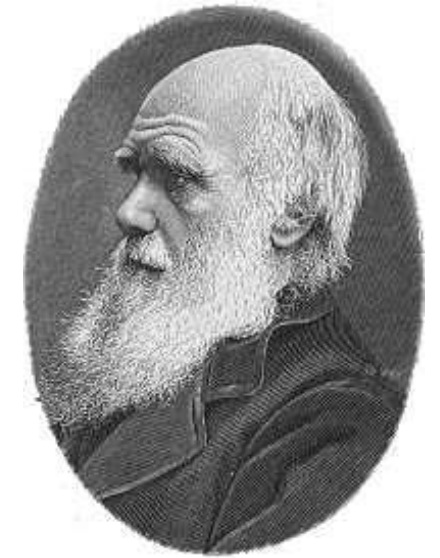


Negative Examples for
Sequential Importance
Sampling of
Binary Contingency Tables

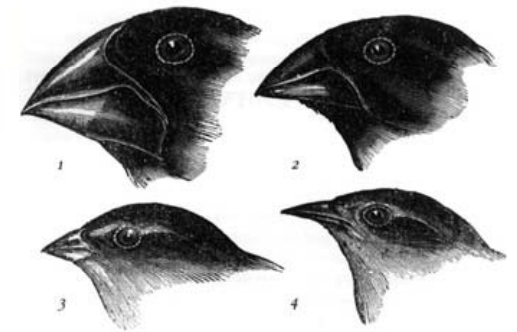
Ivona Bezáková (RIT)
Daniel Štefankovič (Rochester)
Alistair Sinclair (Berkeley)
Eric Vigoda (Gatech)

Darwin's Finches



The Voyage of the Beagle

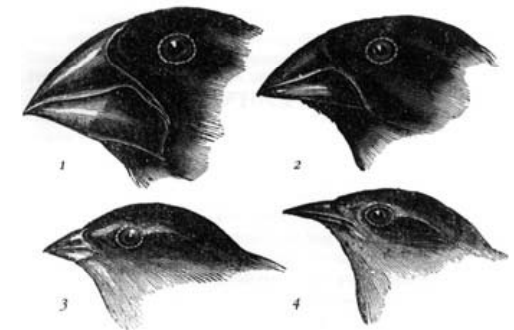
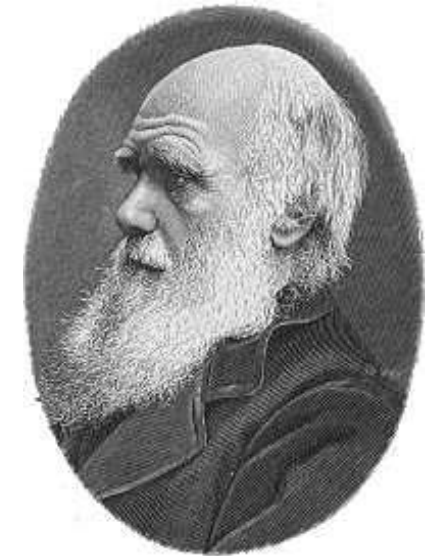
Galápagos archipelago (1835)



Darwin's Finches

Distribution of Darwin's Finches on Visitor Islands

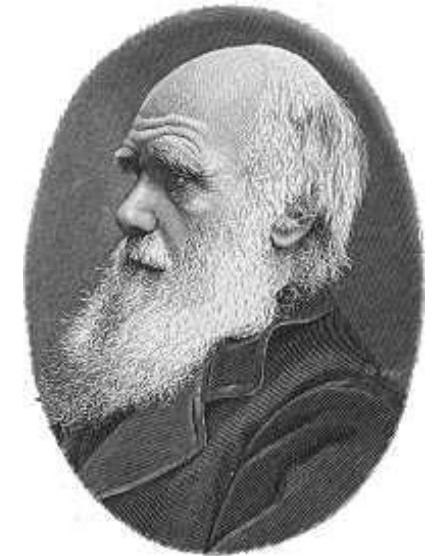
| | Santa Cruz | Plaza | Santa Fe | San Cristobal | Espanola | Floreana | Isabela | Fernandina | Santiago | Rabida |
|---------------------------|------------|-------|----------|---------------|----------|----------|---------|------------|----------|--------|
| Small Ground Finch | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Medium Ground Finch | ● | ● | ● | ● | | ● | ● | ● | ● | ● |
| Large Ground Finch | ● | | | | | | ● | ● | ● | ● |
| Cactus Ground Finch | ● | ● | ● | ● | | ● | ● | | ● | ● |
| Large Cactus Ground Finch | | | | | ● | | | | | |
| Sharp-beaked Ground Finch | | | | | | | | ● | ● | |
| Vegetarian Finch | ● | | | ● | | ● | ● | ● | ● | ● |
| Small Tree Finch | ● | | ● | ● | | ● | ● | ● | ● | ● |
| Medium Tree Finch | | | | | | ● | | | | |
| Large Tree Finch | ● | | | | | ● | ● | ● | ● | ● |
| Woodpecker Finch | ● | | | ● | | | ● | | ● | |
| Mangrove Finch | | | | | | | ● | ● | | |
| Warbler Finch | ● | | ● | ● | ● | ● | ● | ● | ● | ● |



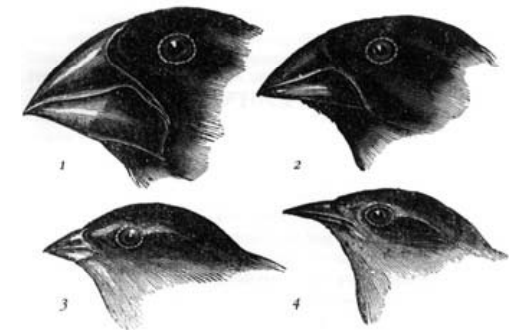
Darwin's Finches

Distribution of Darwin's Finches on Visitor Islands

| | Santa Cruz | Plaza | Santa Fe | San Cristobal | Espanola | Floreana | Isabela | Fernandina | Santiago | Rabida |
|---------------------------|------------|-------|----------|---------------|----------|----------|---------|------------|----------|--------|
| Small Ground Finch | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Medium Ground Finch | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Large Ground Finch | ● | | | | | | ● | ● | ● | ● |
| Cactus Ground Finch | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Large Cactus Ground Finch | | | | | ● | | | | | |
| Sharp-beaked Ground Finch | | | | | | | | ● | ● | |
| Vegetarian Finch | ● | | | ● | | ● | ● | ● | ● | ● |
| Small Tree Finch | ● | | ● | ● | | ● | ● | ● | ● | ● |
| Medium Tree Finch | | | | | | ● | | | | |
| Large Tree Finch | ● | | | | | ● | ● | ● | ● | ● |
| Woodpecker Finch | ● | | | ● | | | ● | | ● | |
| Mangrove Finch | | | | | | | ● | ● | | |
| Warbler Finch | ● | | ● | ● | ● | ● | ● | ● | ● | ● |



8



10

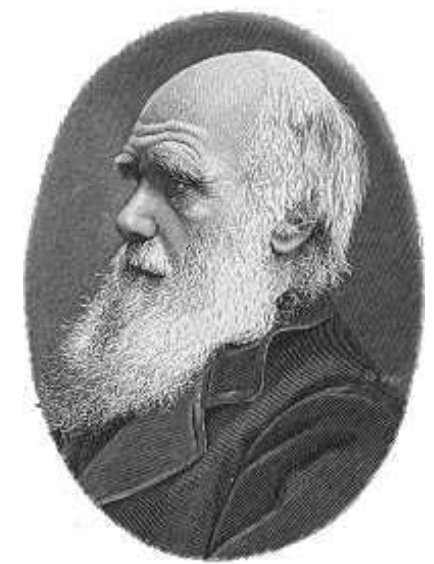
Darwin's Finches

Distribution of Darwin's Finches on Visitor Islands

| | Santa Cruz | Plaza | Santa Fe | San Cristobal | Espanola | Floreana | Isabela | Fernandina | Santiago | Rabida |
|---------------------------|------------|-------|----------|---------------|----------|----------|---------|------------|----------|--------|
| Small Ground Finch | ● | ● | ● | ● | ● | ● | ● | ● | ● | ● |
| Medium Ground Finch | ● | ● | ● | ● | | ● | ● | ● | ● | ● |
| Large Ground Finch | ● | | | | | | ● | ● | ● | ● |
| Cactus Ground Finch | ● | ● | ● | ● | | ● | ● | | ● | ● |
| Large Cactus Ground Finch | | | | | ● | | | | | |
| Sharp-beaked Ground Finch | | | | | | | | ● | ● | |
| Vegetarian Finch | ● | | | ● | | ● | ● | ● | ● | ● |
| Small Tree Finch | ● | | ● | ● | | ● | ● | ● | ● | ● |
| Medium Tree Finch | | | | | | ● | | | | |
| Large Tree Finch | ● | | | | | ● | ● | ● | ● | ● |
| Woodpecker Finch | ● | | | ● | | | ● | | ● | |
| Mangrove Finch | | | | | | | ● | ● | | |
| Warbler Finch | ● | | ● | ● | ● | ● | ● | ● | ● | ● |

9 3 5 7 3 8 10 9 10 8

10
9
6
8
2
3
7
8
1
6
4
2
10



chance
OR
competitive pressures
?

Binary Contingency Tables

Given: marginals (**row sums**, **column sums**)

Goal:

- sample tables uniformly at random
- count tables

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | 4 |
| | | | | | | | 2 |
| | | | | | | | 3 |
| | | | | | | | 5 |
| | | | | | | | 3 |
| 3 | 4 | 2 | 1 | 2 | 2 | 3 | |

Binary Contingency Tables

Given: marginals (**row sums**, **column sums**)

Goal:

- sample tables uniformly at random
- count tables

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | ■ | | ■ | ■ | ■ | | 4 |
| | | | | | ■ | ■ | 2 |
| ■ | ■ | | | | | ■ | 3 |
| ■ | ■ | ■ | | ■ | | ■ | 5 |
| ■ | ■ | ■ | | | | | 3 |
| 3 | 4 | 2 | 1 | 2 | 2 | 3 | |

Binary Contingency Tables

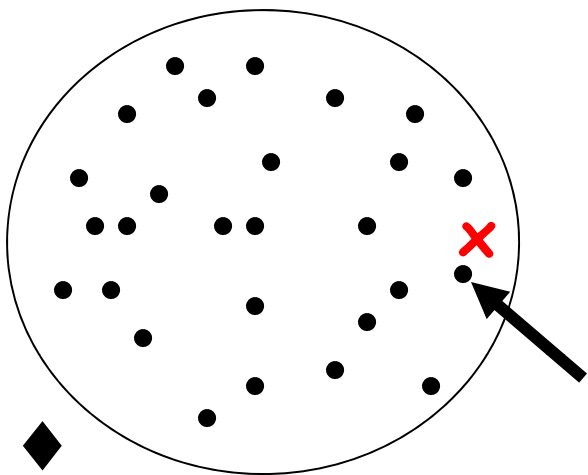
Given: marginals (**row sums**, **column sums**)

Goal:

- sample tables uniformly at random
- count tables

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ■ | ■ | □ | □ | ■ | □ | ■ | 4 |
| □ | ■ | ■ | □ | □ | □ | □ | 2 |
| ■ | ■ | □ | □ | □ | ■ | □ | 3 |
| □ | ■ | ■ | □ | ■ | ■ | ■ | 5 |
| ■ | □ | □ | ■ | □ | □ | ■ | 3 |
| 3 | 4 | 2 | 1 | 2 | 2 | 3 | |

Importance Sampling for counting problems



Probability distribution σ
on the points + \diamond

with **positive** probability $\sigma(x) > 0$

Random variable

$$\eta(s) = \begin{cases} 1/\sigma(s) & \text{if } s \text{ in the set} \\ 0 & \text{if } s \text{ is } \diamond \end{cases}$$

Unbiased estimator

$$\mathbf{E}[\eta] = \sum \sigma(x) \cdot 1/\sigma(x) = \text{size of the set}$$

Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

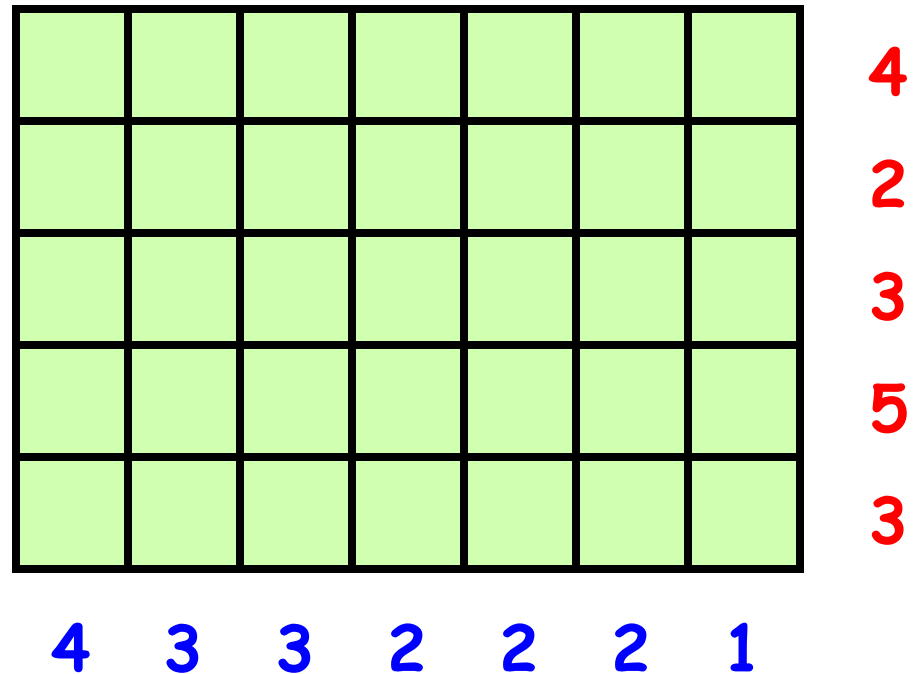
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | | | 4 |
| | | | | | | | 2 |
| | | | | | | | 3 |
| | | | | | | | 5 |
| | | | | | | | 3 |
| 3 | 4 | 2 | 1 | 2 | 2 | 3 | |

Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

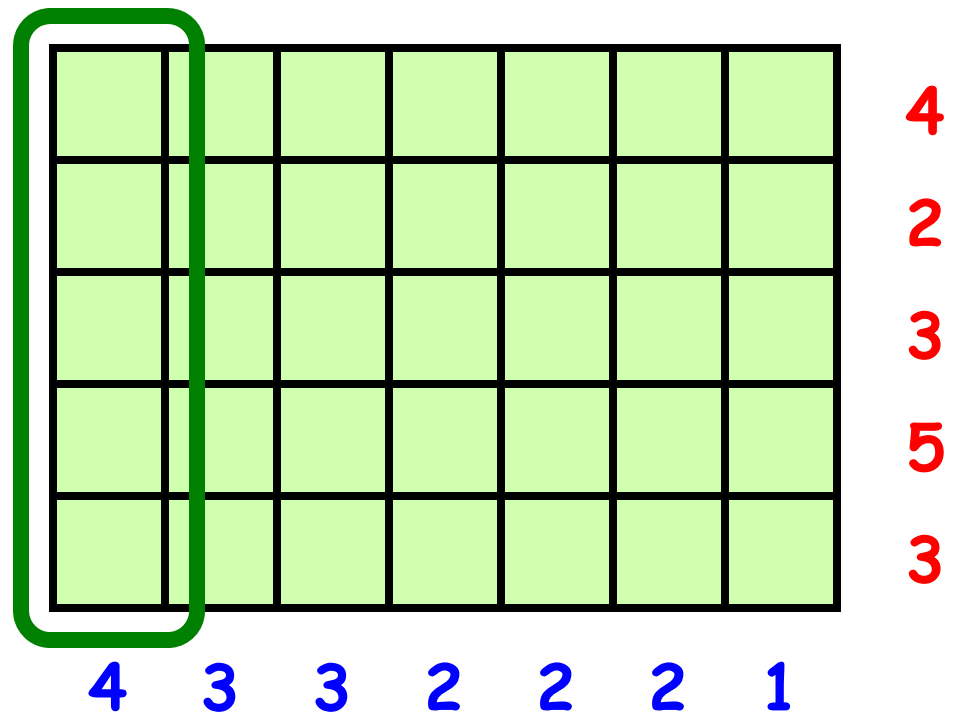


Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

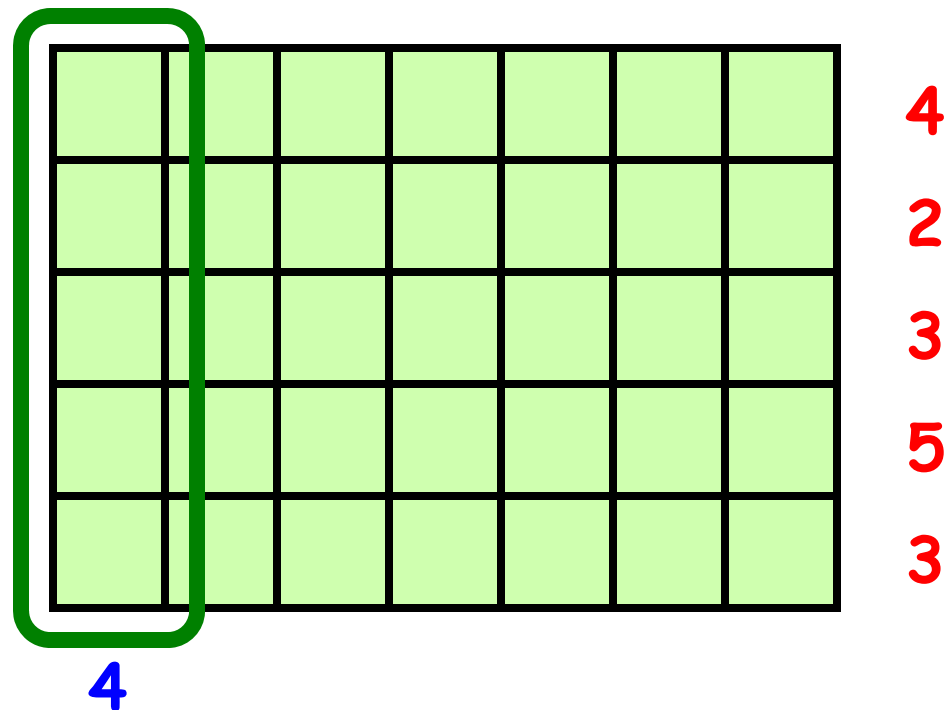


Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

a specific σ

- fill table column-by-column
- assign each column ignoring other column sums



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

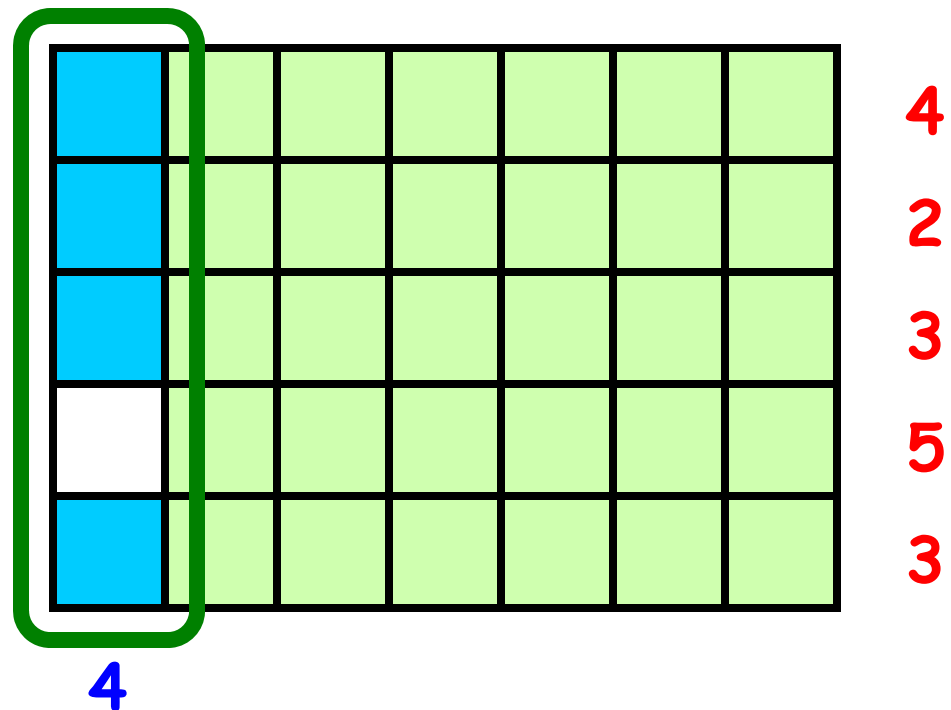
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

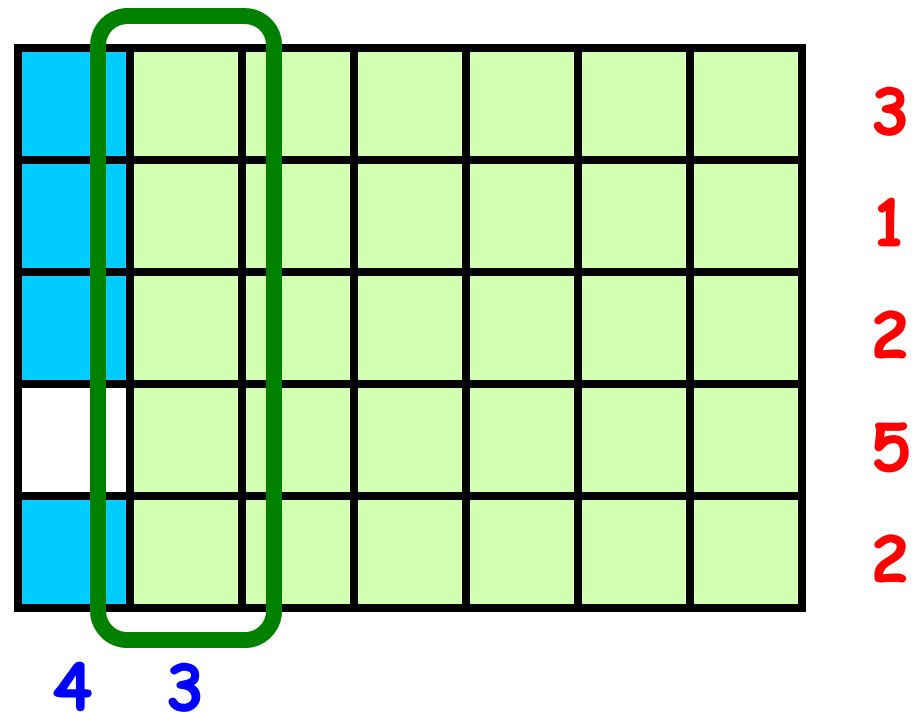
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

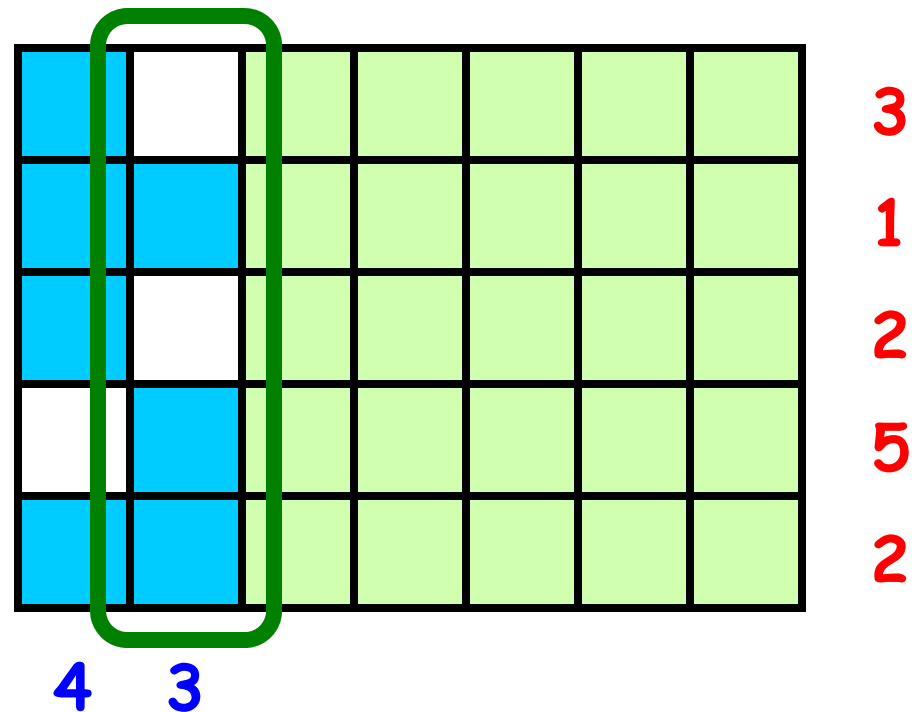
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

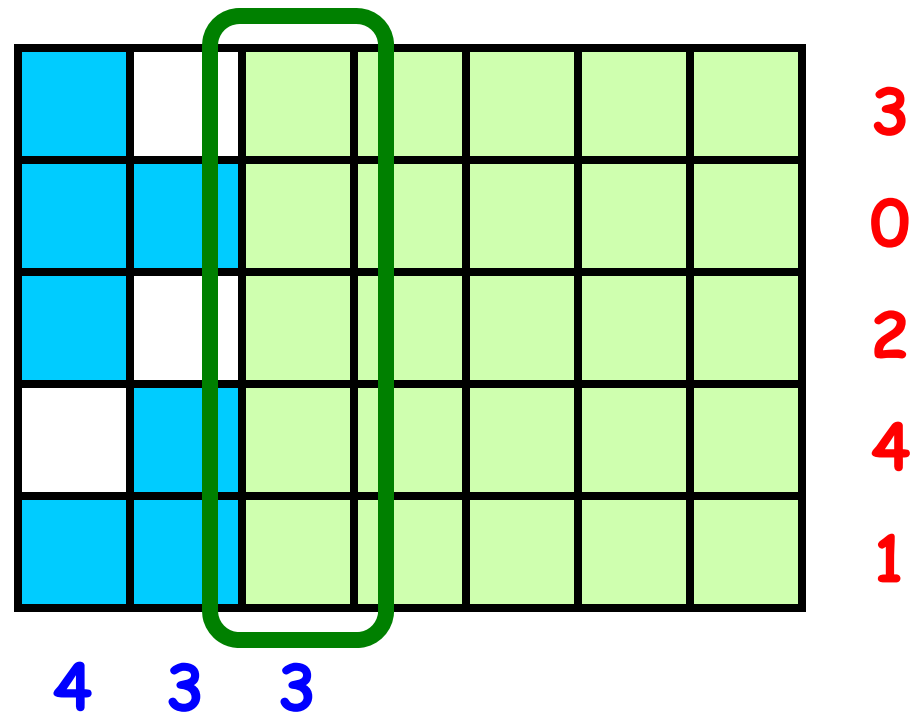
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

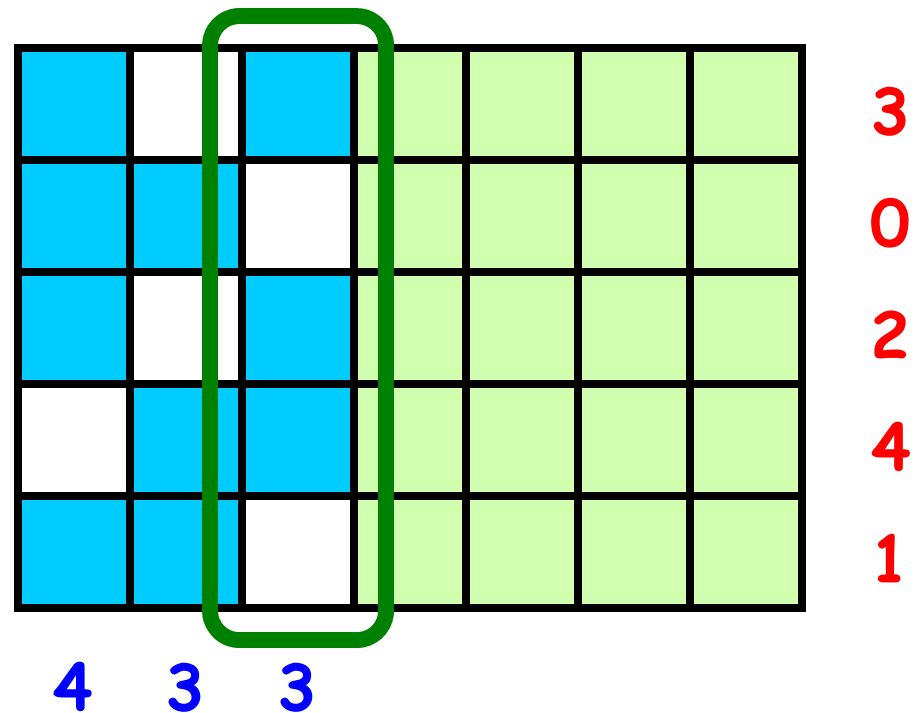
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

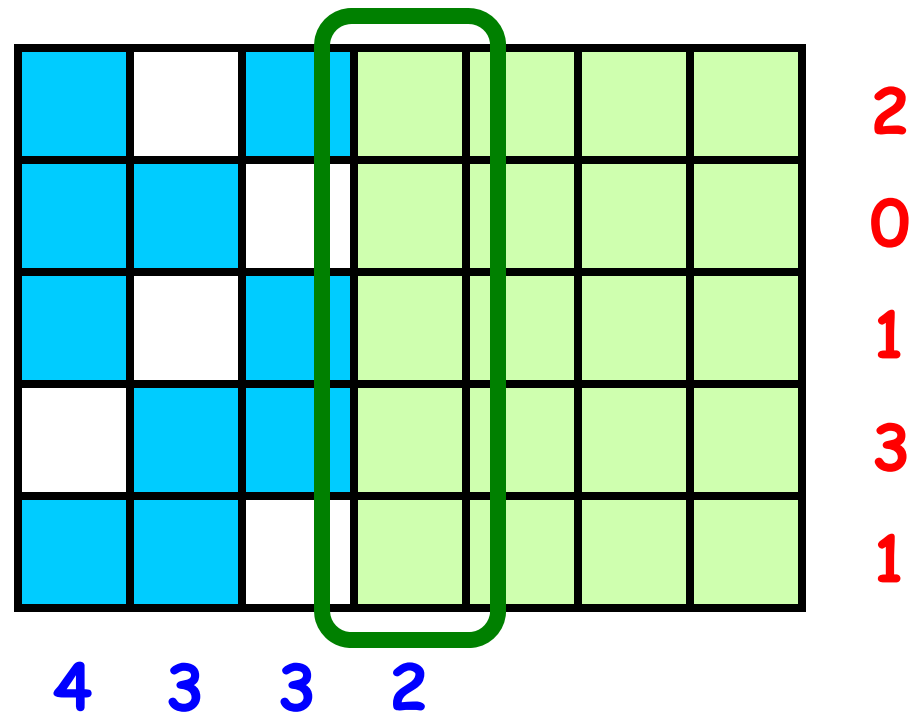
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

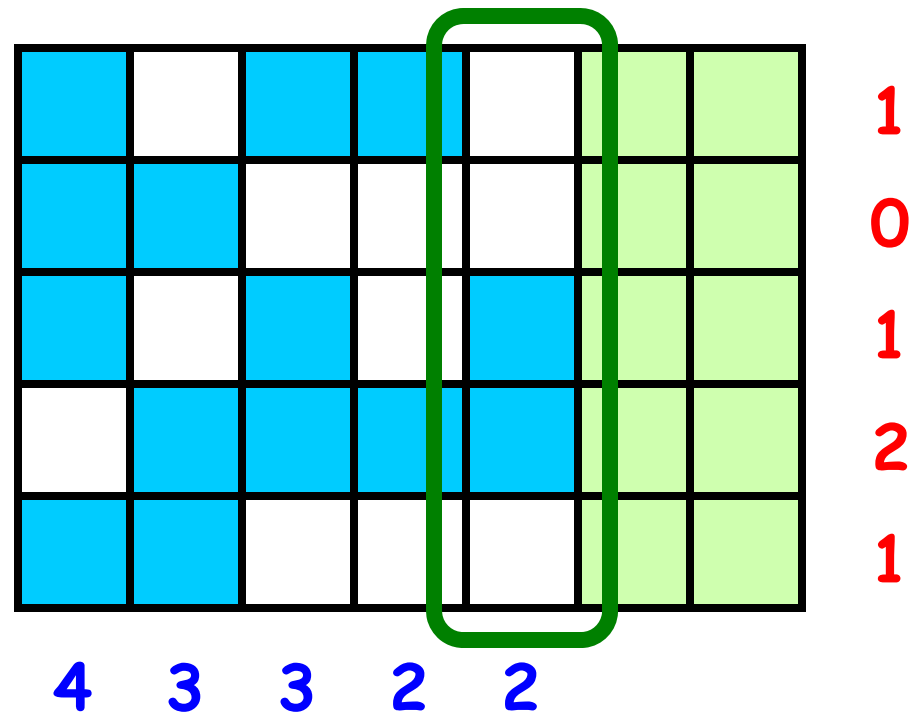
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

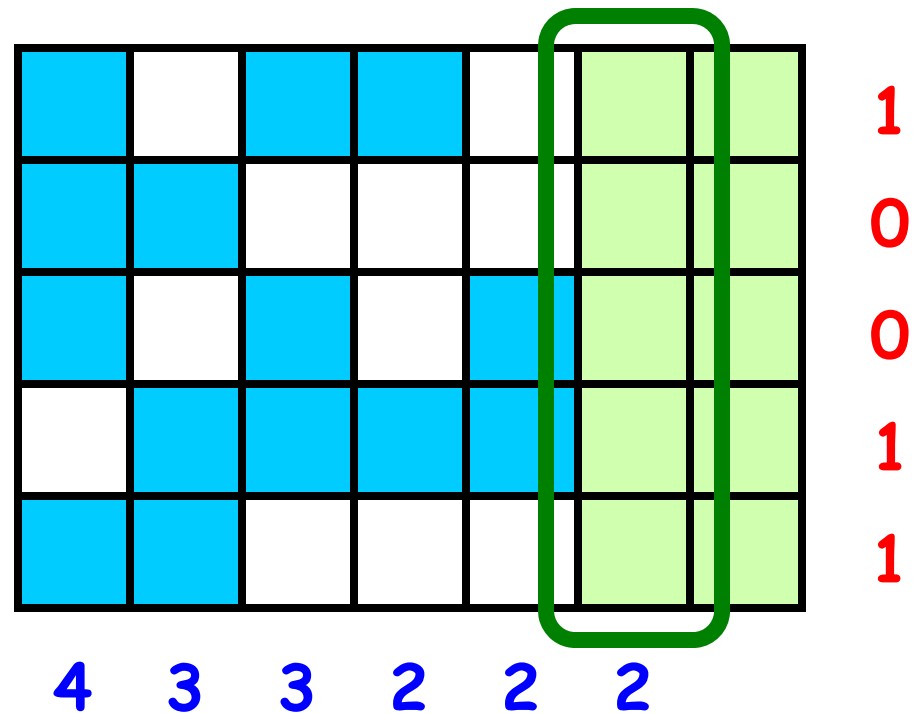
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

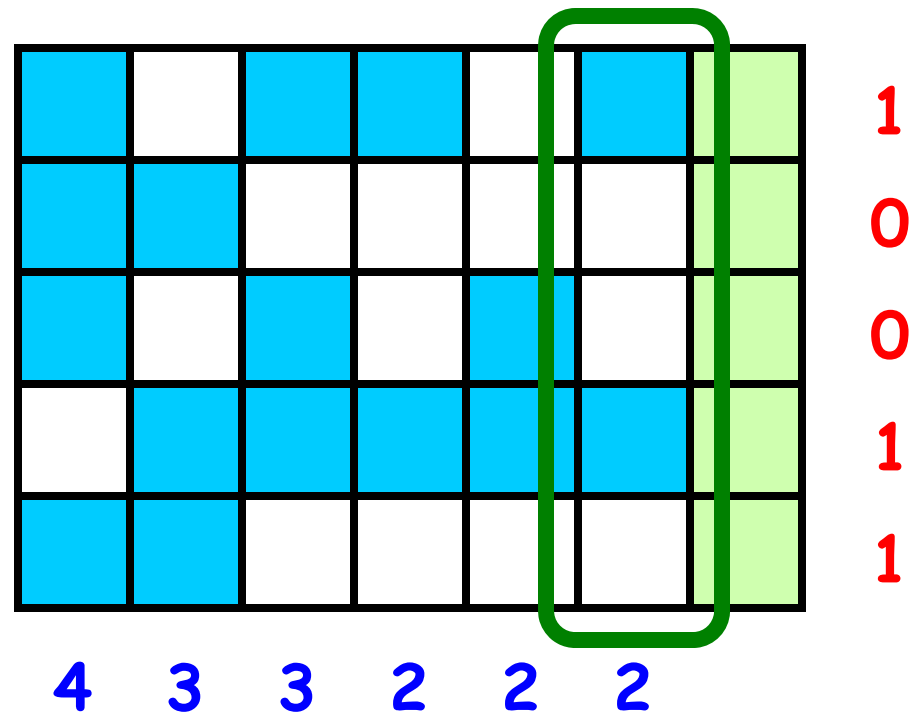
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

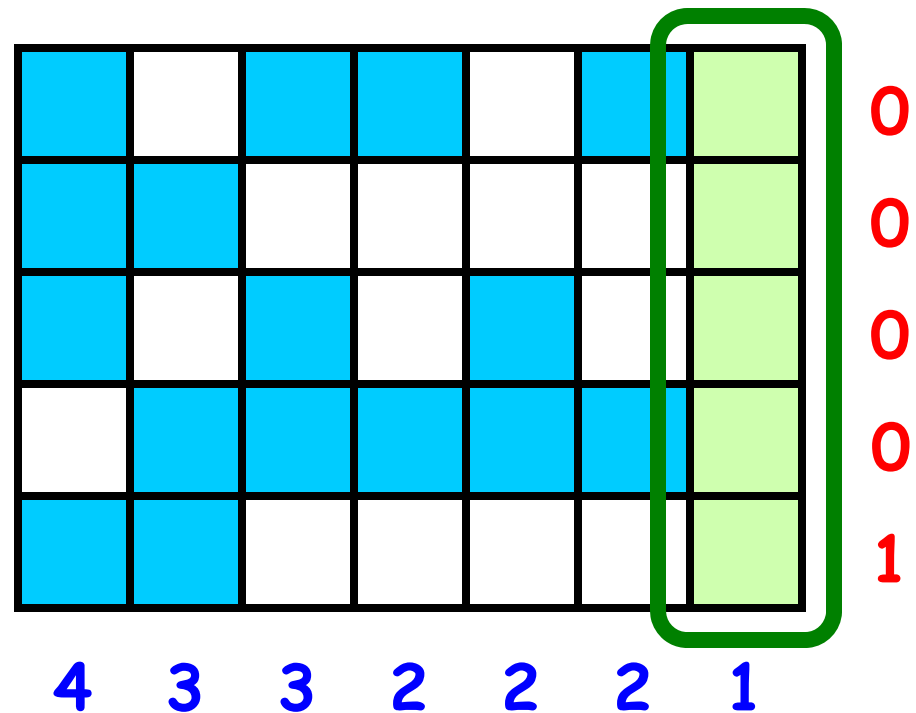
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

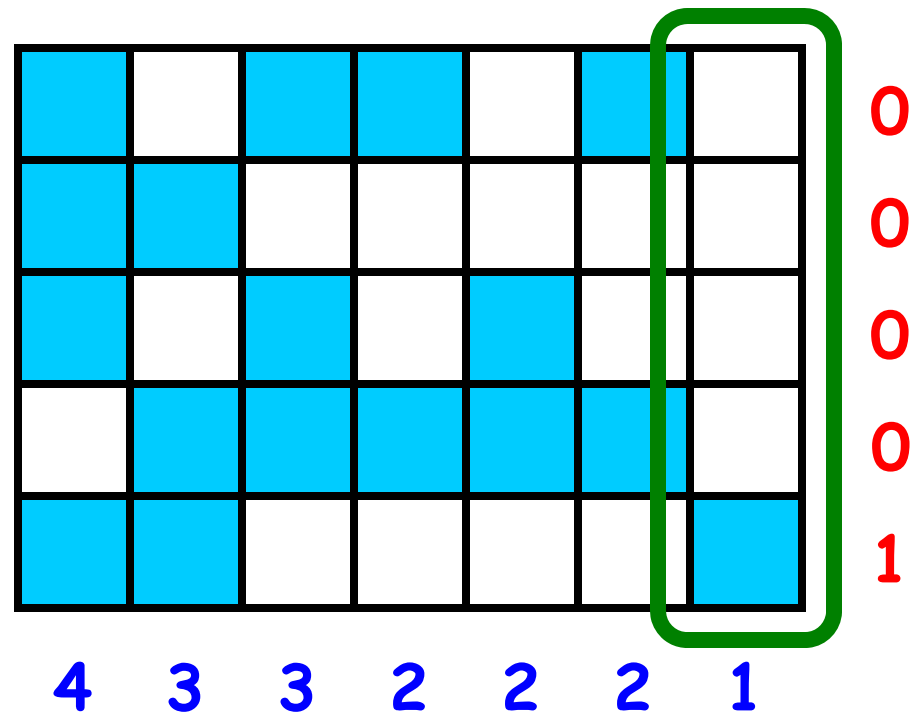
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

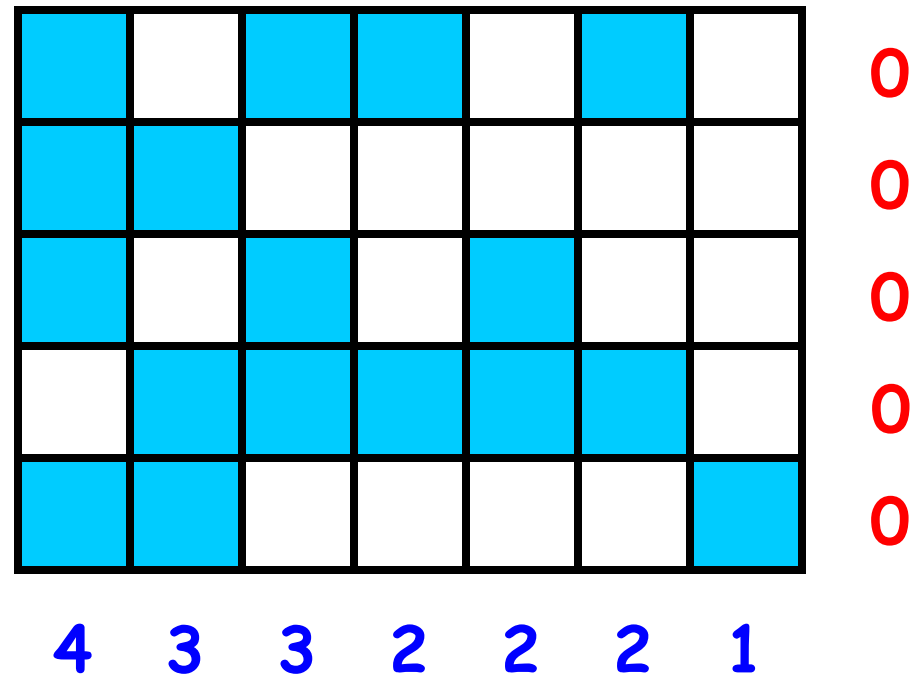
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

where product
ranges over i : rows
with assignment 1



Sequential Importance Sampling for BCT

[Chen-Diaconis-Holmes-Liu '05]

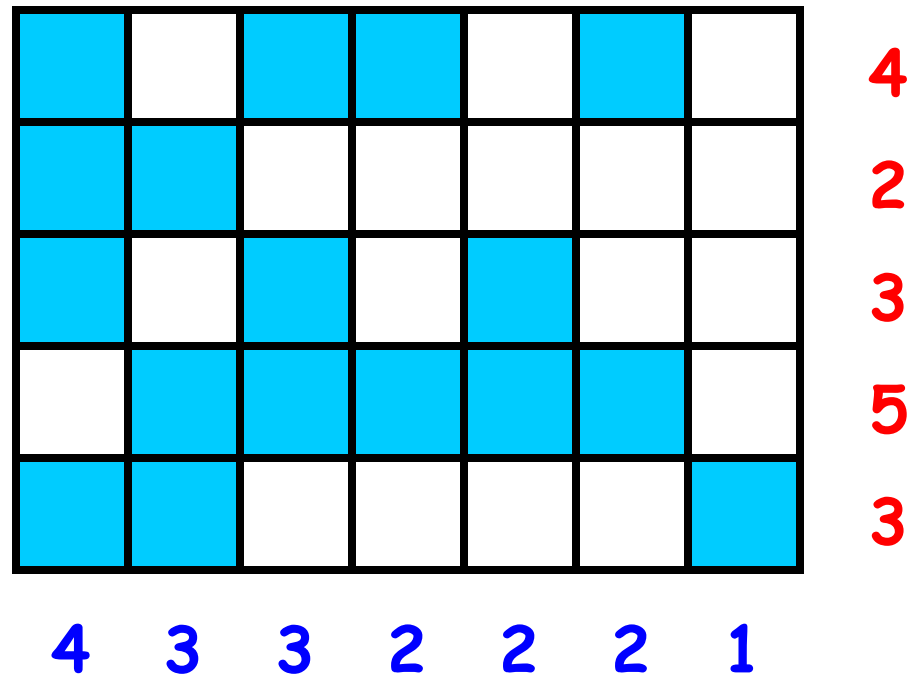
a specific σ

- fill table column-by-column
- assign each column ignoring other column sums

assign the column
with probability
proportional to

$$\prod r_i / (n - r_i)$$

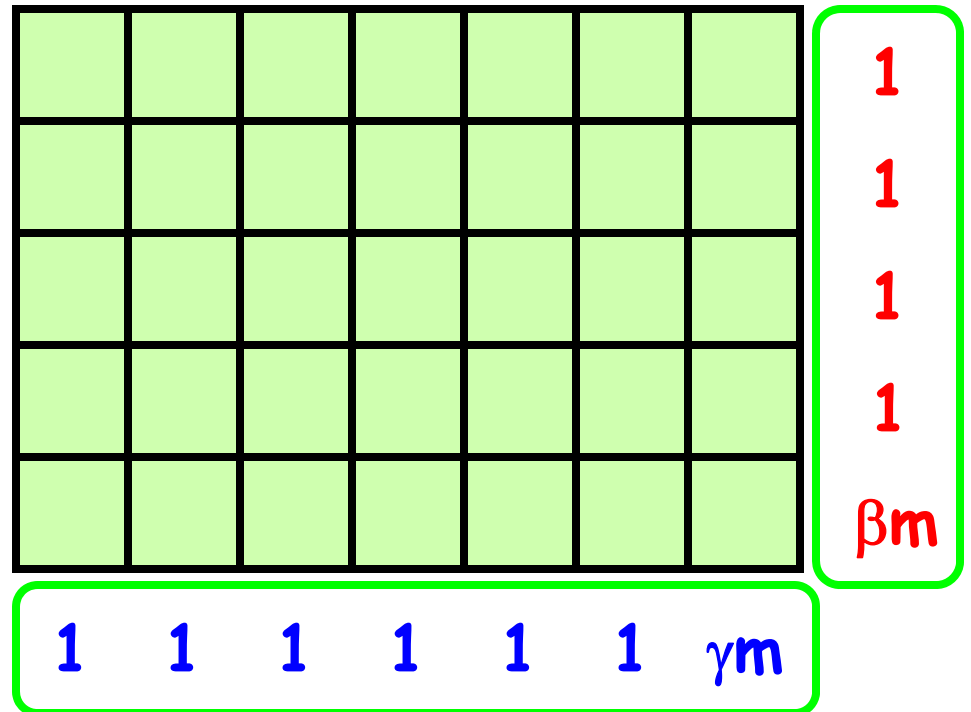
where product
ranges over i : rows
with assignment 1



A Counterexample for SIS

Thm [Bezáková-Sinclair-Štefankovič-Vigoda '06]:

For any $\beta \neq \gamma$, SIS output after any subexponential number of trials is **off by an exponential factor** (with high probability).

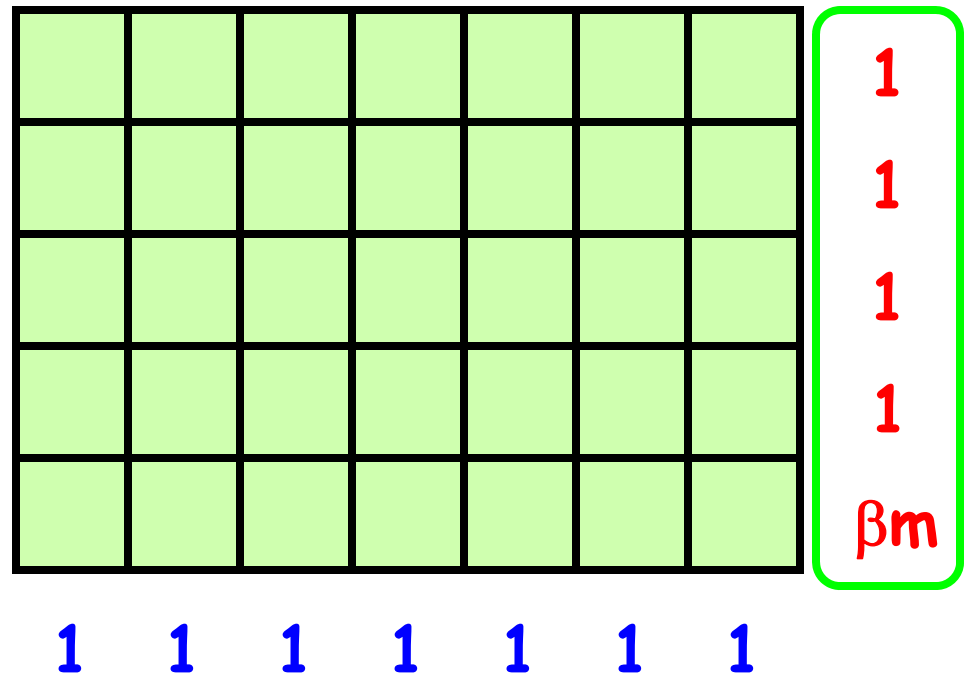


A Counterexample for SIS

Thm [Bezáková-Sinclair-Štefankovič-Vigoda '06]:

For any $\beta \neq \gamma$, SIS output after any subexponential number of trials is **off by an exponential factor** (with high probability).

Intuition



A Counterexample for SIS

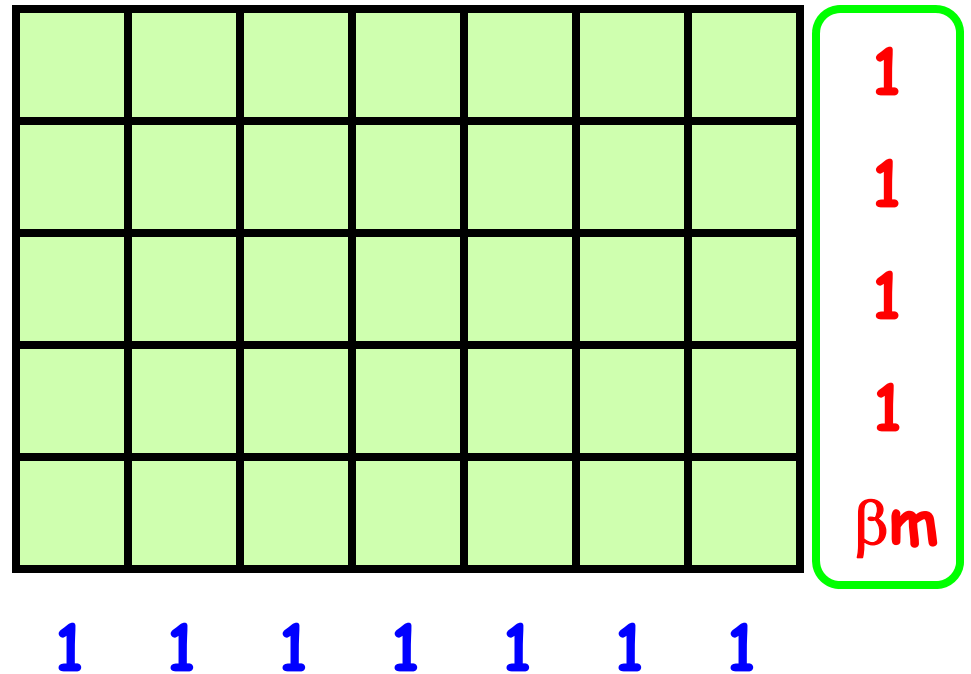
Thm [Bezáková-Sinclair-Štefankovič-Vigoda '06]:

For any $\beta \neq \gamma$, SIS output after any subexponential number of trials is **off by an exponential factor** (with high probability).

Intuition

Random table:

- randomly choose βm ones



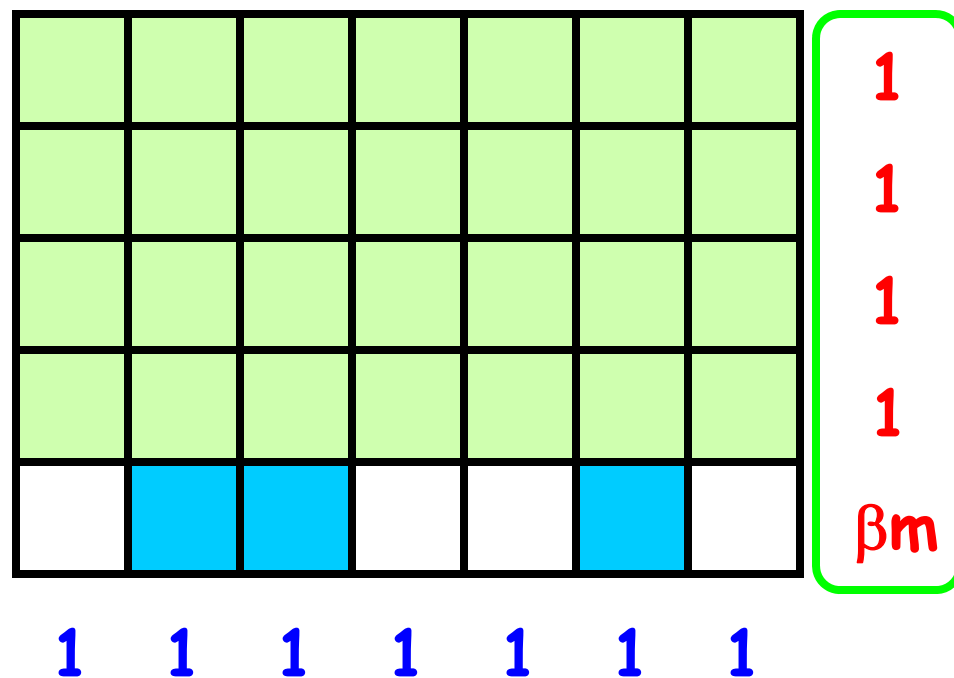
A Counterexample for SIS

Thm [Bezáková-Sinclair-Štefankovič-Vigoda '06]:

For any $\beta \neq \gamma$, SIS output after any subexponential number of trials is **off by an exponential factor** (with high probability).

Intuition

Random table:
- randomly choose βm ones



A Counterexample for SIS

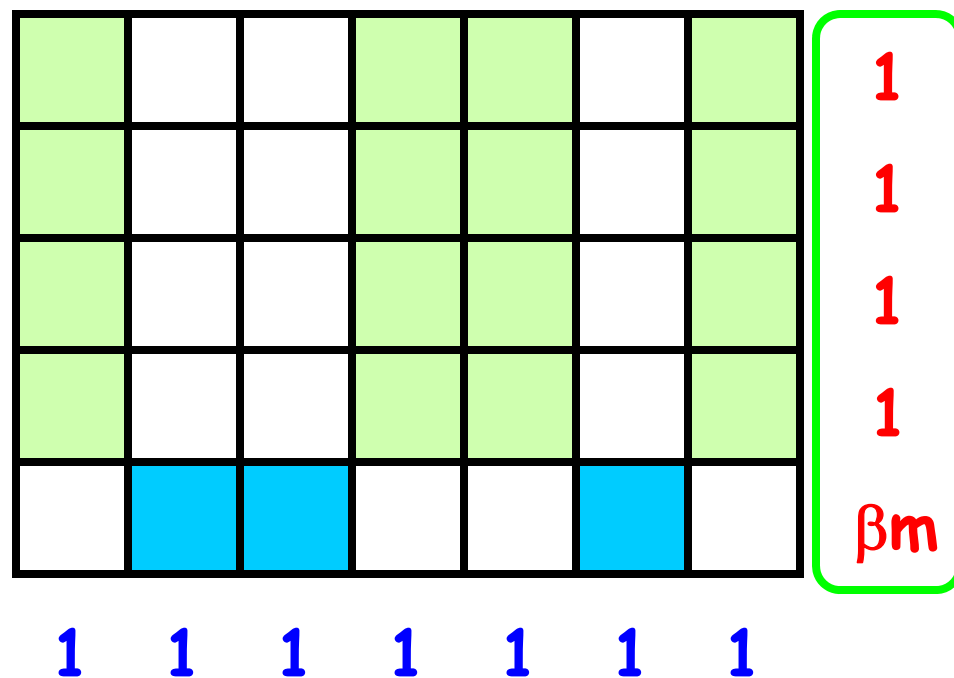
Thm [Bezáková-Sinclair-Štefankovič-Vigoda '06]:

For any $\beta \neq \gamma$, SIS output after any subexponential number of trials is **off by an exponential factor** (with high probability).

Intuition

Random table:

- randomly choose βm ones



A Counterexample for SIS

Thm [Bezáková-Sinclair-Štefankovič-Vigoda '06]:

For any $\beta \neq \gamma$, SIS output after any subexponential number of trials is **off by an exponential factor** (with high probability).

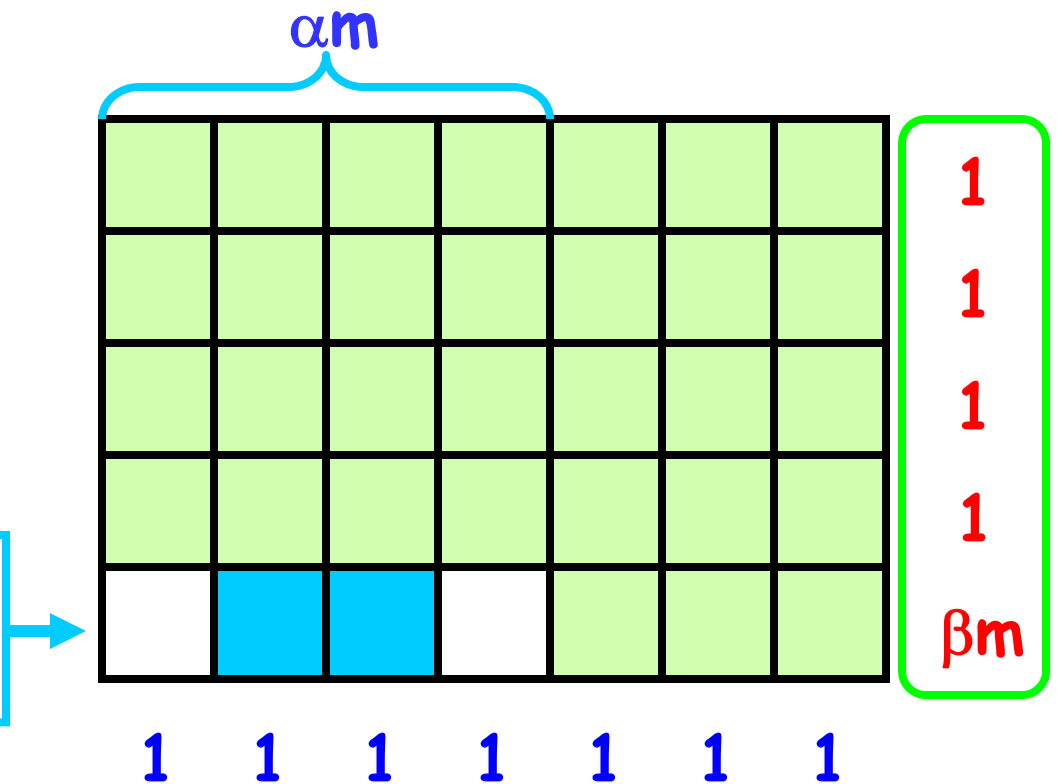
Intuition

Expect: $\alpha\beta m$ ones

SIS: asymptotically fewer

Random table:

- randomly choose βm ones



A Counterexample for SIS

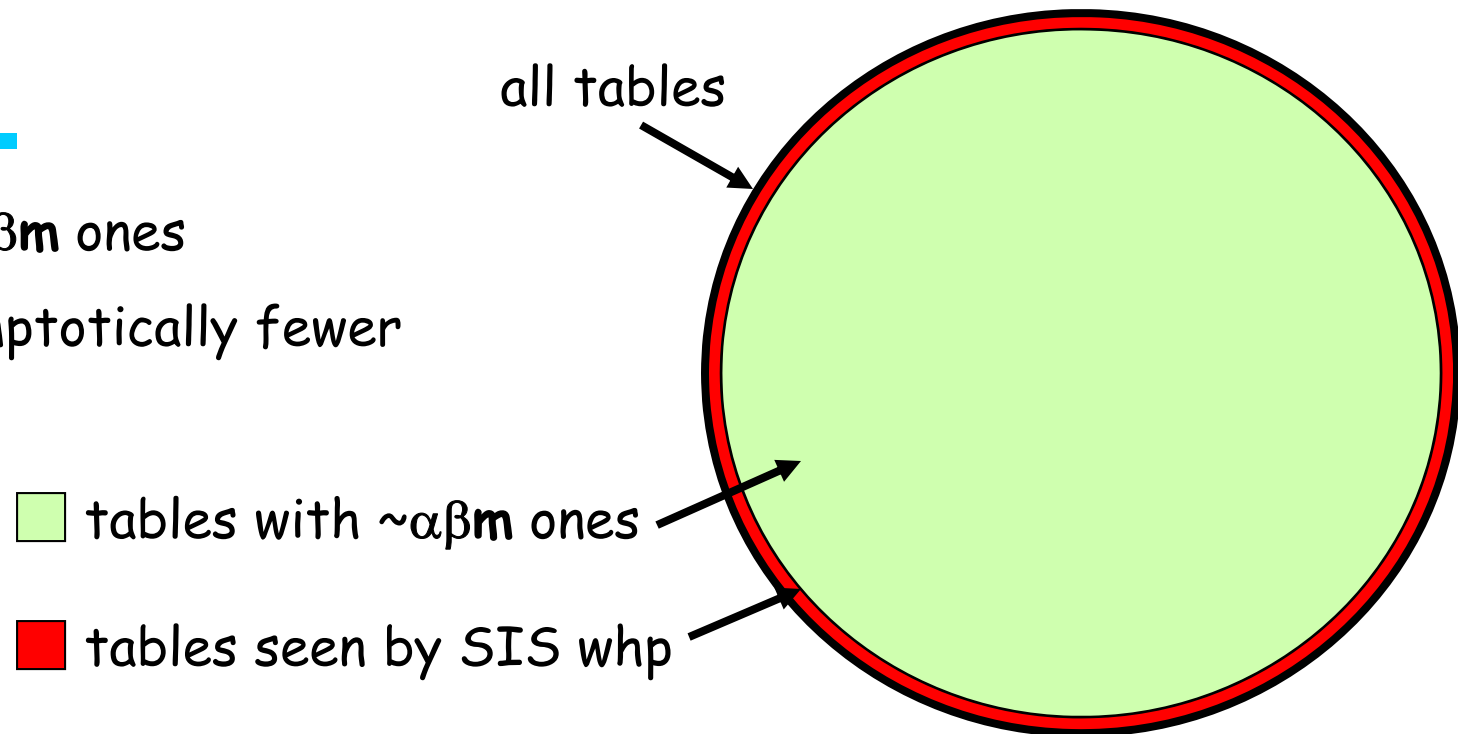
Thm [Bezáková-Sinclair-Štefankovič-Vigoda '06]:

For any $\beta \neq \gamma$, SIS output after any subexponential number of trials is **off by an exponential factor** (with high probability).

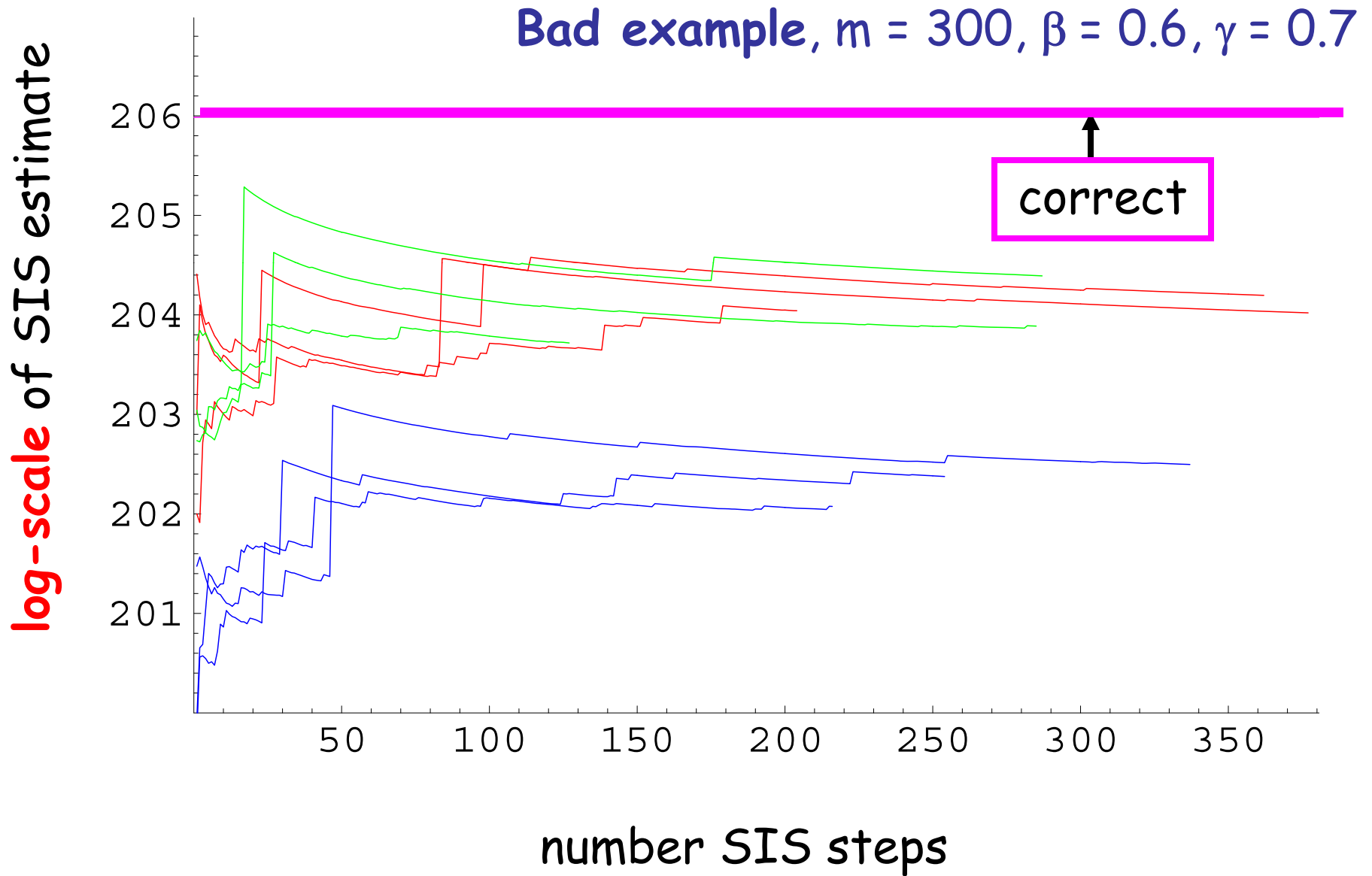
Intuition

Expect: $\alpha\beta m$ ones

SIS: asymptotically fewer



SIS - Experimental Results



SIS - Experimental Results

Regular marginals: $m=50$, marginals 5

