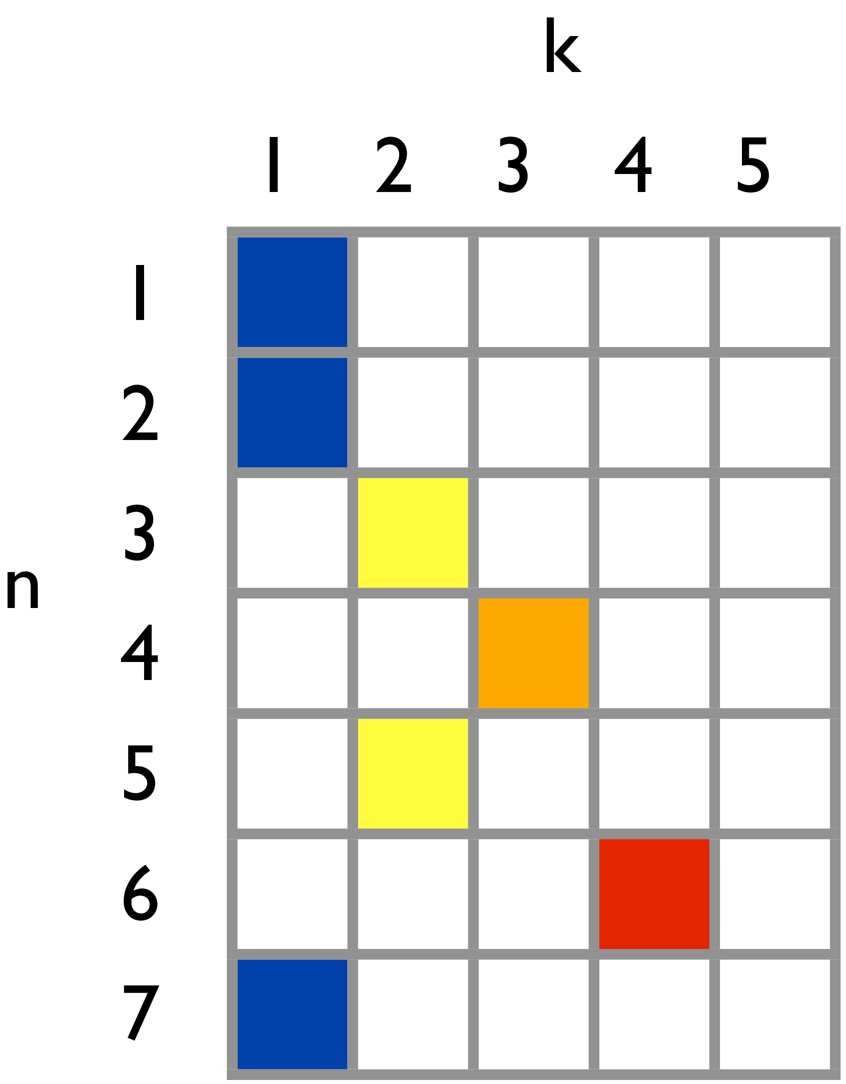




Order of appearance labels	Random labels
$Z_1 = 1$	$Z'_1 = \phi_1$
$Z_2 = 1$	$Z'_2 = \phi_1$
$Z_3 = 2$	$Z'_3 = \phi_2$
$Z_4 = 3$	$Z'_4 = \phi_3$
$Z_5 = 2$	$Z'_5 = \phi_2$
$Z_6 = 4$	
$Z_7 = 1$	

Cluster labels





Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$

$Z_6 = 4$

$Z_7 = 1$

Random labels

$Z'_1 = \phi_1$

$Z'_2 = \phi_1$

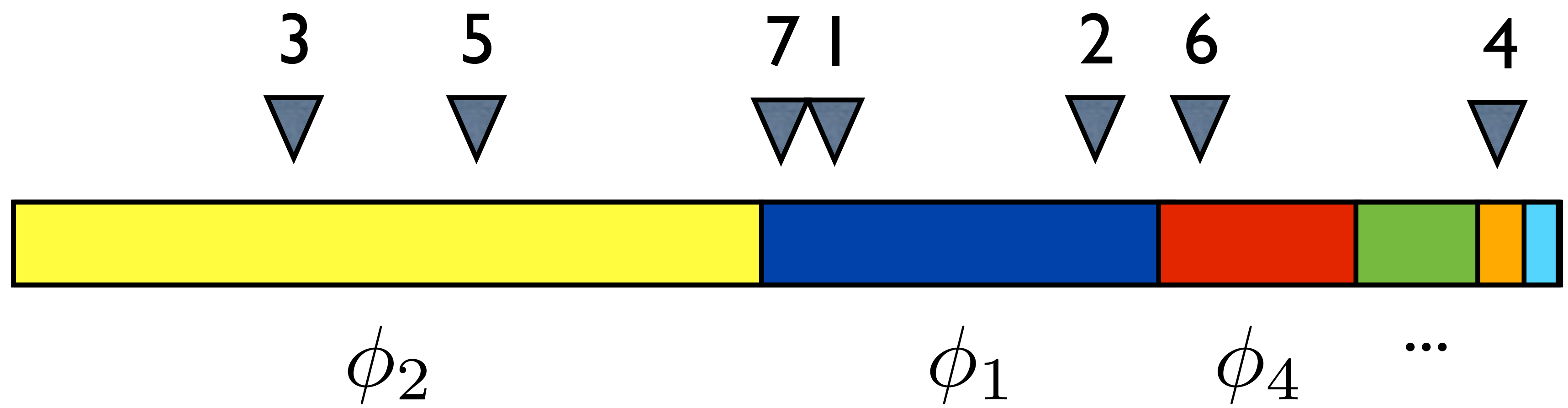
$Z'_3 = \phi_2$

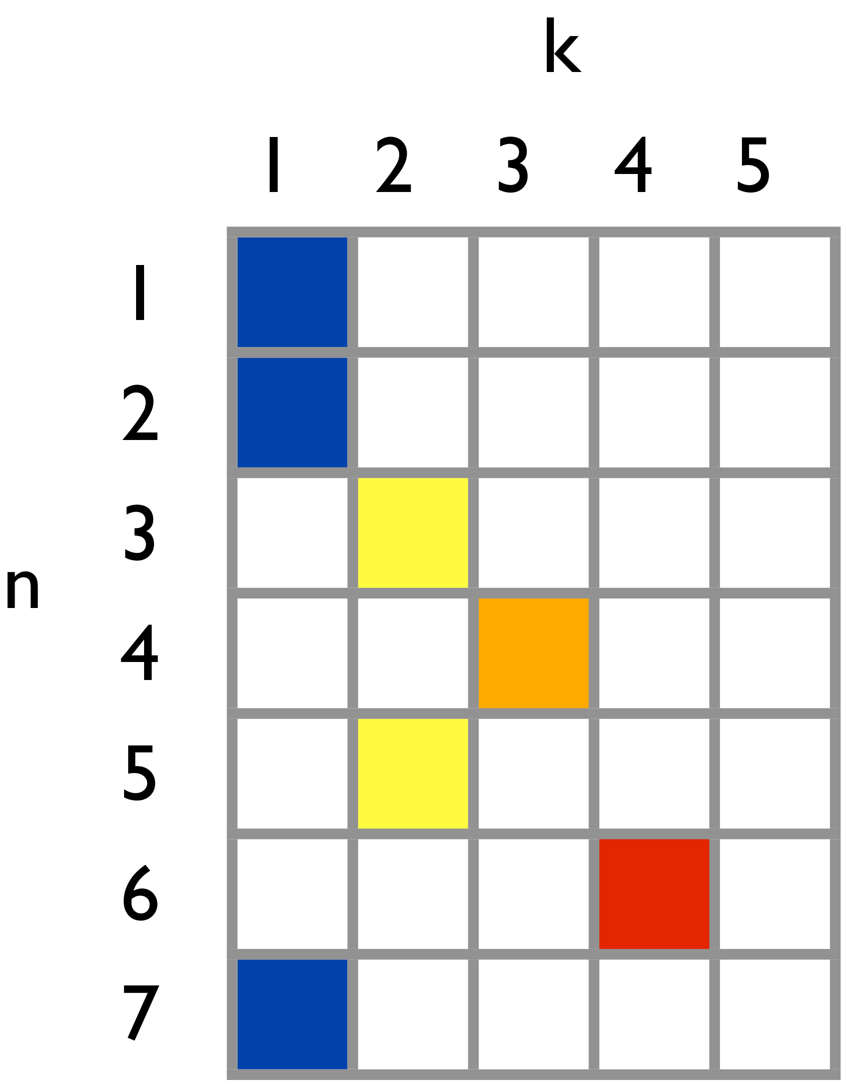
$Z'_4 = \phi_3$

$Z'_5 = \phi_2$

$Z'_6 = \phi_4$

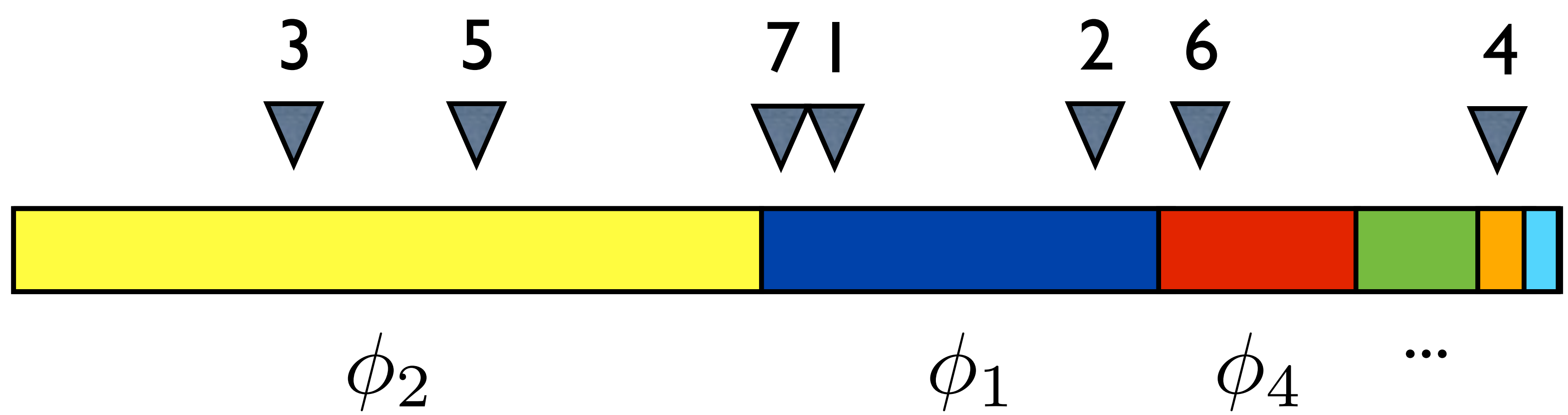
Cluster labels

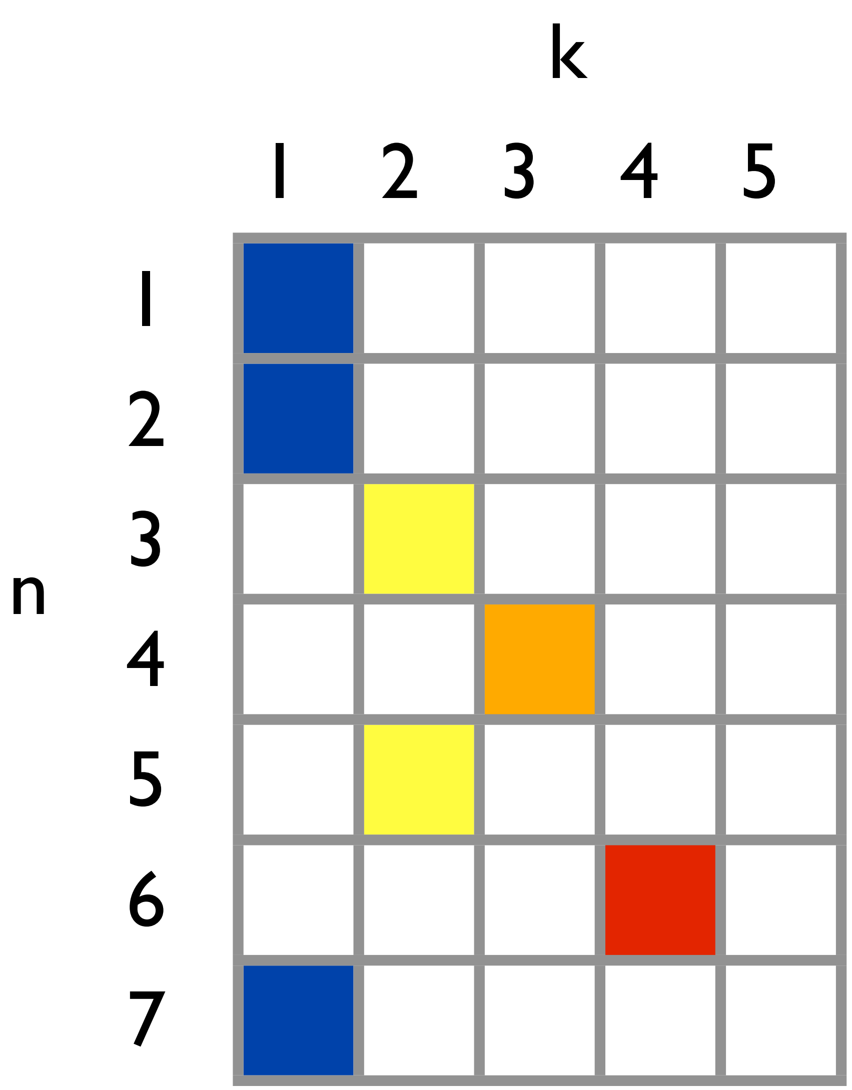




Order of appearance labels	Random labels
$Z_1 = 1$	$Z'_1 = \phi_1$
$Z_2 = 1$	$Z'_2 = \phi_1$
$Z_3 = 2$	$Z'_3 = \phi_2$
$Z_4 = 3$	$Z'_4 = \phi_3$
$Z_5 = 2$	$Z'_5 = \phi_2$
$Z_6 = 4$	$Z'_6 = \phi_4$
$Z_7 = 1$	$Z'_7 = \phi_1$

Cluster labels





Order of appearance labels	Random labels
$Z_1 = 1$	$Z'_1 = \phi_1$
$Z_2 = 1$	$Z'_2 = \phi_1$
$Z_3 = 2$	$Z'_3 = \phi_2$
$Z_4 = 3$	$Z'_4 = \phi_3$
$Z_5 = 2$	$Z'_5 = \phi_2$
$Z_6 = 4$	$Z'_6 = \phi_4$
$Z_7 = 1$	$Z'_7 = \phi_1$

$$\phi_k \stackrel{iid}{\sim} H$$

H continuous

Cluster labels

Order of appearance labels

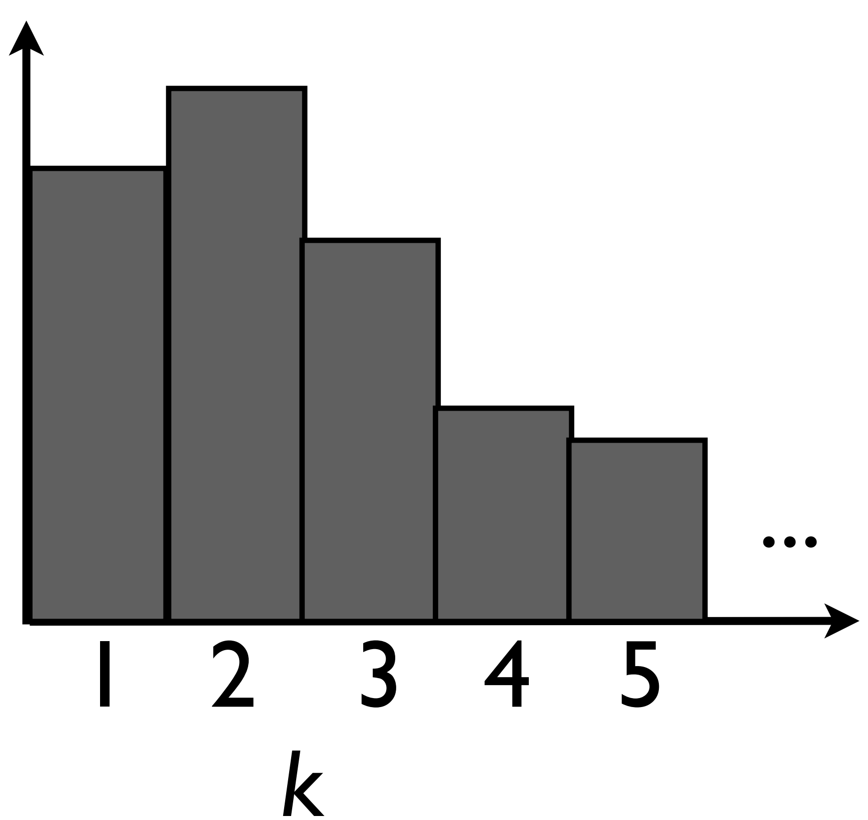
$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$



Random labels

$Z'_1 = \phi_1$

$Z'_2 = \phi_1$

$Z'_3 = \phi_2$

$Z'_4 = \phi_3$

$Z'_5 = \phi_2$

?

Cluster labels

Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$

Random labels

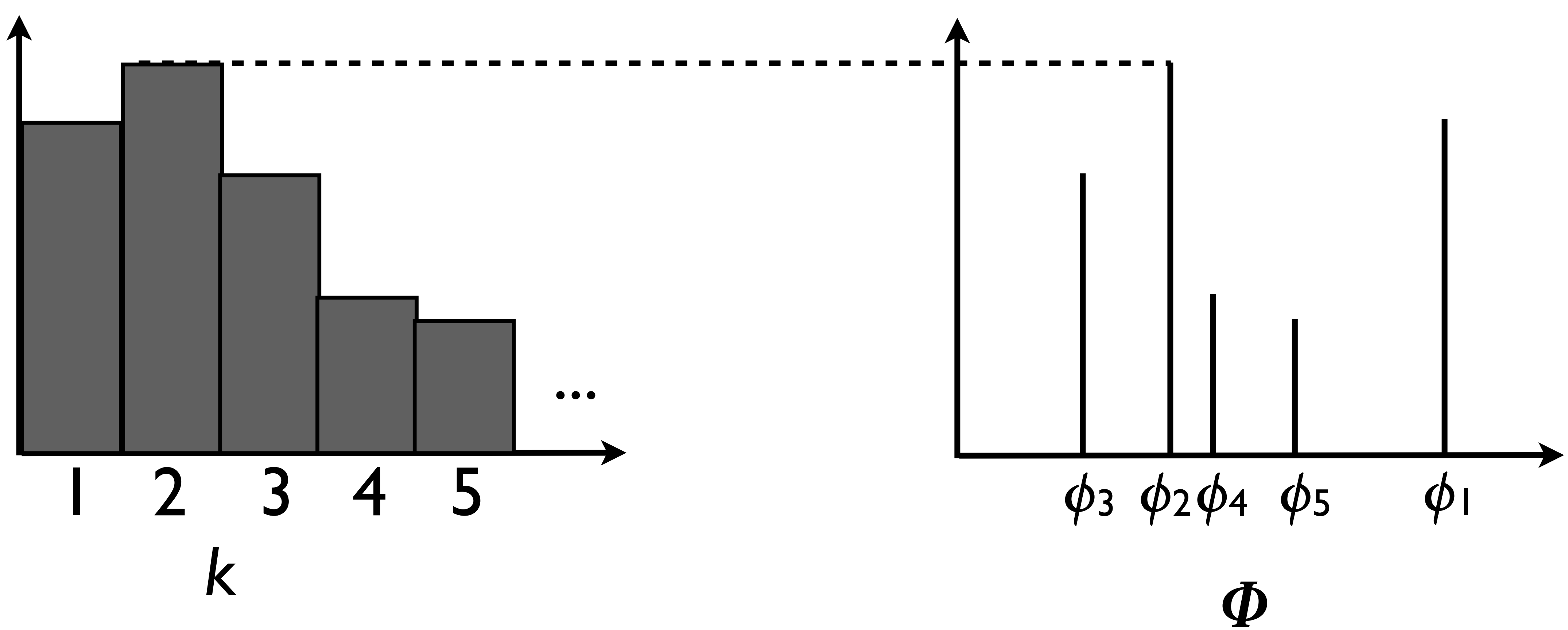
$Z'_1 = \phi_1$

$Z'_2 = \phi_1$

$Z'_3 = \phi_2$

$Z'_4 = \phi_3$

$Z'_5 = \phi_2$



Cluster labels

Order of appearance labels

$Z_1 = 1$

$Z_2 = 1$

$Z_3 = 2$

$Z_4 = 3$

$Z_5 = 2$

Random labels

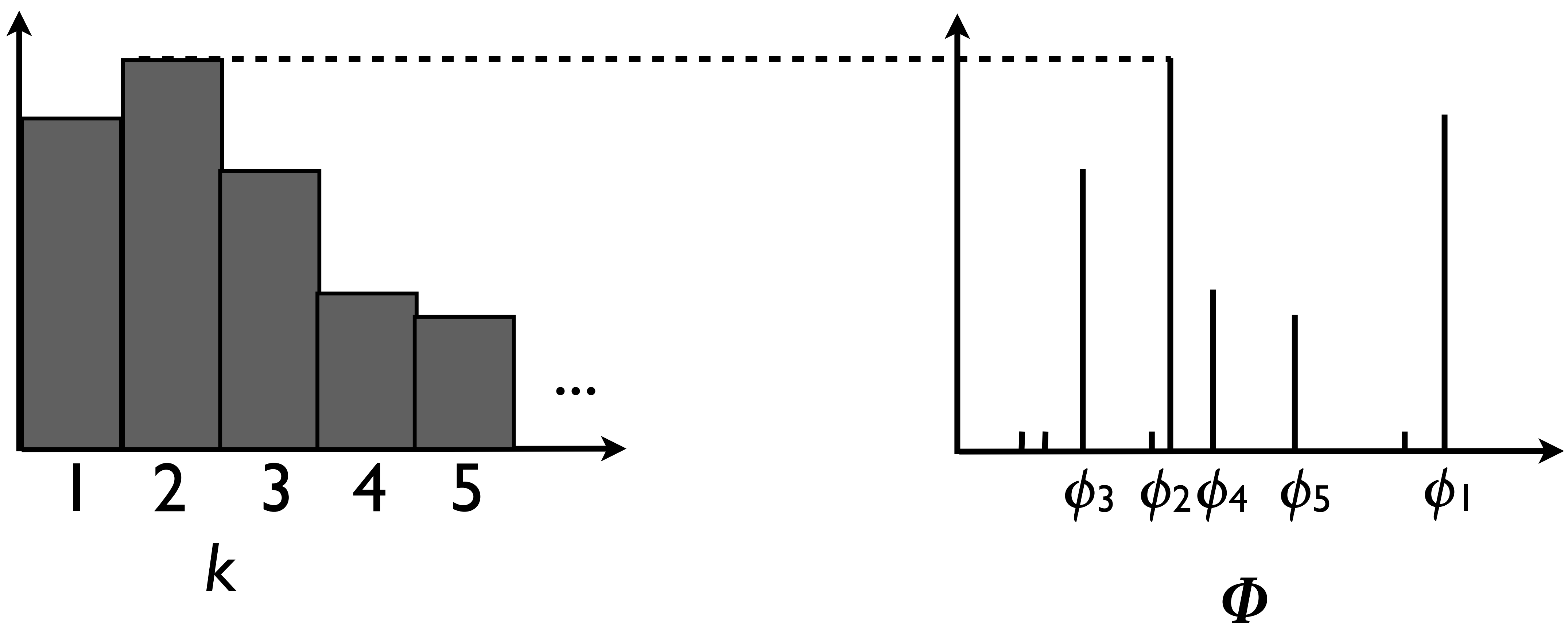
$Z'_1 = \phi_1$

$Z'_2 = \phi_1$

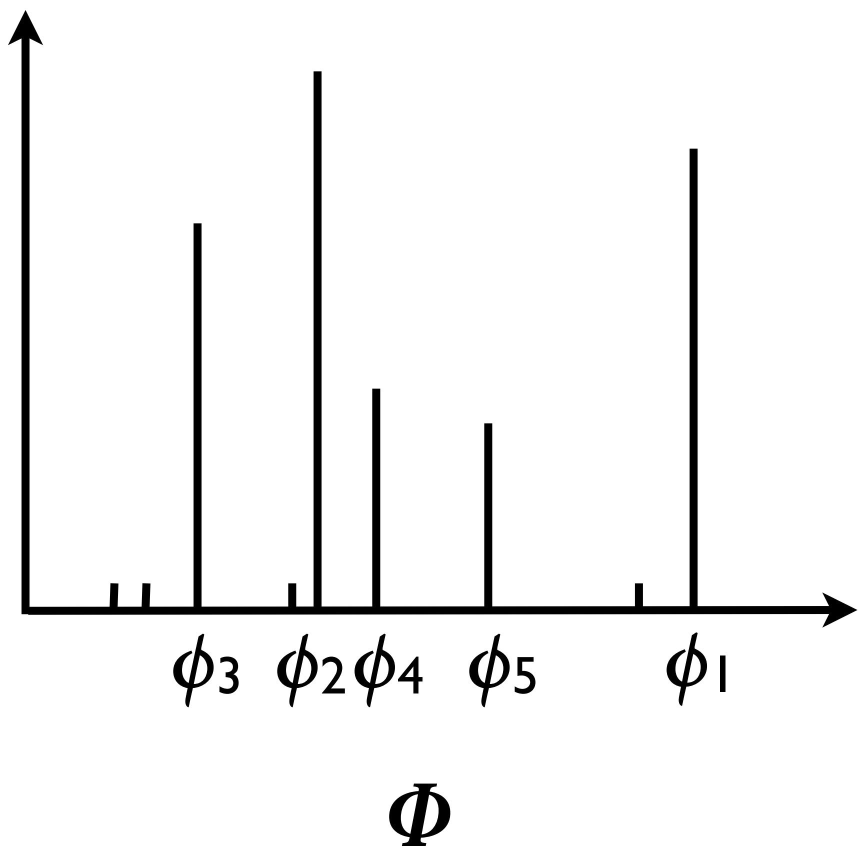
$Z'_3 = \phi_2$

$Z'_4 = \phi_3$

$Z'_5 = \phi_2$

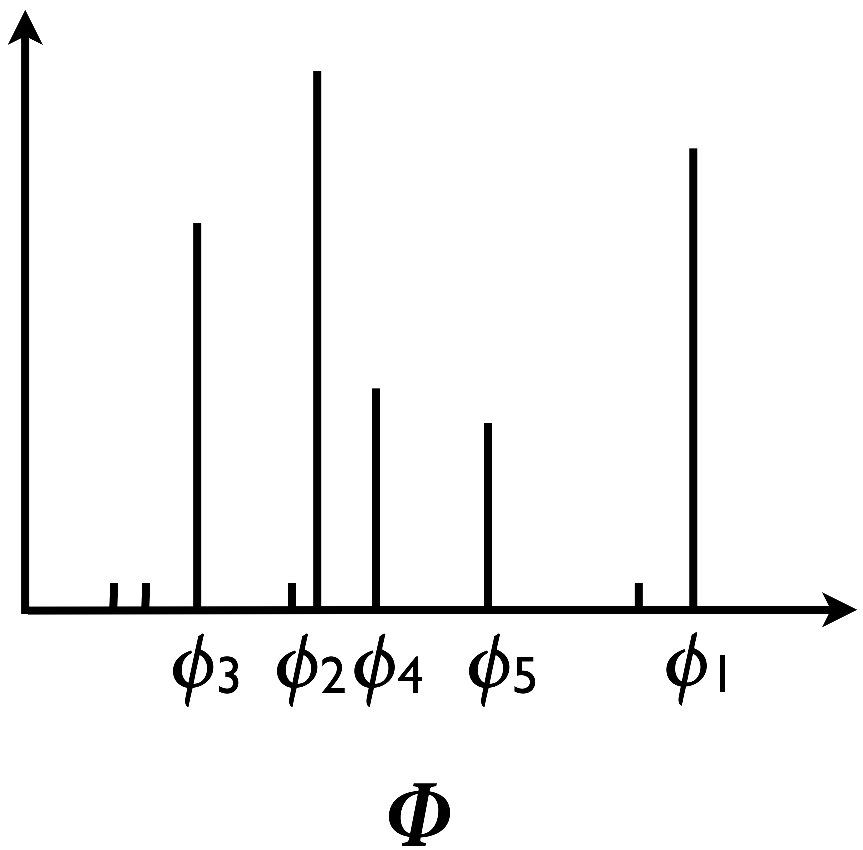


Random probability measure



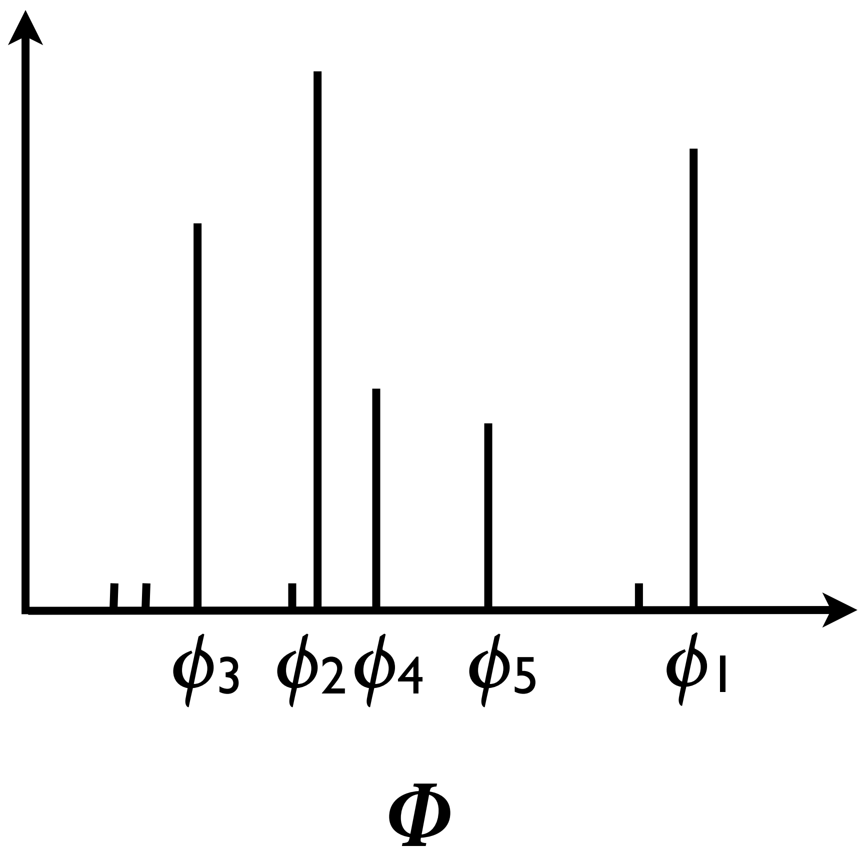
Random probability measure

- Def: Random measure with total mass one



Random probability measure

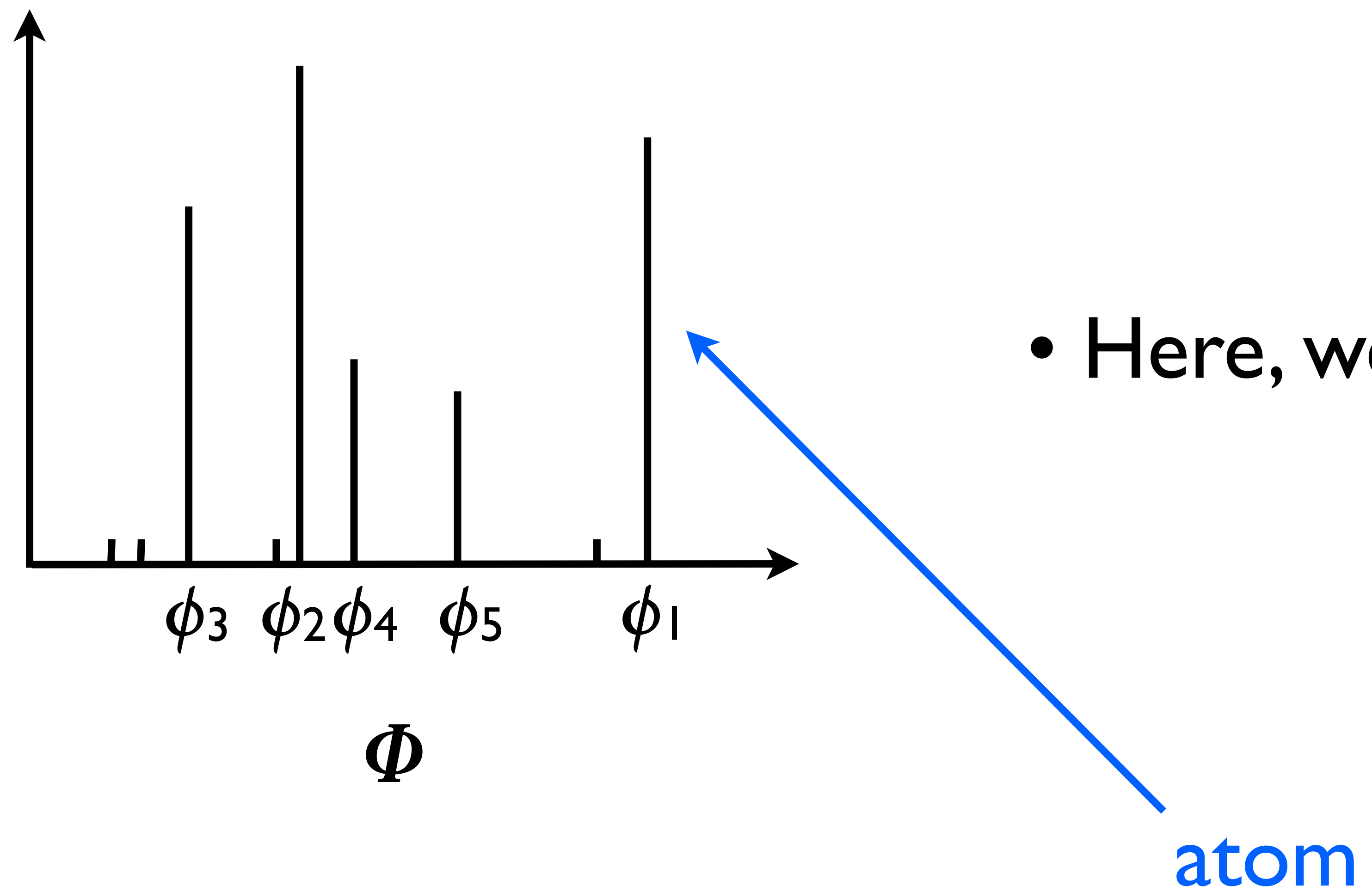
- Def: Random measure with total mass one



- Here, we also have point process

Random probability measure

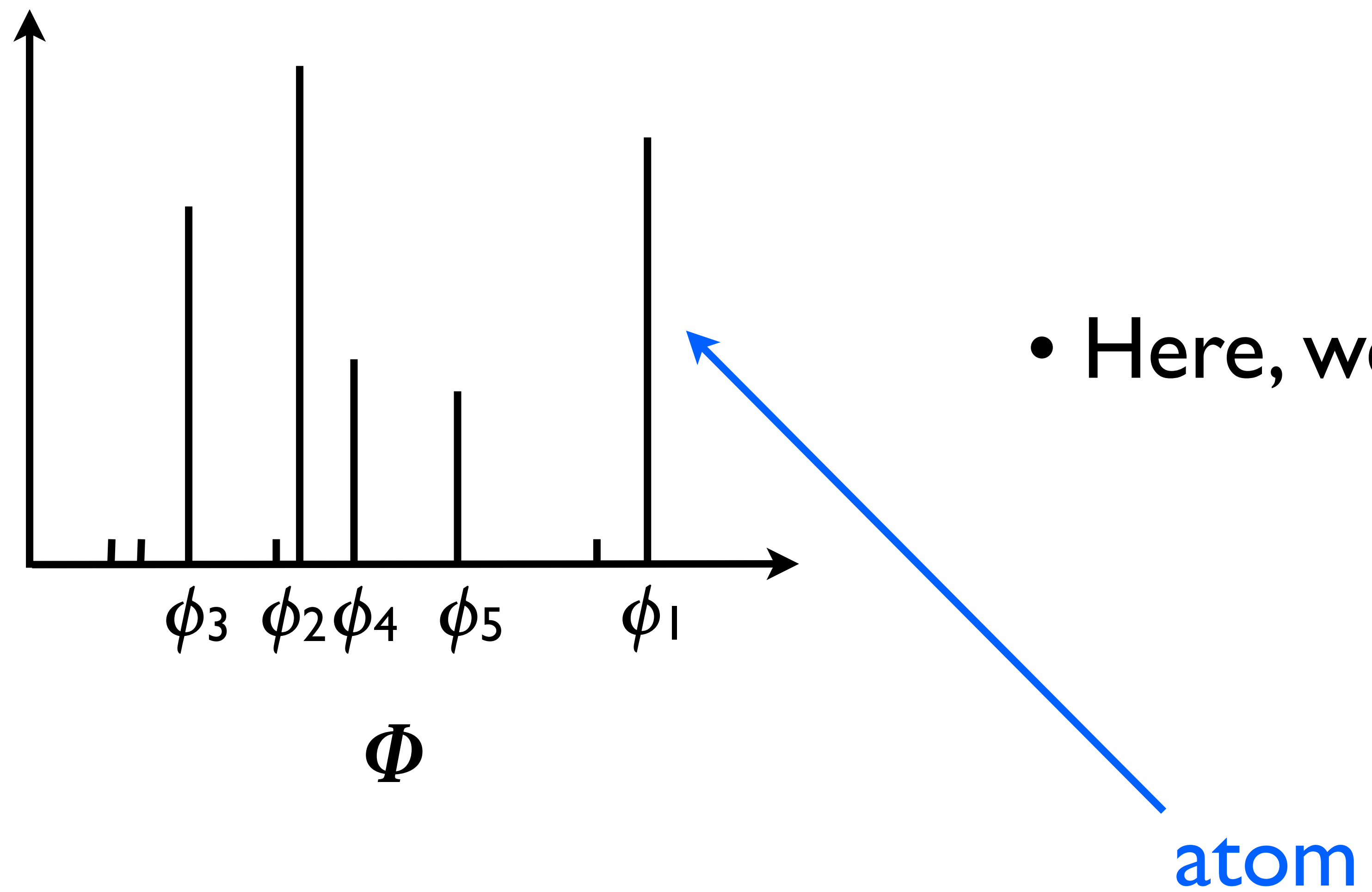
- Def: Random measure with total mass one



- Here, we also have point process

Random probability measure

- Def: Random measure with total mass one



- Here, we also have point process

Example: Dirichlet process

- The random probability measure with CRP stick-breaking atom sizes

Clusters: augmentation

Clusters: augmentation

random partition
& EPPF

Clusters: augmentation

random partition
& EPPF
CRP

Clusters: augmentation

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

random partition
& EPPF
CRP

Clusters: augmentation

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

random partition
& EPPF
CRP
(continuous-valued)
random cluster labels

Clusters: augmentation

$$\pi_9 = \{\{9,2,7,1\}, \{8,4,6\},\{5,3\}\}$$

random partition & EPPF

CRP

(continuous-valued) random cluster labels

CRP with cluster means

Clusters: augmentation

$$\pi_9 = \{\{9,2,7,1\}, \{8,4,6\},\{5,3\}\}$$

$Z'_1 = \phi_1, Z'_2 = \phi_1, Z'_3 = \phi_2, \ldots, Z'_9 = \phi_1$

random partition & EPPF

CRP

(continuous-valued) random cluster labels

CRP with cluster means

Clusters: augmentation

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

$Z'_1 = \phi_1, Z'_2 = \phi_1, Z'_3 = \phi_2, \ldots, Z'_9 = \phi_1$

random partition & EPPF
CRP

(continuous-valued) random cluster labels
CRP with cluster means

cluster proportions/ Kingman paintbox

Clusters: augmentation

$$\pi_9 = \{\{9,2,7,1\},$$
$$\{8,4,6\},\{5,3\}\}$$

$Z'_1 = \phi_1, Z'_2 = \phi_1, Z'_3 = \phi_2, \ldots, Z'_9 = \phi_1$

random partition & EPPF
- CRP

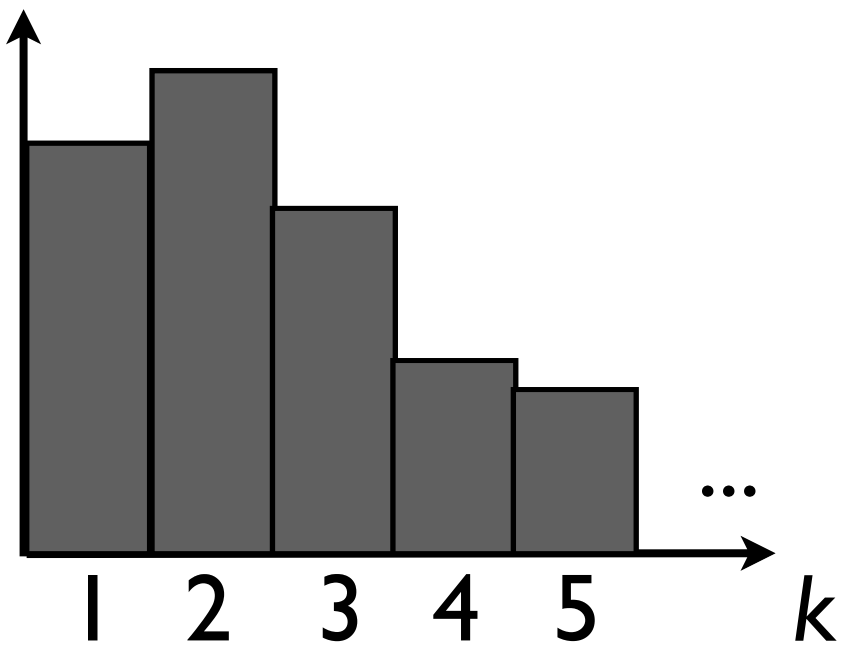
(continuous-valued) random cluster labels
- CRP with cluster means

cluster proportions/ Kingman paintbox
- CRP stick-breaking

Clusters: augmentation

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

Z'$_1$ = ϕ_1, Z'$_2$ = ϕ_1, Z'$_3$ = ϕ_2, ..., Z'$_9$ = ϕ_1



random partition & EPPF

CRP

(continuous-valued) random cluster labels

CRP with cluster means

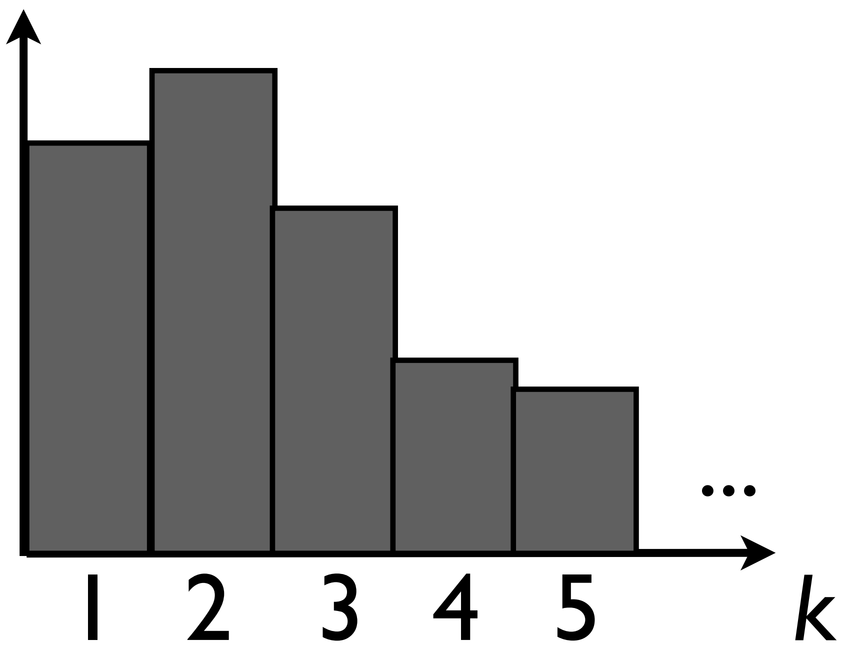
cluster proportions/ Kingman paintbox

CRP stick-breaking

Clusters: augmentation

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$

Z'$_1$ = ϕ_1, Z'$_2$ = ϕ_1, Z'$_3$ = ϕ_2, ..., Z'$_9$ = ϕ_1



random partition & EPPF

CRP

(continuous-valued) random cluster labels

CRP with cluster means

cluster proportions/ Kingman paintbox

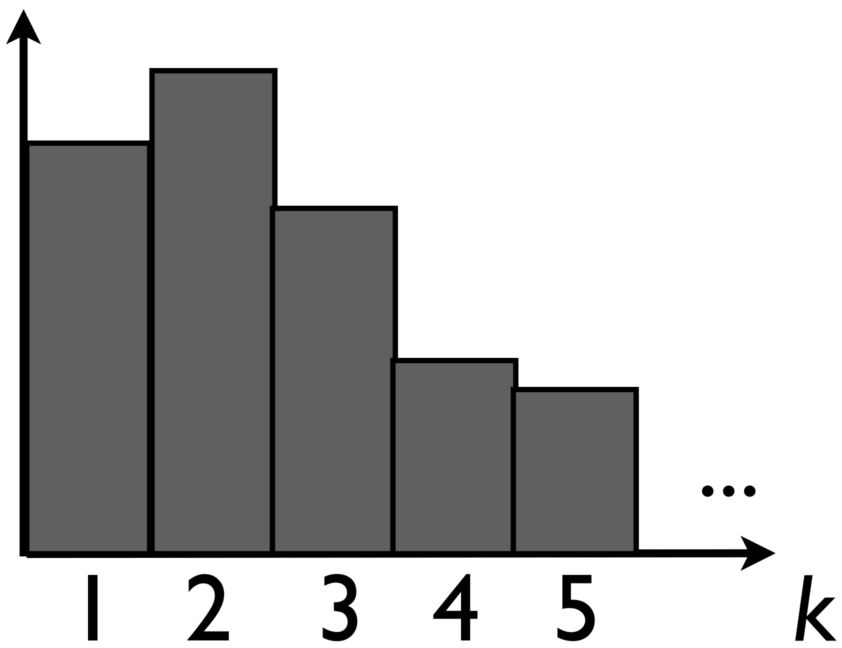
CRP stick-breaking

random, discrete probability measure

Clusters: augmentation

$$\pi_9 = \{\{9,2,7,1\},\ \{8,4,6\},\{5,3\}\}$$

$Z'_1 = \phi_1, Z'_2 = \phi_1, Z'_3 = \phi_2, \ldots, Z'_9 = \phi_1$



random partition & EPPF

CRP

(continuous-valued) random cluster labels

CRP with cluster means

cluster proportions/ Kingman paintbox

CRP stick-breaking

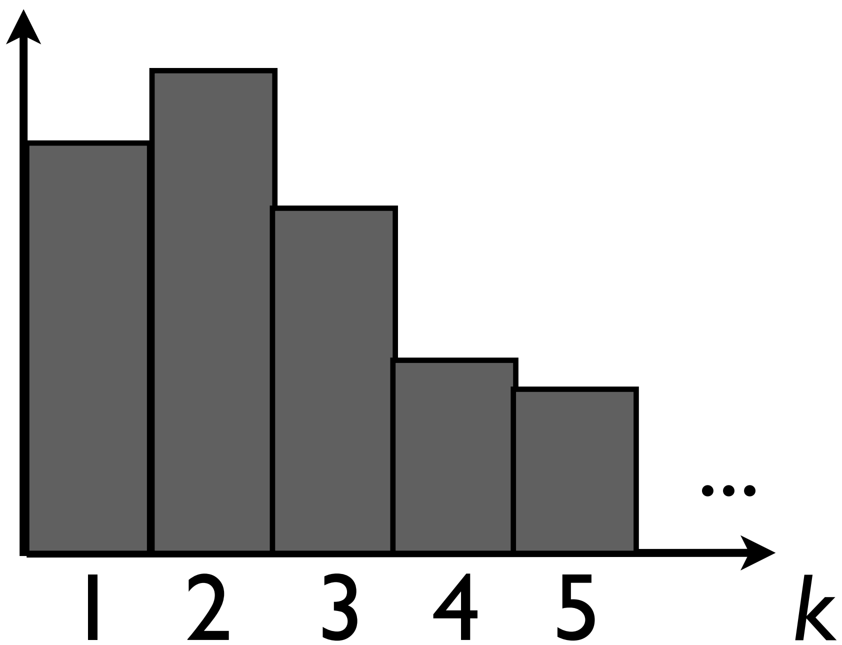
random, discrete probability measure

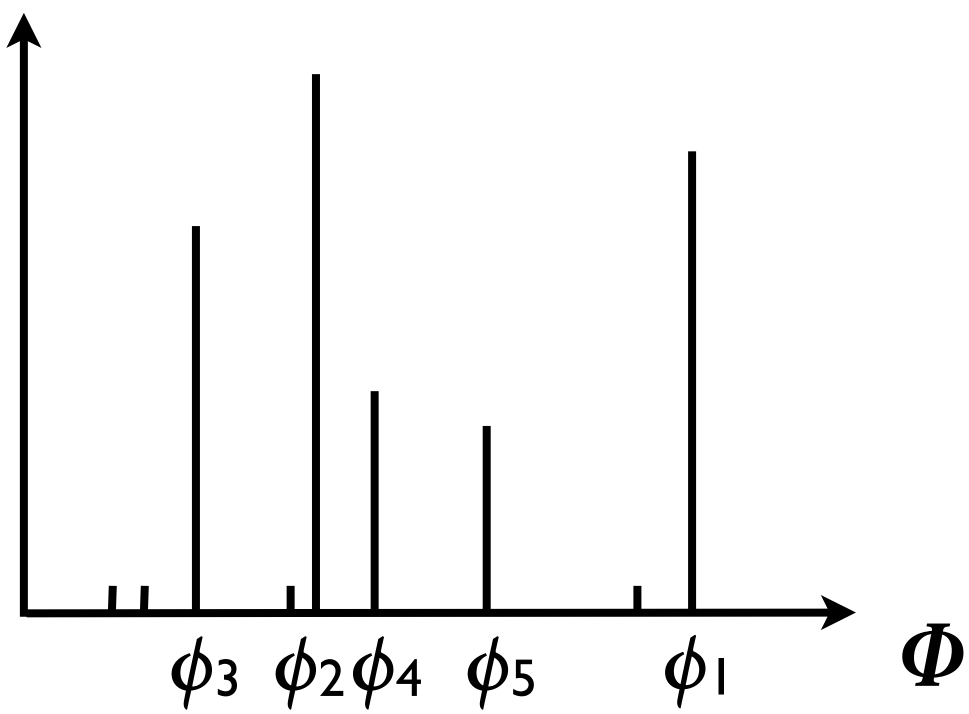
Dirichlet process

Clusters: augmentation

$$\pi_9 = \{\{9,2,7,1\},\ \{8,4,6\},\{5,3\}\}$$

$Z'_1 = \phi_1, Z'_2 = \phi_1, Z'_3 = \phi_2, ..., Z'_9 = \phi_1$





random partition & EPPF

CRP

(continuous-valued) random cluster labels

CRP with cluster means

cluster proportions/ Kingman paintbox

CRP stick-breaking

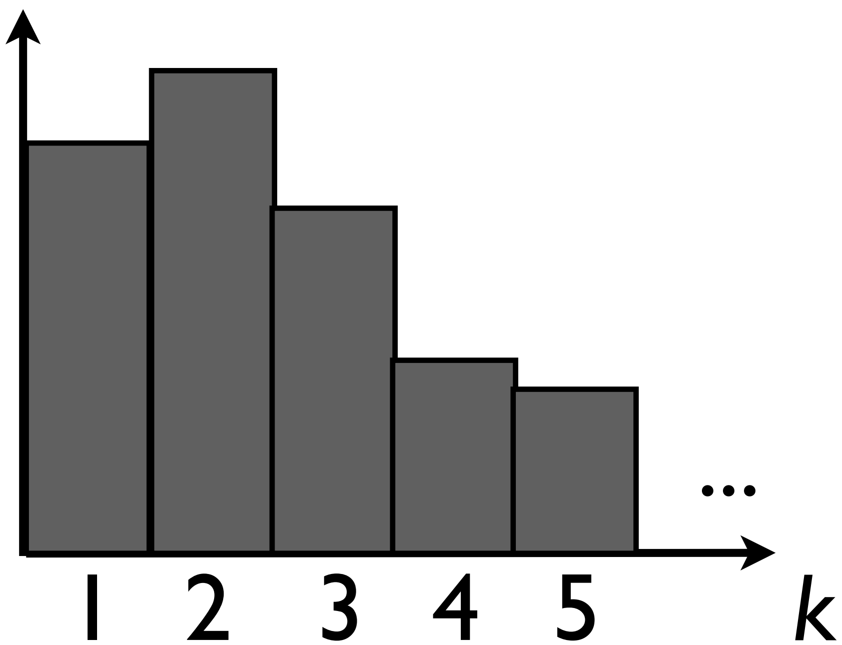
random, discrete probability measure

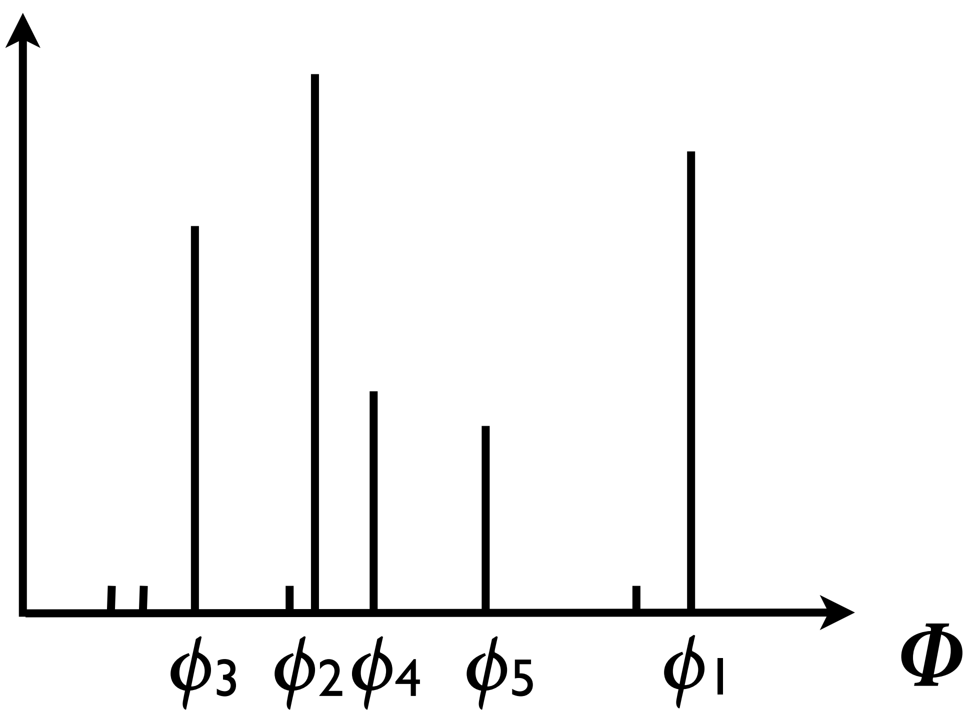
Dirichlet process

Clusters: augmentation

$$\pi_9 = \{\{9, 2, 7, 1\},\ \{8, 4, 6\}, \{5, 3\}\}$$

$Z'_1 = \phi_1, Z'_2 = \phi_1, Z'_3 = \phi_2, \ldots, Z'_9 = \phi_1$





random partition & EPPF
CRP

(continuous-valued) random cluster labels
CRP with cluster means

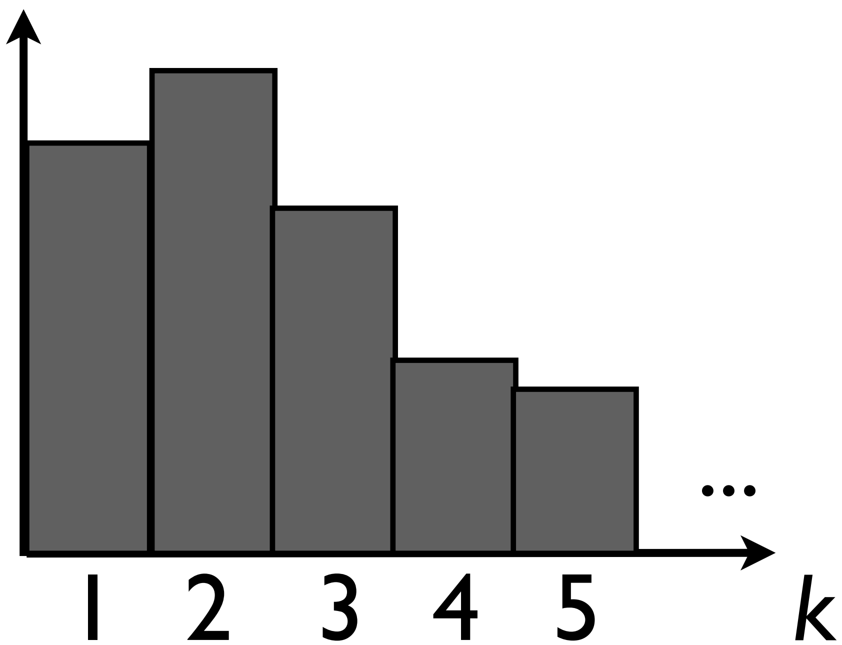
cluster proportions/ Kingman paintbox
CRP stick-breaking

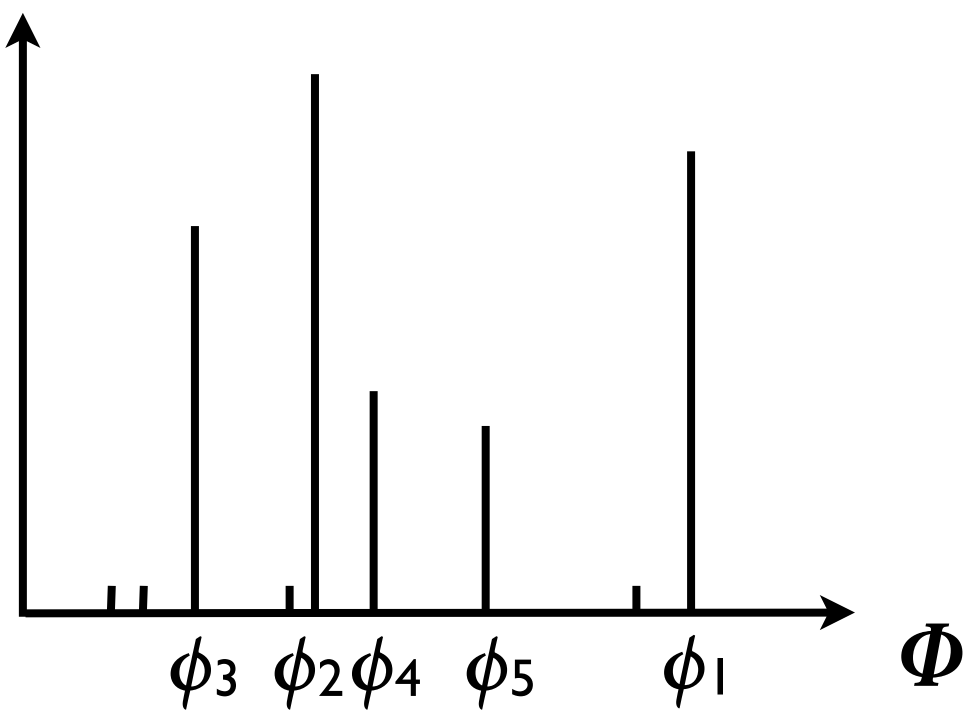
random, discrete probability measure
Dirichlet process

Clusters: integrating out

$$\pi_9 = \{\{9, 2, 7, 1\},$$
$$\{8, 4, 6\}, \{5, 3\}\}$$

$Z'_1 = \phi_1, Z'_2 = \phi_1, Z'_3 = \phi_2, ..., Z'_9 = \phi_1$





random partition & EPPF

CRP

(continuous-valued) random cluster labels

CRP with cluster means

cluster proportions/ Kingman paintbox

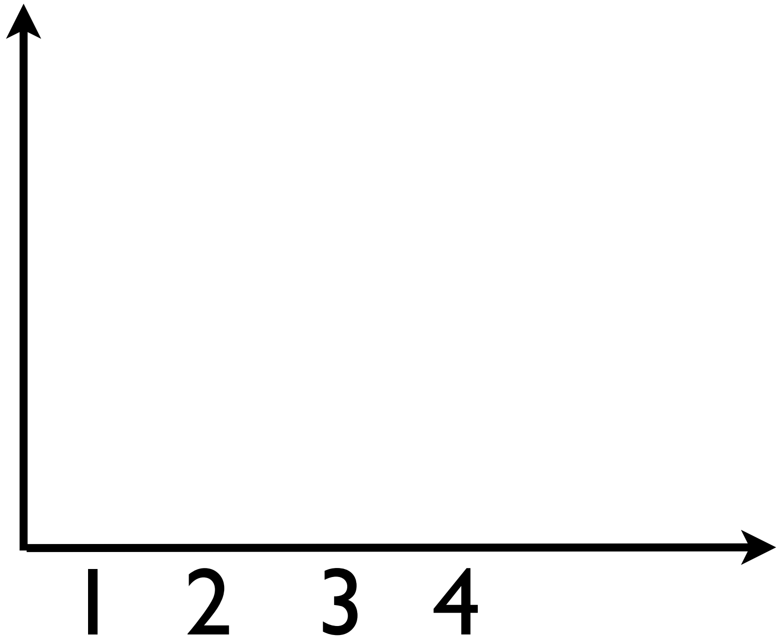
CRP stick-breaking

random, discrete probability measure

Dirichlet process

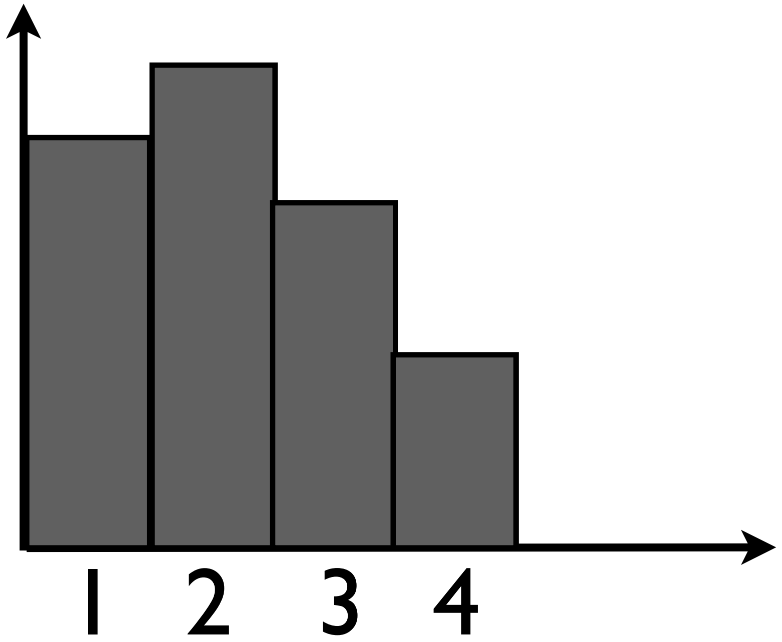
Why the CRP?

Finite, fixed number of clusters



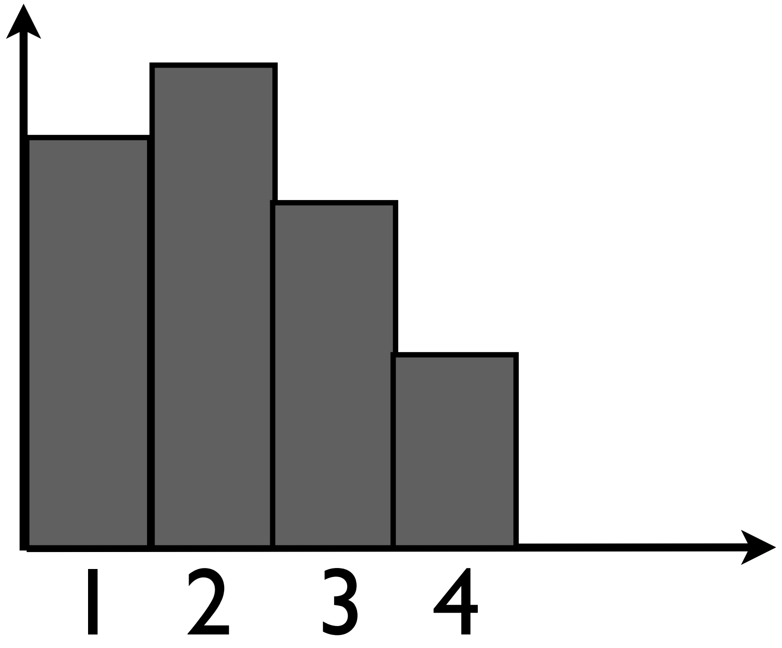
Why the CRP?

Finite, fixed number of clusters



Why the CRP?

Finite, fixed number of clusters

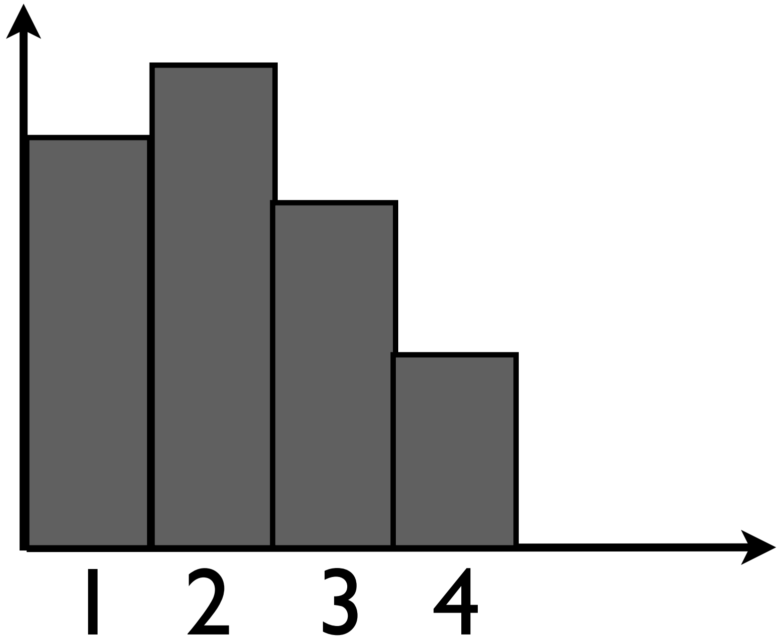


$$(q_k)_{k=1}^{K} \sim \mathrm{Dirichlet}(K, \theta)$$

Why the CRP?

Finite, fixed number of clusters

$$\pi_9 = \{\{9,2,7,1\},\ \{8,4,6\},\{5,3\}\}$$

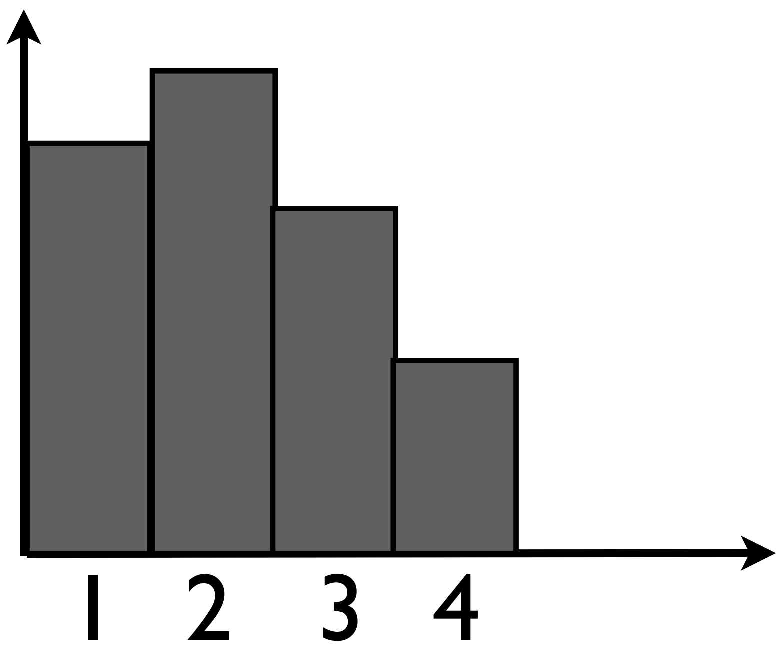


$$(q_k)_{k=1}^{K} \sim \text{Dirichlet}(K,\theta)$$

Why the CRP?

Finite, fixed number of clusters

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$



$$Z_n \overset{iid}{\sim} q$$

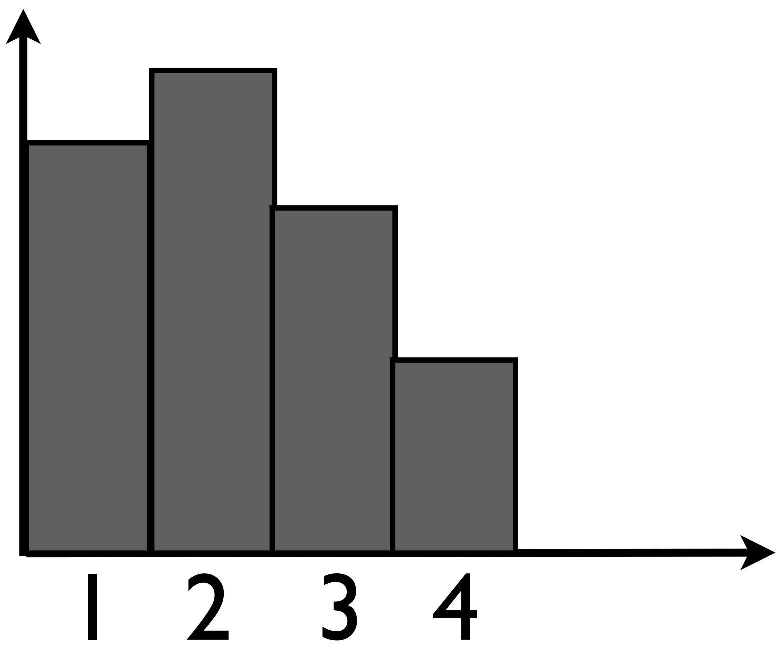
$$k = 1, \ldots, K$$

$$(q_k)_{k=1}^{K} \sim \text{Dirichlet}(K, \theta)$$

Why the CRP?

Finite, fixed number of clusters

$$\pi_9 = \{\{9, 2, 7, 1\}, \{8, 4, 6\}, \{5, 3\}\}$$



$$Z_n \stackrel{iid}{\sim} q$$

$$k = 1, \ldots, K$$

$$(q_k)_{k=1}^K \sim \text{Dirichlet}(K, \theta)$$

$$\phi_k \stackrel{iid}{\sim} H$$

$$k = 1, \ldots, K$$

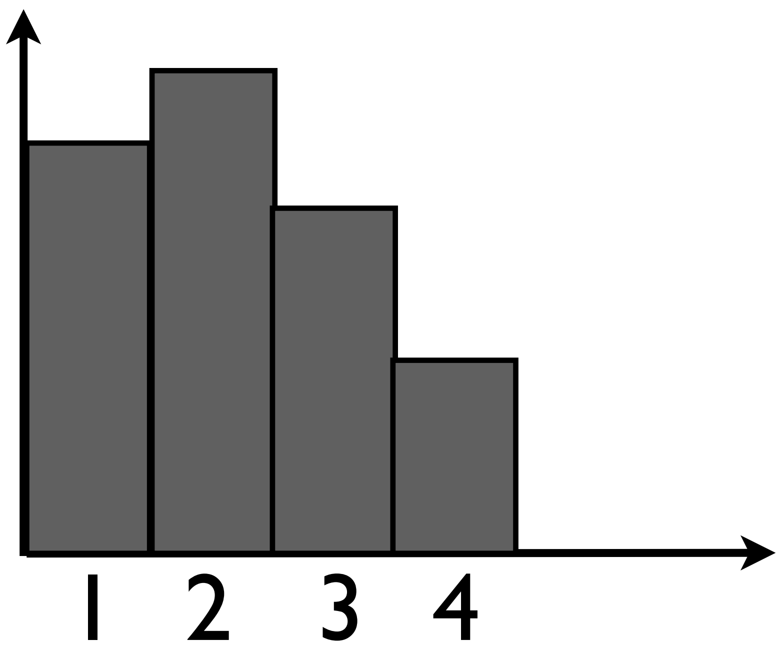
Why the CRP?

Finite, fixed number of clusters

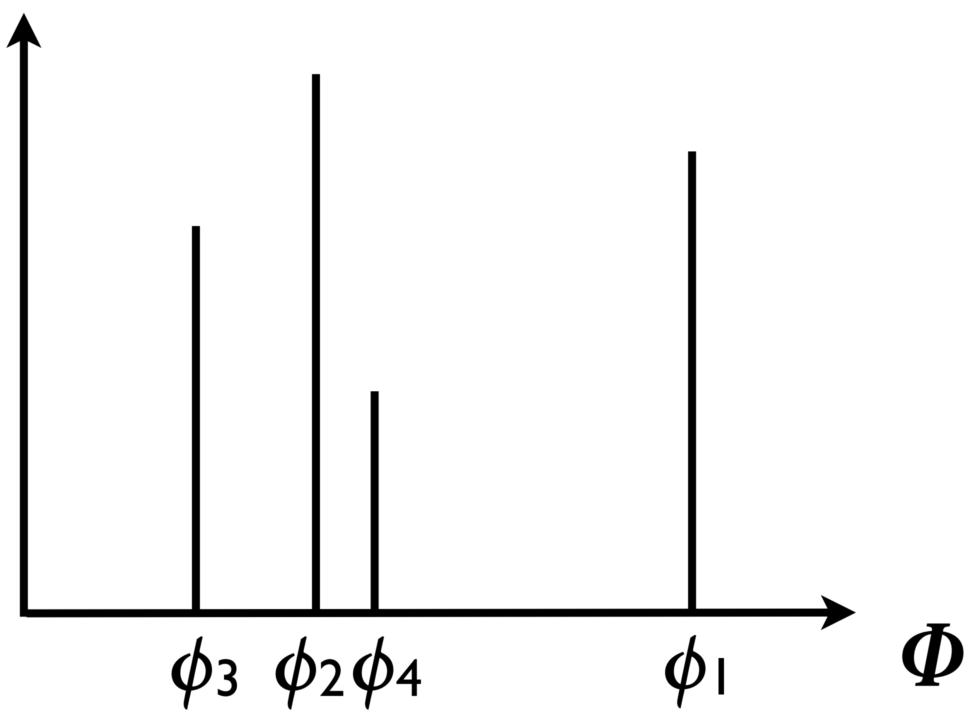
$$\pi_9 = \{\{9, 2, 7, 1\},\ \{8, 4, 6\}, \{5, 3\}\}$$

$$Z_n \overset{iid}{\sim} q$$

$$k = 1, \ldots, K$$



$$(q_k)_{k=1}^{K} \sim \text{Dirichlet}(K, \theta)$$



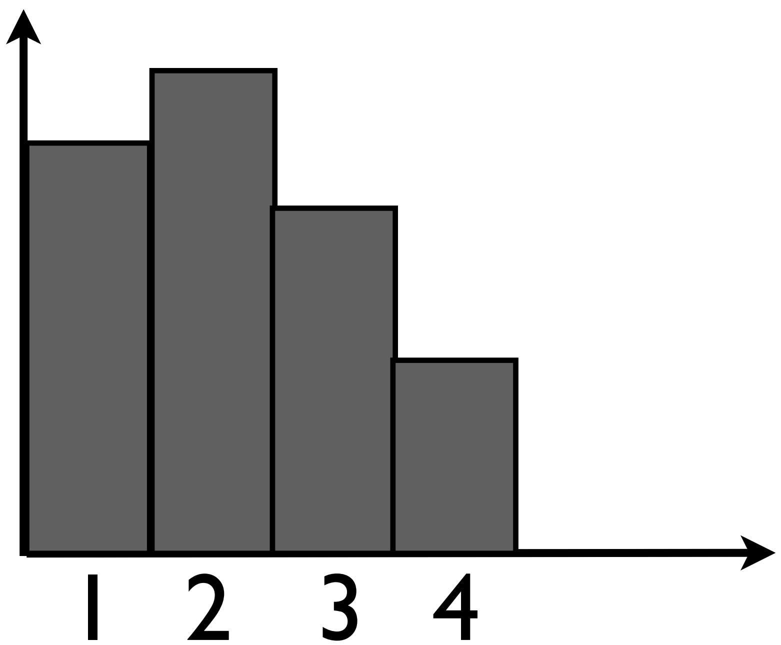
$$\phi_k \overset{iid}{\sim} H$$

$$k = 1, \ldots, K$$

Why the CRP?

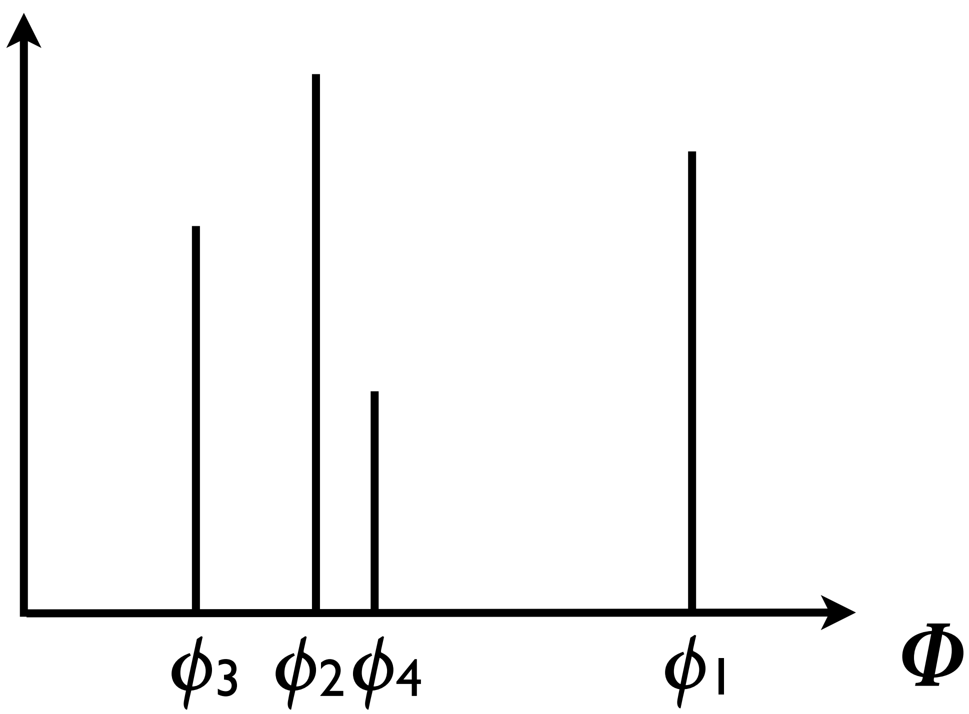
Unbounded number of clusters

$$\pi_9 = \{\{9,2,7,1\},\\ \{8,4,6\},\{5,3\}\}$$



$$Z_n \stackrel{iid}{\sim} q \qquad k = 1,\ldots,K$$

$$(q_k)_{k=1}^{K} \sim \mathrm{Dirichlet}(K,\theta)$$



$$\phi_k \stackrel{iid}{\sim} H \qquad k = 1,\ldots,K$$

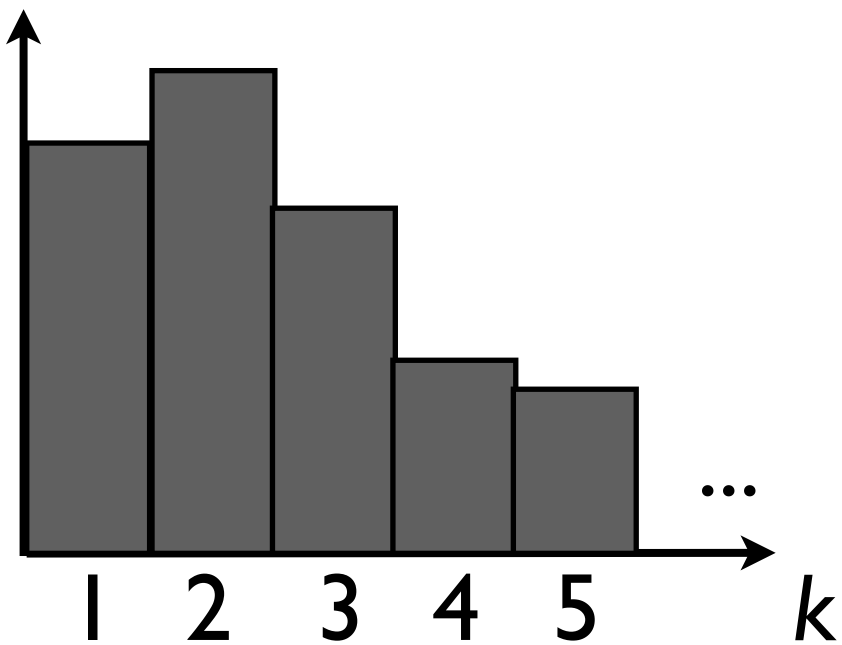
Why the CRP?

Unbounded number of clusters

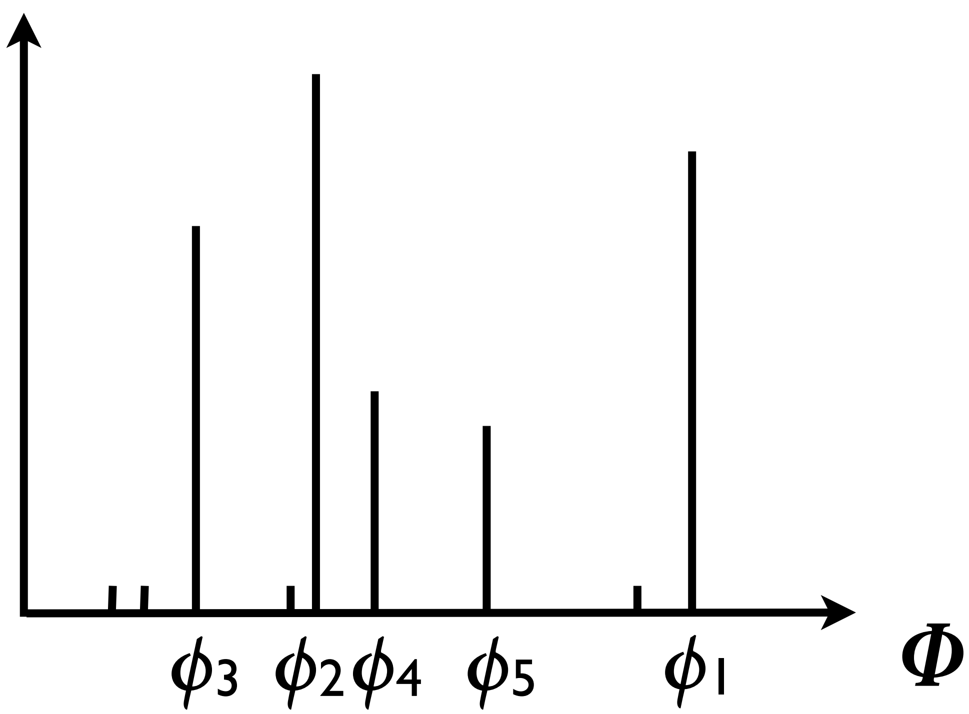
$$\pi_9 = \{\{9,2,7,1\}, \{8,4,6\},\{5,3\}\}$$

$$Z_n \overset{iid}{\sim} q$$

$$k = 1, 2, \ldots$$



$$(q_k)_{k=1}^{\infty} \sim \text{atom weights of Dirichlet Process}(\theta)$$



$$\phi_k \overset{iid}{\sim} H$$

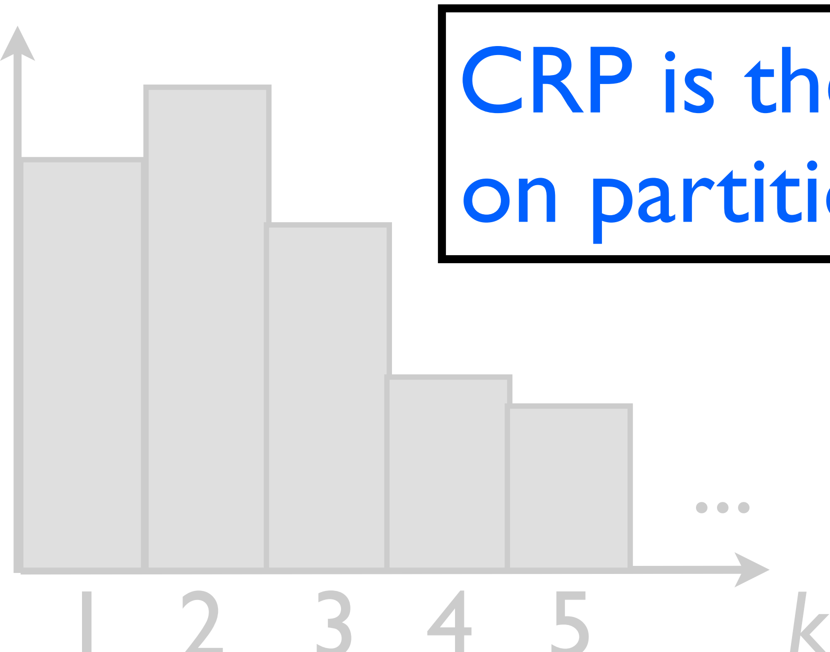
$$k = 1, 2, \ldots$$

Why the CRP?

Unbounded number of clusters

$$\pi_9 = \{\{9,2,7,1\},\\ \{8,4,6\},\{5,3\}\}$$

$$Z_n \overset{iid}{\sim} q \qquad k = 1, \ldots, K$$



CRP is the marginal distribution on partitions of the data indices

$(q_k)_{k=1}^{\infty} \sim$ atom weights of Dirichlet Process(θ)

$$\phi_k \overset{iid}{\sim} H \qquad k = 1, \ldots, K$$