Lecture ~~11~~ 12 pre-video

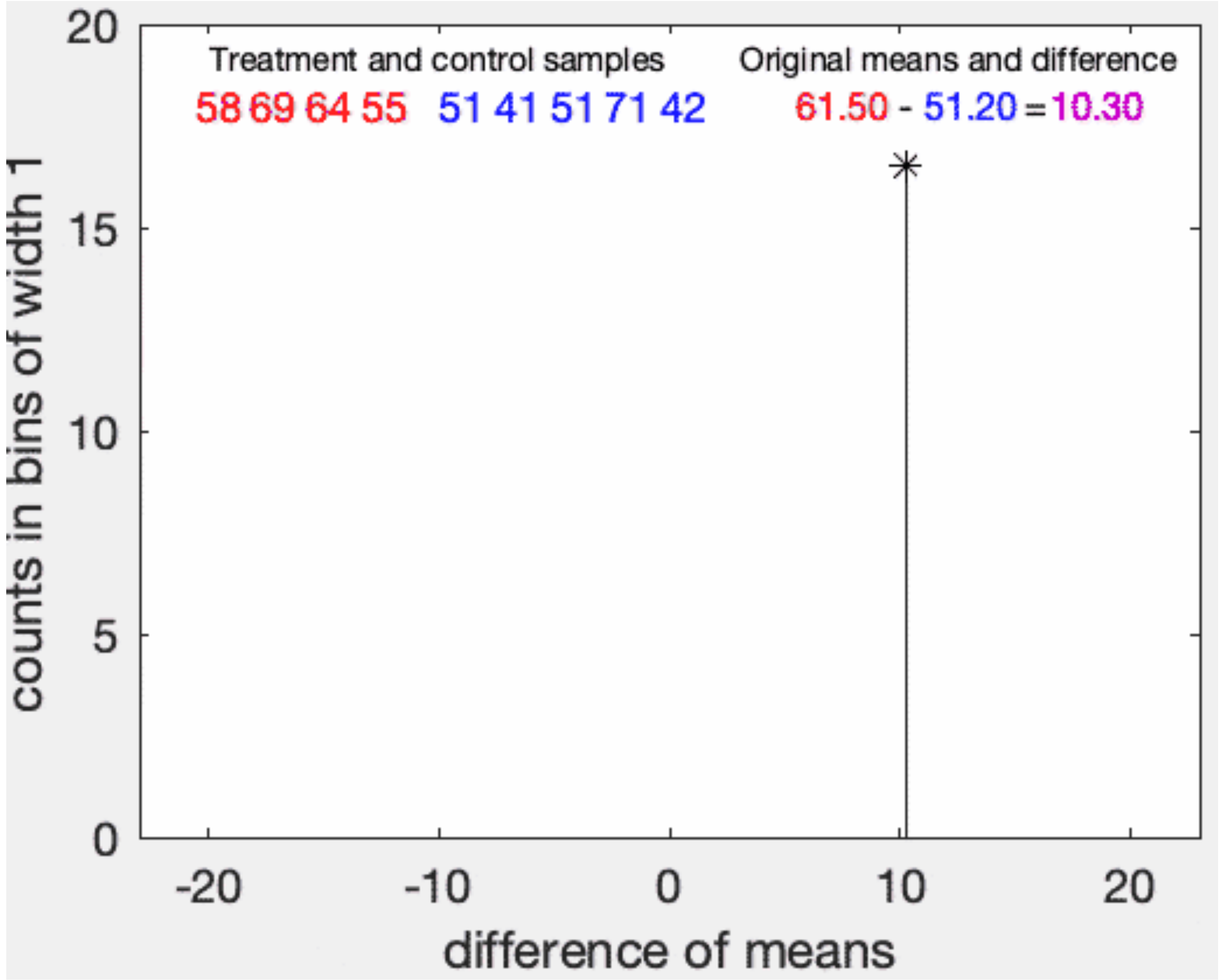
Resampling methods

Resampling

- Why do it?
    - More accurate estimation of population from sample measurements
    - Make better use of limited sample size
    - A static technique: don’t need more experiments
- Why not?
    - Could get better results by doing experiments and actively choosing new samples
    - Computationally intensive

Resampling techniques

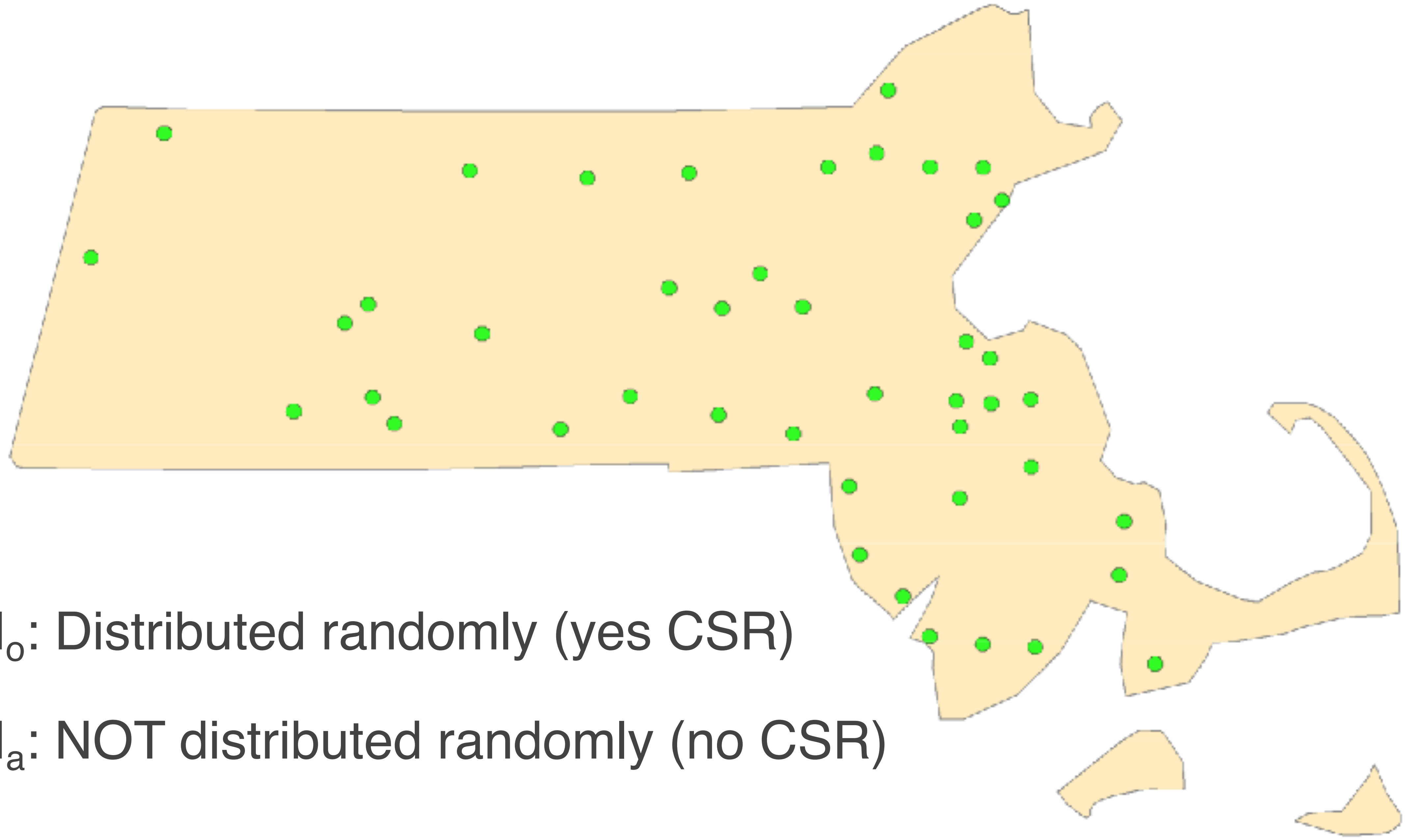
Permutation tests using Monte Carlo simulation



Resampling techniques

Permutation tests using Monte Carlo simulation

Is this distribution of Walmarts in MA the result of CSR?



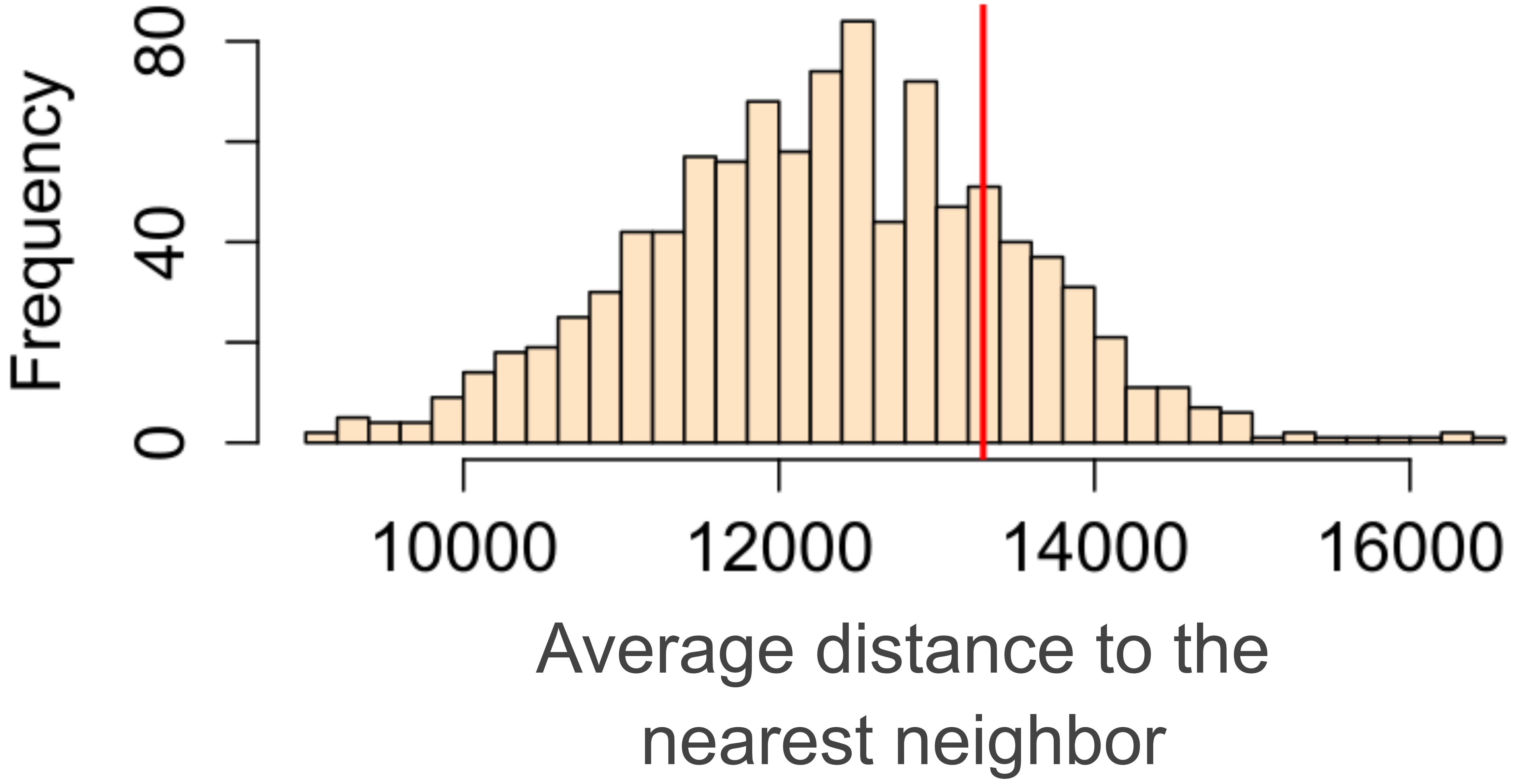
H_o: Distributed randomly (yes CSR)

H_a: NOT distributed randomly (no CSR)

p = # more extreme / # simulations

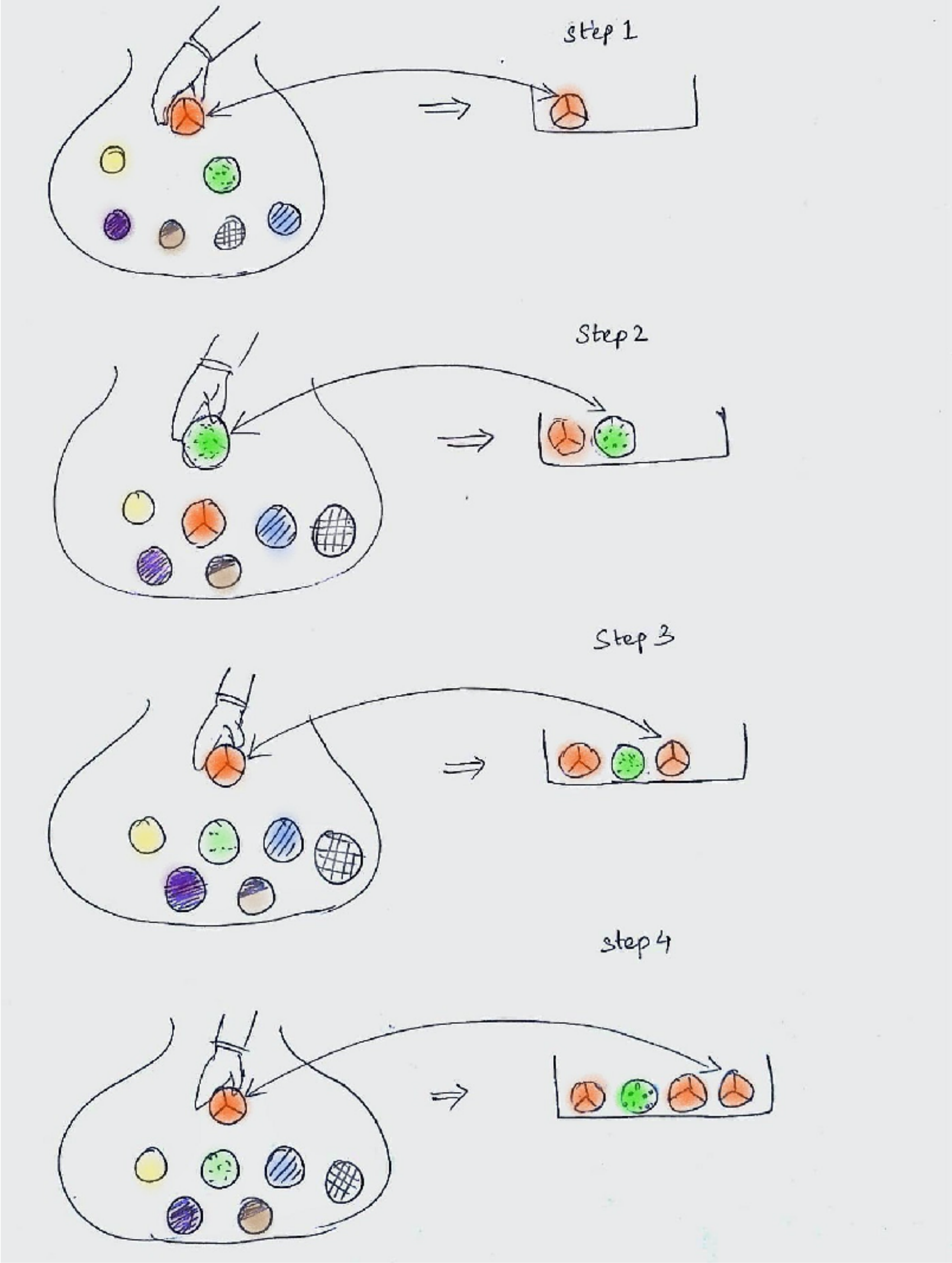
p = (319 + 1) / (1000 + 1)

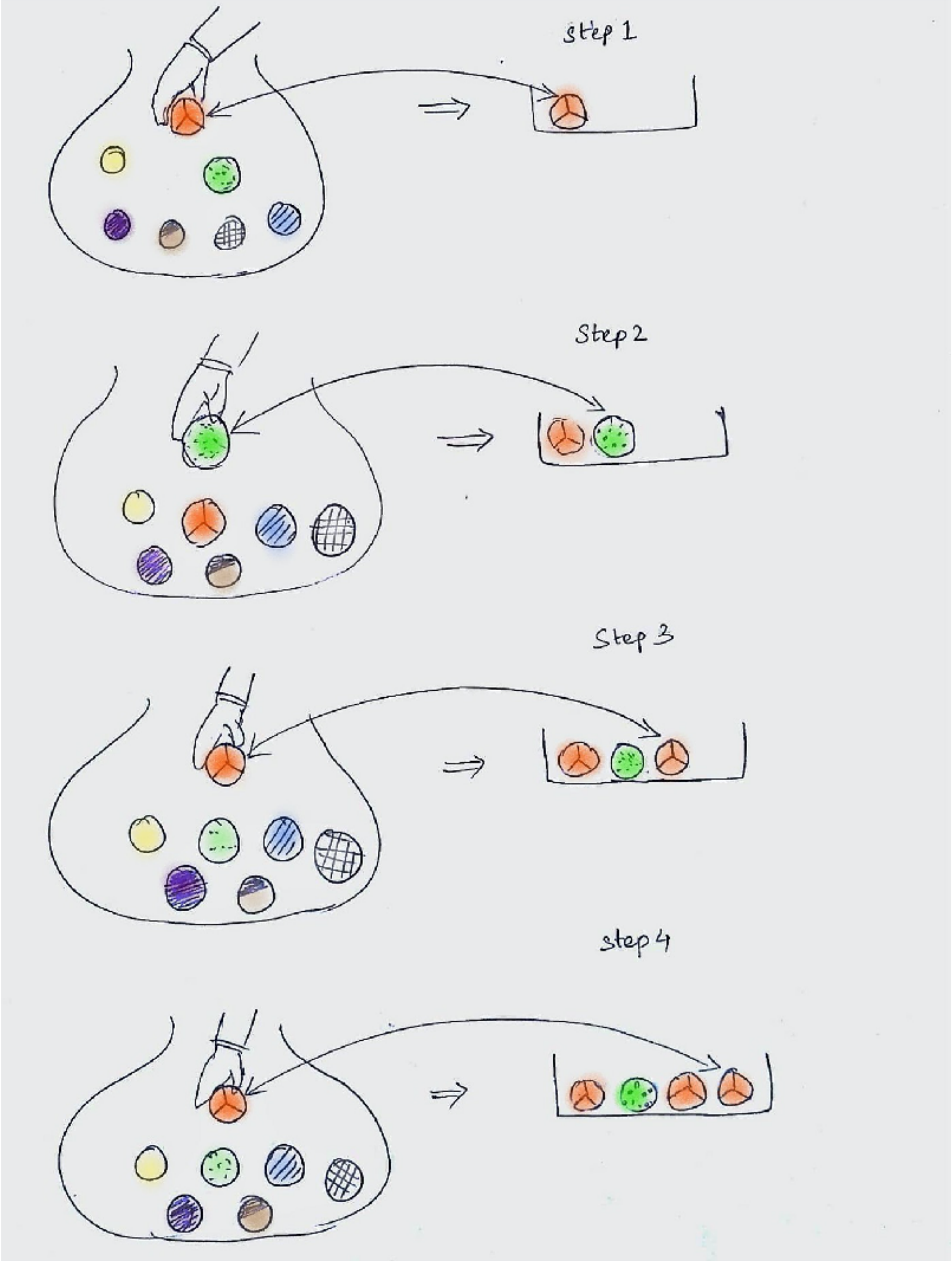
p = 0.32



Resampling techniques

Bootstrap aka
resample with replacement





Bootstrap Sampling

$$P(\textbf{not chosen}) = \left(1 - \frac{1}{n}\right)^n,$$

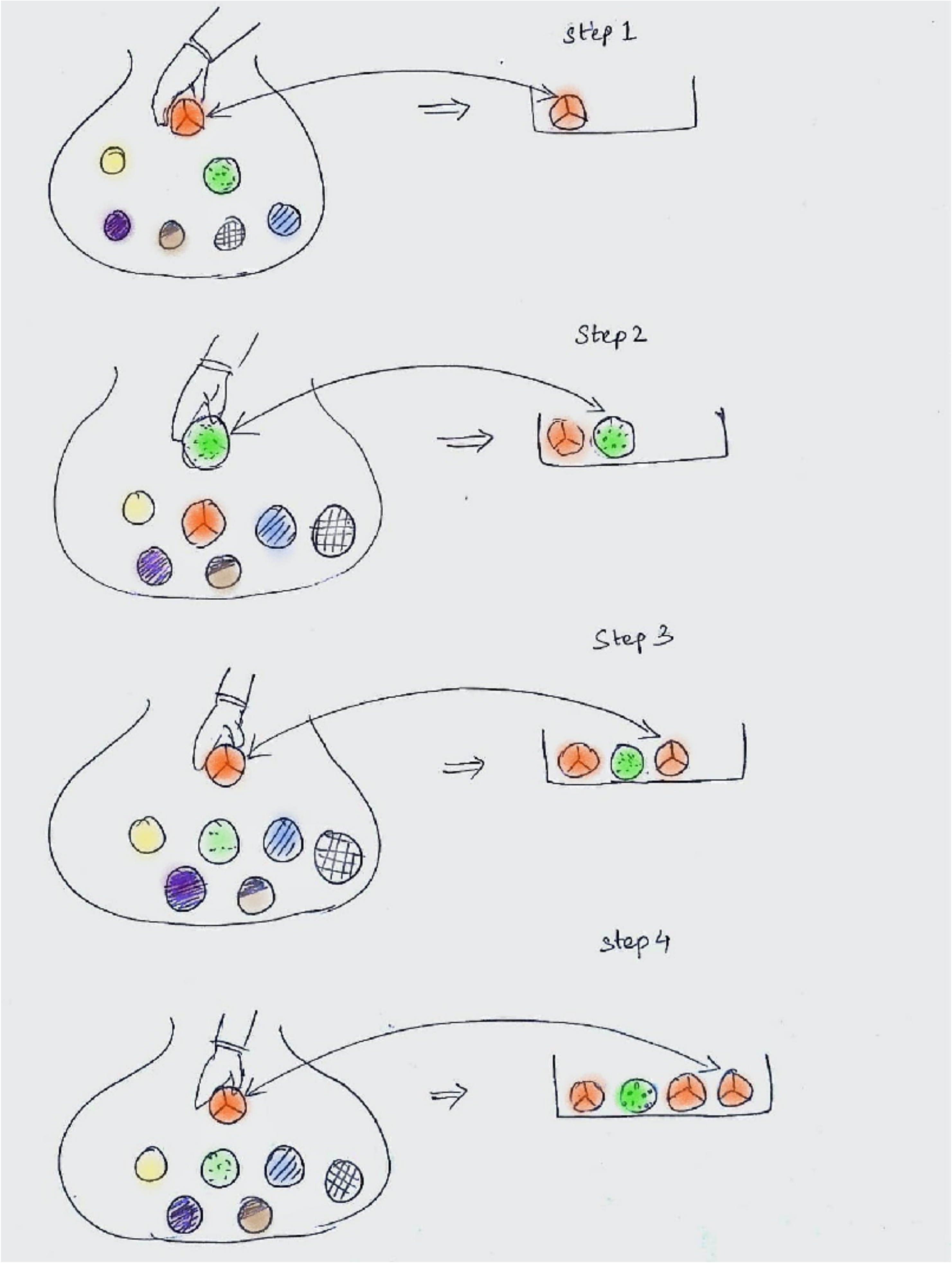
$$\frac{1}{e} \approx 0.368, \quad n \to \infty\,.$$

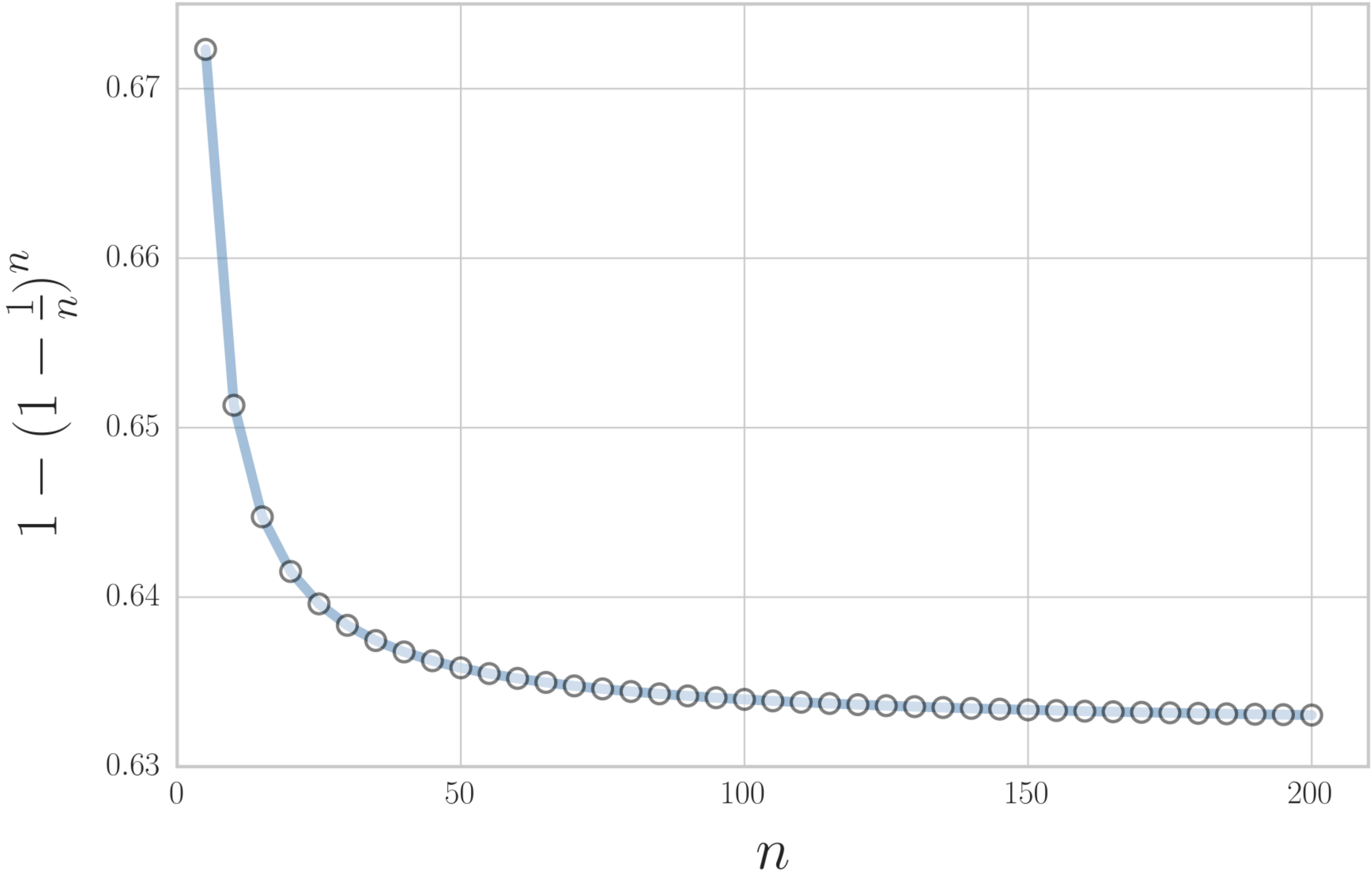
Bootstrap sampling

$$P(\textbf{not chosen}) = \left(1 - \frac{1}{n}\right)^n,$$

$$\frac{1}{e} \approx 0.368, \quad n \to \infty.$$

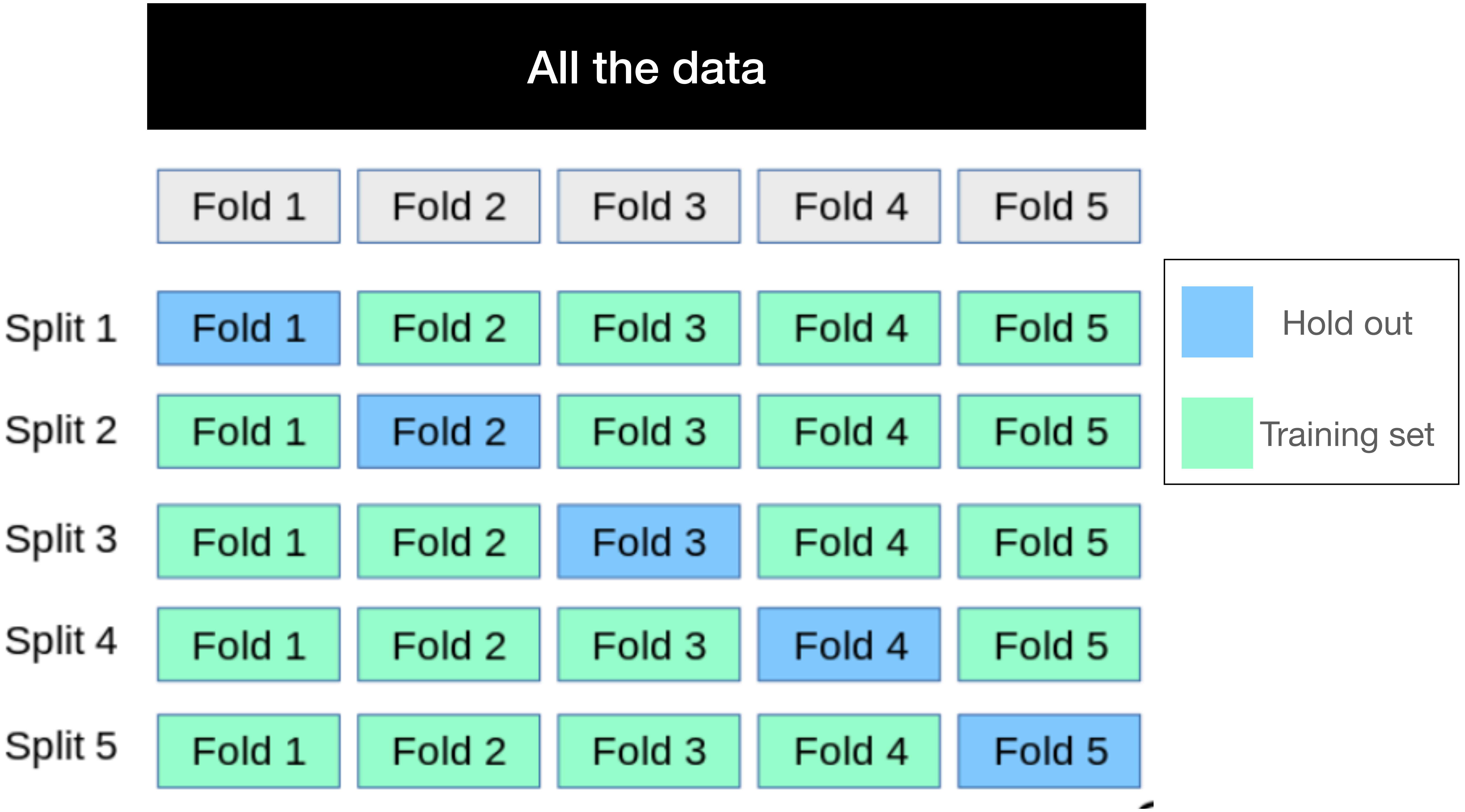
$$P(\textbf{chosen}) = 1 - \left(1 - \frac{1}{n}\right)^n \approx 0.632$$





Resampling techniques

Cross validation



Use the mean of the hold out set performances to estimate either test or validation error

Resampling techniques

- Permutation tests using Monte Carlo simulation
  - Estimate if model #1 is really better than model #2 or if its just by chance
- Bootstrap sampling, Cross validation:
  - Reuse limited data to get a better estimate of a metric

Lecture #8

Cross validation

Jason G. Fleischer, Ph.D.
Asst. Teaching Professor
Department of Cognitive Science, UC San Diego

jfleischer@ucsd.edu

@jasongfleischer

https://jgfleischer.com

Slides in this presentation are from material kindly provided by Shannon Ellis and Sebastian Rashka

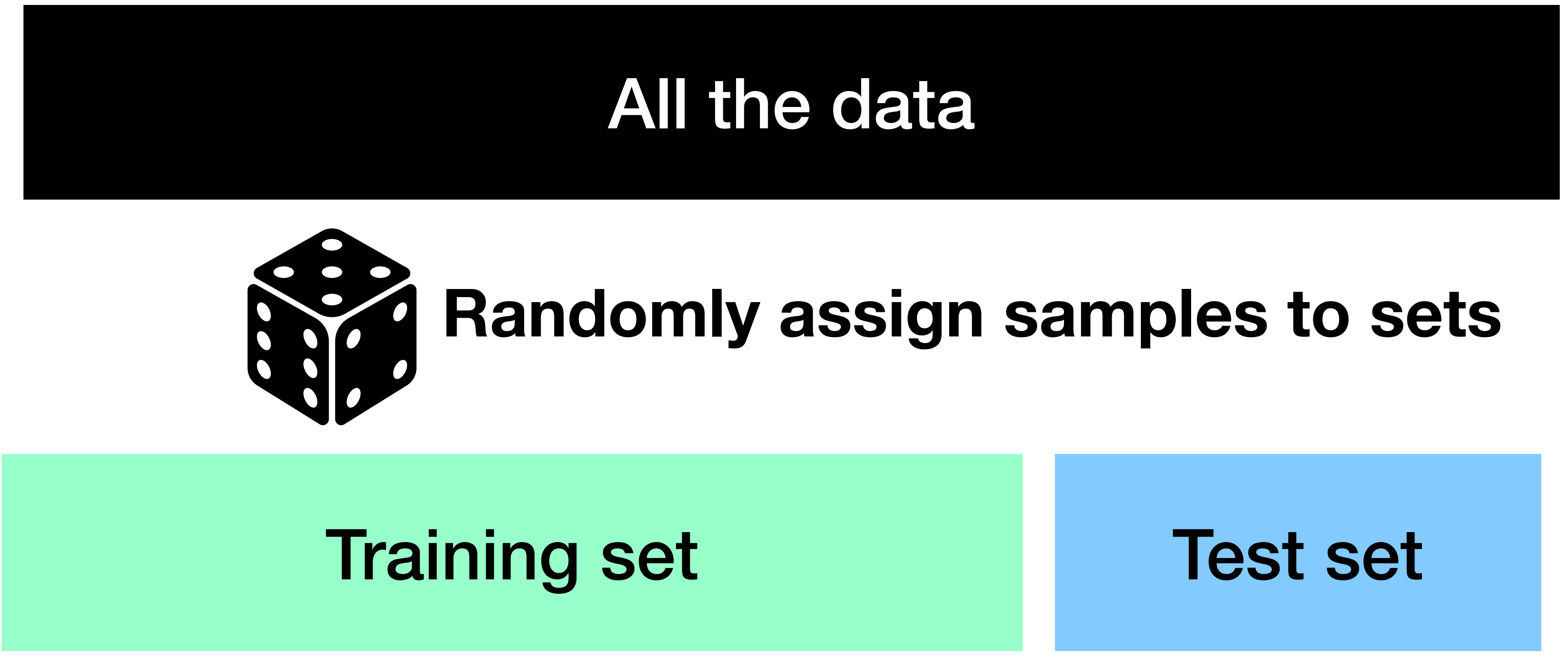
Drowning in data?

Image borrowed from
https://medium.com/@tmaoconnor/marketing-is-drowning-in-data-yet-its-starving-for-answers-f15ee890a732

Limited data?

Evaluating generalization via train - test

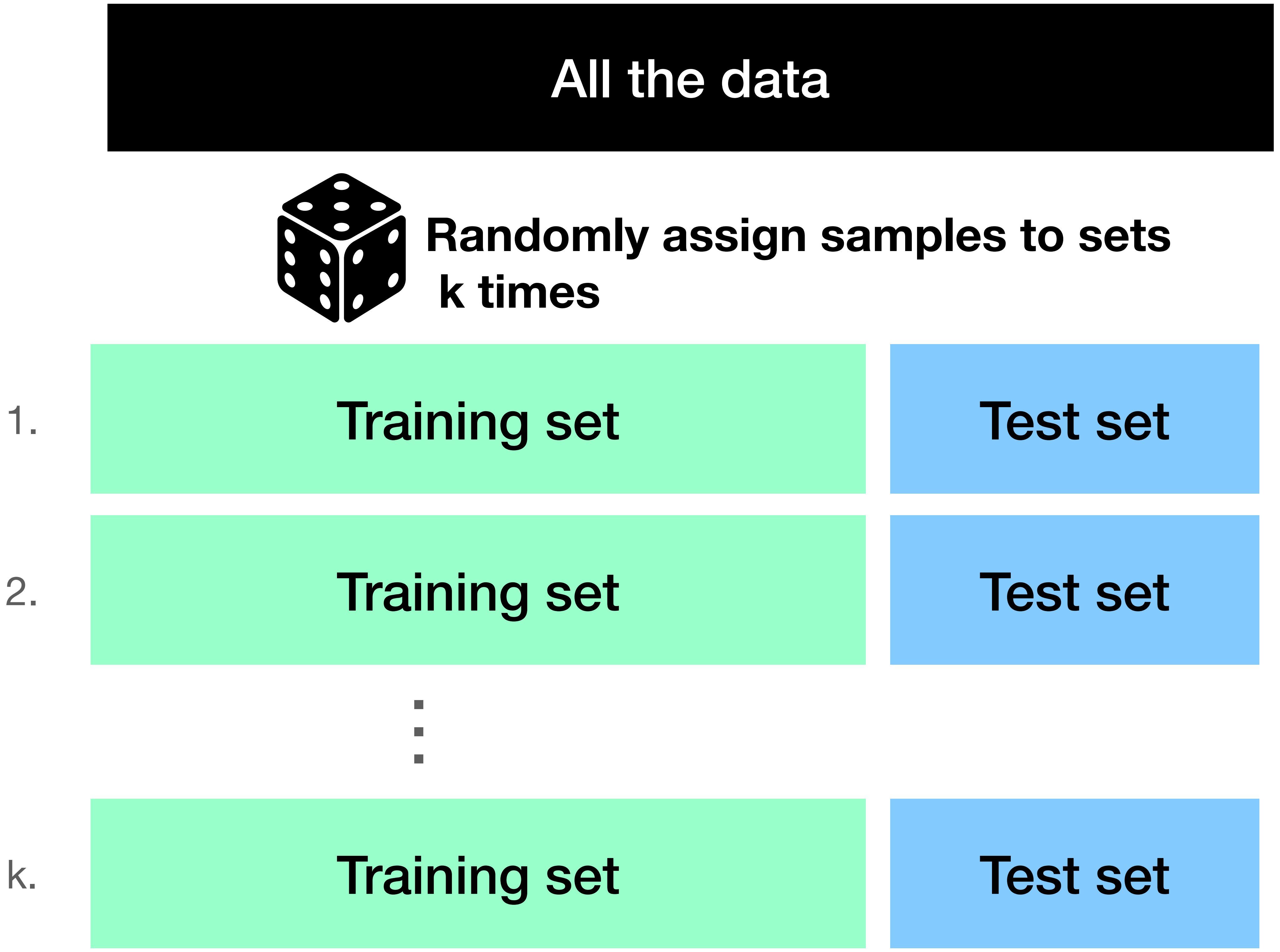
Huge data technique



$$\epsilon_{\text{testing}} = \epsilon_{\text{training}} + \epsilon_{\text{generalization}}$$

Evaluating generalization via k shuffle splits

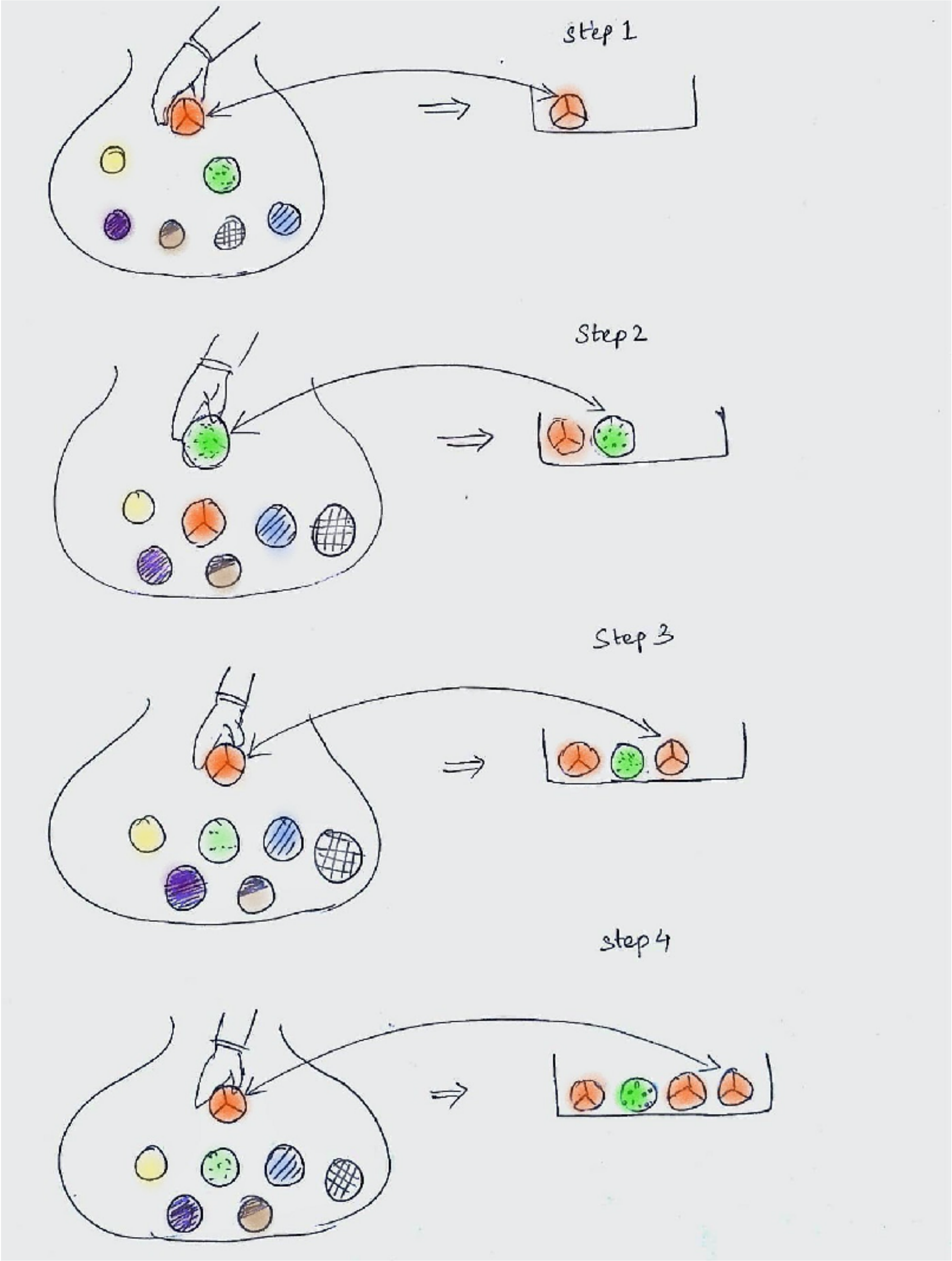
Limited data technique



Use the mean of the k test set performances

$$\bar{\epsilon}_{\text{testing}} = \epsilon_{\text{training}} + \epsilon_{\text{generalization}}$$

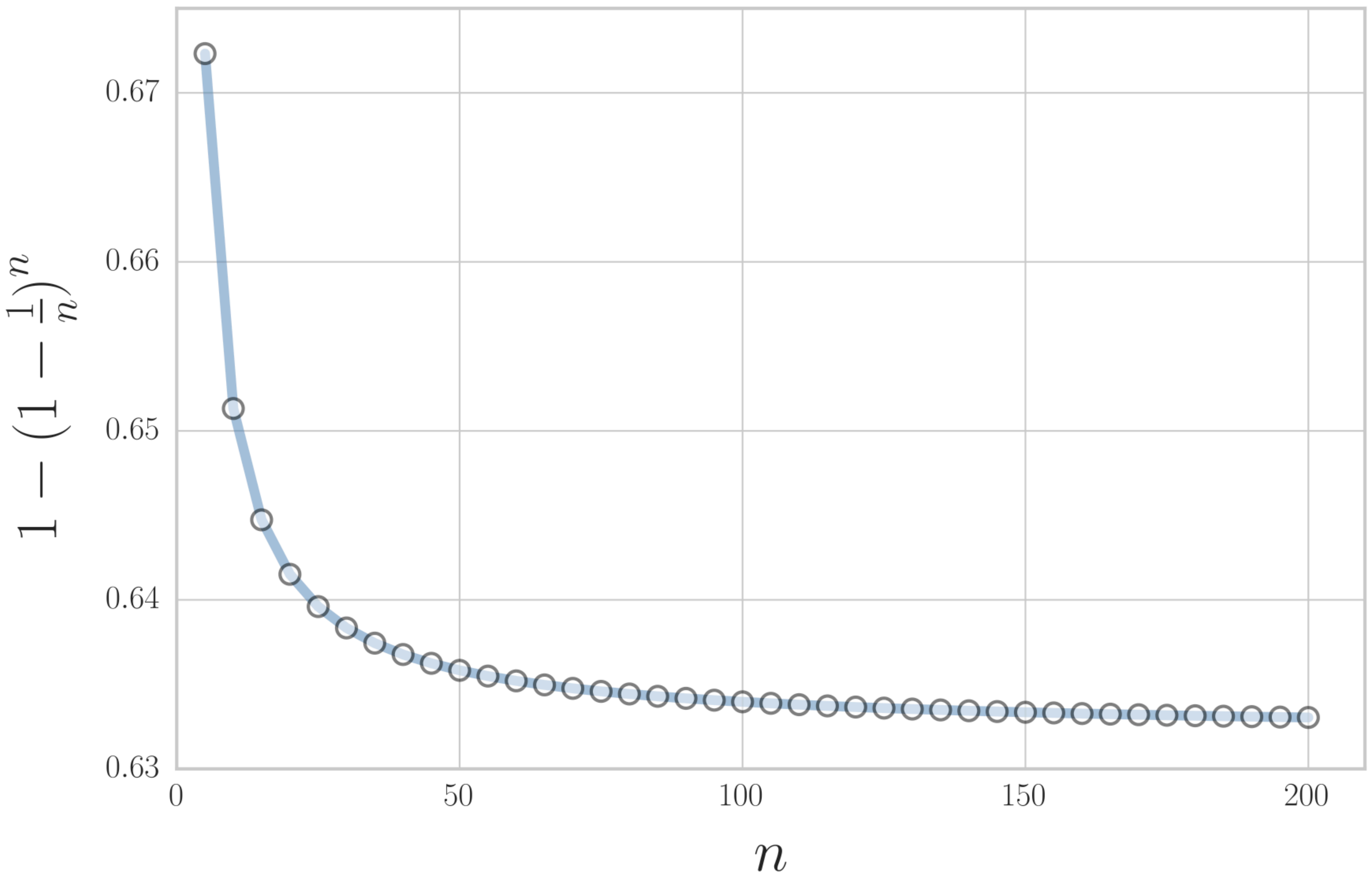
$$\bar{\epsilon}_{\text{testing}} = 1/k \sum_{i \in [0,k]} \epsilon_{\text{test}_i}$$



$$P(\textbf{not chosen}) = \left(1 - \frac{1}{n}\right)^n,$$

$$\frac{1}{e} \approx 0.368, \quad n \to \infty \,.$$

$$P(\textbf{chosen}) = 1 - \left(1 - \frac{1}{n}\right)^n \approx 0.632$$



Evaluating generalization via bootstrap sampling

Limited data technique

All the data

Build k bootstrap samples

1. Training set | Out of bag set
2. Training set | Out of bag set

⋮

k. Training set | Out of bag set

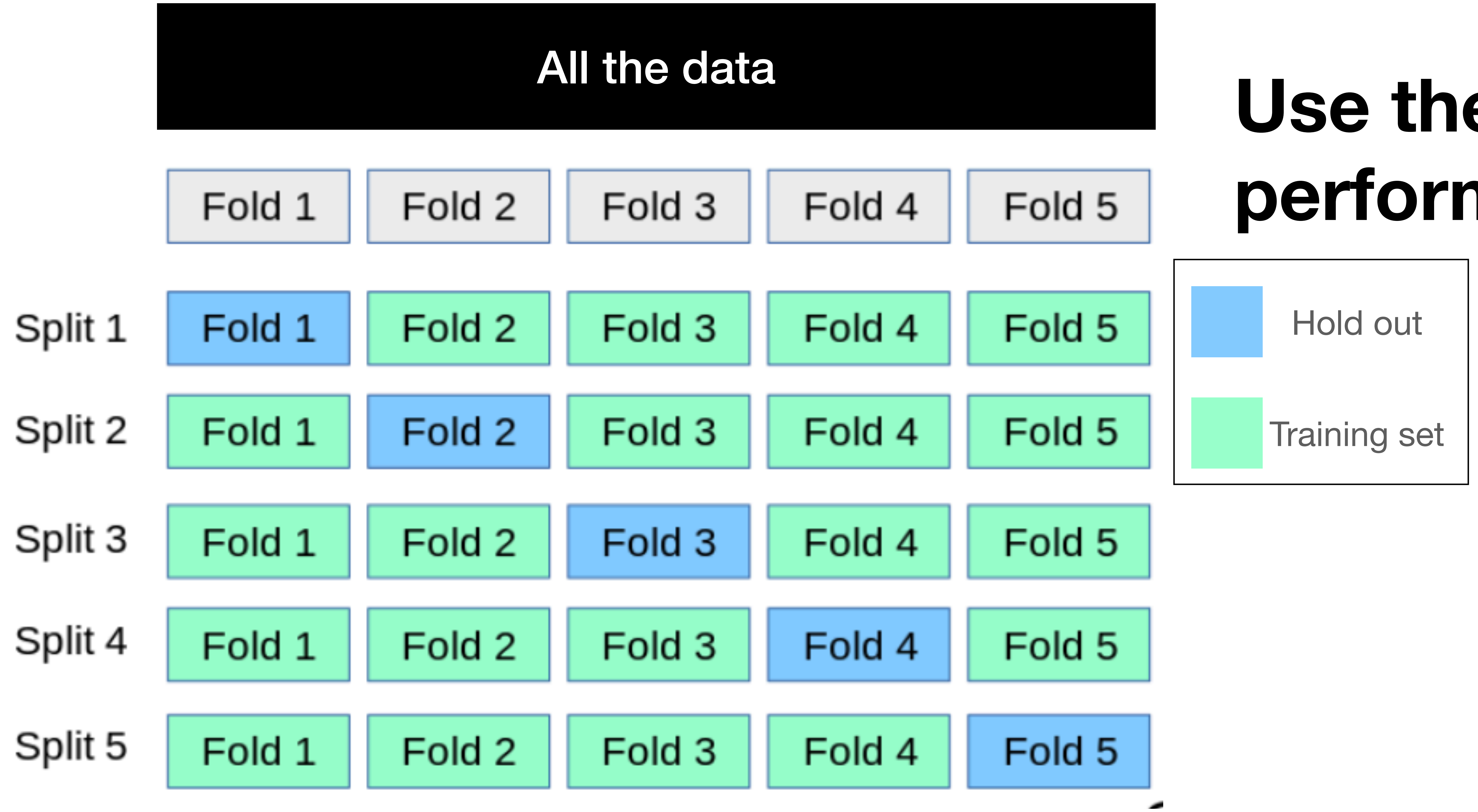
Use the a corrected mean of the n performances, combining a known overestimate with a known underestimate

$$\bar{\epsilon}_{.632+} = 1/k \sum_{i \in [0,k]} \left(\omega * \epsilon_{\mathrm{OOB}_i} + (1-\omega) * \epsilon_{\mathrm{training}_i} \right)$$

$$\omega = \frac{.632}{(1-.368)R}, R = -\frac{\epsilon_{\mathrm{OOB}_i} - \epsilon_{\mathrm{training}_i}}{\gamma - (1-\epsilon_{\mathrm{OOB}_i})},$$ where γ is a constant calulated on the dataset

Evaluating generalization via k-folds cross-validation

Limited data technique

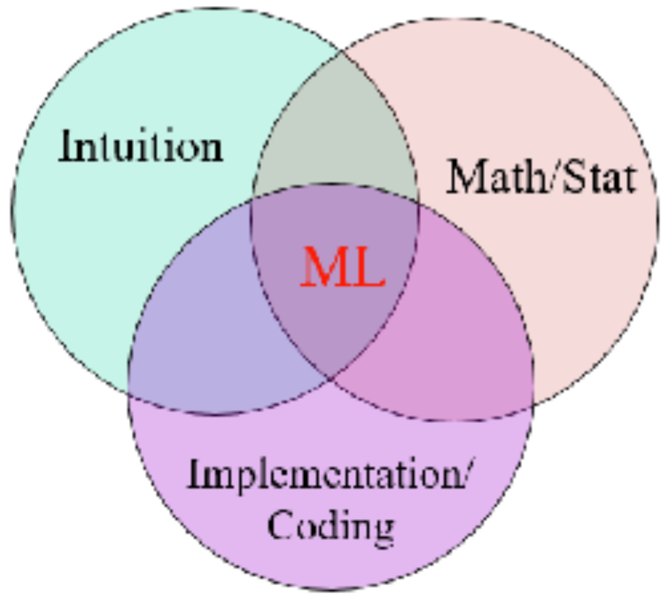


Use the mean of the k hold out set performances

$$\bar{\epsilon}_{\text{testing}} = \epsilon_{\text{training}} + \epsilon_{\text{generalization}}$$

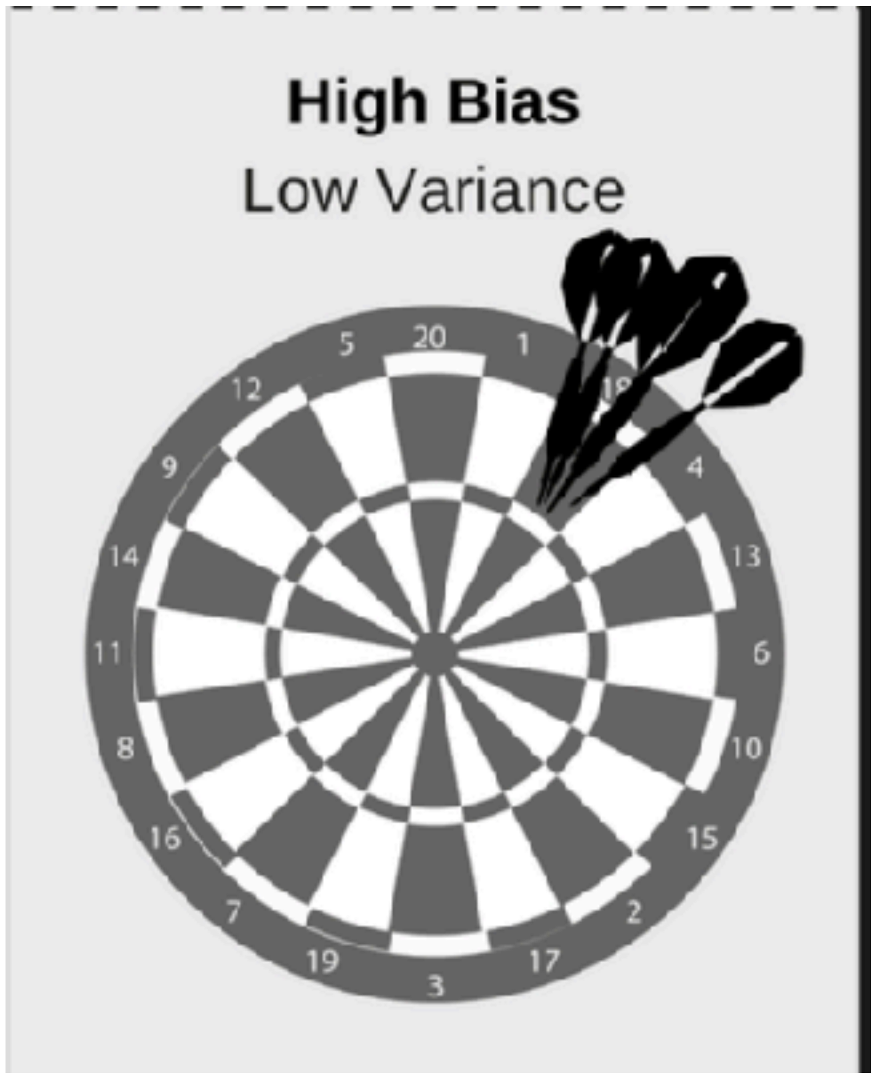
$$\bar{\epsilon}_{\text{testing}} = 1/k \sum_{i \in [0,k]} \epsilon_{\text{hold out}_i}$$

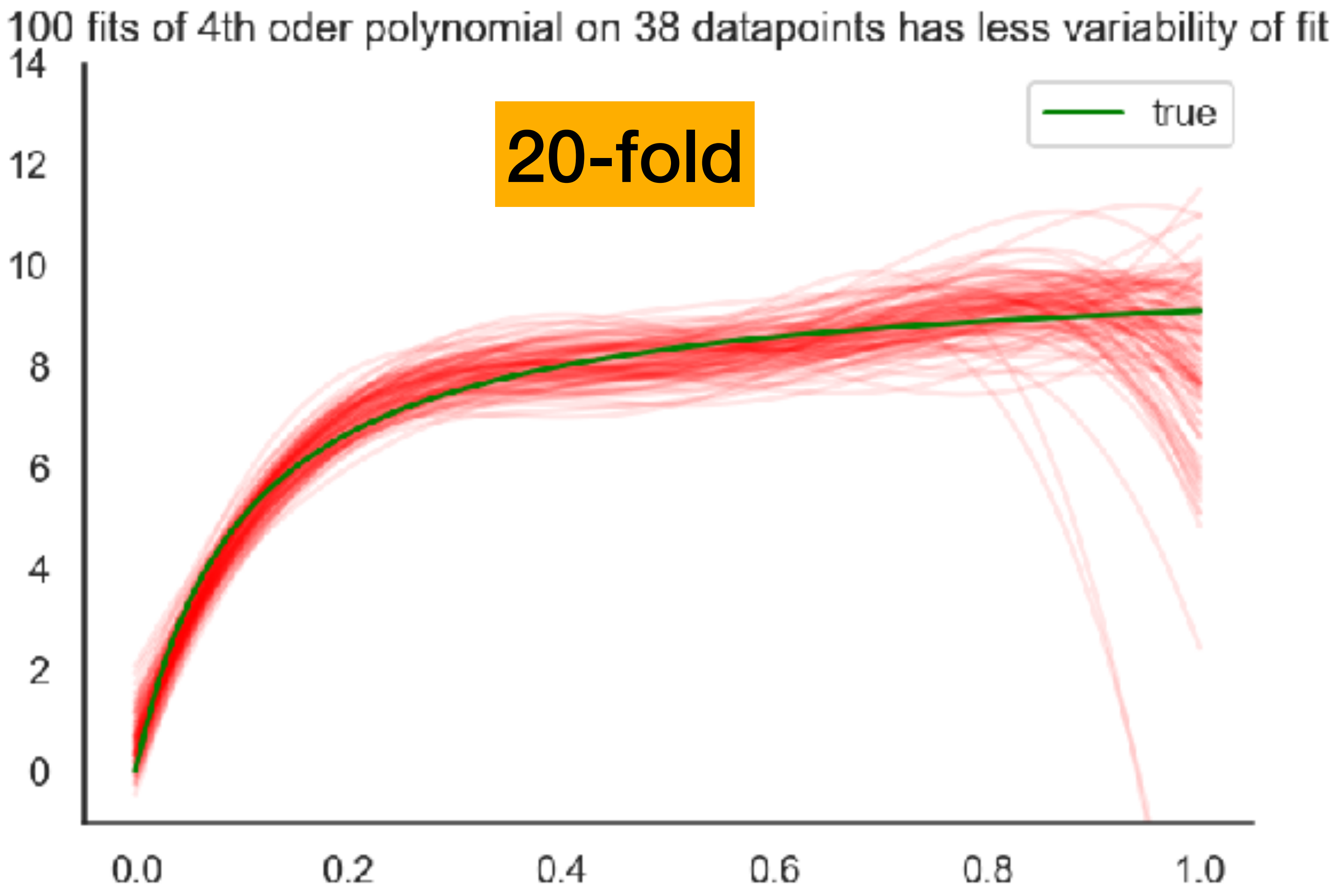
When k = # samples, this is Leave One Out Cross Validation



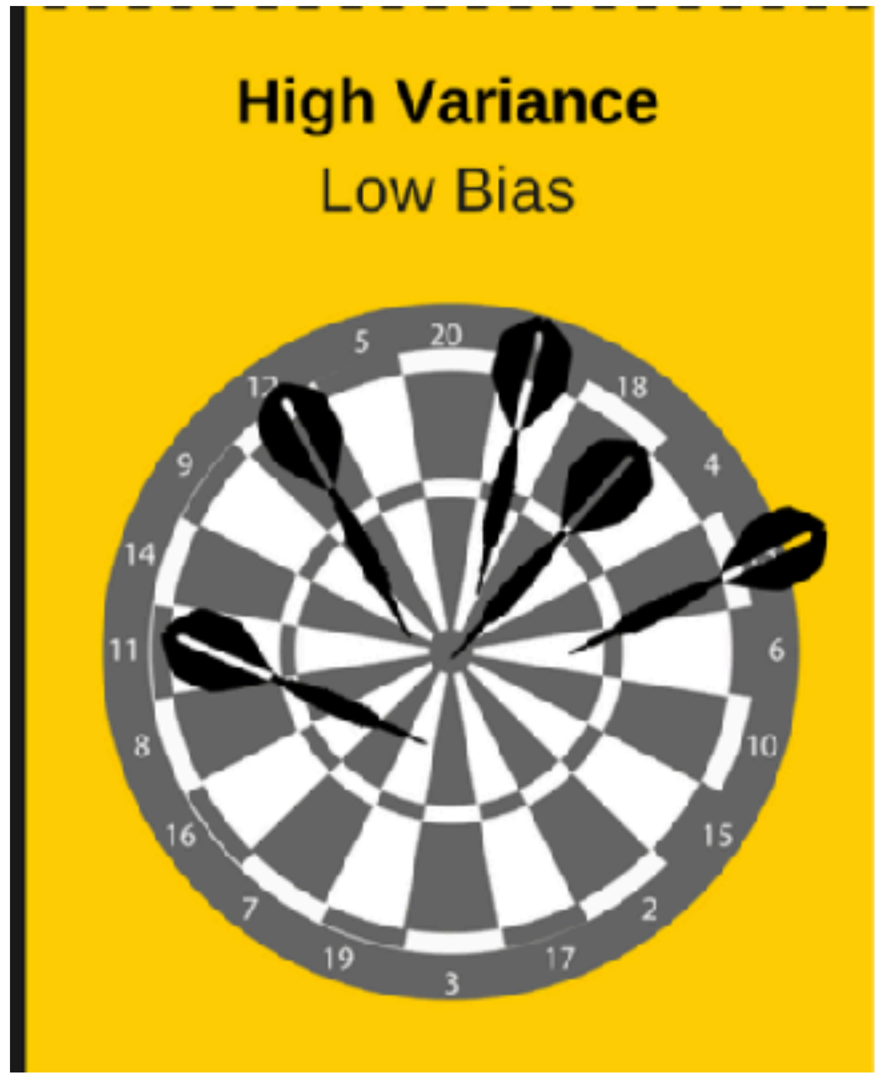
Bias-variance tradeoff in cross validation

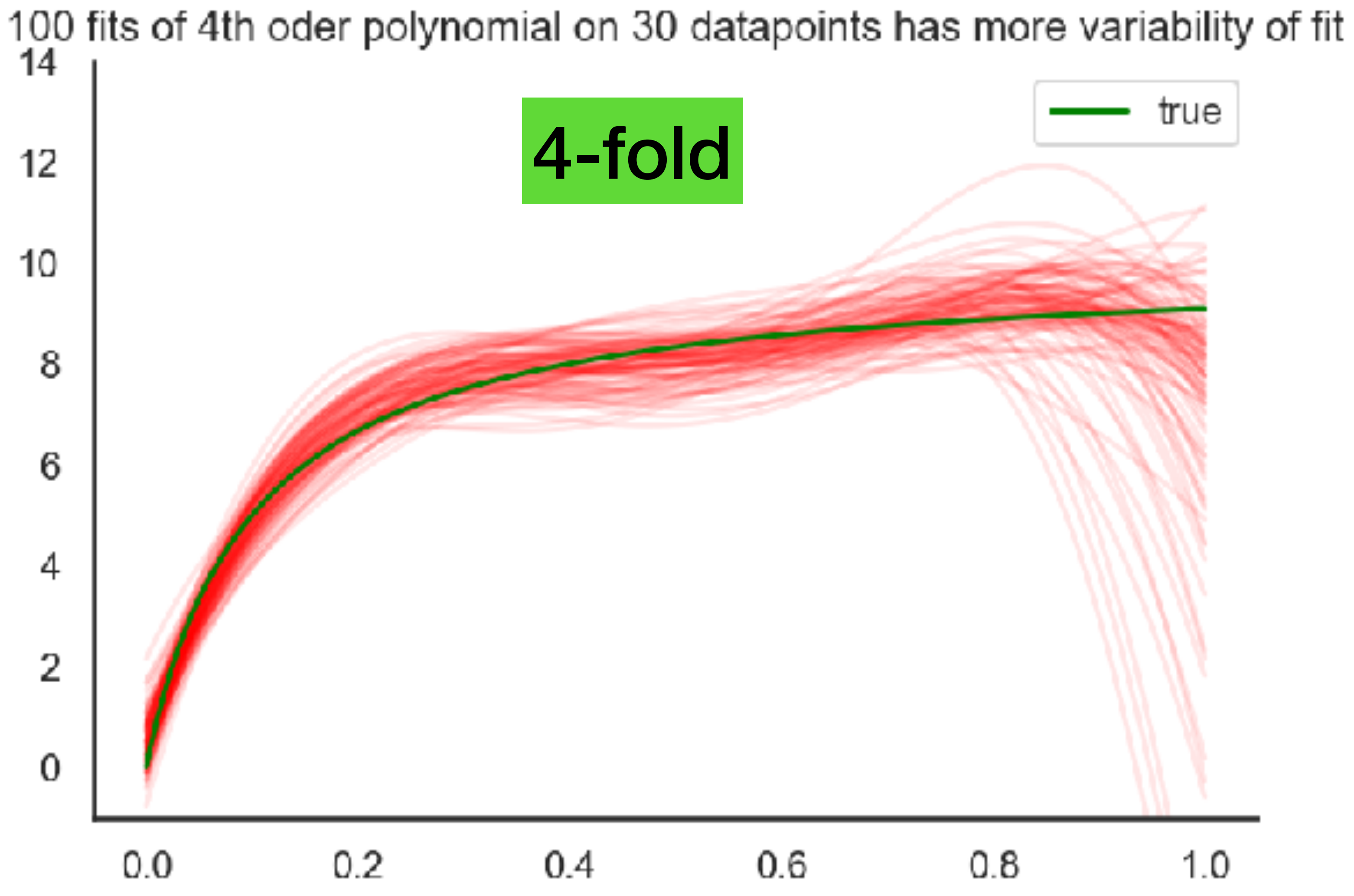
In terms of sample size effects on fit



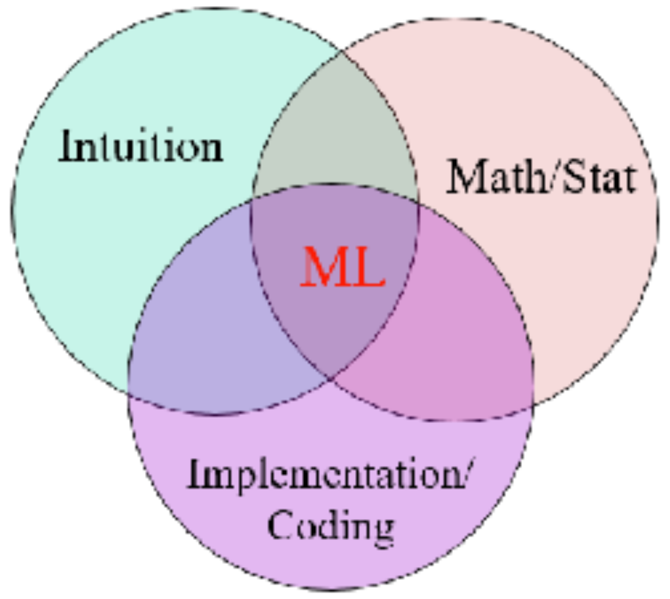


Larger training set size
(K-folds as k gets large -> LOOCV)
more uniformity of fit





Smaller training set size
(K-folds as k gets small)
more variety of fit



Bias-variance tradeoff in cross validation

In terms of error estimate its the opposite, ⬆️ fit variability == ⬇️ estimate var