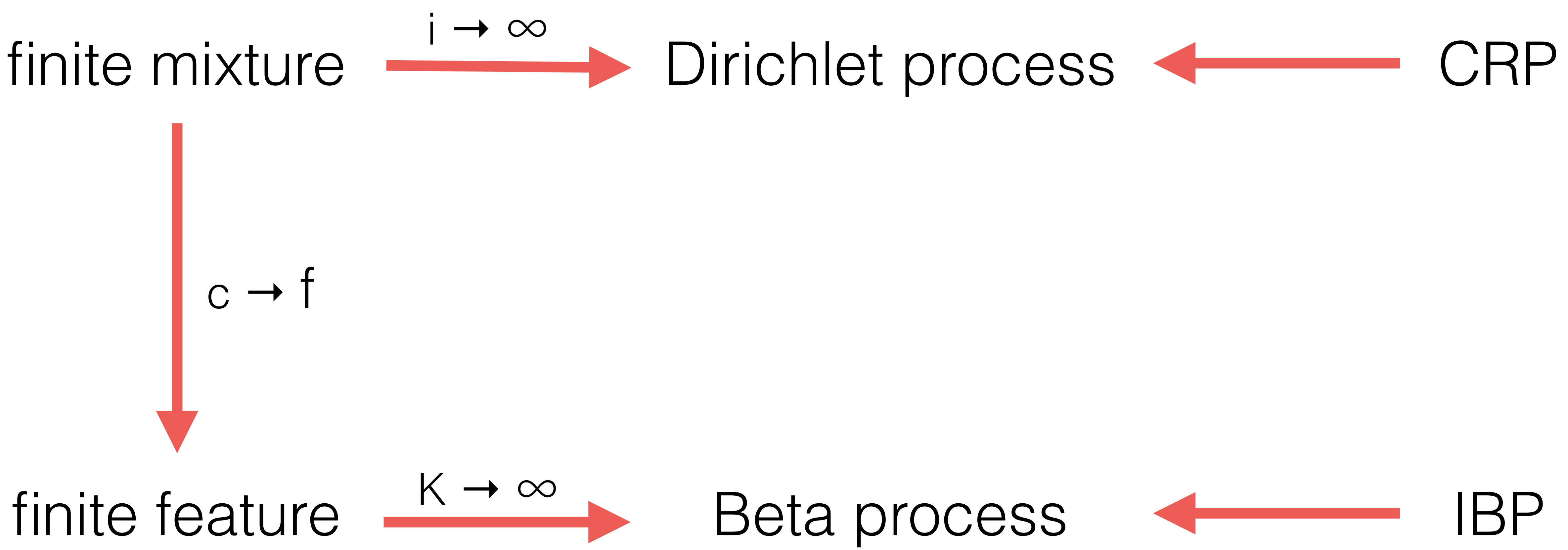
Infinite Latent Feature Models and the Indian Buffet Process

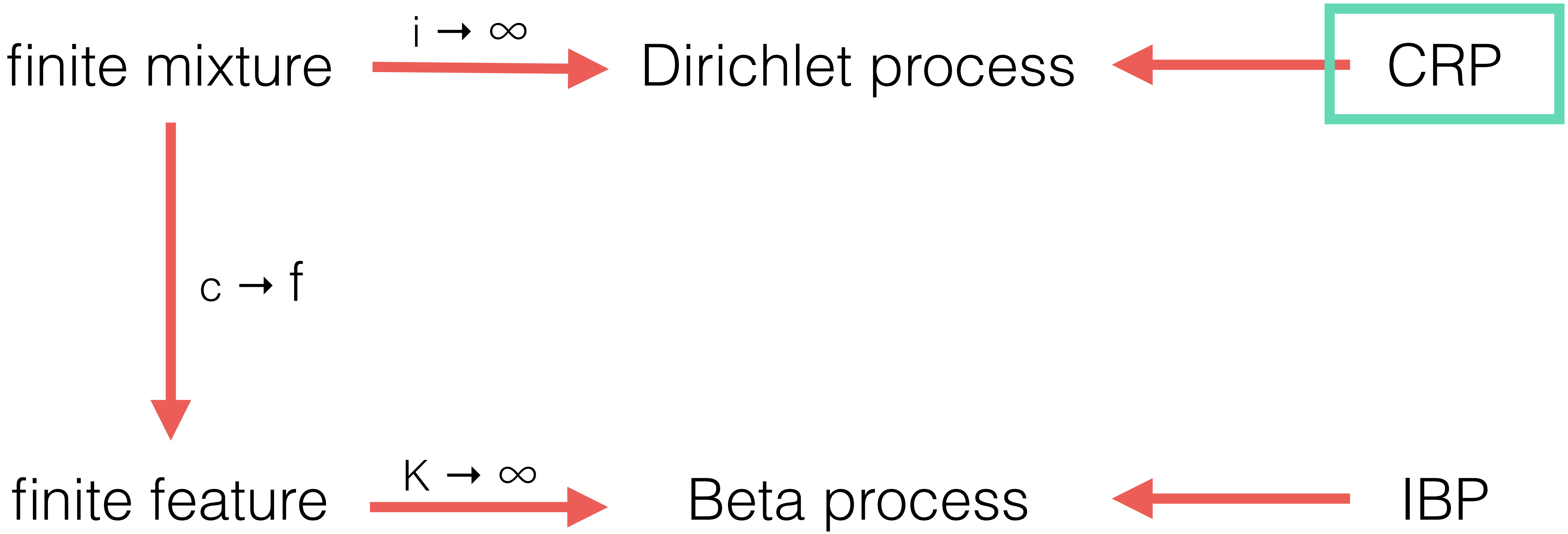
Tutorial on Bayesian nonparametrics

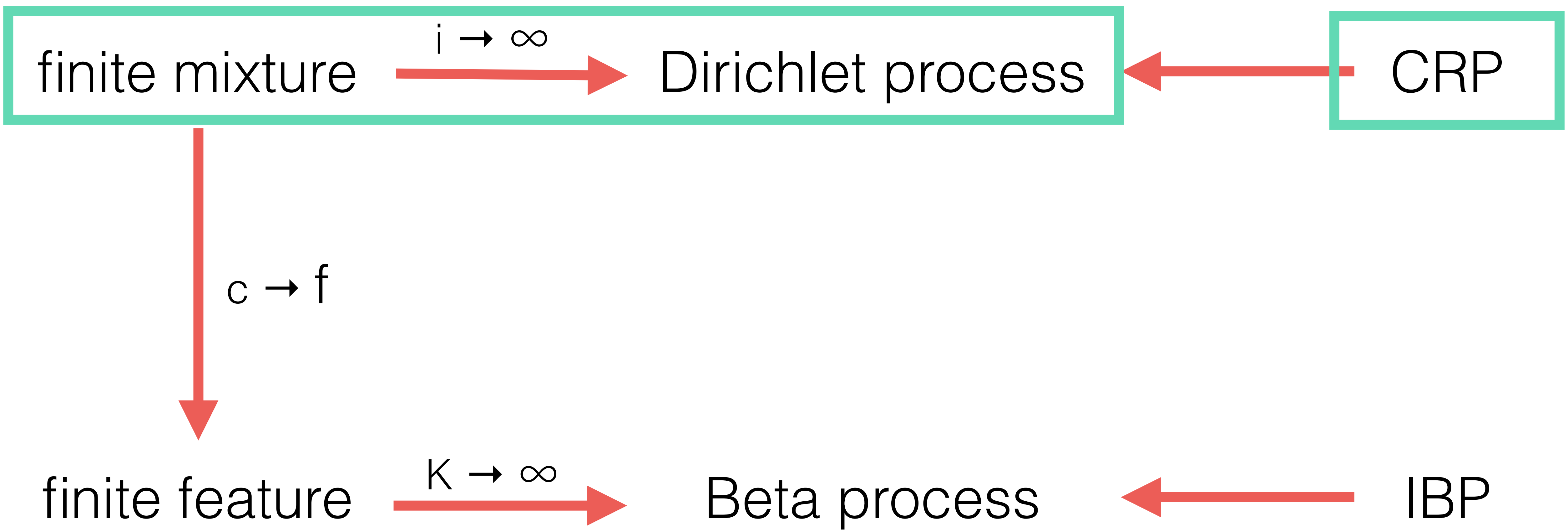
David Nagy

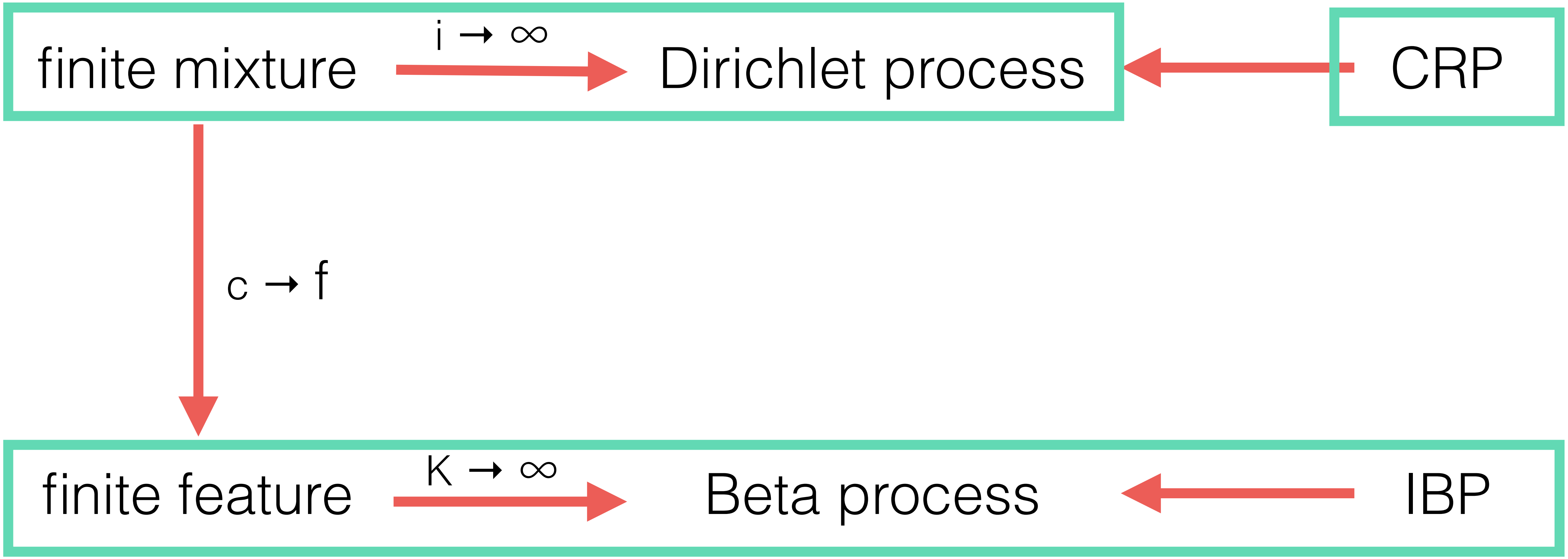
WIGNER RESEARCH CENTRE FOR PHYSICS

Mátrafüred, 2014


finite mixture
i → ∞
Dirichlet process
CRP
c → f
finite feature
K → ∞
Beta process
IBP


finite mixture
i → ∞
Dirichlet process
CRP
c → f
finite feature
K → ∞
Beta process
IBP


finite mixture
i → ∞
Dirichlet process
CRP
c → f
finite feature
K → ∞
Beta process
IBP


finite mixture
i → ∞
Dirichlet process
CRP
c → f
finite feature
K → ∞
Beta process
IBP

FMM

$$\mathbf{x}_i \longrightarrow C_i$$

FMM

$$\mathbf{x}_i \longrightarrow c_i$$

FFM

$$\mathbf{x}_i \longrightarrow \mathbf{f}_i$$

FMM

$$\mathbf{x}_i \longrightarrow c_i$$

$$P(\mathbf{X}) <- P(\mathbf{X}|\mathbf{c})\ P(\mathbf{c})$$

FFM

$$\mathbf{x}_i \longrightarrow \mathbf{f}_i$$

$$P(\mathbf{X}) <- P(\mathbf{X}|\mathbf{F})\ P(\mathbf{F})$$

FMM

$\mathbf{x}_i \longrightarrow c_i$

$P(\mathbf{X}) \leftarrow P(\mathbf{X}|\mathbf{c})\, P(\mathbf{c})$

FFM

$\mathbf{x}_i \longrightarrow \mathbf{f}_i$

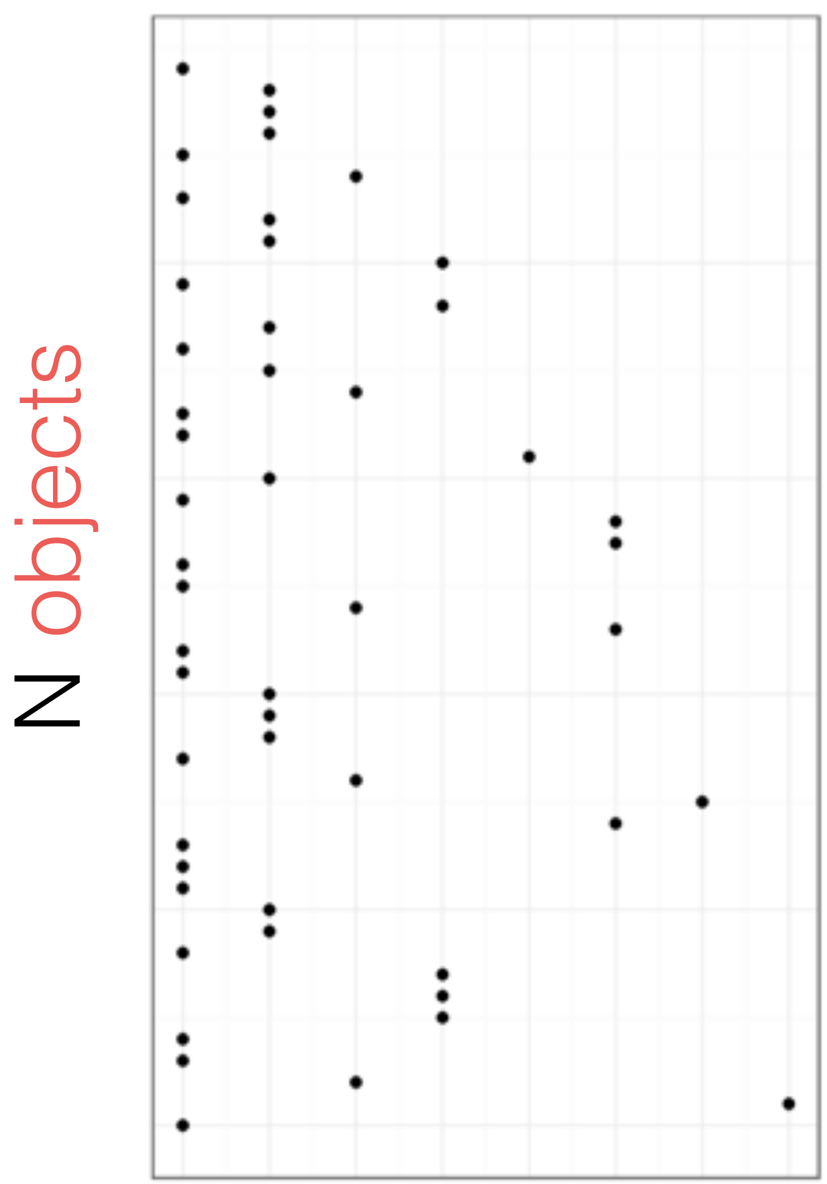
$P(\mathbf{X}) \leftarrow P(\mathbf{X}|\mathbf{F})\, \boxed{P(\mathbf{F})}$

FMM

$\mathbf{x}_i \longrightarrow c_i$

$P(\mathbf{X}) \leftarrow P(\mathbf{X}|\mathbf{c})\, P(\mathbf{c})$

K classes

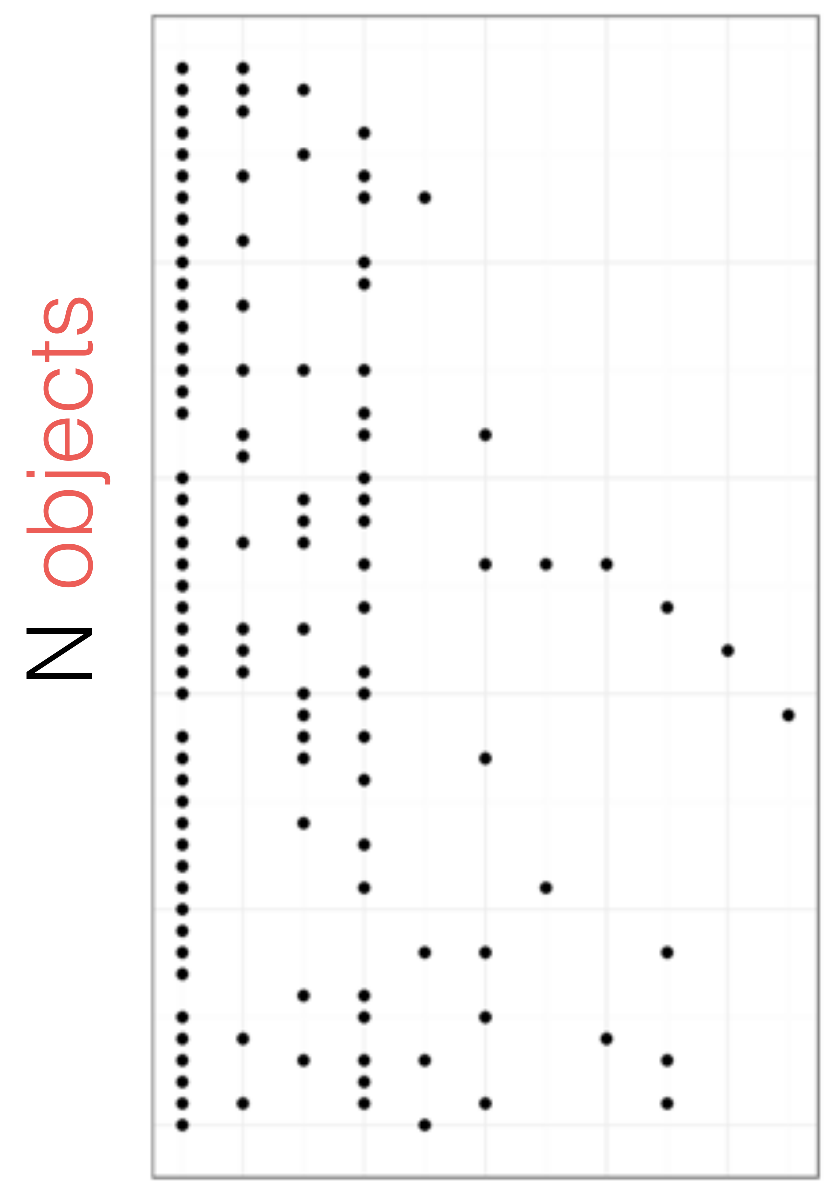


FFM

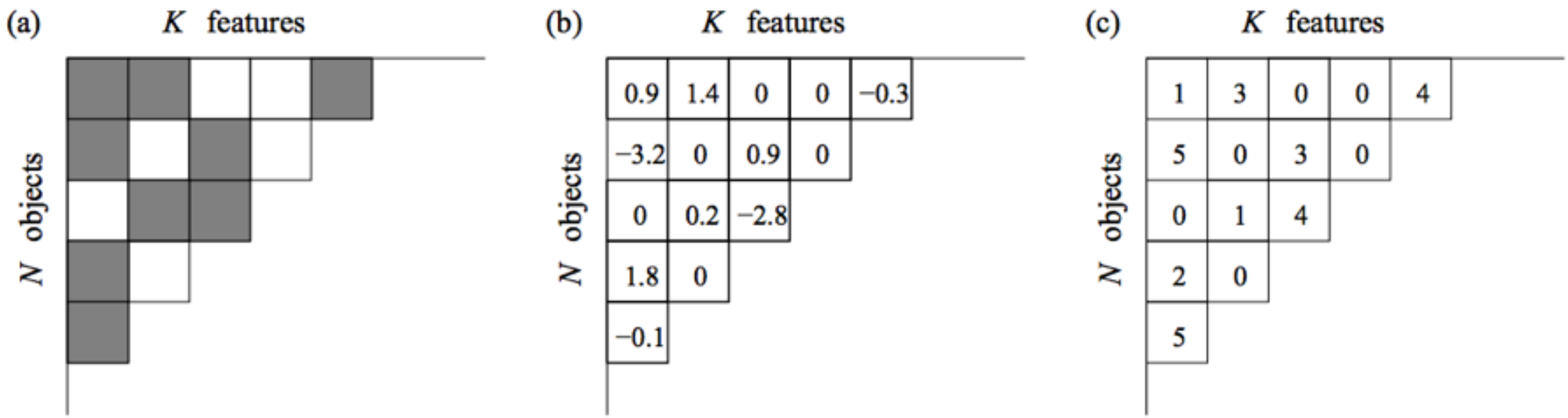
$\mathbf{x}_i \longrightarrow \mathbf{f}_i$

$P(\mathbf{X}) \leftarrow P(\mathbf{X}|\mathbf{F})\, P(\mathbf{F})$

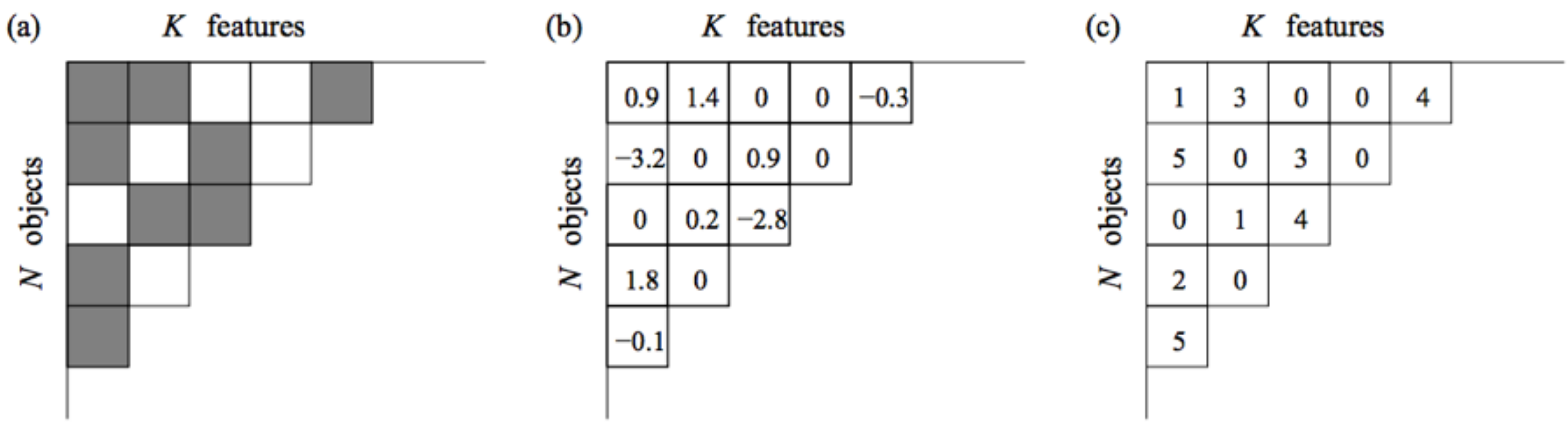
K features



non-binary features

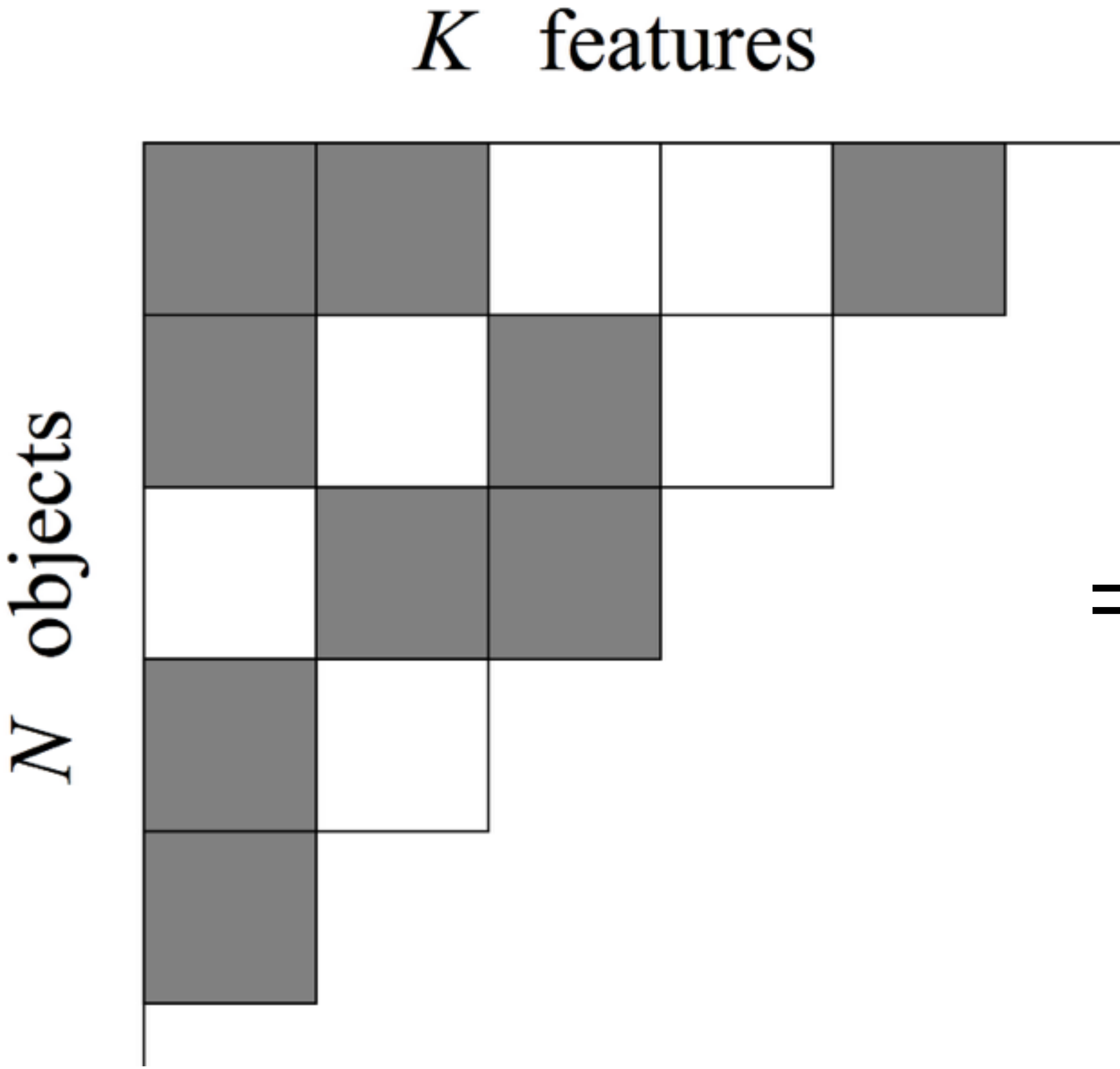


non-binary features

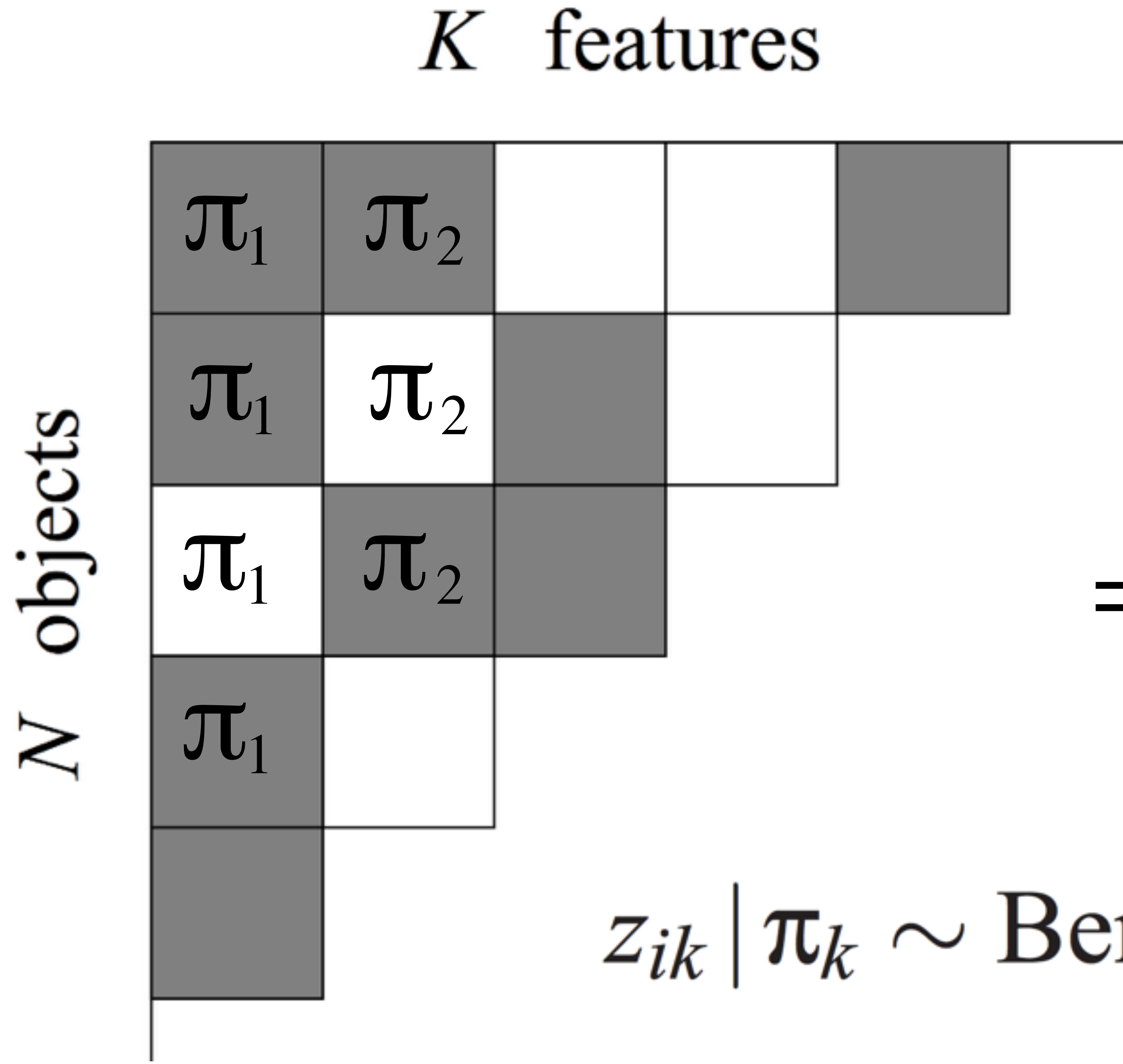


$$\mathbf{F} = \mathbf{Z} \otimes \mathbf{V}$$

$$p(\mathbf{F}) = P(\mathbf{Z})\,p(\mathbf{V})$$

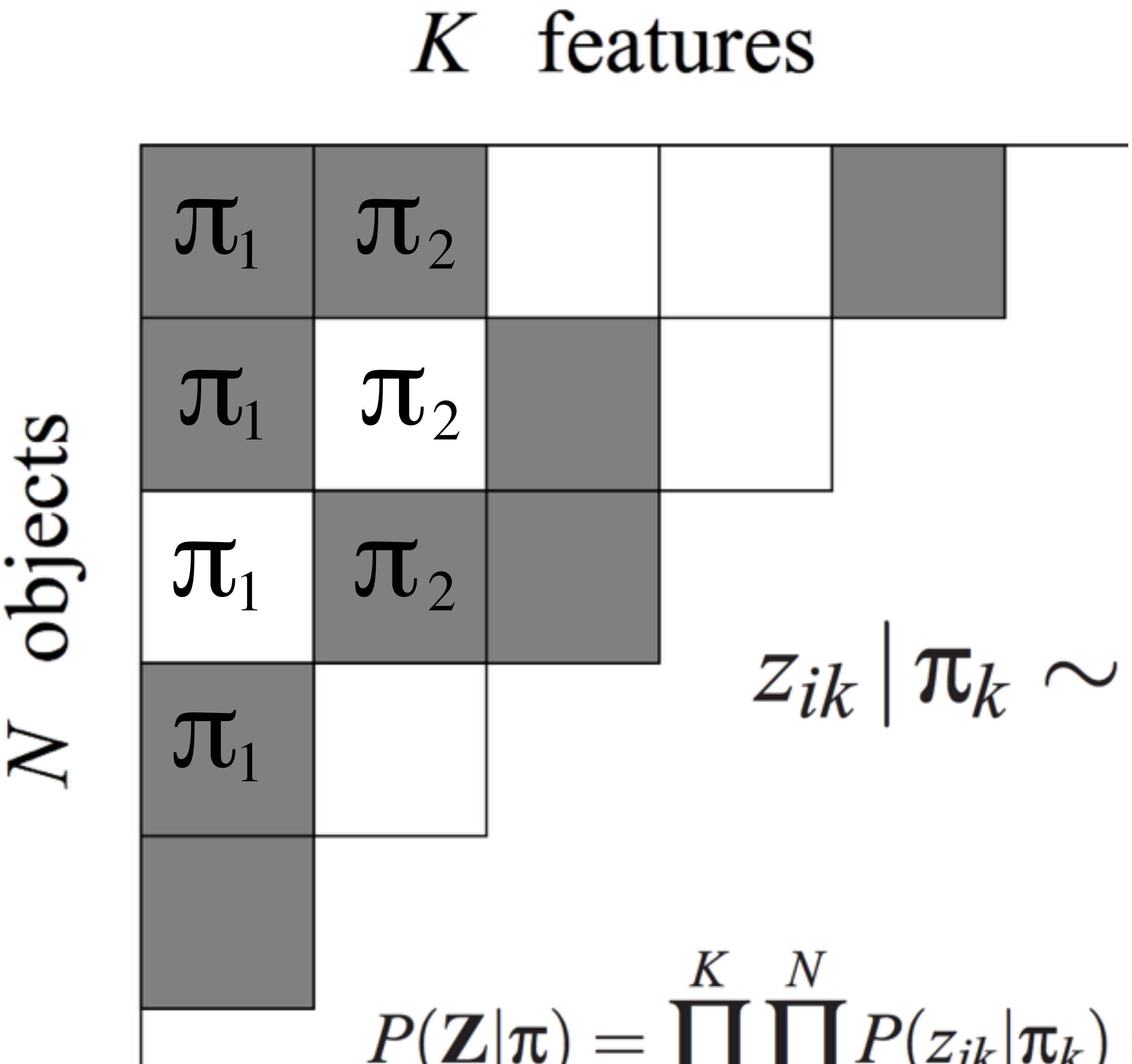


$= \mathbf{Z}$



$= \mathbf{Z}$

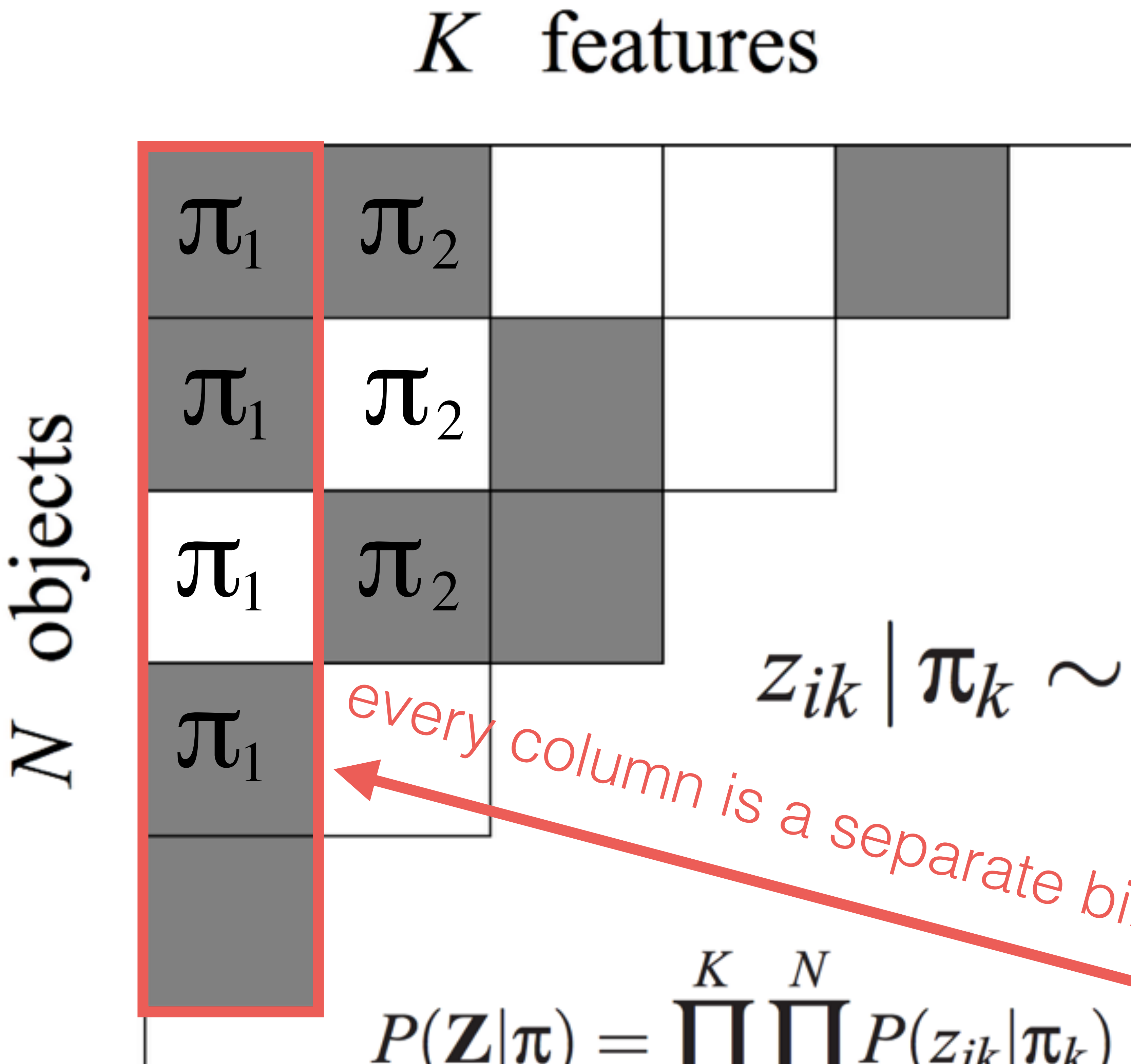
$$z_{ik} \mid \pi_k \sim \text{Bernoulli}(\pi_k)$$



$$= \mathbf{Z}$$

$$z_{ik} \mid \pi_k \sim \text{Bernoulli}(\pi_k)$$

$$P(\mathbf{Z}|\pi) = \prod_{k=1}^{K} \prod_{i=1}^{N} P(z_{ik}|\pi_k) = \prod_{k=1}^{K} \pi_k^{m_k}(1-\pi_k)^{N-m_k}$$



$$= \mathbf{Z}$$

$$z_{ik} \mid \pi_k \sim \text{Bernoulli}(\pi_k)$$

every column is a separate binomial

$$P(\mathbf{Z}|\pi) = \prod_{k=1}^{K}\prod_{i=1}^{N} P(z_{ik}|\pi_k) = \prod_{k=1}^{K} \pi_k^{m_k}(1-\pi_k)^{N-m_k}$$

prior on π : $\quad \boldsymbol{\pi}_k \,|\, \alpha \sim \mathrm{Beta}(\frac{\alpha}{K}, 1)$

prior on π : $\pi_k \mid \alpha \sim \mathrm{Beta}(\frac{\alpha}{K}, 1)$

$$p(\pi_k) = \frac{\pi_k^{r-1}(1-\pi_k)^{s-1}}{B(r,s)}$$

prior on π : $\pi_k \mid \alpha \sim \text{Beta}(\frac{\alpha}{K}, 1)$

$$p(\pi_k) = \frac{\pi_k^{r-1}(1-\pi_k)^{s-1}}{B(r,s)}$$

beta function

$$B(r,s) = \int_0^1 \pi_k^{r-1}(1-\pi_k)^{s-1}\, d\pi_k = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

prior on π : $\pi_k \mid \alpha \sim \text{Beta}(\frac{\alpha}{K}, 1)$

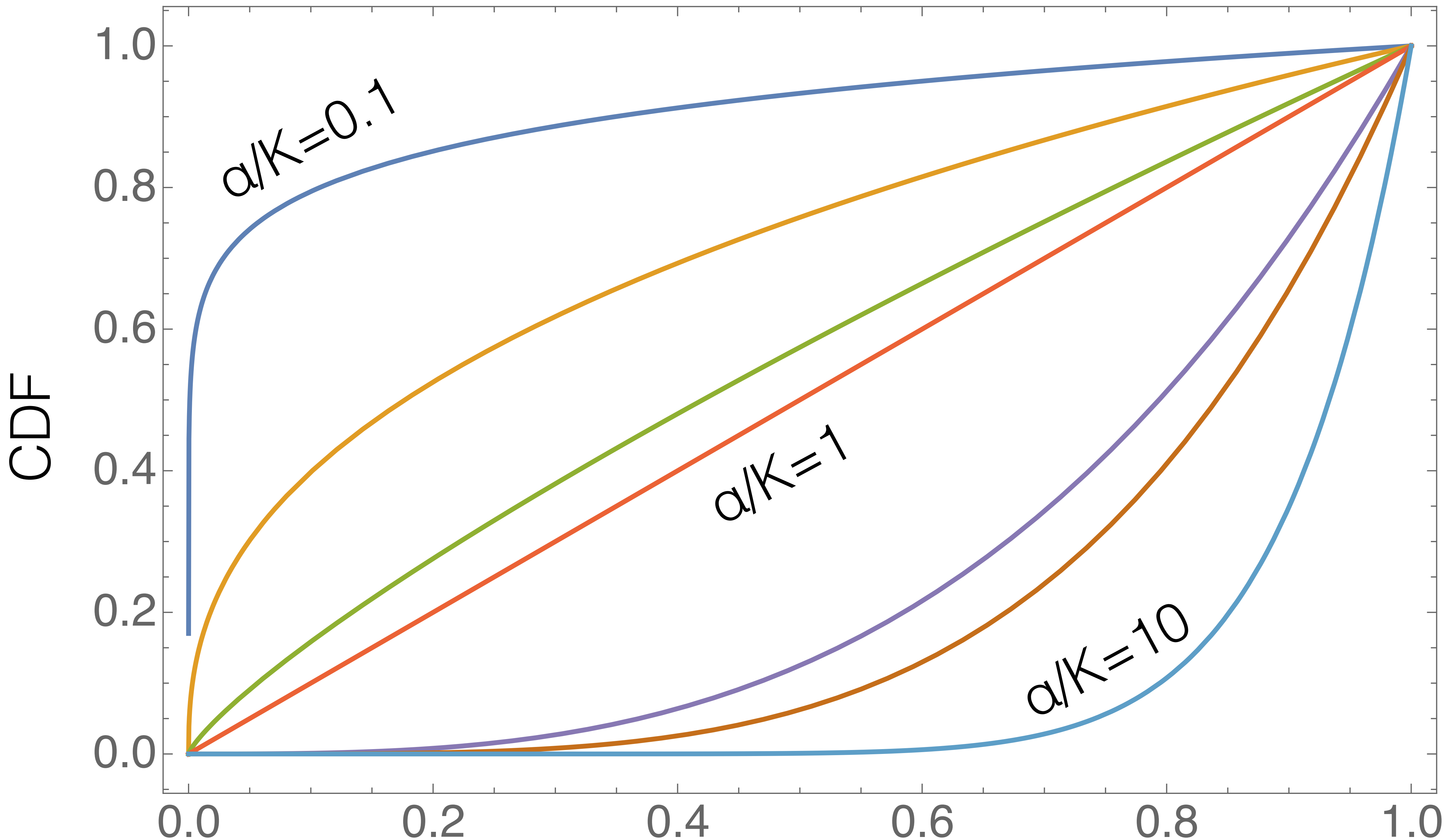
$$p(\pi_k) = \frac{\pi_k^{r-1}(1-\pi_k)^{s-1}}{B(r,s)}$$

beta function

$$B(r,s) = \int_0^1 \pi_k^{r-1}(1-\pi_k)^{s-1}\, d\pi_k = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

$$B(\tfrac{\alpha}{K}, 1) = \frac{\Gamma(\frac{\alpha}{K})}{\Gamma(1+\frac{\alpha}{K})} = \frac{K}{\alpha}$$

$$\pi_k \mid \alpha \sim \mathrm{Beta}\left(\frac{\alpha}{K}, 1\right)$$



$$P(\mathbf{Z}) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik}|\pi_k) \right) p(\pi_k)\, d\pi_k$$

binomial

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik} | \pi_k) \right) p(\pi_k)\, d\pi_k$$

binomial beta

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik} | \pi_k) \right) p(\pi_k) \, d\pi_k$$

marginalize π out | binomial | beta

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik}|\pi_k) \right) p(\pi_k)\, d\pi_k$$

marginalize π out

binomial

beta

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik} | \pi_k) \right) p(\pi_k)\, d\pi_k$$

over columns

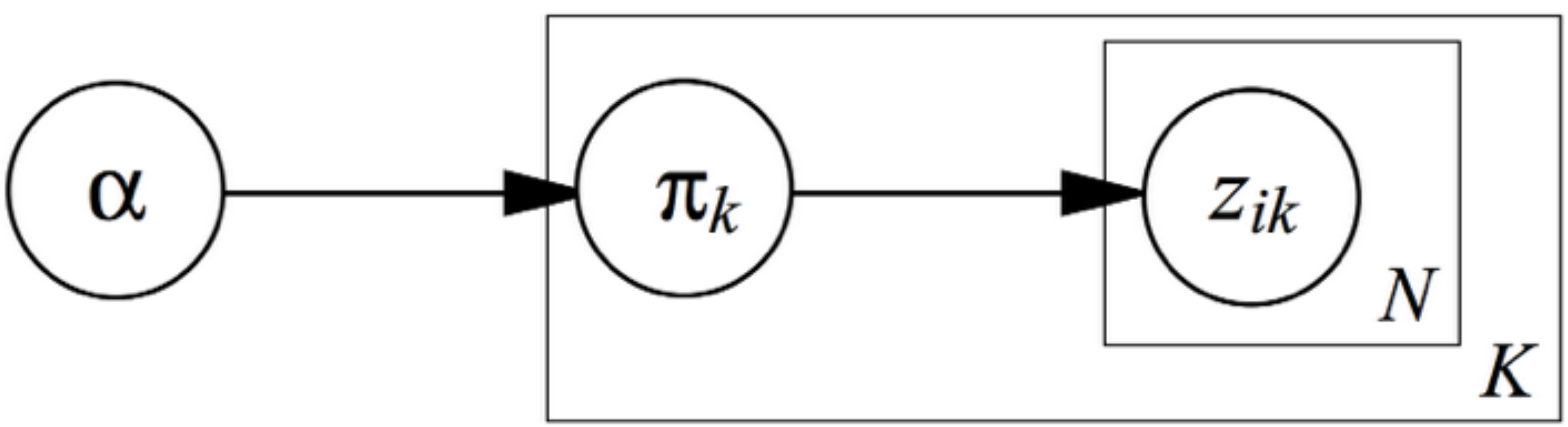
$$
\begin{aligned}
P(\mathbf{Z}) &= \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik}|\pi_k) \right) p(\pi_k)\, d\pi_k \\
&= \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.
\end{aligned}
$$

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik}|\pi_k) \right) p(\pi_k)\, d\pi_k$$

$$= \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.$$

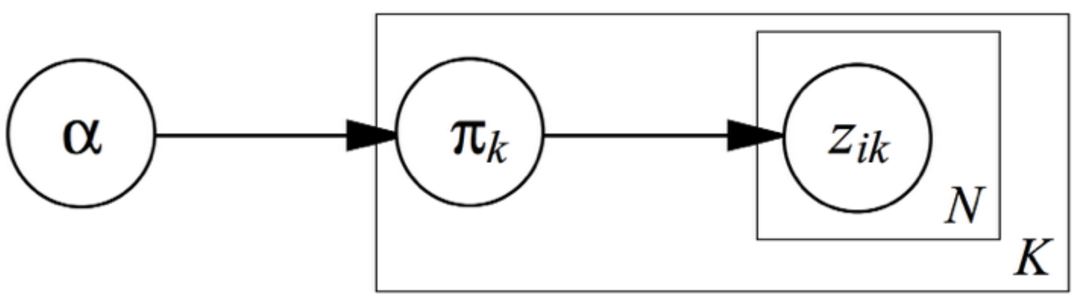
# of objects with feature *k*

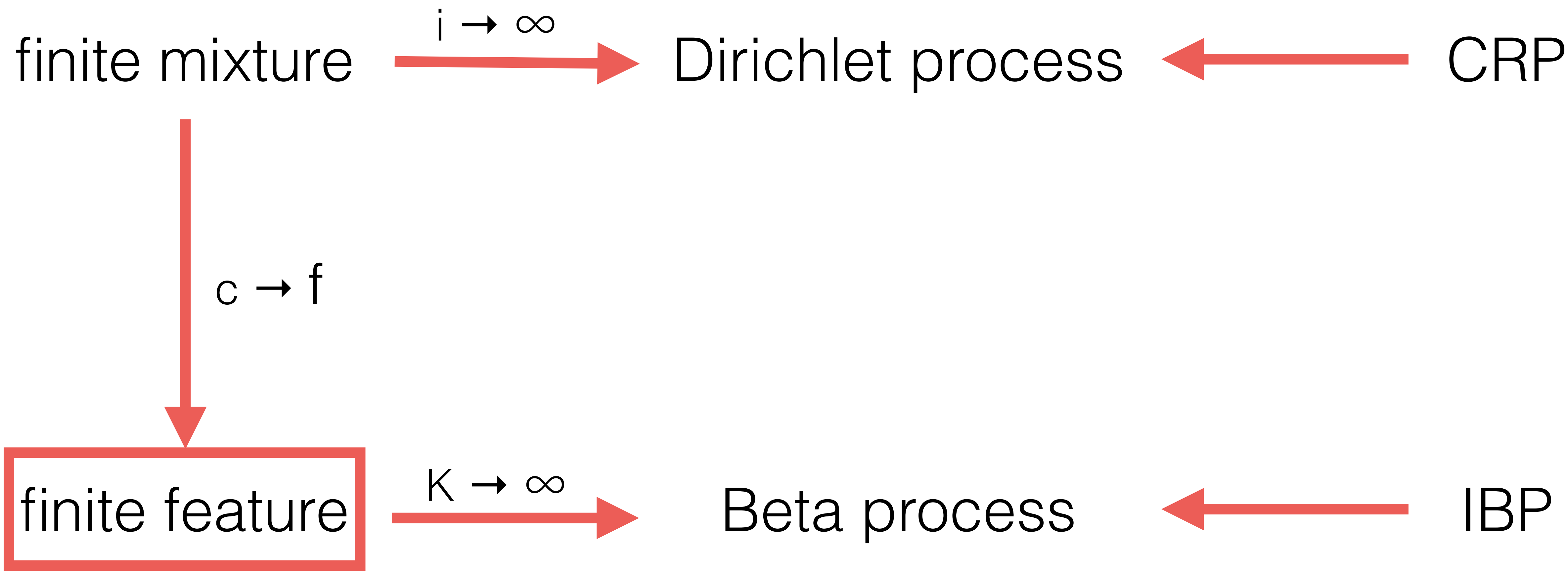
$$
\begin{aligned}
P(\mathbf{Z}) &= \prod_{k=1}^{K} \int \left(\prod_{i=1}^{N} P(z_{ik} | \pi_k) \right) p(\pi_k)\, d\pi_k \\
&= \prod_{k=1}^{K} \frac{\frac{\alpha}{K} \Gamma(m_k + \frac{\alpha}{K}) \Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.
\end{aligned}
$$



finite feature model

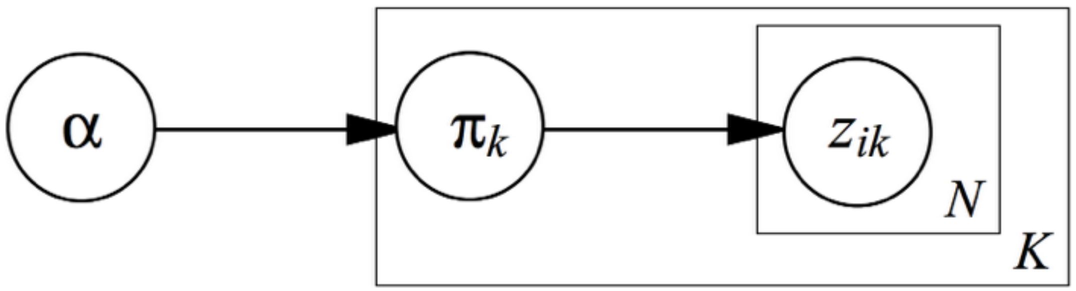
$$P(\mathbf{Z}) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.$$




finite mixture
i → ∞
Dirichlet process
CRP
c → f
finite feature
K → ∞
Beta process
IBP

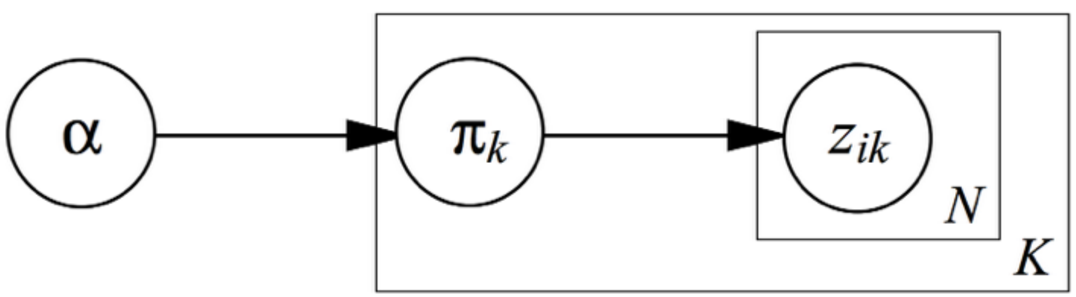
finite feature model

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.$$



finite feature model

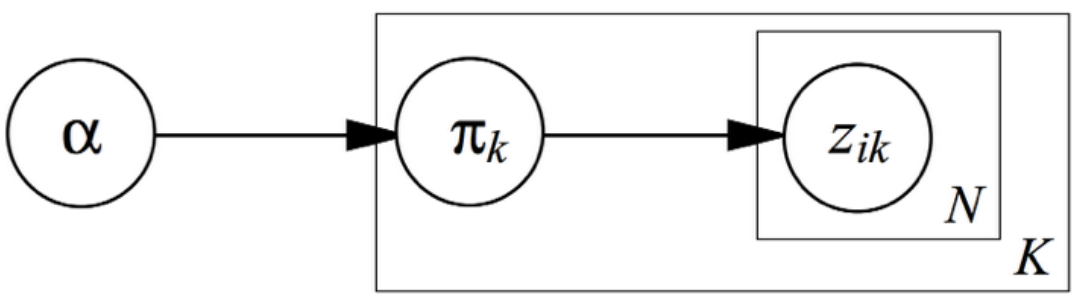
$$P(\mathbf{Z}) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.$$



K → ∞ limit

finite feature model

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.$$

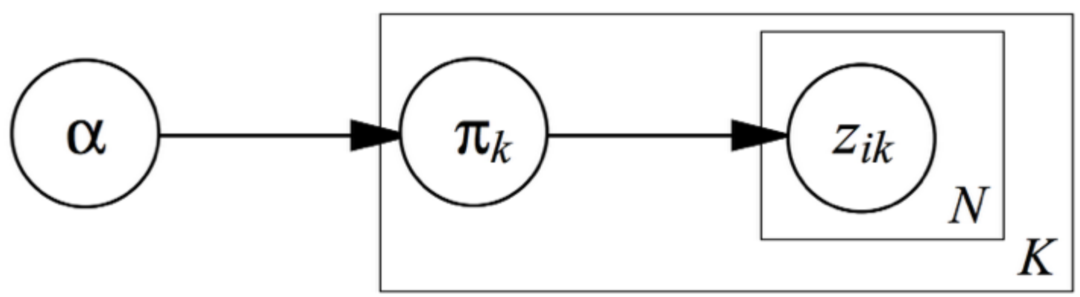


$K \to \infty$ limit

$P(\mathbf{Z}) \to 0$

finite feature model

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}.$$



$K \to \infty$ limit

P(**Z**) → 0

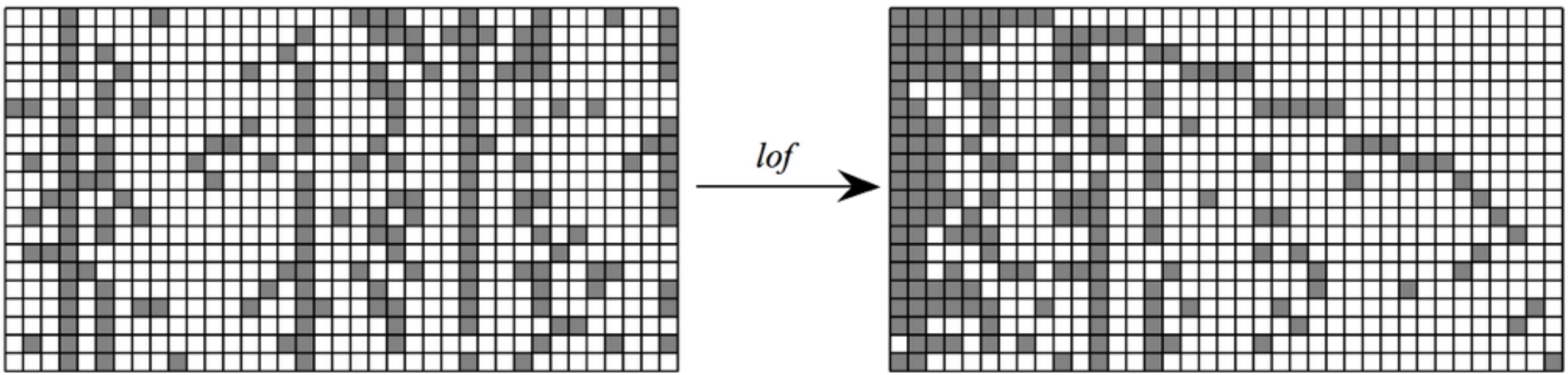
P([**Z**]) instead of P(**Z**)

equivalence class

$$[\mathbf{Z}] = \{\ \forall \mathbf{Y} \mid \mathit{lof}(\mathbf{Y}) = \mathit{lof}(\mathbf{Z})\ \}$$

equivalence class

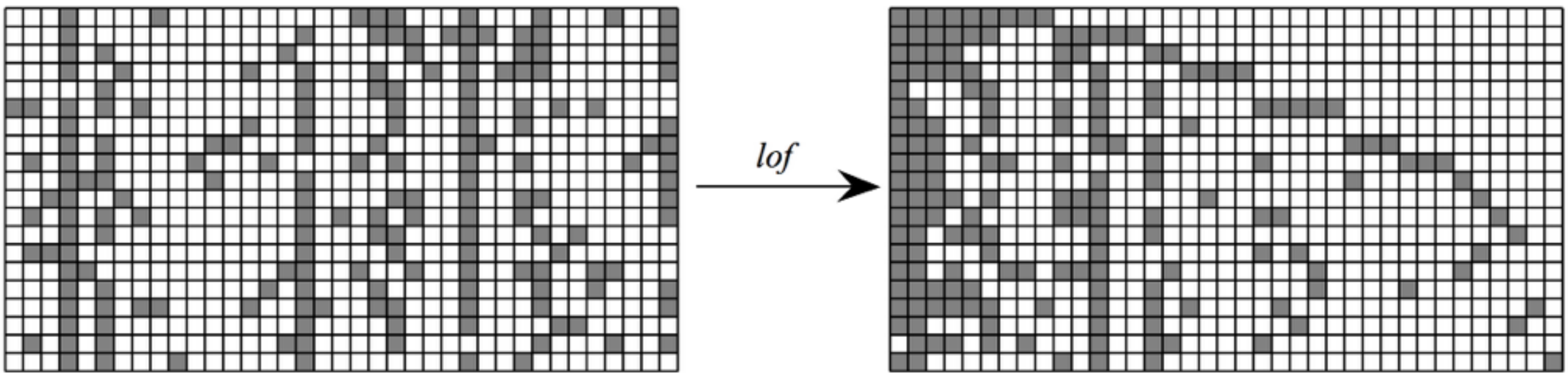
$$[\mathbf{Z}] = \{ \forall \mathbf{Y} \mid lof(\mathbf{Y}) = lof(\mathbf{Z}) \}$$



lof : order columns interpreted as binary numbers by value

equivalence class

$[\mathbf{Z}] = \{ \forall \mathbf{Y} \mid lof(\mathbf{Y}) = lof(\mathbf{Z}) \}$



lof : order columns interpreted as binary numbers by value

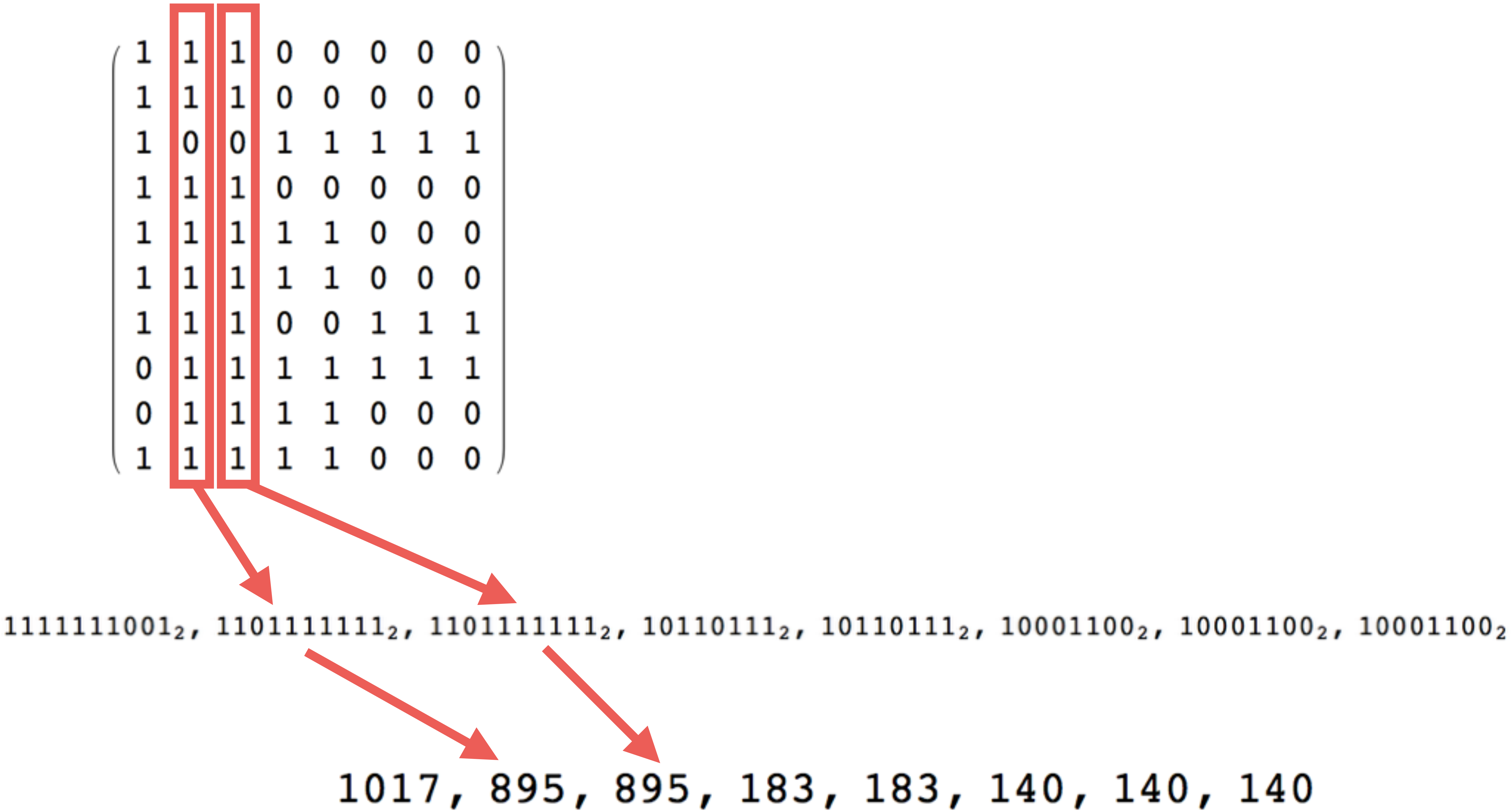
cardinality $\frac{K!}{\prod_{h=0}^{2^N-1} K_h!}$

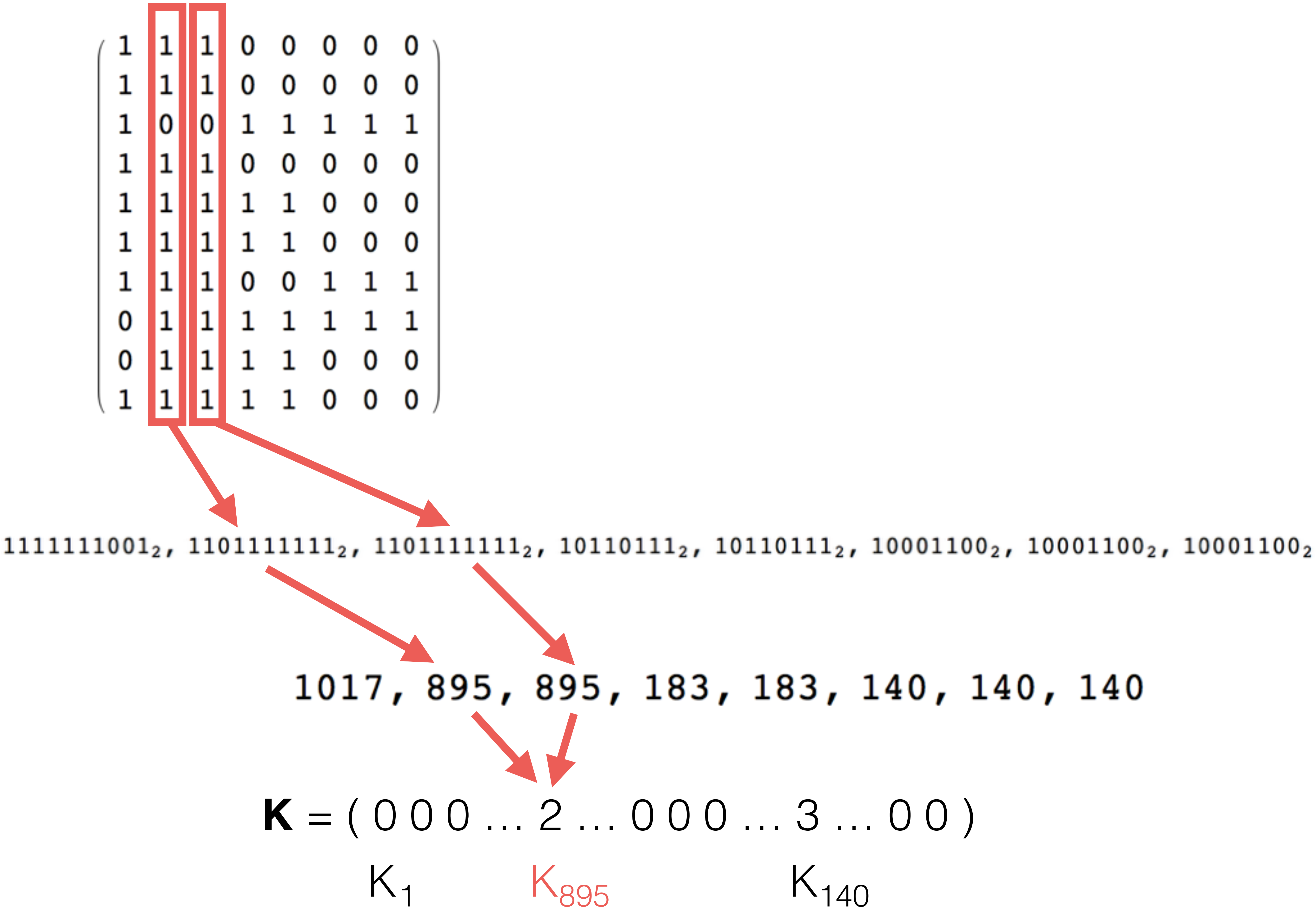
$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

K_h

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$1111111001_2,\ 1101111111_2,\ 1101111111_2,\ 10110111_2,\ 10110111_2,\ 10001100_2,\ 10001100_2,\ 10001100_2$


1 1 1 0 0 0 0 0
1 1 1 0 0 0 0 0
1 0 0 1 1 1 1 1
1 1 1 0 0 0 0 0
1 1 1 1 1 0 0 0
1 1 1 1 1 0 0 0
1 1 1 0 0 1 1 1
0 1 1 1 1 1 1 1
0 1 1 1 1 0 0 0
1 1 1 1 1 0 0 0
1111111001₂, 1101111111₂, 1101111111₂, 10110111₂, 10110111₂, 10001100₂, 10001100₂, 10001100₂
1017, 895, 895, 183, 183, 140, 140, 140



1 1 1 0 0 0 0 0
1 1 1 0 0 0 0 0
1 0 0 1 1 1 1 1
1 1 1 0 0 0 0 0
1 1 1 1 1 0 0 0
1 1 1 1 1 0 0 0
1 1 1 0 0 1 1 1
0 1 1 1 1 1 1 1
0 1 1 1 1 0 0 0
1 1 1 1 1 0 0 0
1111111001_2, 1101111111_2, 1101111111_2, 10110111_2, 10110111_2, 10001100_2, 10001100_2, 10001100_2
1017, 895, 895, 183, 183, 140, 140, 140
K = (0 0 0 … 2 … 0 0 0 … 3 … 0 0)
K_1
K_{895}
K_{140}

$$
\begin{aligned}
P([\mathbf{Z}]) &= \sum_{\mathbf{Z}\in[\mathbf{Z}]} P(\mathbf{Z}) \\
&= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k+\frac{\alpha}{K})\Gamma(N-m_k+1)}{\Gamma(N+1+\frac{\alpha}{K})}
\end{aligned}
$$

FFM over equivalence classes

$$\begin{aligned} P([\mathbf{Z}]) &= \sum_{\mathbf{Z} \in [\mathbf{Z}]} P(\mathbf{Z}) \\ &= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})} \end{aligned}$$

FFM over equivalence classes

$$P([\mathbf{Z}]) = \sum_{\mathbf{Z}\in[\mathbf{Z}]} P(\mathbf{Z})$$

$$= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})}$$

$K \rightarrow \infty$

FFM over equivalence classes

$$P([\mathbf{Z}]) = \sum_{\mathbf{Z}\in[\mathbf{Z}]} P(\mathbf{Z})$$

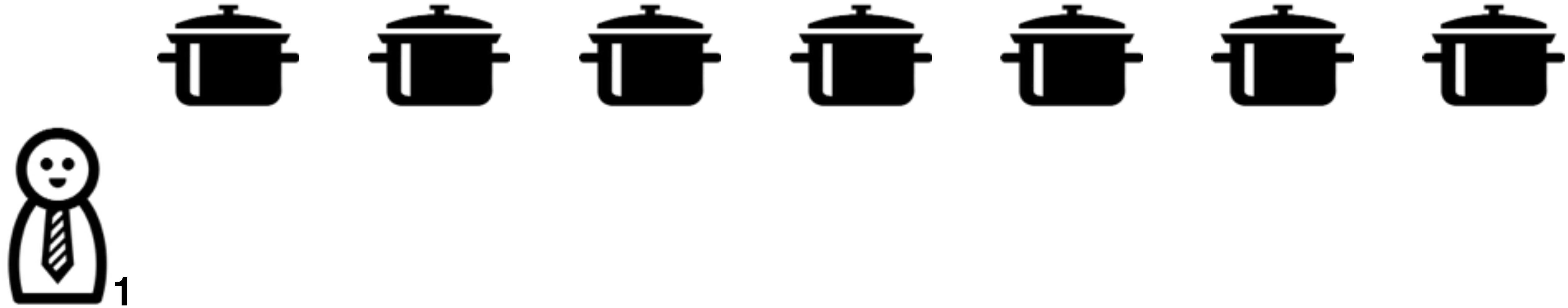
$$= \frac{K!}{\prod_{h=0}^{2^N-1} K_h!} \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k+\frac{\alpha}{K})\Gamma(N-m_k+1)}{\Gamma(N+1+\frac{\alpha}{K})}$$

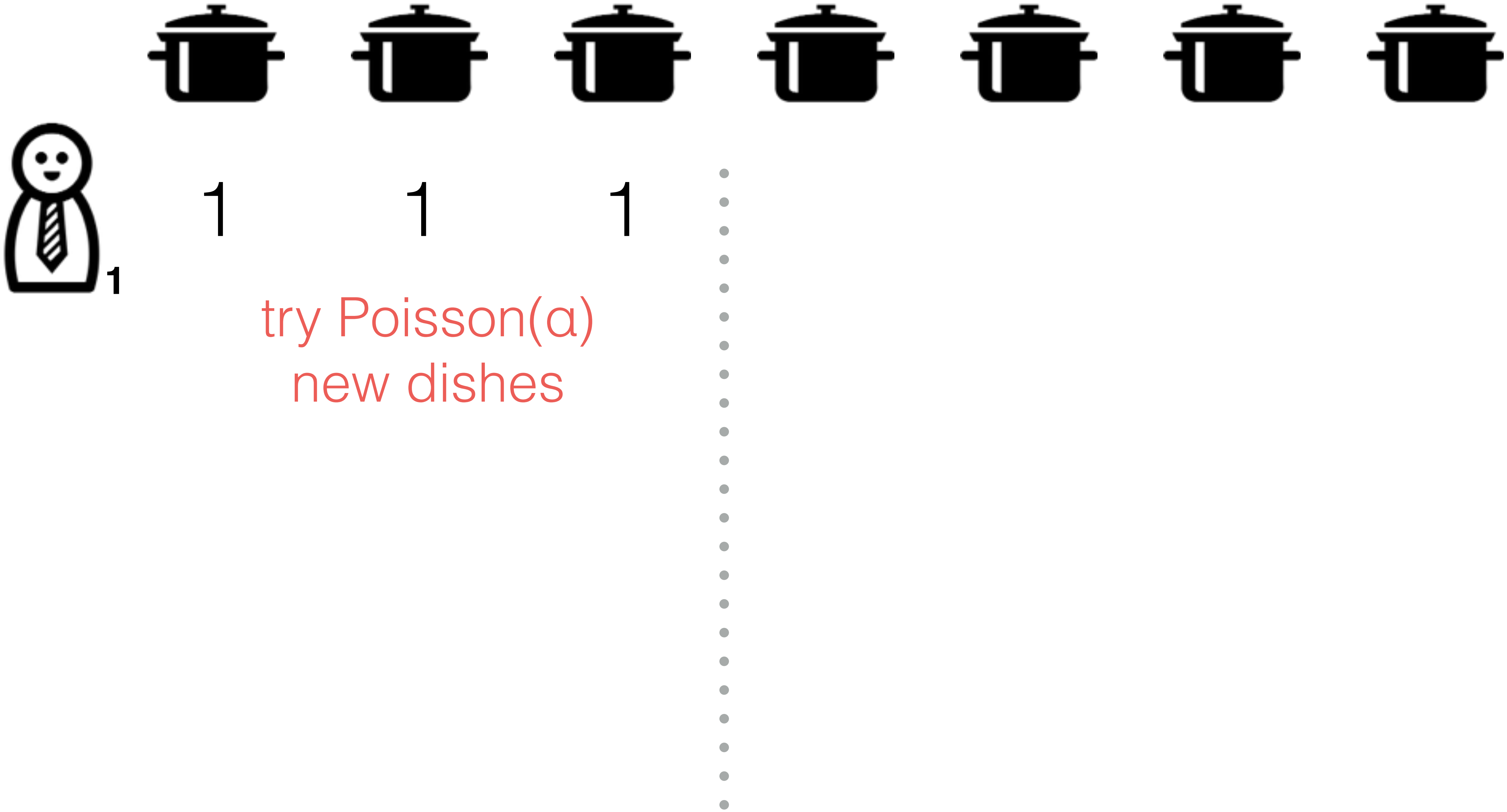
$K \rightarrow \infty$

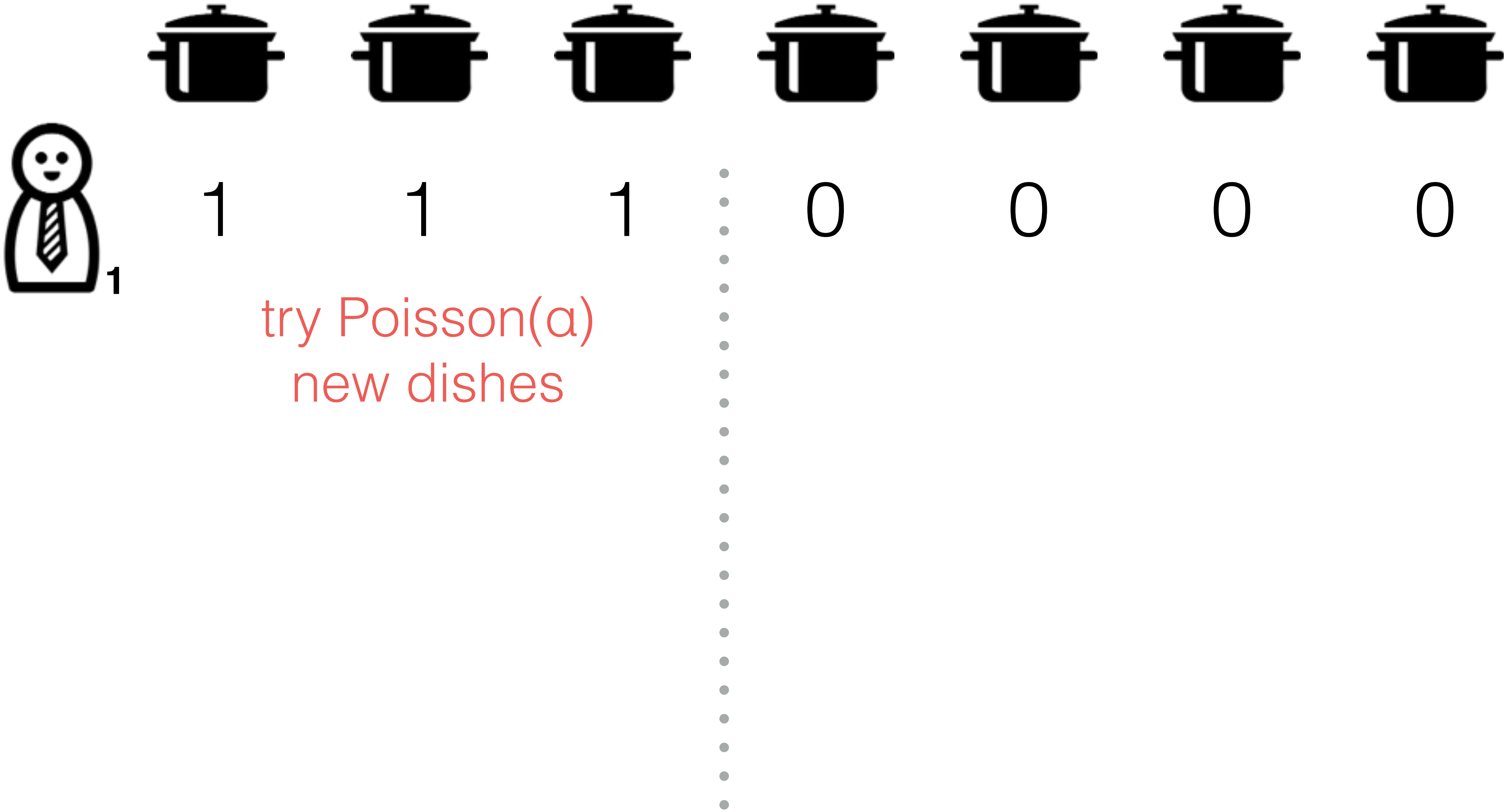
$$\lim_{K\to\infty} P([\mathbf{Z}]) = \frac{\alpha^{K_+}}{\prod_{h=1}^{2^N-1} K_h!} \cdot \exp\{-\alpha H_N\} \cdot \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

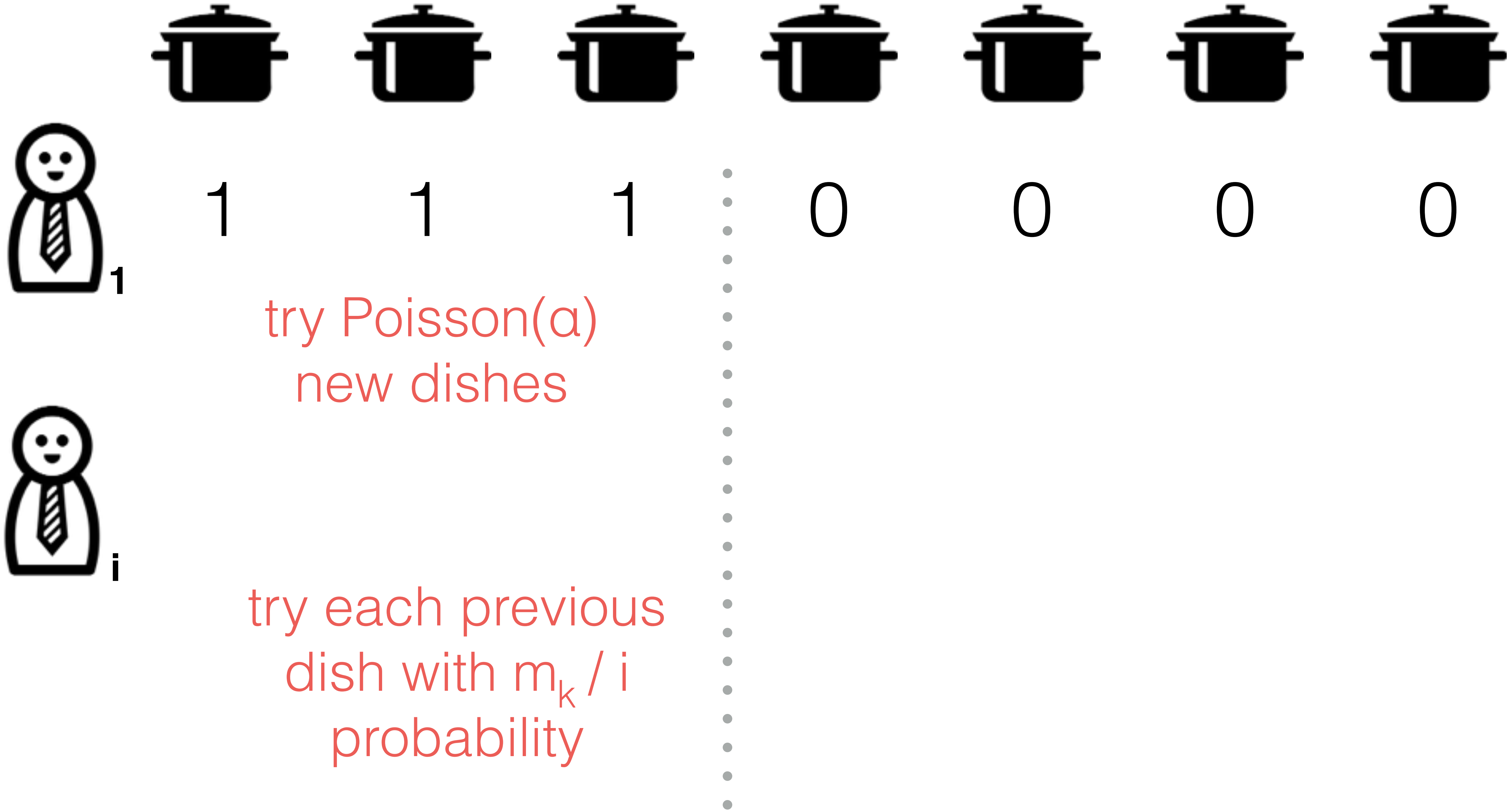
infinite latent feature model

Indian Buffet Process

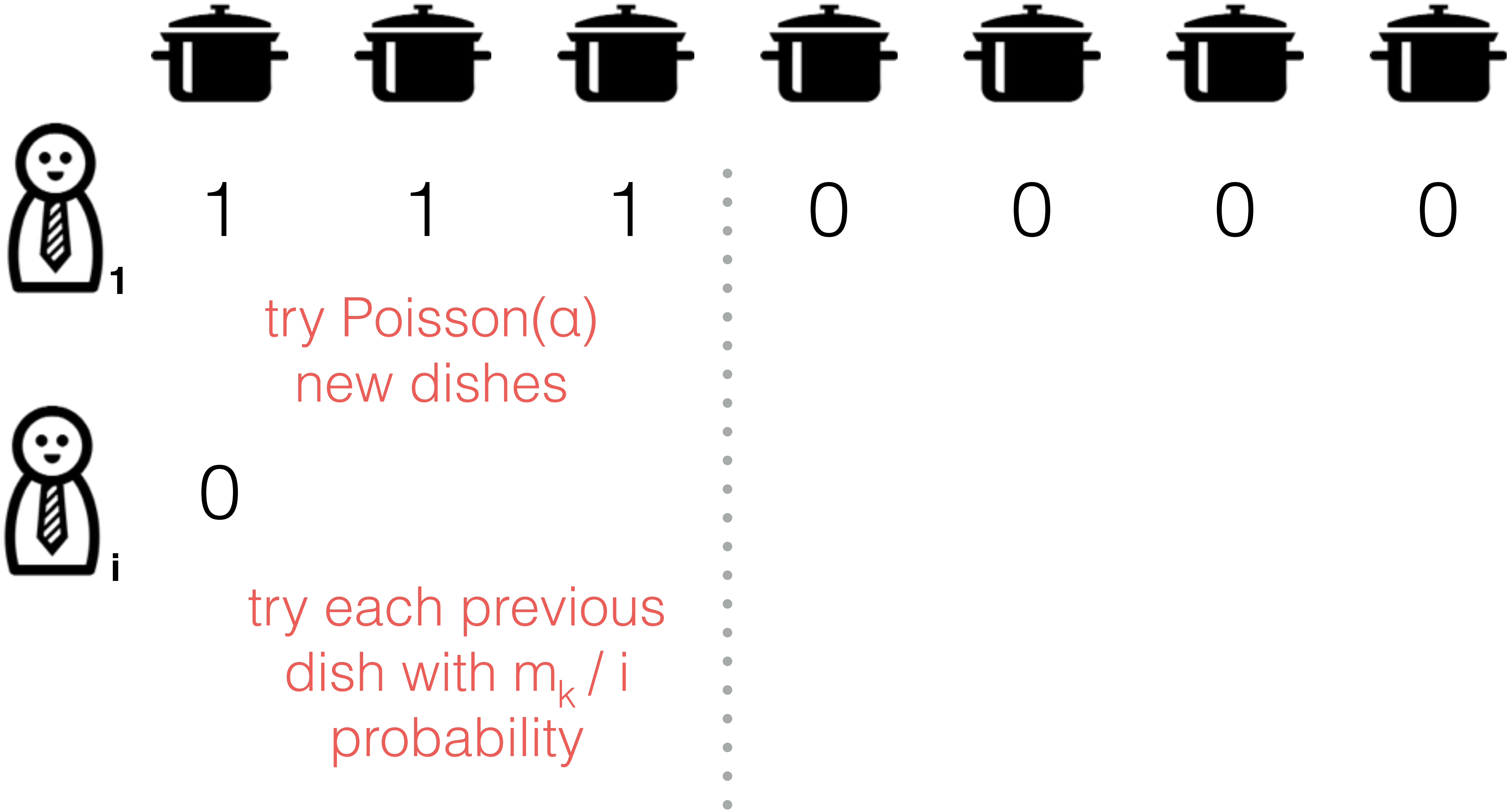

1

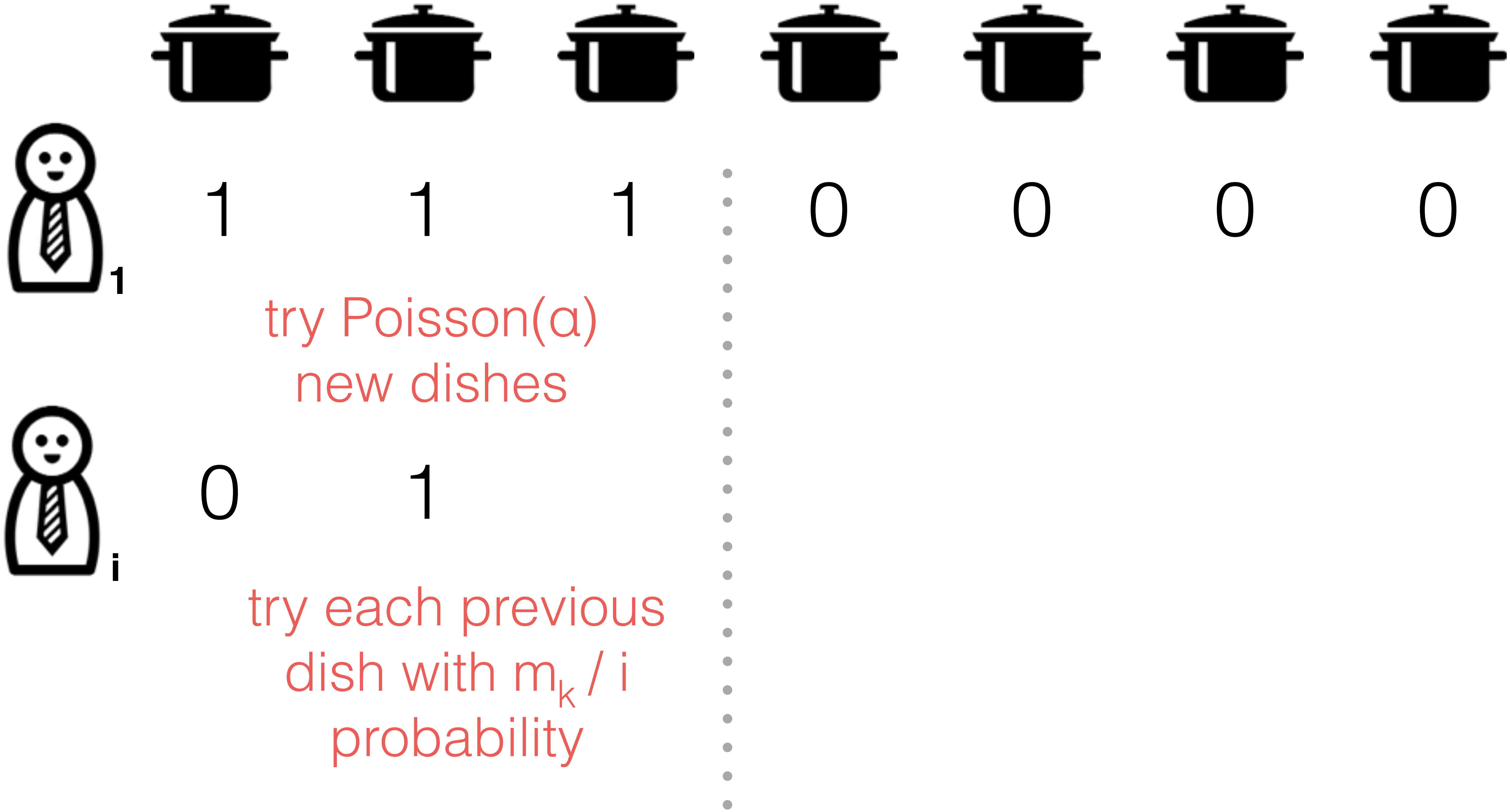

1
1
1
1
try Poisson(α)
new dishes

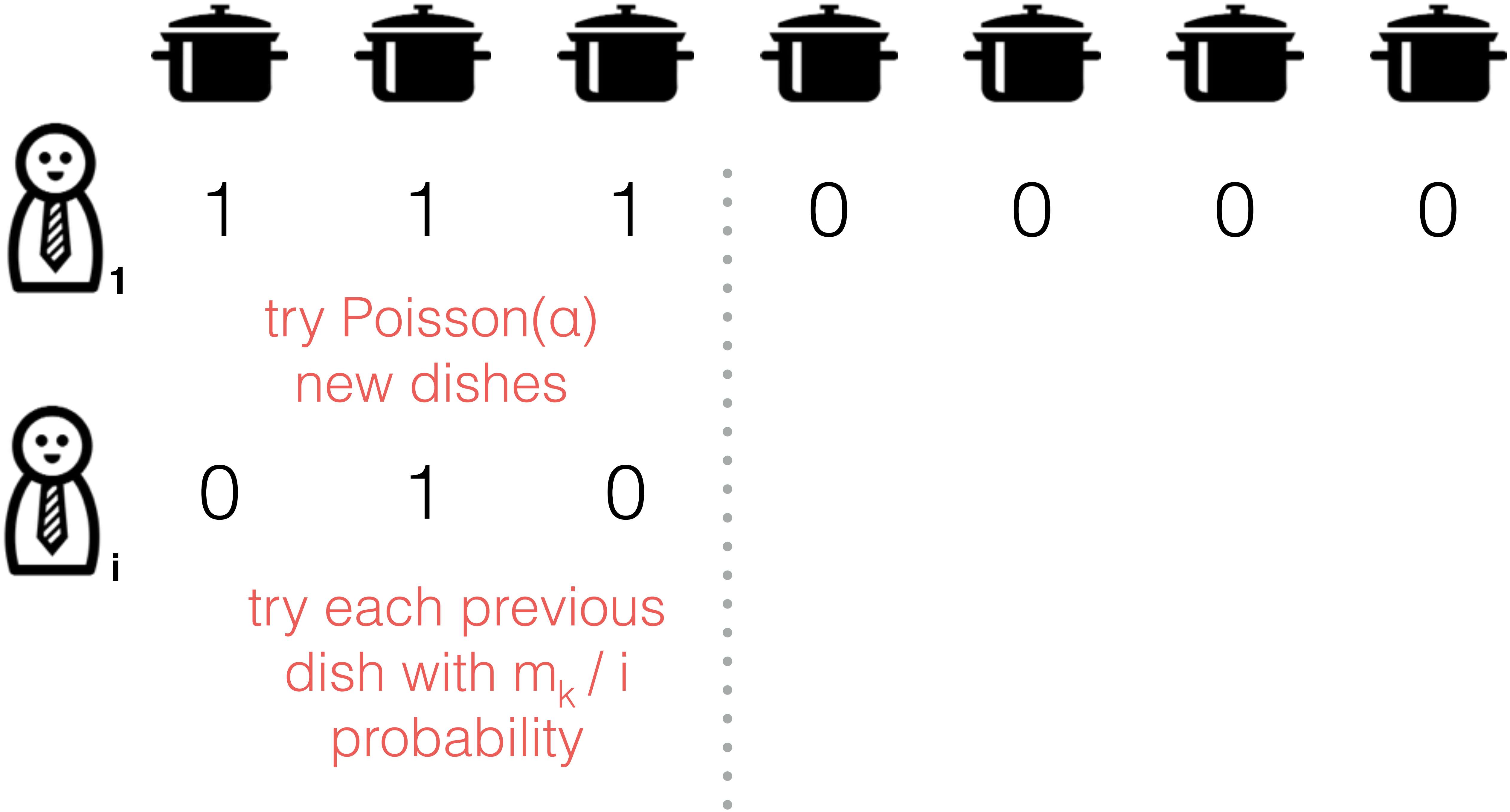

1
1 1 1 0 0 0 0
try Poisson(α)
new dishes

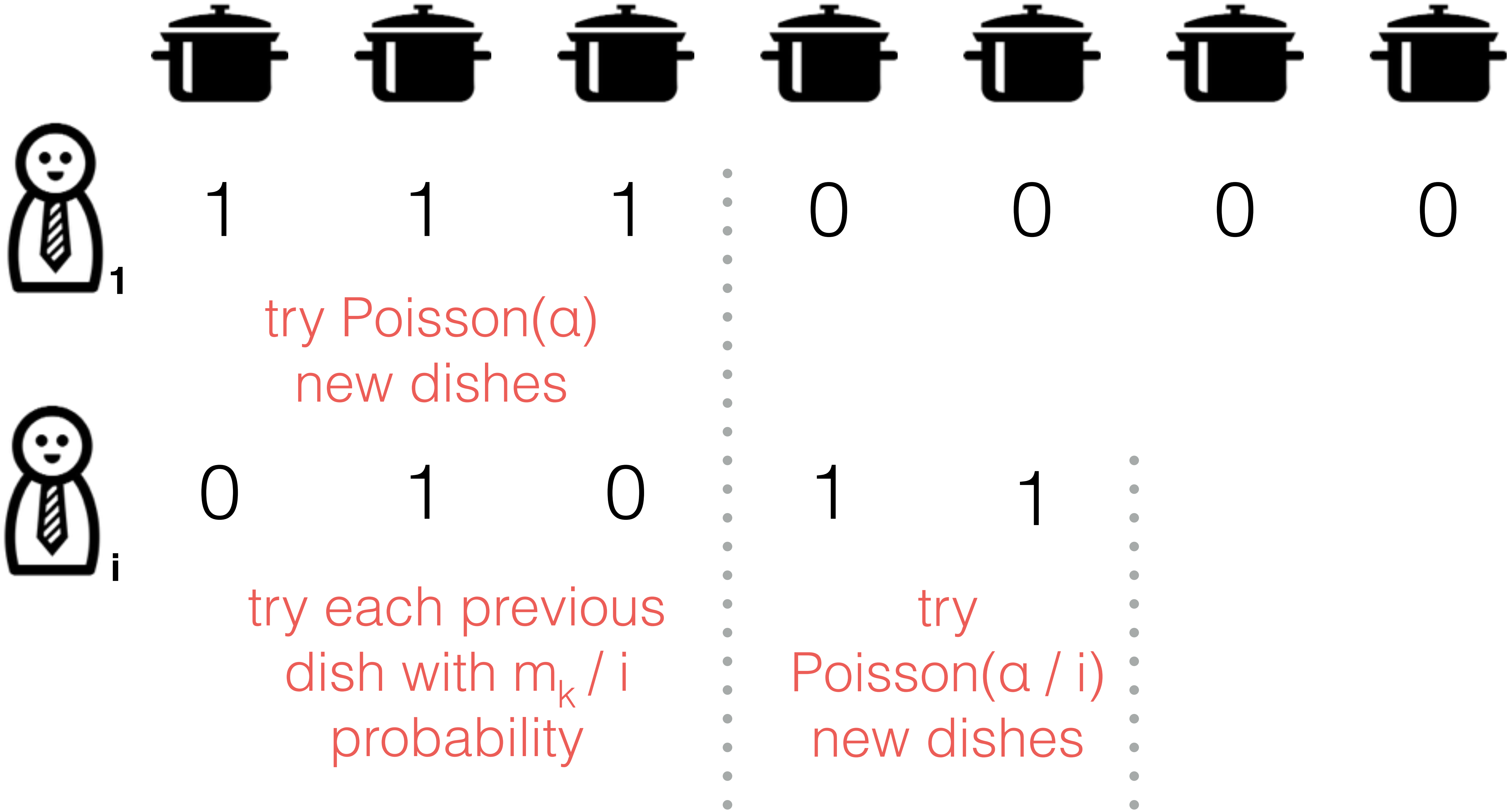


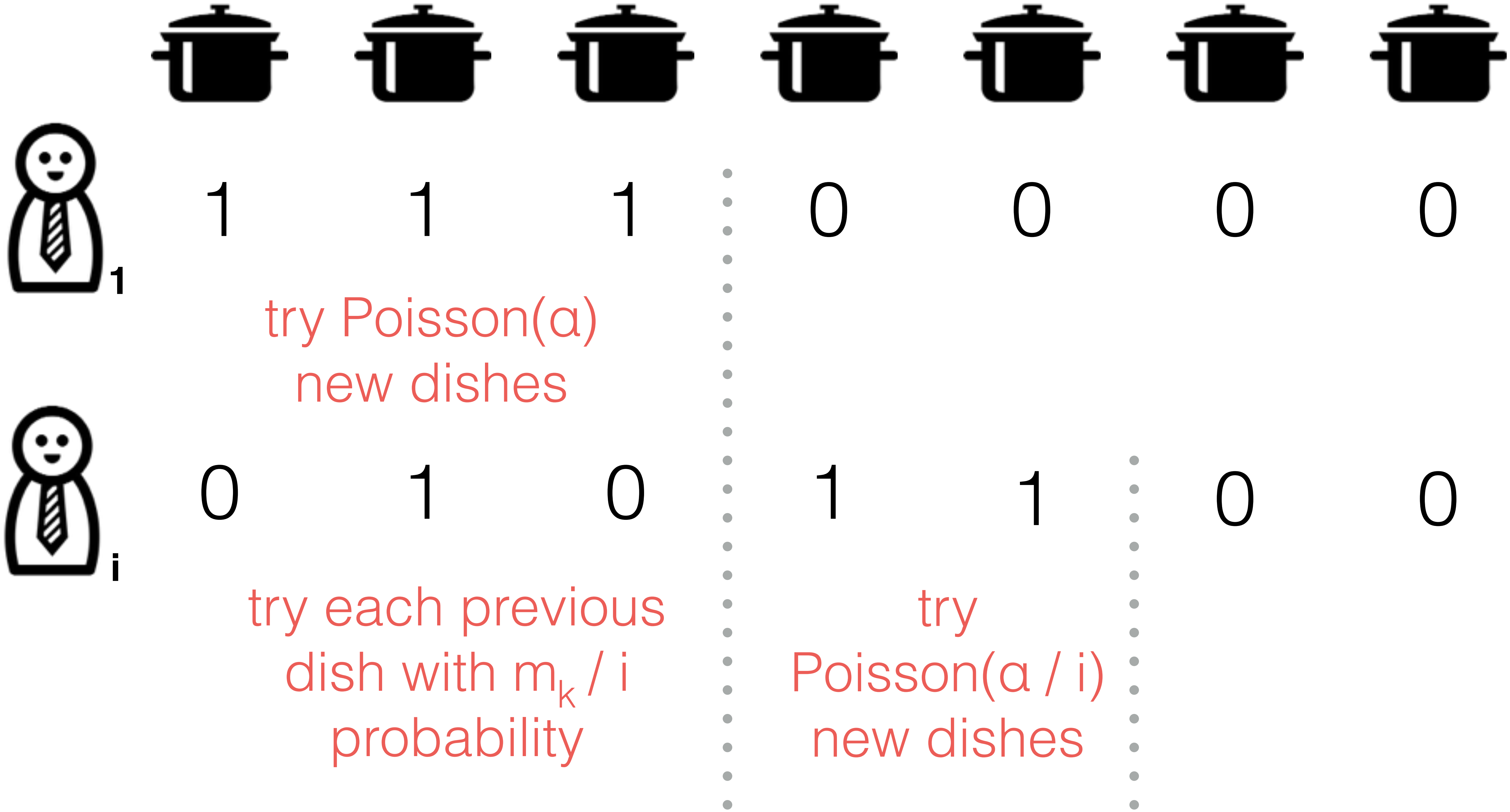
1 1 1 0 0 0 0
1
try Poisson(α) new dishes
i
try each previous dish with m_k / i probability

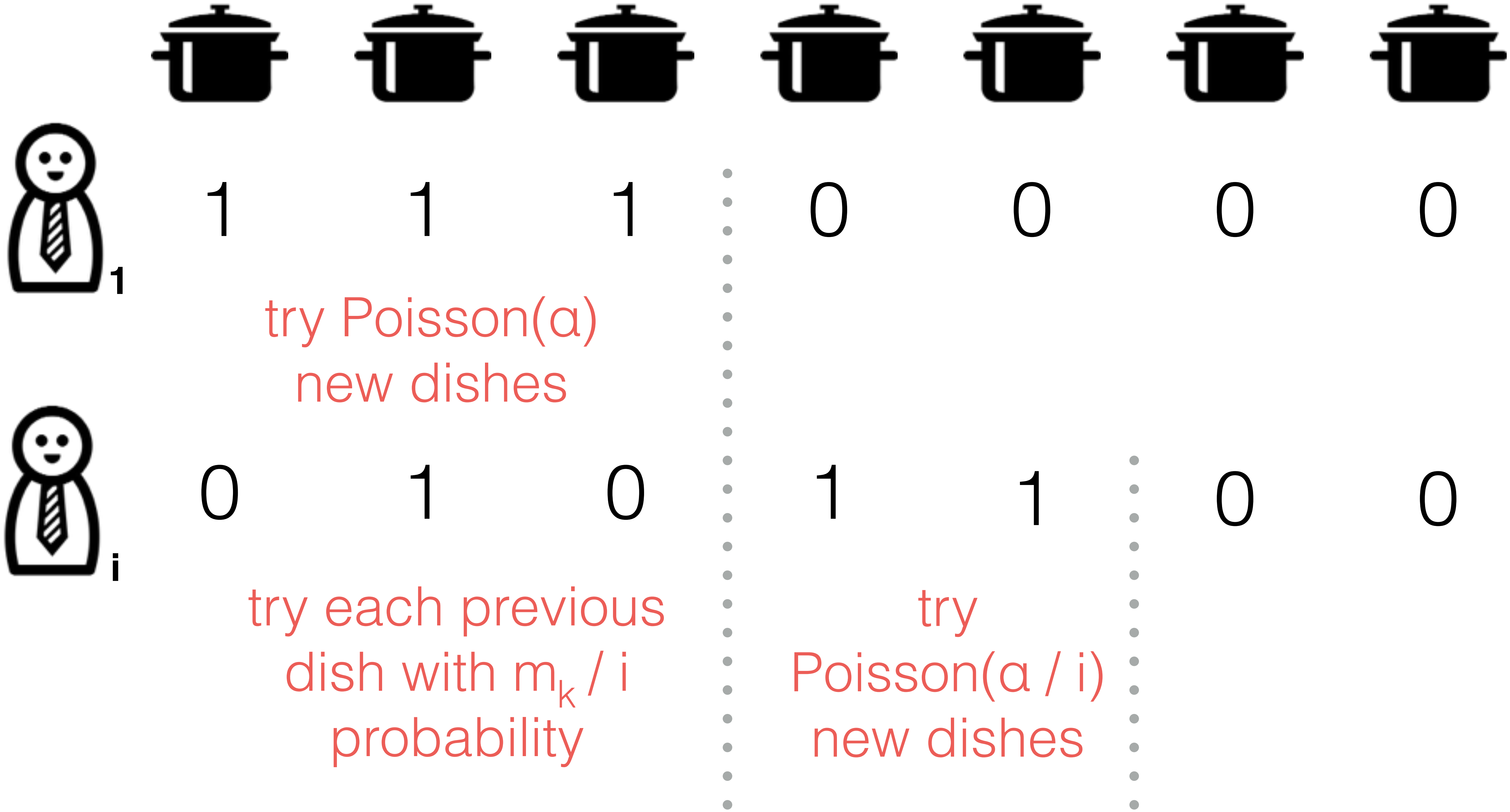

1 1 1 0 0 0 0
1
try Poisson(α) new dishes
0
i
try each previous dish with m_k / i probability


1 1 1 0 0 0 0
1
try Poisson(α)
new dishes
0 1
i
try each previous
dish with m_k / i
probability

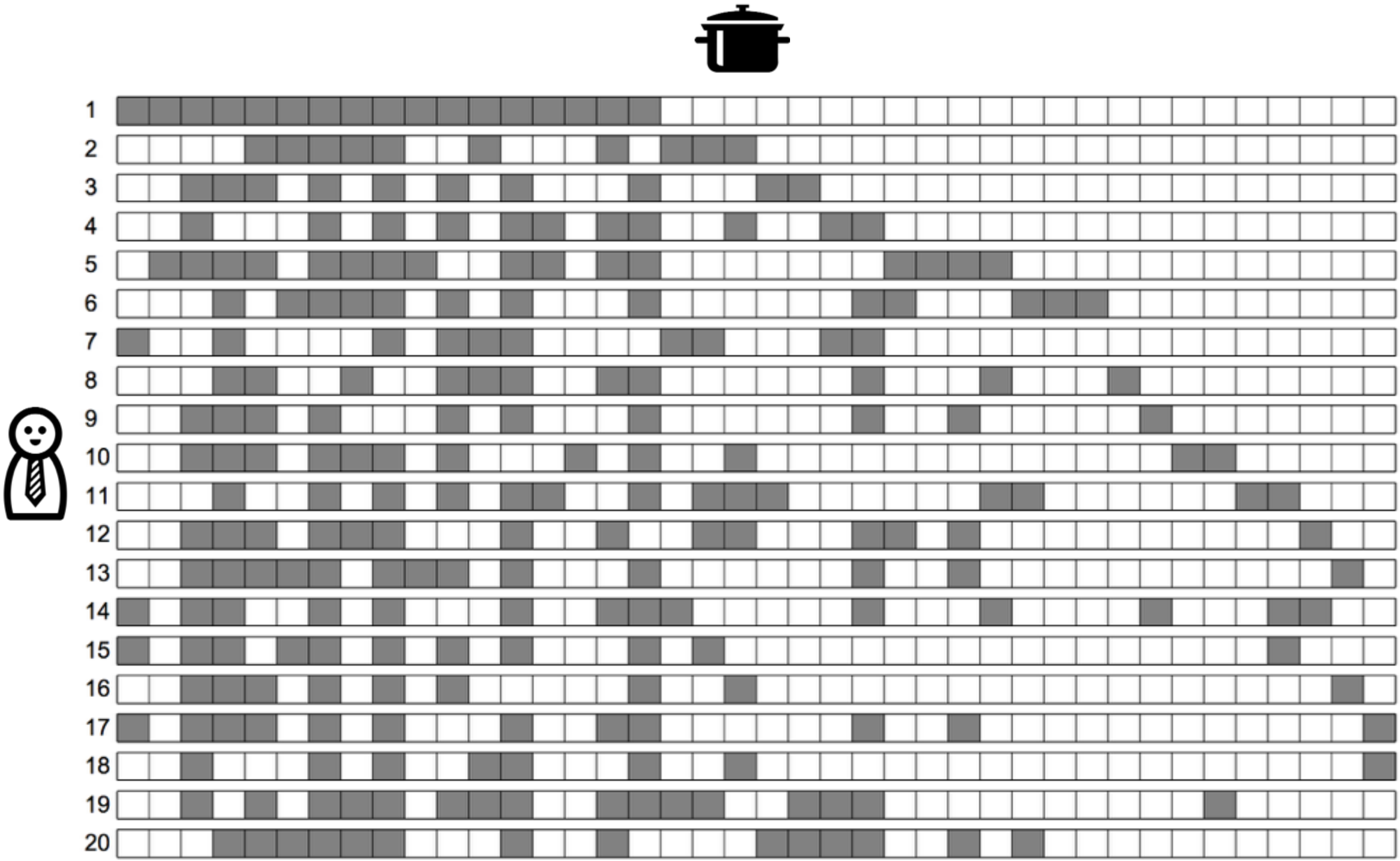

1 1 1 0 0 0 0
1
try Poisson(α)
new dishes
0 1 0
i
try each previous
dish with m_k / i
probability


1
1 1 1 0 0 0 0
try Poisson(α) new dishes
i
0 1 0 1 1
try each previous dish with m_k / i probability
try Poisson(α / i) new dishes


1
1 1 1 0 0 0 0
try Poisson(α)
new dishes
i
0 1 0 1 1 0 0
try each previous
dish with m_k / i
probability
try
Poisson(α / i)
new dishes



$$P(\mathbf{Z}) = \frac{\alpha^{K_+}}{\prod_{i=1}^{N} K_1^{(i)}!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N - m_k)!(m_k - 1)!}{N!}$$



$\alpha = 10$

ILFM

$$\lim_{K \to \infty} P([\mathbf{Z}]) = \frac{\alpha^{K_+}}{\prod_{h=1}^{2^N-1} K_h!} \cdot \exp\{-\alpha H_N\} \cdot \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

IBP

$$P(\mathbf{Z}) = \frac{\alpha^{K_+}}{\prod_{i=1}^{N} K_1^{(i)}!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

ILFM

$$\lim_{K\to\infty} P([\mathbf{Z}]) \quad = \quad \frac{\alpha^{K_+}}{\prod_{h=1}^{2^N-1} K_h!} \cdot \exp\{-\alpha H_N\} \cdot \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

IBP

$$P(\mathbf{Z}) = \frac{\alpha^{K_+}}{\prod_{i=1}^{N} K_1^{(i)}!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

- resulting matrices are not in left ordered form
- customers are not exchangeable

ILFM

$$\lim_{K\to\infty} P([\mathbf{Z}]) = \frac{\alpha^{K_+}}{\prod_{h=1}^{2^N-1} K_h!} \cdot \exp\{-\alpha H_N\} \cdot \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

IBP

$$P(\mathbf{Z}) = \frac{\alpha^{K_+}}{\prod_{i=1}^{N} K_1^{(i)}!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

- resulting matrices are not in left ordered form
- customers are not exchangeable

consider only the equivalence class!

ILFM

$$\lim_{K\to\infty} P([\mathbf{Z}]) = \frac{\alpha^{K_+}}{\prod_{h=1}^{2^N-1} K_h!} \cdot \exp\{-\alpha H_N\} \cdot \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

IBP

$$P(\mathbf{Z}) = \frac{\alpha^{K_+}}{\prod_{i=1}^{N} K_1^{(i)}!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!}$$

- resulting matrices are not in left ordered form
- customers are not exchangeable

consider only the equivalence class!

or, a slightly more complex process can be constructed, called the **exchangeable IBP**

collection of histories
representation

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

lof binary matrices can be encoded as collections of histories

$1111111001_2, 1101111111_2, 1101111111_2, 10110111_2, 10110111_2, 10001100_2, 10001100_2, 10001100_2$

1017, 895, 895, 183, 183, 140, 140, 140

$\mathbf{K} = (\, 0\ 0\ 0 \ldots 2 \ldots 0\ 0\ 0 \ldots 3 \ldots 0\ 0 \,)$

K_1 K_{895} K_{140}

$$\mathbf{K} = (0\ 0\ 0 \ldots 2 \ldots 0\ 0\ 0 \ldots 3 \ldots 0\ 0)$$

K_1 K_{895} K_{140}

$P(\mathbf{K})$ instead of $P([\mathbf{Z}])$

$$\mathbf{K} = (\ 0\ 0\ 0\ \dots\ 2\ \dots\ 0\ 0\ 0\ \dots\ 3\ \dots\ 0\ 0\ )$$

K_1 K_{895} K_{140}

each $K_h \sim \text{Poisson}(\ \alpha B(m_h, N - m_h + 1)\ )$

↑

P(**K**) instead of P([**Z**])

↑

K = (0 0 0 … 2 … 0 0 0 … 3 … 0 0)

K_1 K_{895} K_{140}

# of nonzero elements in history *h*

each K_h ~ Poisson($\alpha B(m_h, N - \overbrace{m_h}^{} + 1)$)

P(**K**) instead of P([**Z**])

$\mathbf{K}$ = (0 0 0 … 2 … 0 0 0 … 3 … 0 0)

K_1 K_{895} K_{140}

properties

K_+ total used features ~ Poisson(αH_N)

properties

K_+ total used features ~ Poisson(αH_N)

features per object ~ Poisson(α)

properties

K_+ total used features ~ Poisson(α H_N)

features per object ~ Poisson(α)

Z is sparse

properties

K+ total used features ~ Poisson(α H_N)

features per object ~ Poisson(α)

Z is sparse

inference can be done by Gibbs sampling

sources

The Indian Buffet Process: An Introduction and Review
Griffiths, Ghahramani, 2011

A tutorial on Bayesian nonparametric models
Gershmana, Blei, 2012

icons

from The Noun Project,
by Nate Eul and Francielly Constantin Senra