

Week 8: Mixture Models & Expectation-Maximisation

Matthew Caldwell

COMP0088 Introduction to Machine Learning • UCL Computer Science • Autumn 2024

Admin

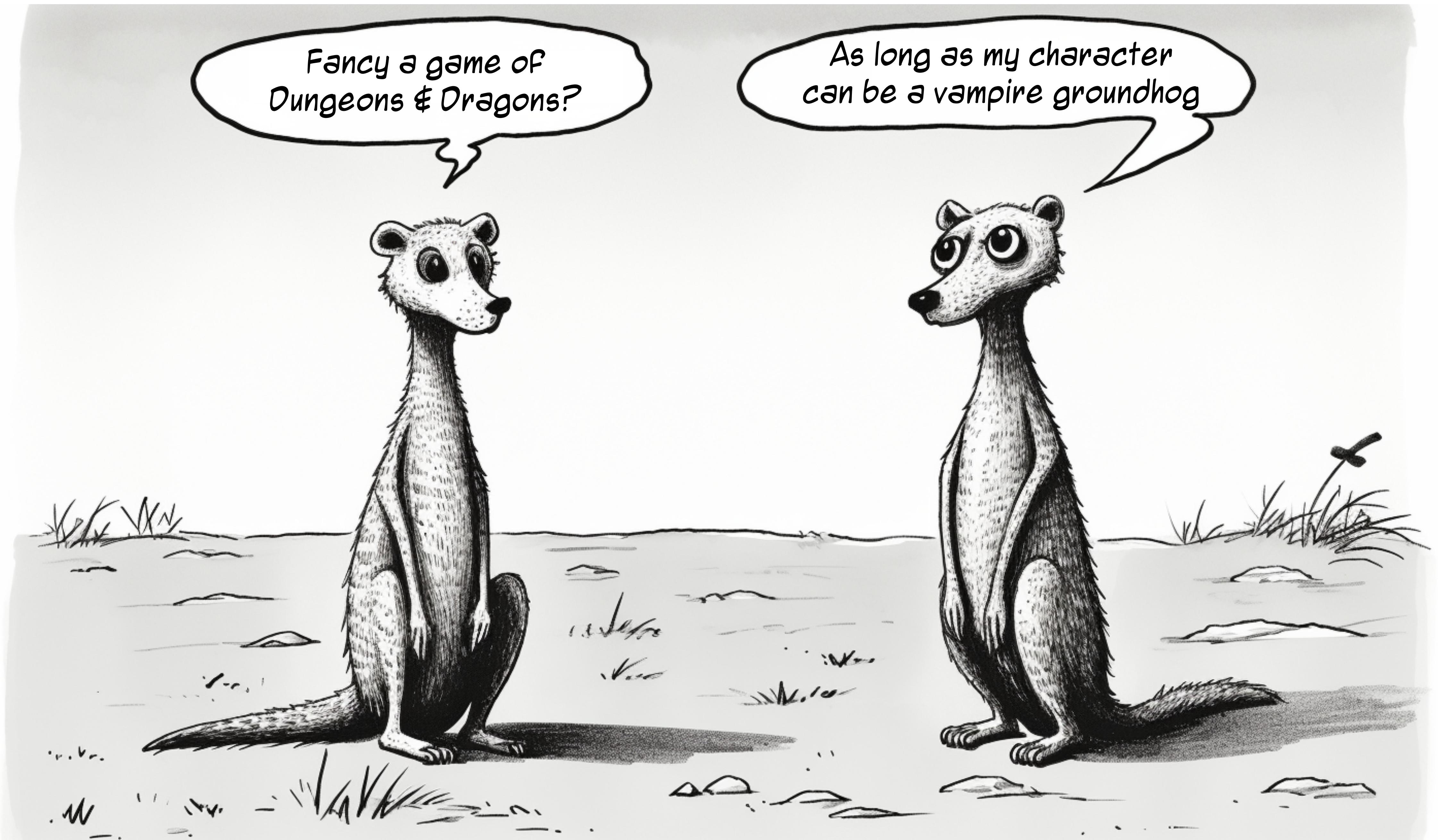
- Good news: no more lab exercises!
- Bad news: exam practice
- Group 1 will be shuffled into Groups 2 & 3
- All sessions will be in 66GS G01

Week 8 Recap

The Invisible Enemy

Fancy a game of
Dungeons & Dragons?

As long as my character
can be a vampire groundhog



d6



d20



Dice Model 1

- **Repeat**
 - **Draw a die**
 - **Roll it and record result**
 - **Return it to the bag**

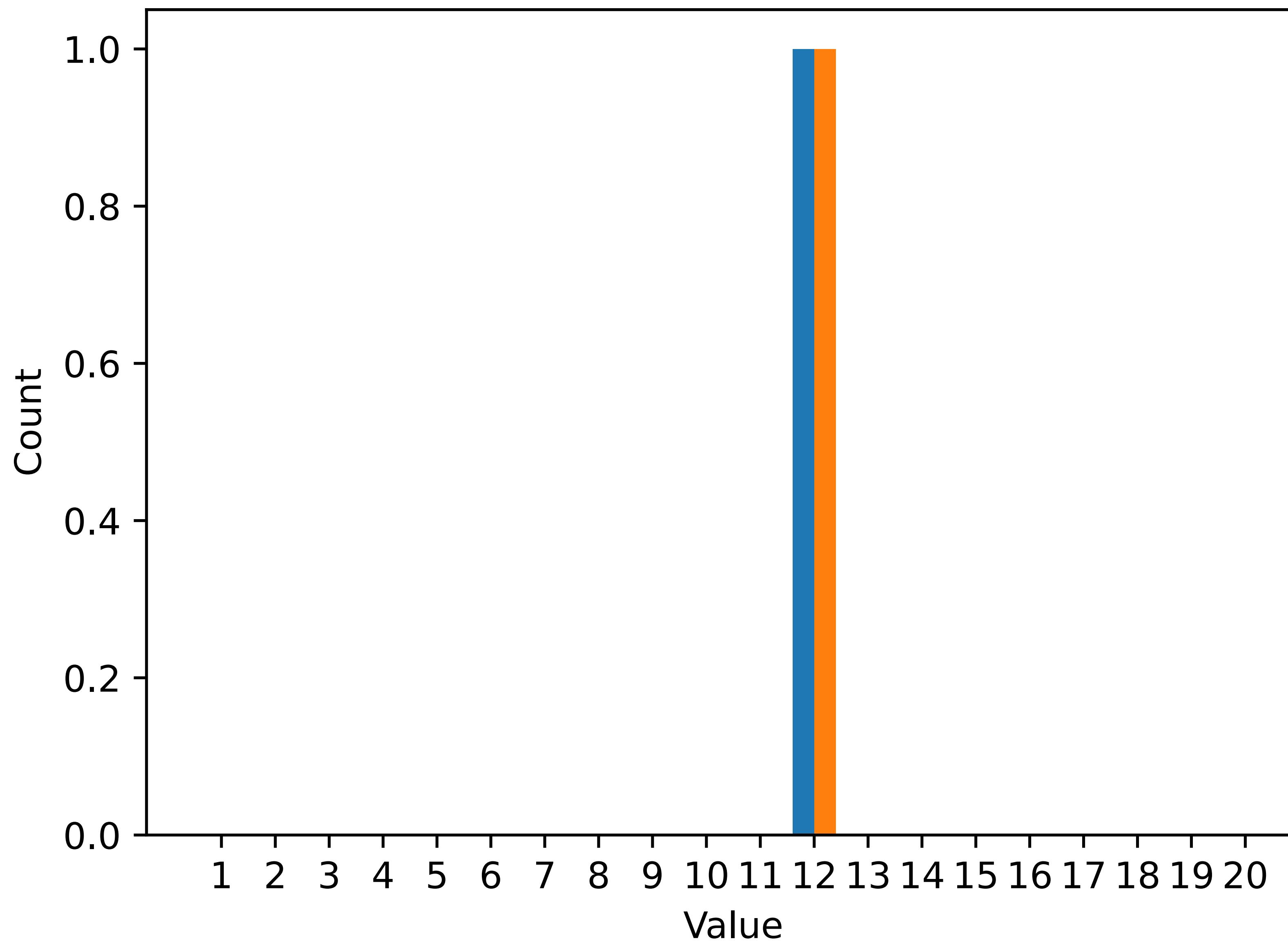
Dice Model 2

- **Draw a die**
- **Repeat**
 - **Roll it and record result**
 - **Roll again and don't record result**
 - **If second roll=1, return die to bag and draw again**

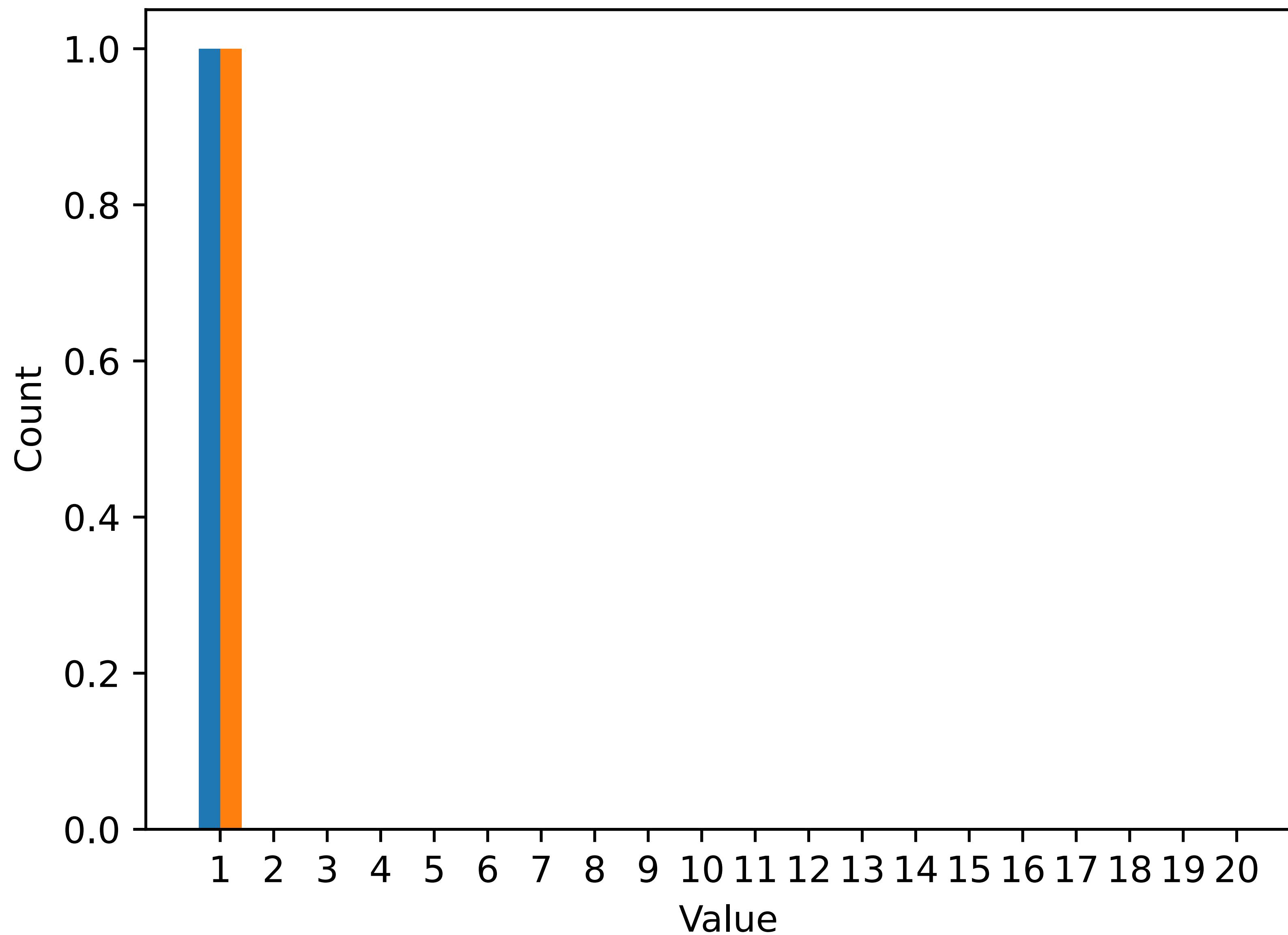
Model 1
Independent Mixture

Model 2
Hidden Markov

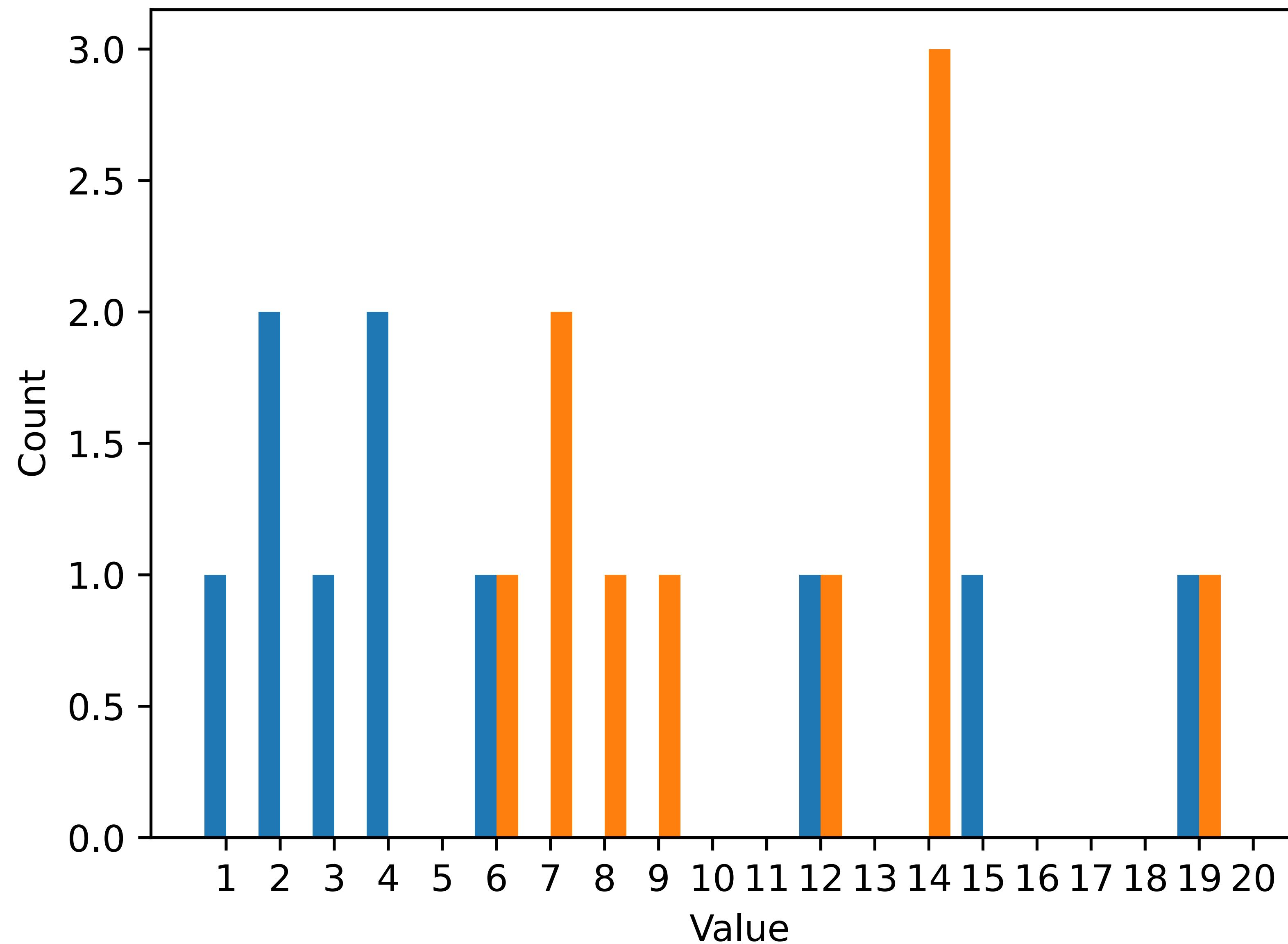
1 Throw



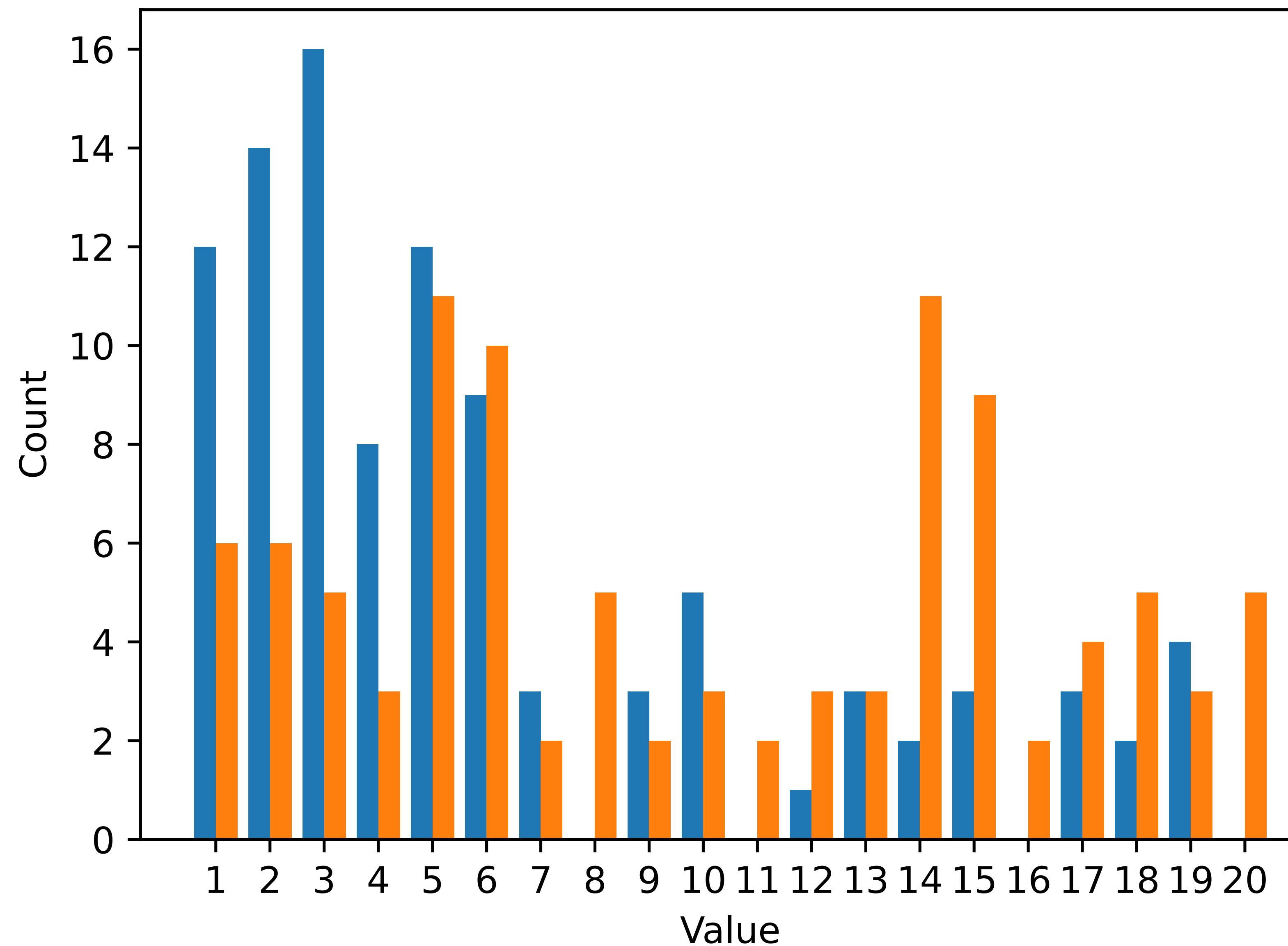
1 Throw



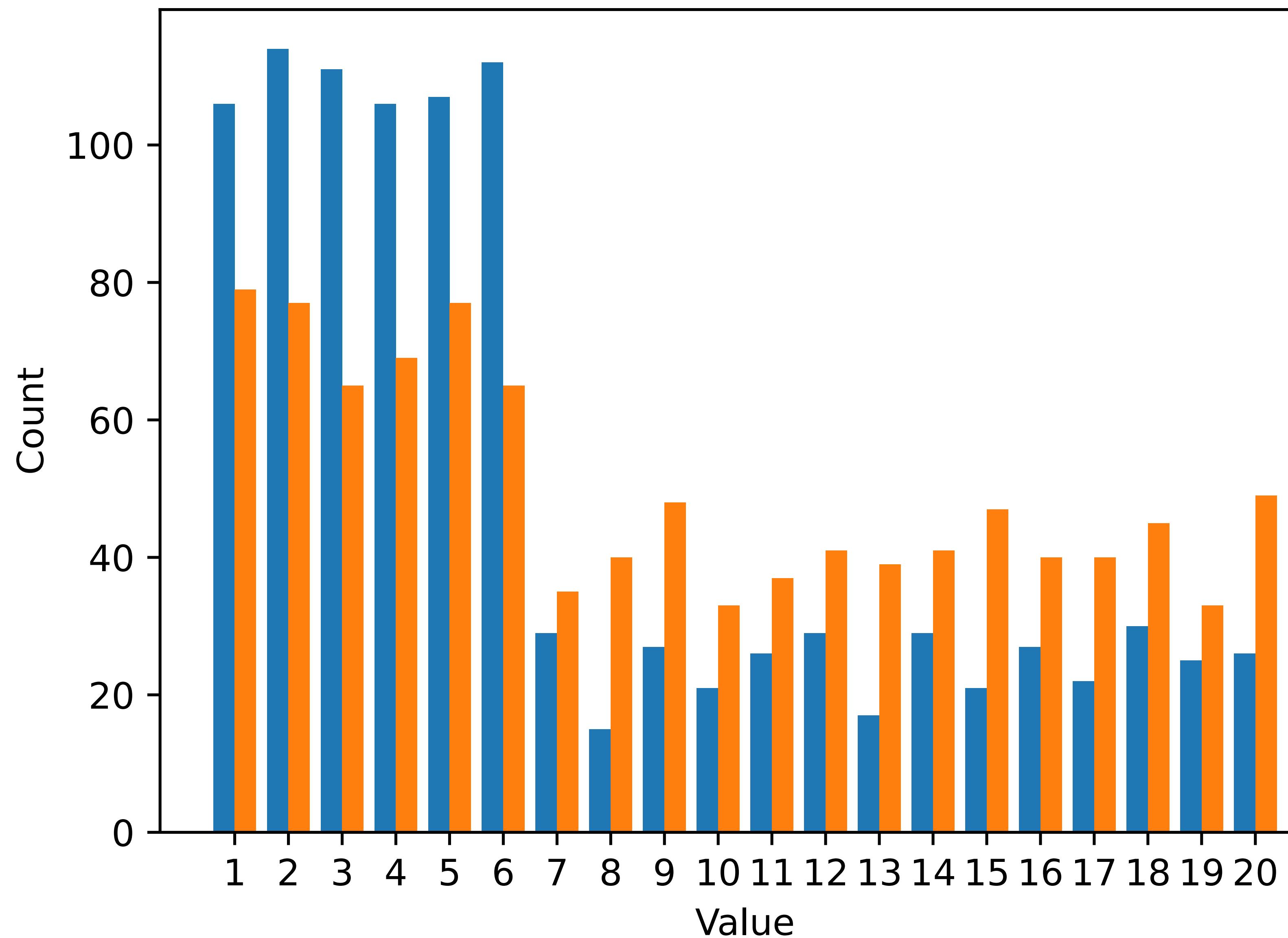
10 Throws



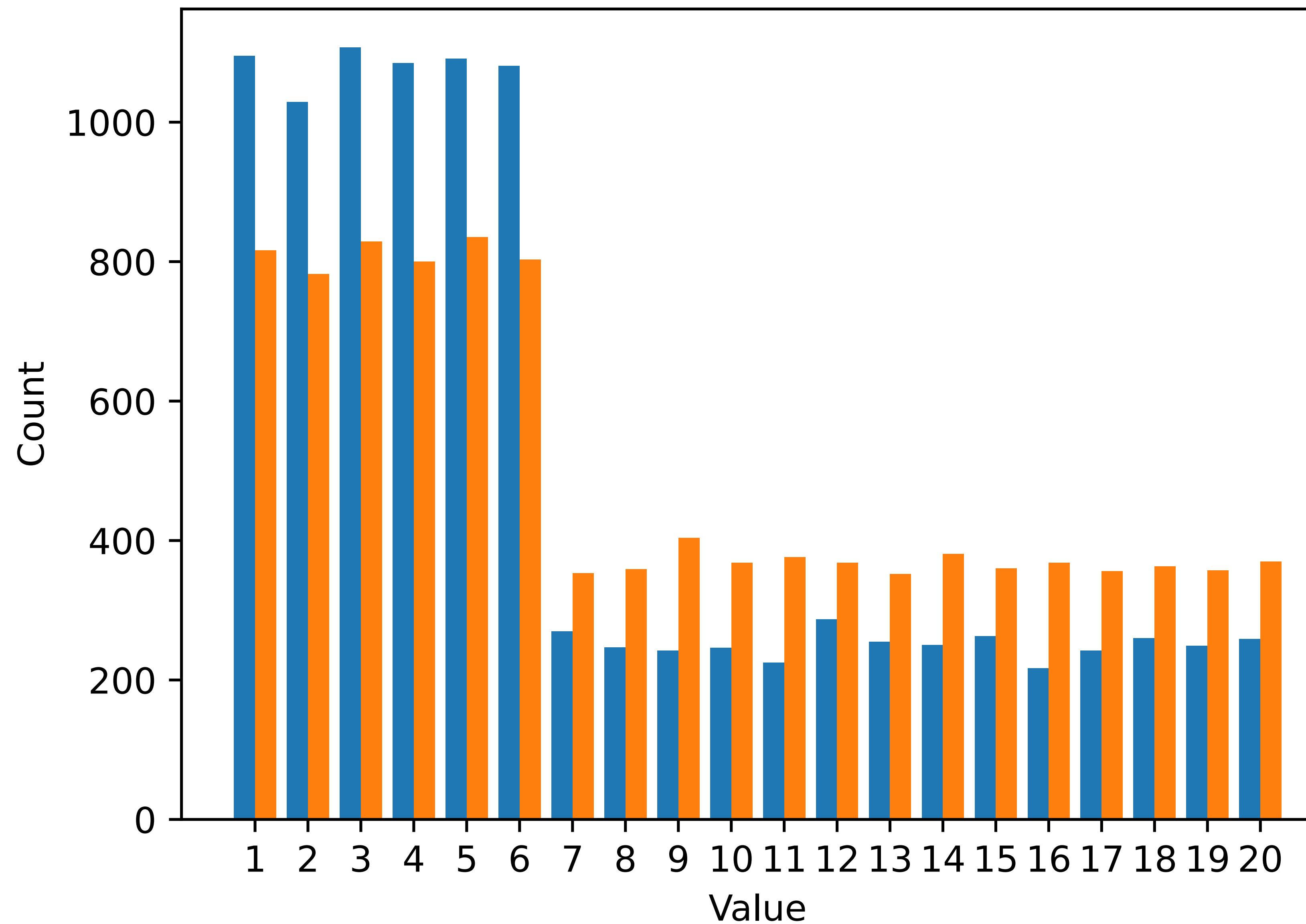
100 Throws



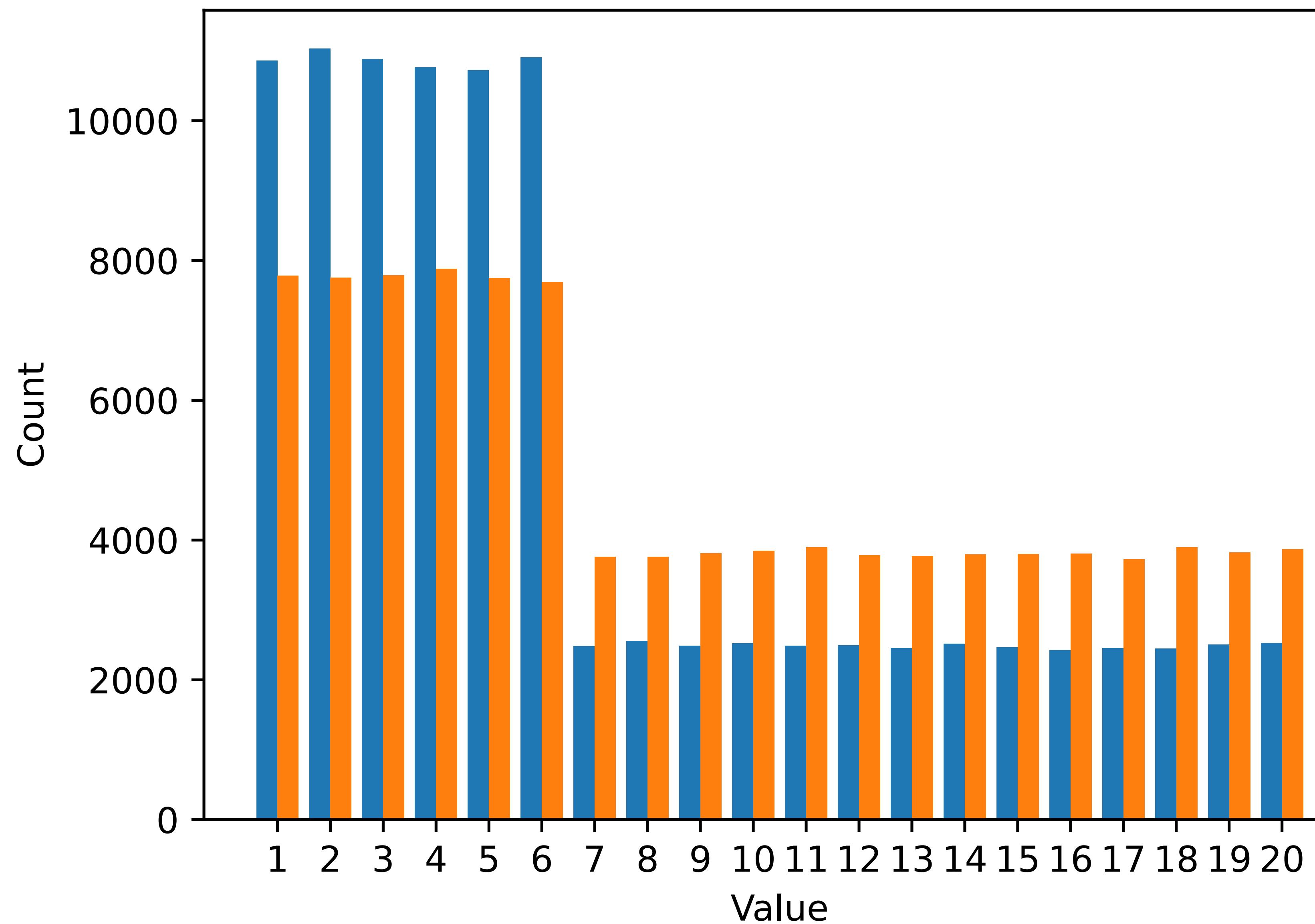
1000 Throws



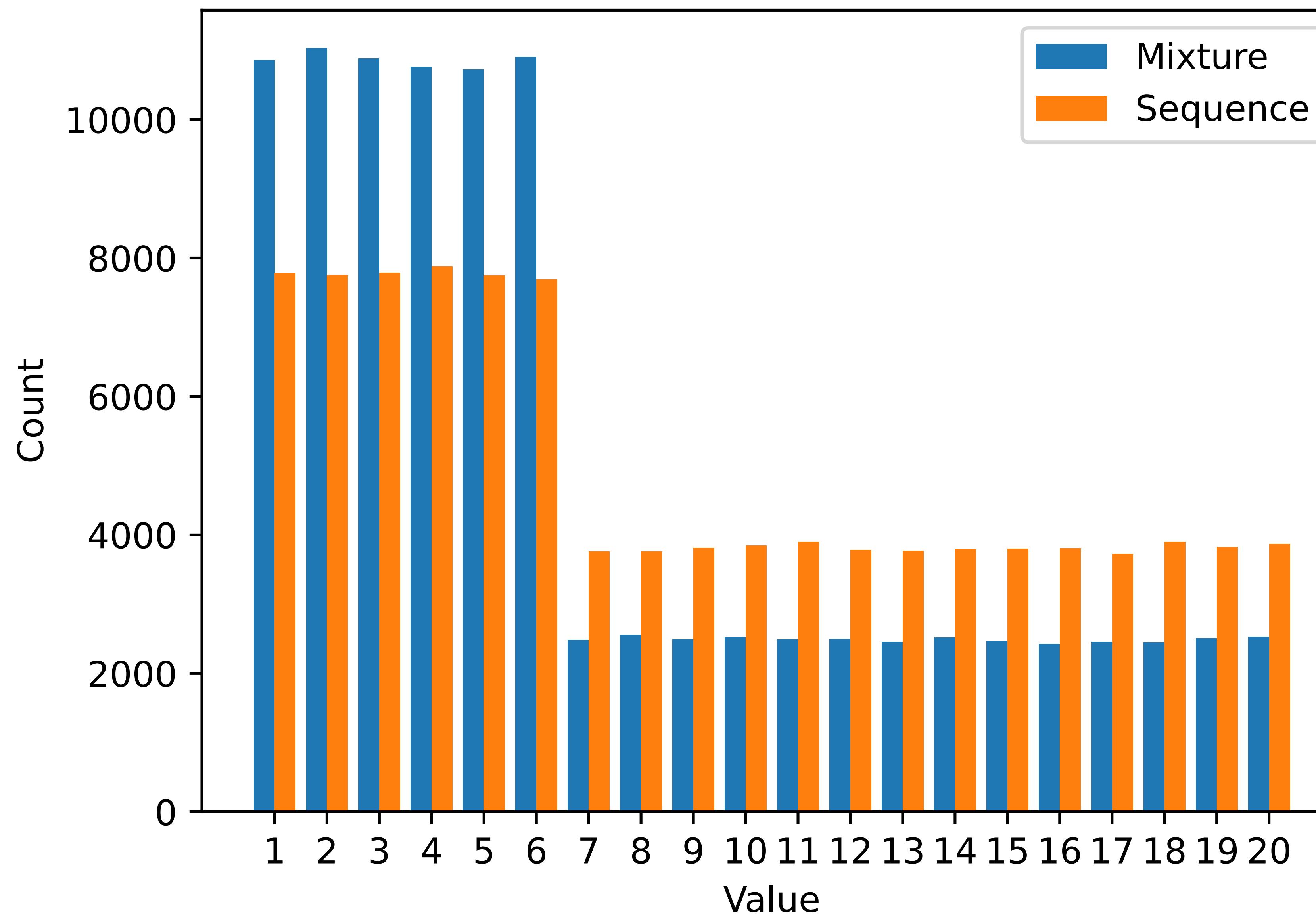
10000 Throws



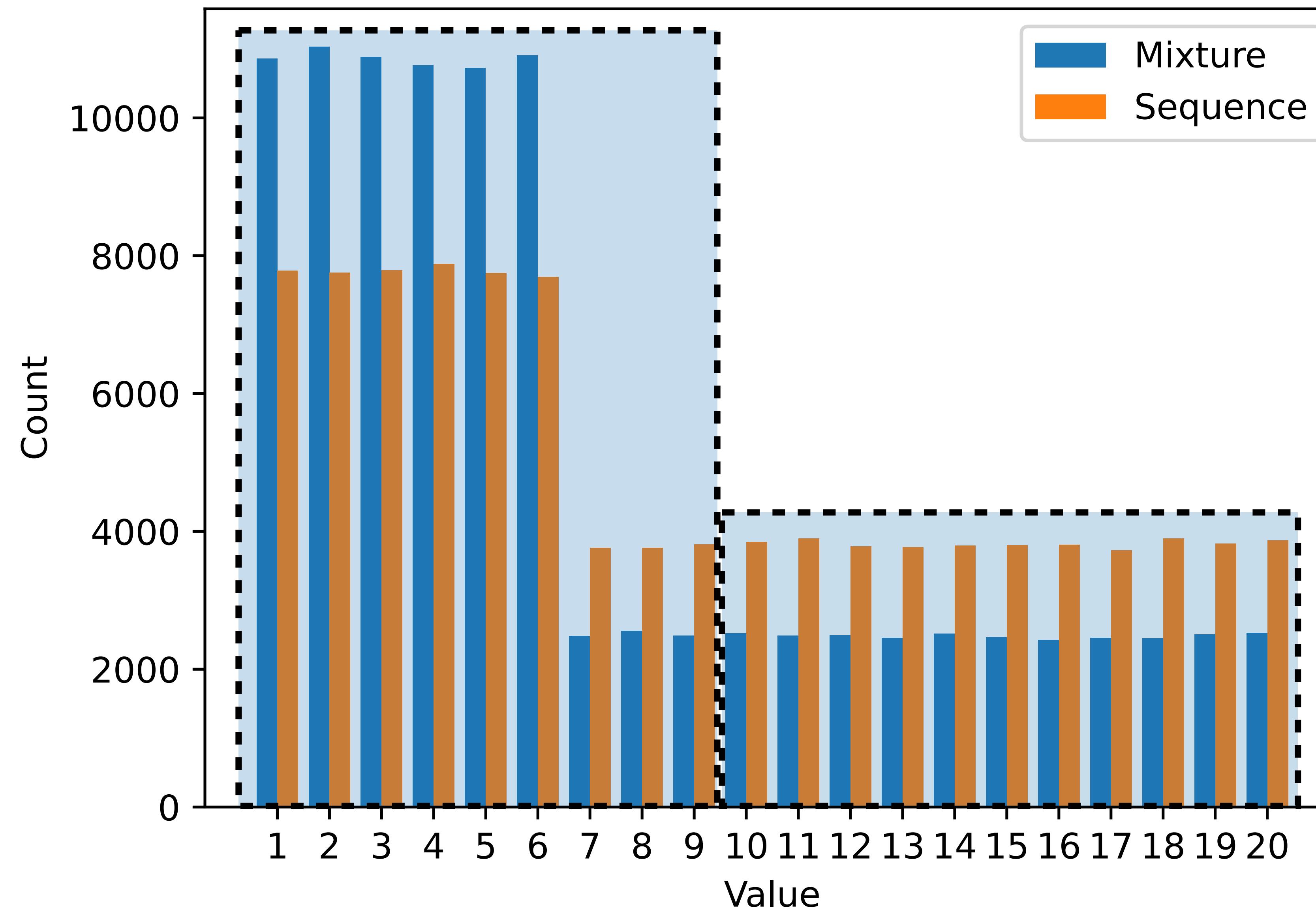
100000 Throws



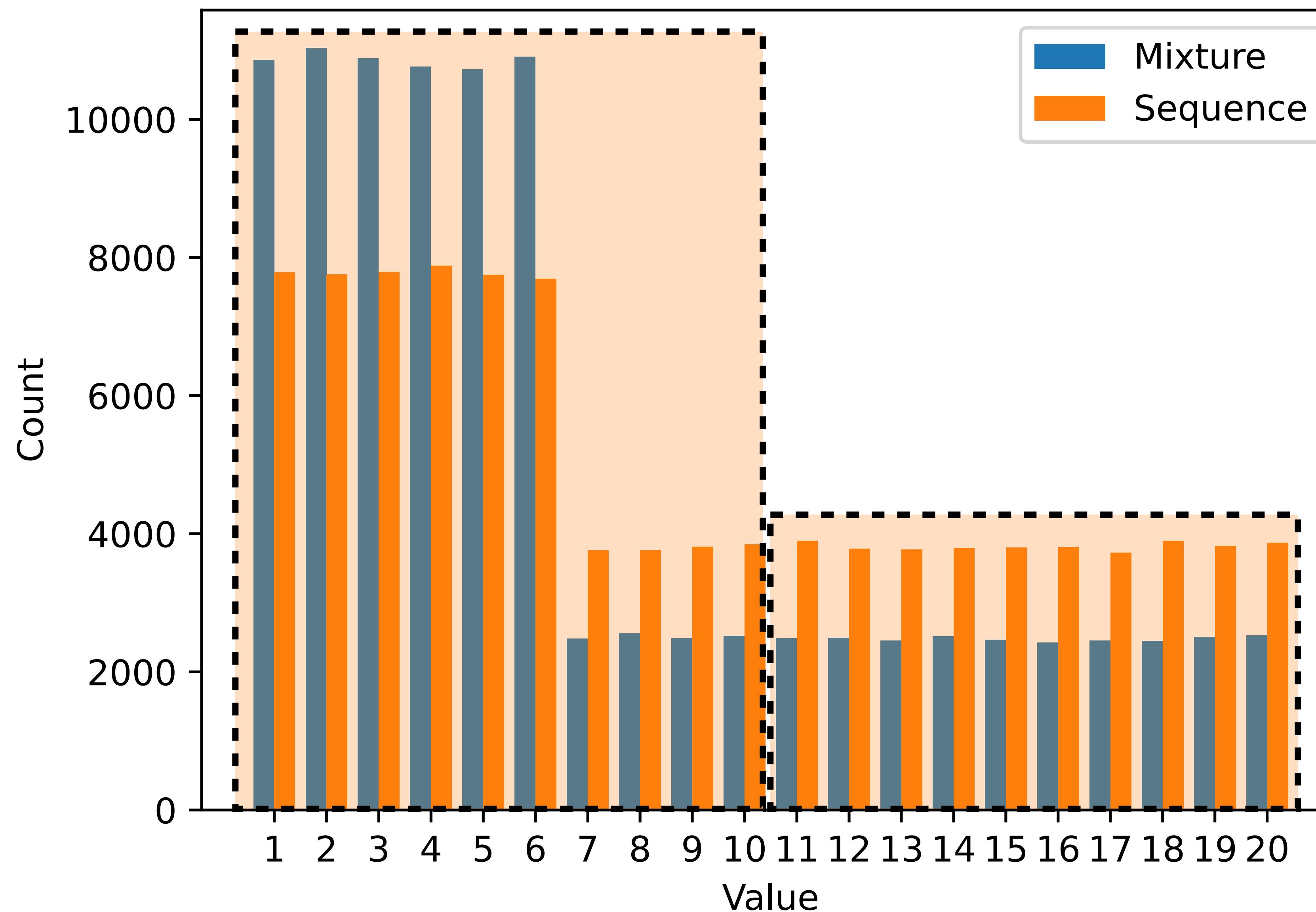
100000 Throws



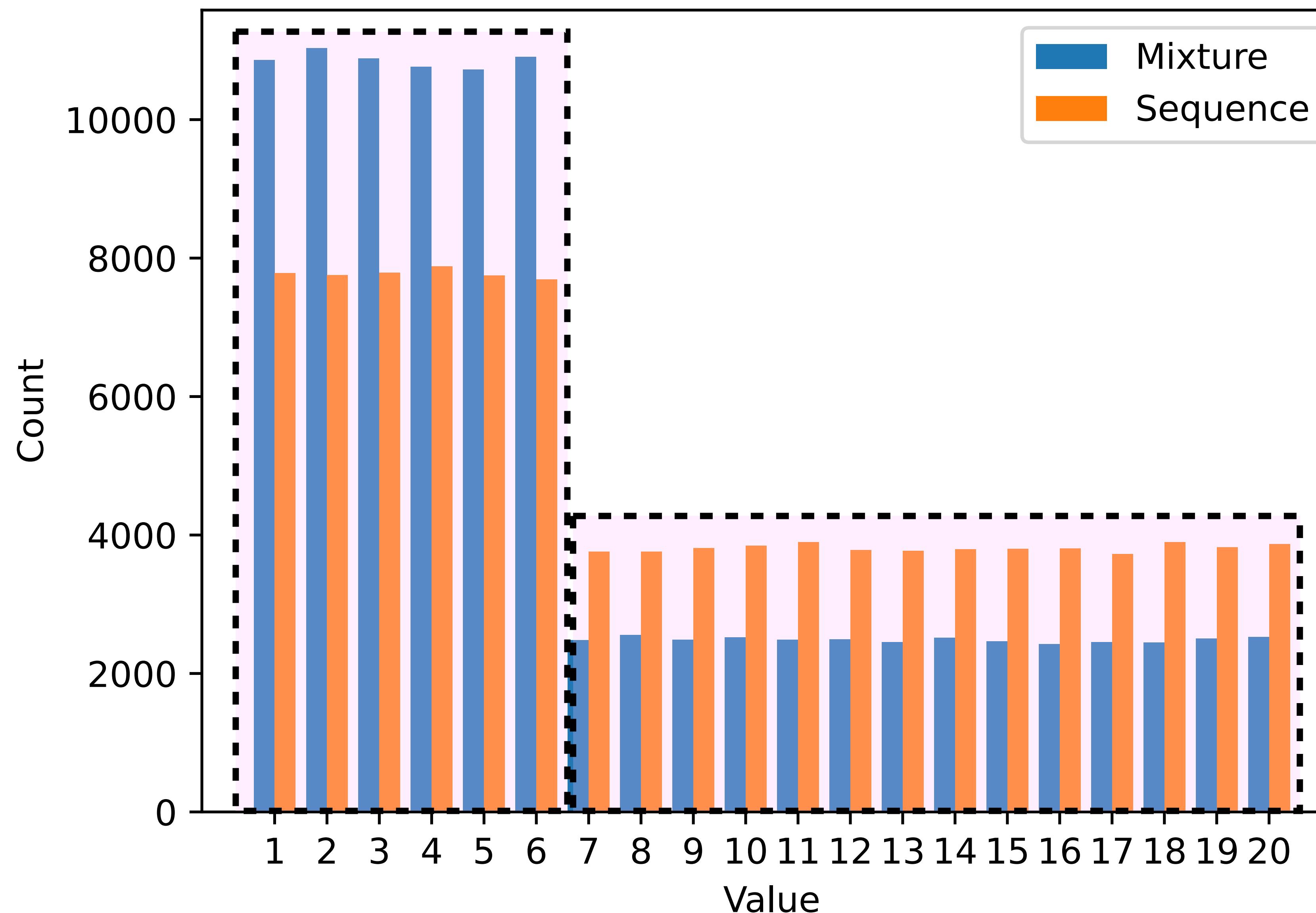
100000 Throws



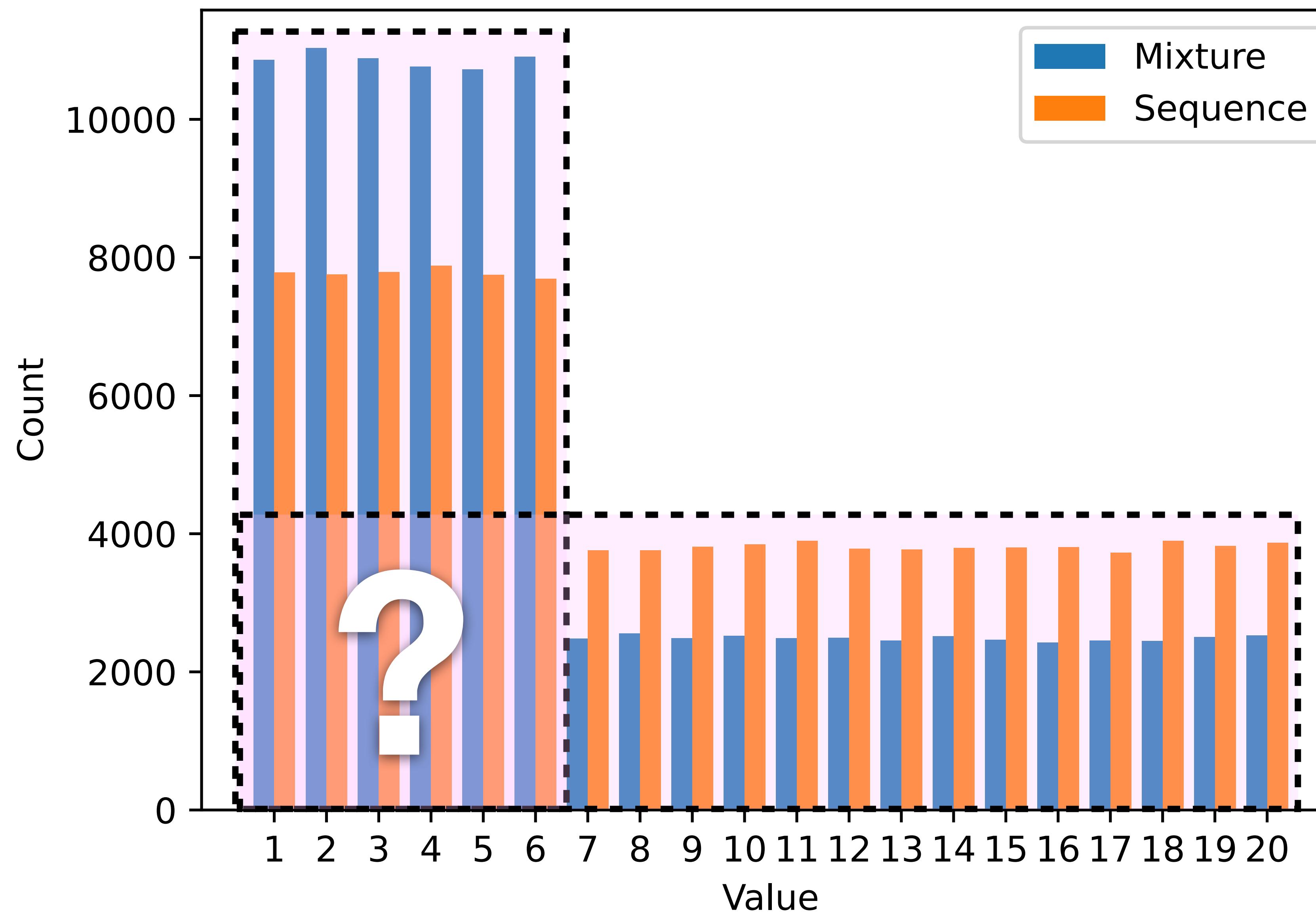
100000 Throws



100000 Throws



100000 Throws



■ ■ ■

13

2

17

■ ■ ■

... ■ ■ ■

13



2



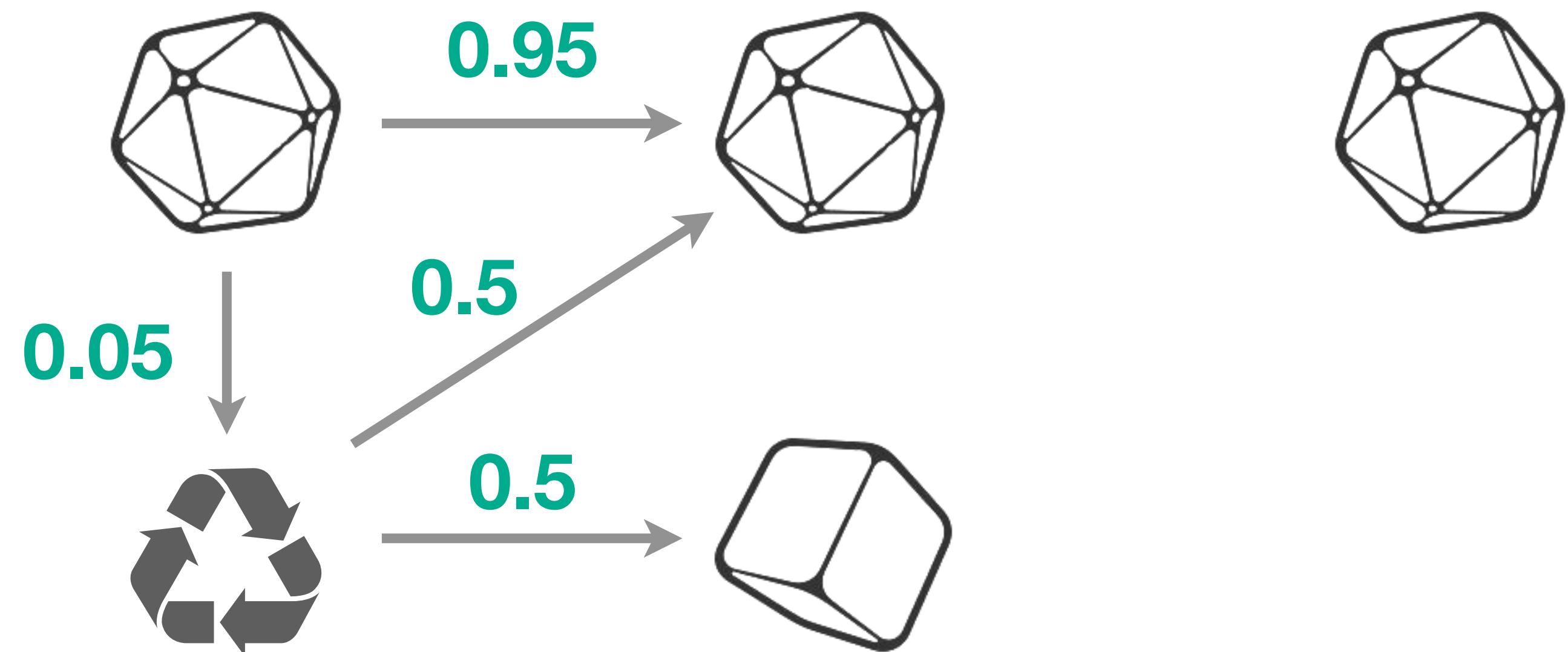
17



■ ■ ■ ...

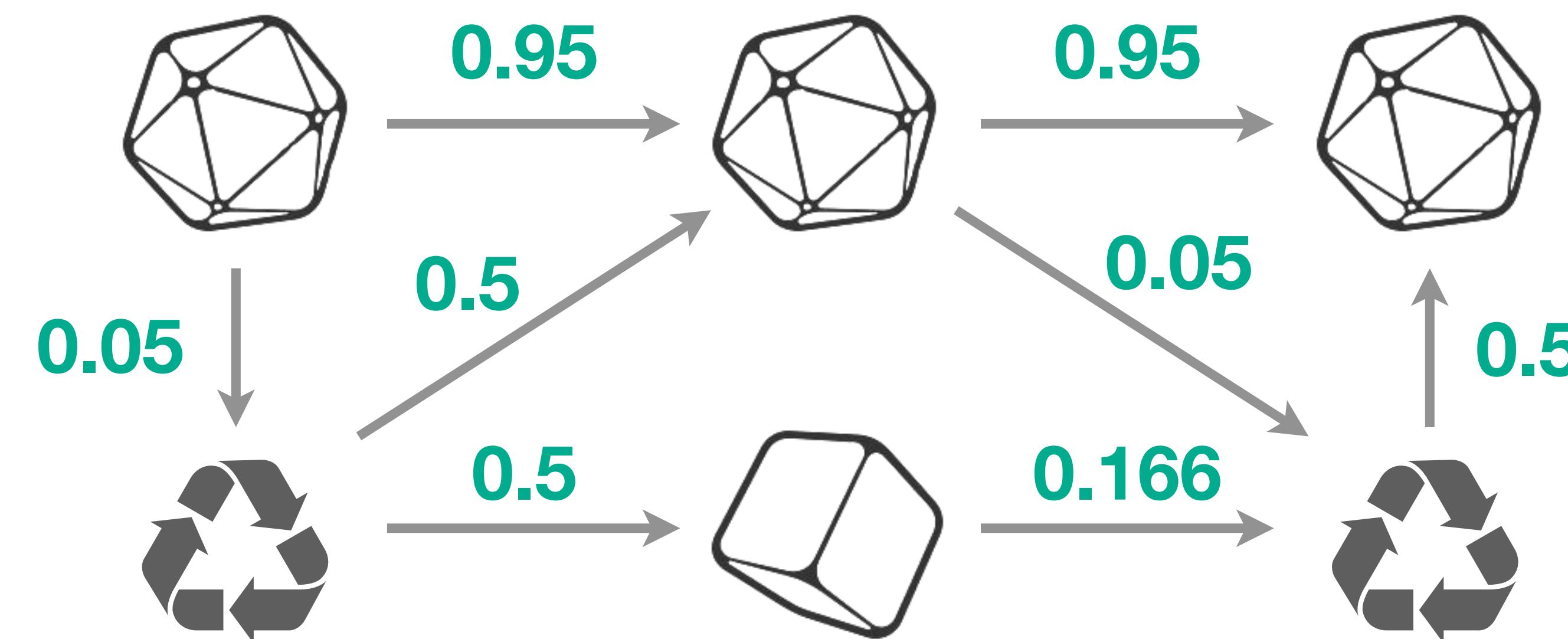


... 13 2 17 ...



“Filtering”

... 13 2 17 ...



“Smoothing”

Gaussian Mixture Models

$$\tilde{\mathbf{x}}_i = \{ \mathbf{x}_i, z_i \}$$

$\mathbf{x}_i \in \mathbb{R}^d \rightarrow$ observed

$z_i \in \{1, 2, \dots, k\} \rightarrow$ invisible

$$z_i \sim \text{Cat}(k, \alpha)$$

$$\mathbf{x}_i | (z_i = j) \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$$

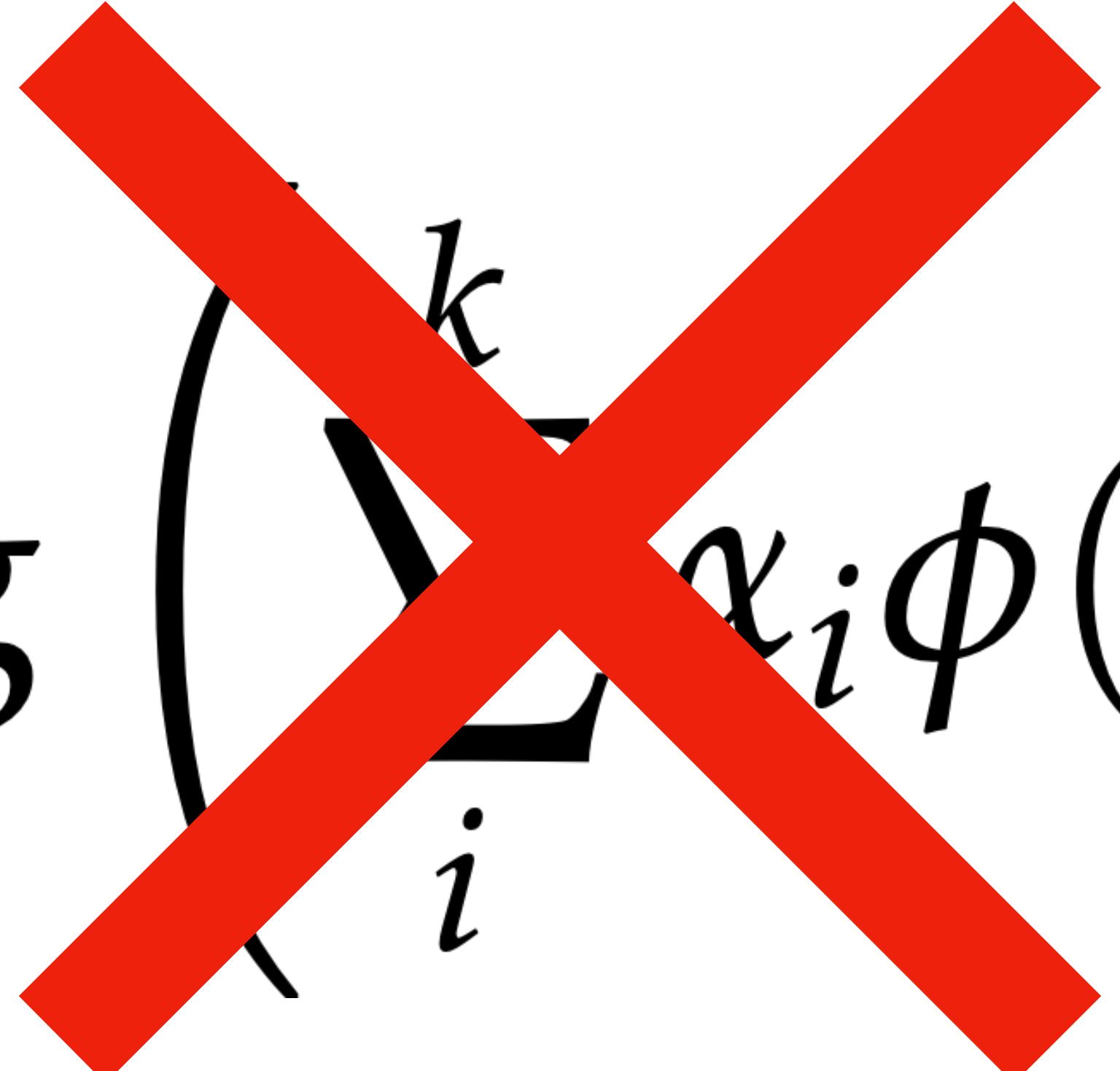
$$P(\mathbf{x}) = \sum_i^k \alpha_i \phi(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

$$\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}}$$

$$P(\mathbf{x}) = \sum_i^k \alpha_i \phi(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

$$L(\theta) = \sum_i^k \alpha_i \phi(x; \mu_i, \Sigma_i)$$

$$\theta = \{\alpha, \mu_1, \dots, \mu_k, \Sigma_1, \dots, \Sigma_k\}$$

$$l(\theta) = \log \left(\alpha_i \phi(x; \mu_i, \Sigma_i) \right)$$


Maximum Likelihood Estimators

Gaussian

μ = sample mean

Σ = sample covariance

Categorical

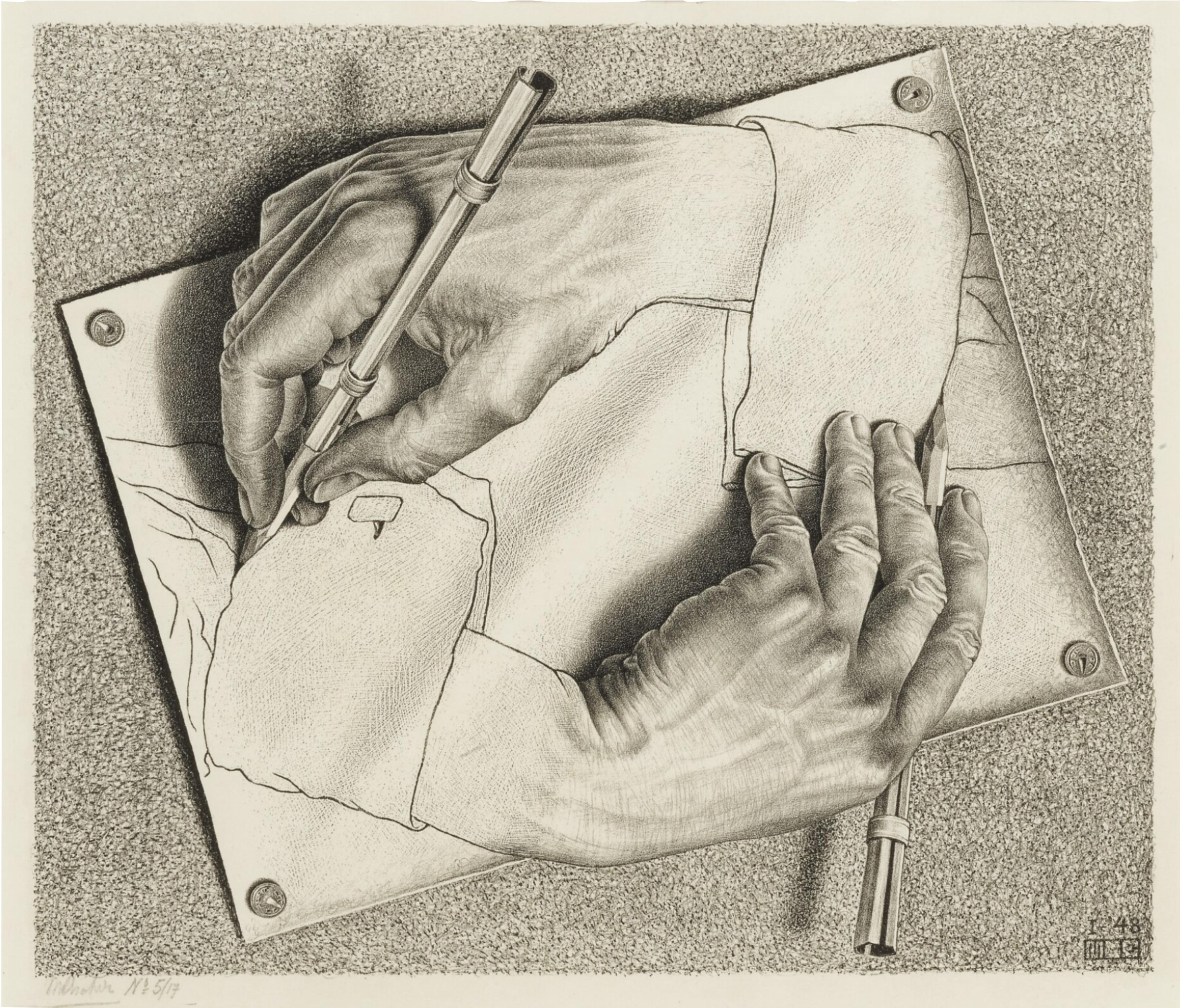
α = sample proportions

$$\alpha_j = \frac{\sum_i 1(z_i = j)}{n}$$

$$\mu_j = \frac{\sum_i 1(z_i = j) x_i}{\sum_i 1(z_i = j)}$$

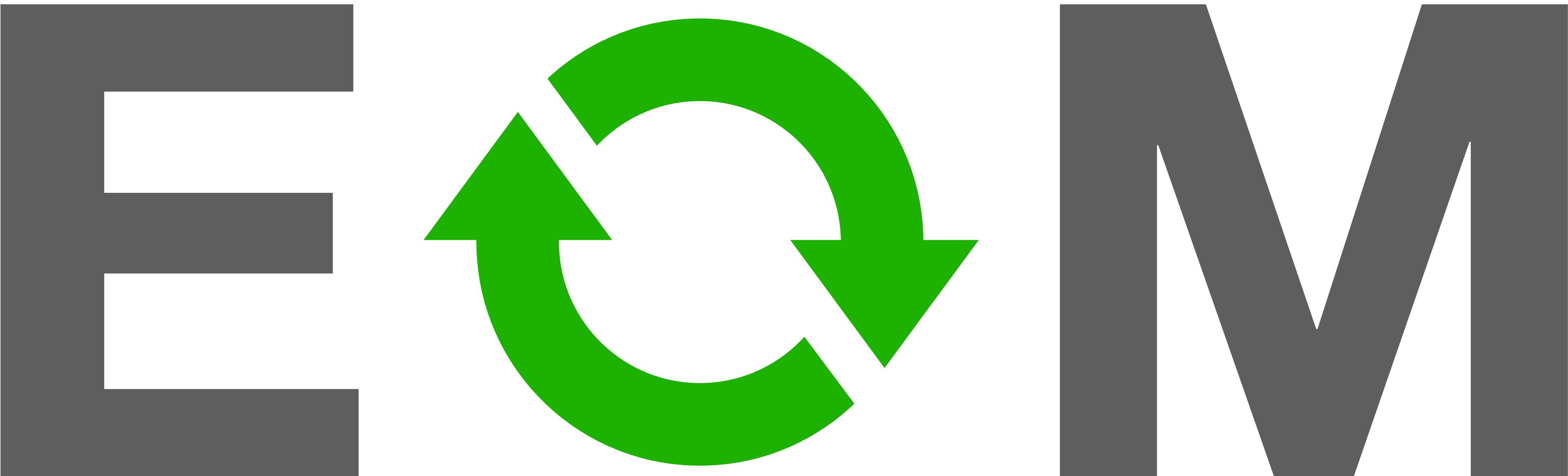
$$\Sigma_j = \frac{\sum_i \mathbf{1}(z_i = j) (\mathbf{x}_i - \mu_j) (\mathbf{x}_i - \mu_j)^T}{\sum_i \mathbf{1}(z_i = j)}$$

$$z_i = \operatorname{argmax}_z P(x_i | z)$$



Expectation-Maximisation

(Gaussian Mixture Models)



E-step

Estimate distribution of Z

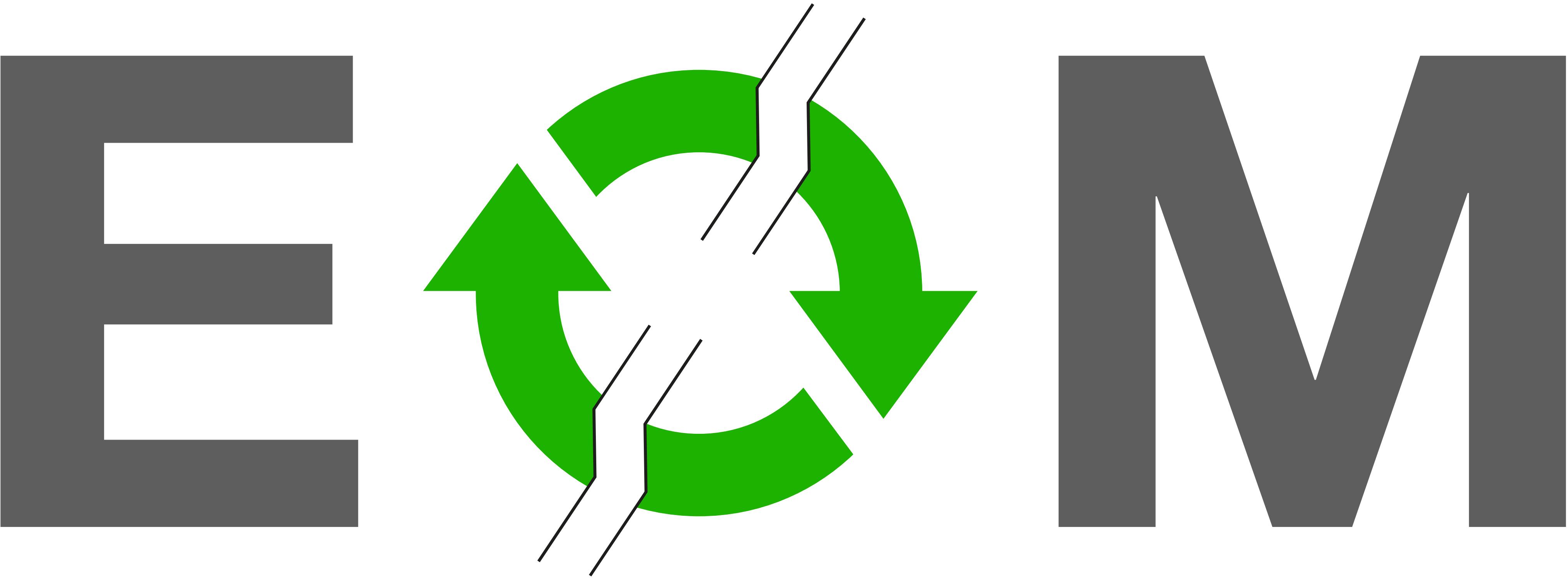
$$\gamma_{i,j} = \frac{P(\mathbf{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \alpha_j}{\sum_l P(\mathbf{x}_i | \boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \alpha_l}$$

M-step
Maximise likelihood of θ

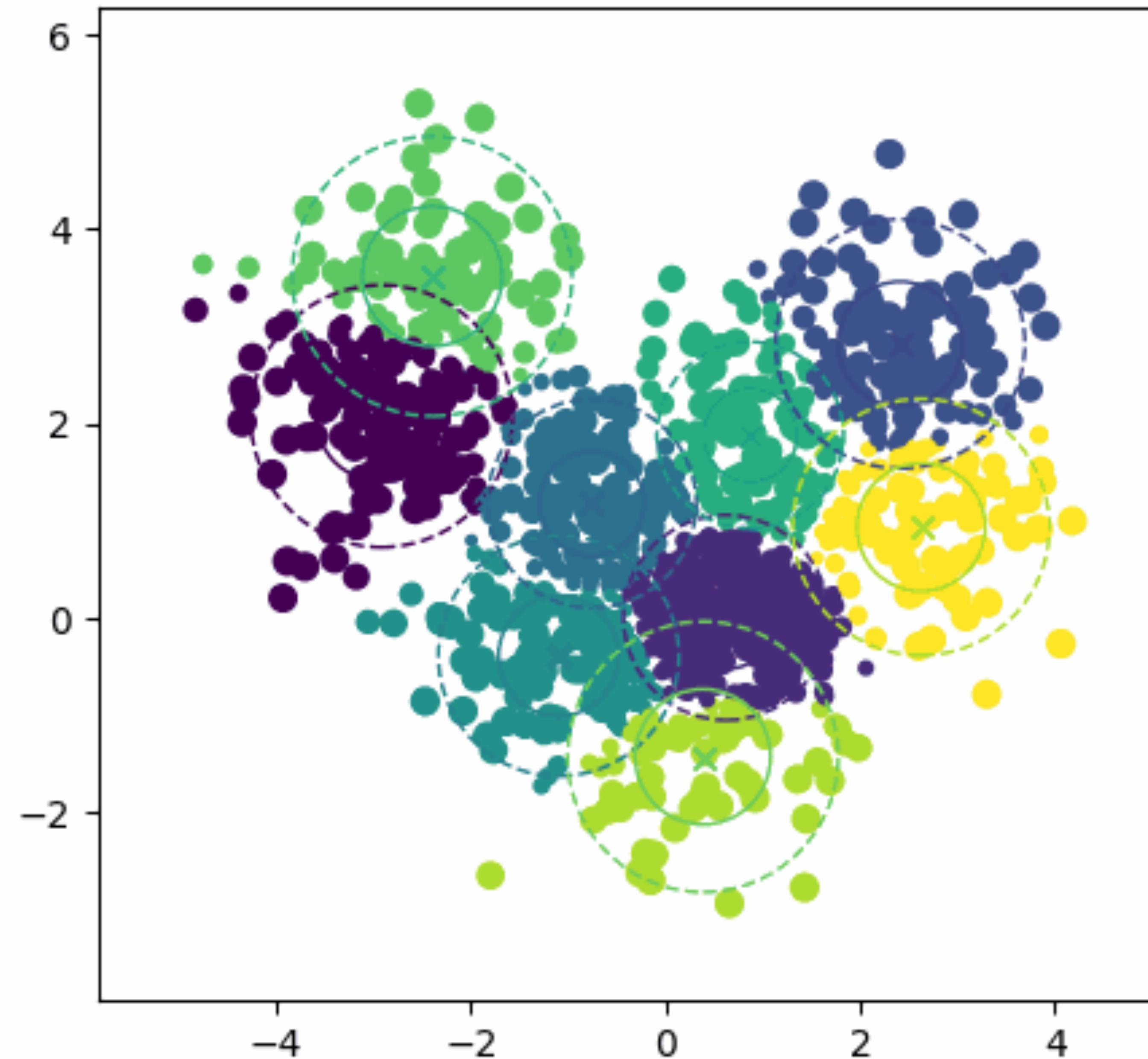
$$\alpha_j = \frac{\sum_i \gamma_{i,j}}{n}$$

$$\mu_j = \frac{\sum_i \gamma_{i,j} \mathbf{x}_i}{\sum_i \gamma_{i,j}}$$

$$\Sigma_j = \frac{\sum_i \gamma_{i,j} (\mathbf{x}_i - \boldsymbol{\mu}_j)(\mathbf{x}_i - \boldsymbol{\mu}_j)^T}{\sum_i \gamma_{i,j}}$$



GMM with 9 clusters and 1000 samples



Expectation-Maximisation

(Latent Variable Models)

$$\tilde{\mathbf{x}}_i = \{ \mathbf{x}_i, \mathbf{z}_i \}$$

$\mathbf{x}_i \in \mathbb{R}^d \rightarrow$ observed

$\mathbf{z}_i \in \mathcal{Z} \rightarrow$ invisible

$$\tilde{\mathbf{X}} = \{\mathbf{X}, \mathbf{z}\}$$

$$\begin{aligned} P(\tilde{\mathbf{X}}; \theta) &= P(\mathbf{X}, \mathbf{Z}; \theta) \\ &= P(\mathbf{X}|\mathbf{Z}; \theta)P(\mathbf{Z}; \theta) \end{aligned}$$

$$L(\theta) = P(\mathbf{X}|\mathbf{Z}; \theta)P(\mathbf{Z}; \theta)$$

$$P(\mathbf{Z}|\mathbf{X}; \theta) = \frac{P(\mathbf{X}|\mathbf{Z}; \theta)P(\mathbf{Z}; \theta)}{P(\mathbf{X}; \theta)}$$

E-step

Estimate distribution of Z

$$Q^{(t)}(\mathbf{Z}) = P(\mathbf{Z}|\mathbf{X}; \boldsymbol{\theta}^{(t-1)})$$

M-step
Maximise likelihood of θ

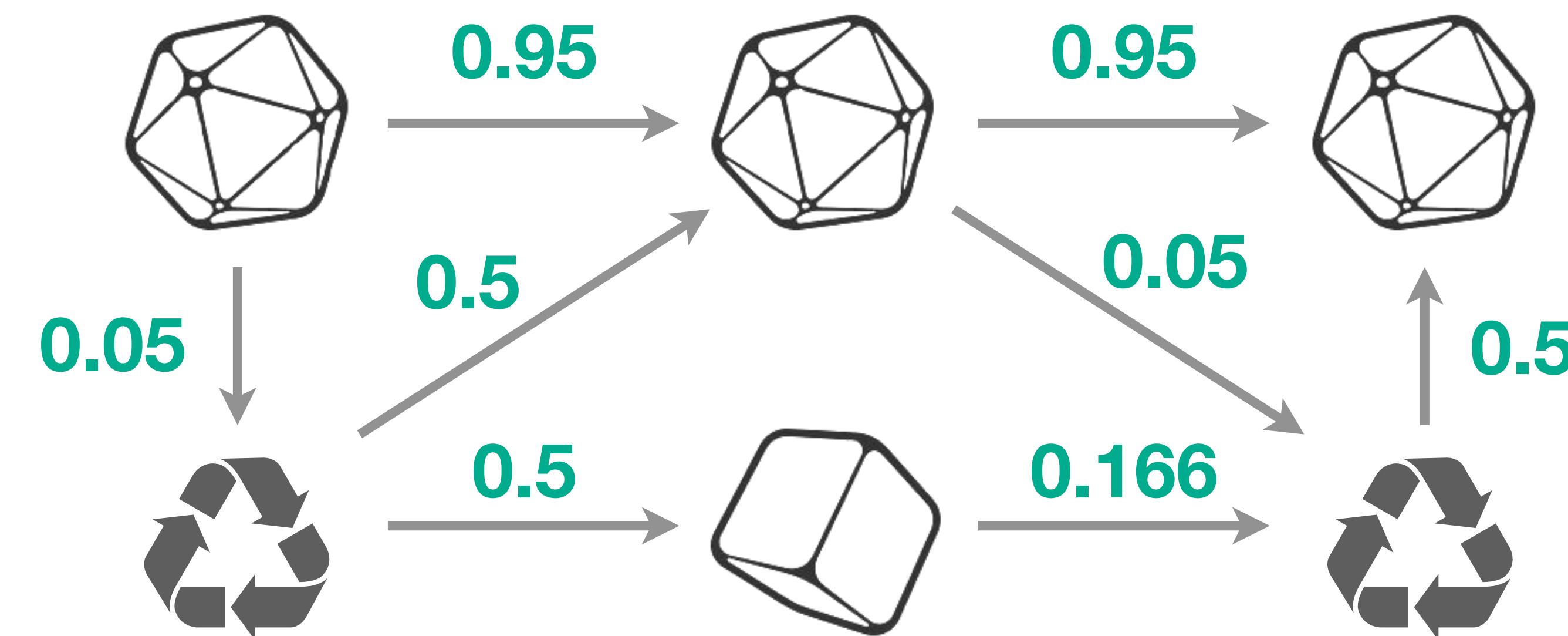
$$l^{(t)}(\theta) = \sum_{\mathbf{z} \in \mathcal{Z}} Q^{(t)}(\mathbf{z}) \log P(\mathbf{x}, \mathbf{z}; \theta^{(t-1)})$$

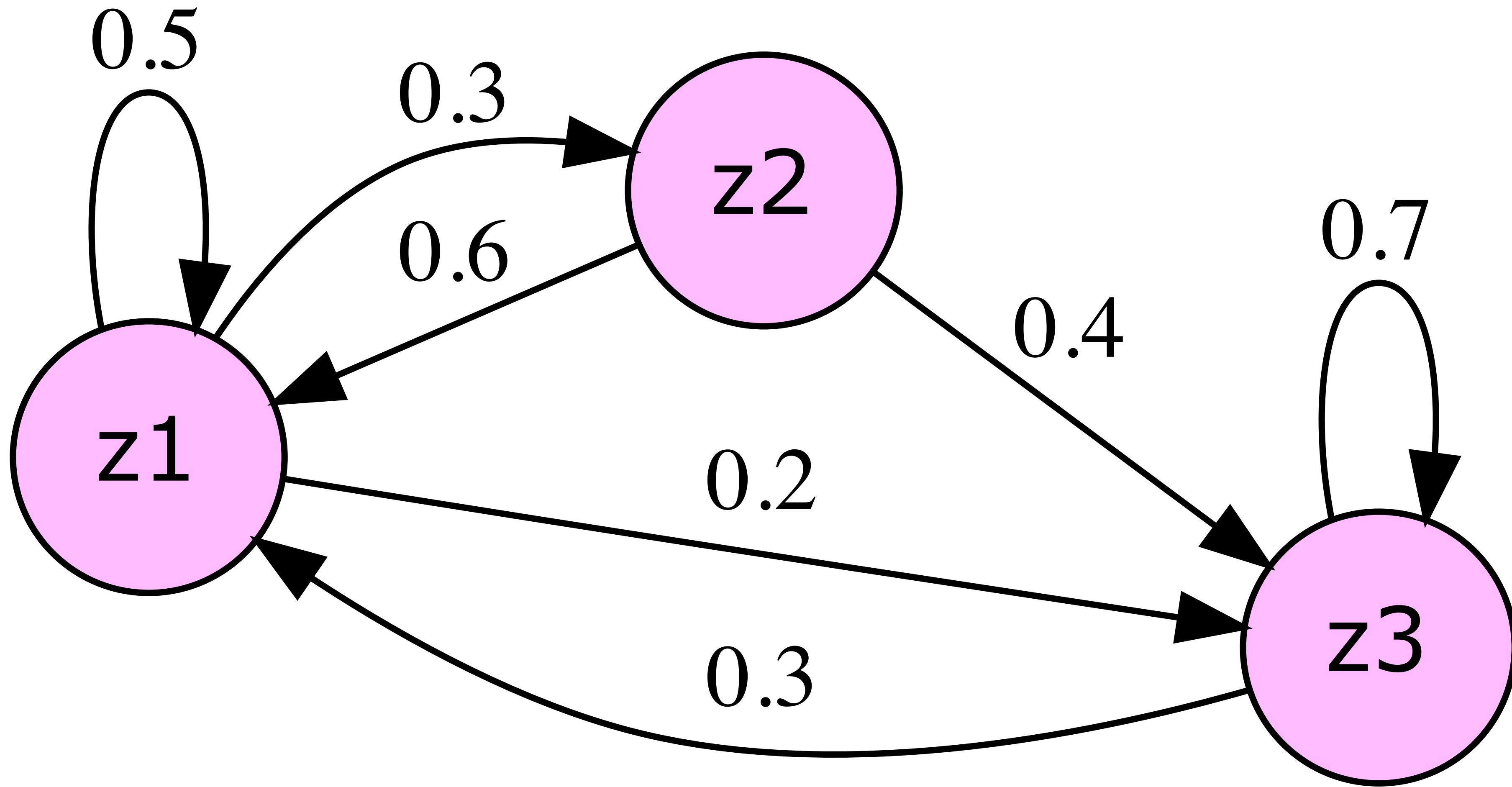
$$\theta^{(t)} = \operatorname{argmax}_{\theta} l^{(t)}(\theta)$$

$$\theta^{(0)} = ???$$

Hidden Markov Models

... 13 2 17 ...





state at $t+1$

	z1	z2	z3
z1	0.5	0.3	0.2
z2	0.6	0	0.4
z3	0.3	0	0.7

state at $t+1$

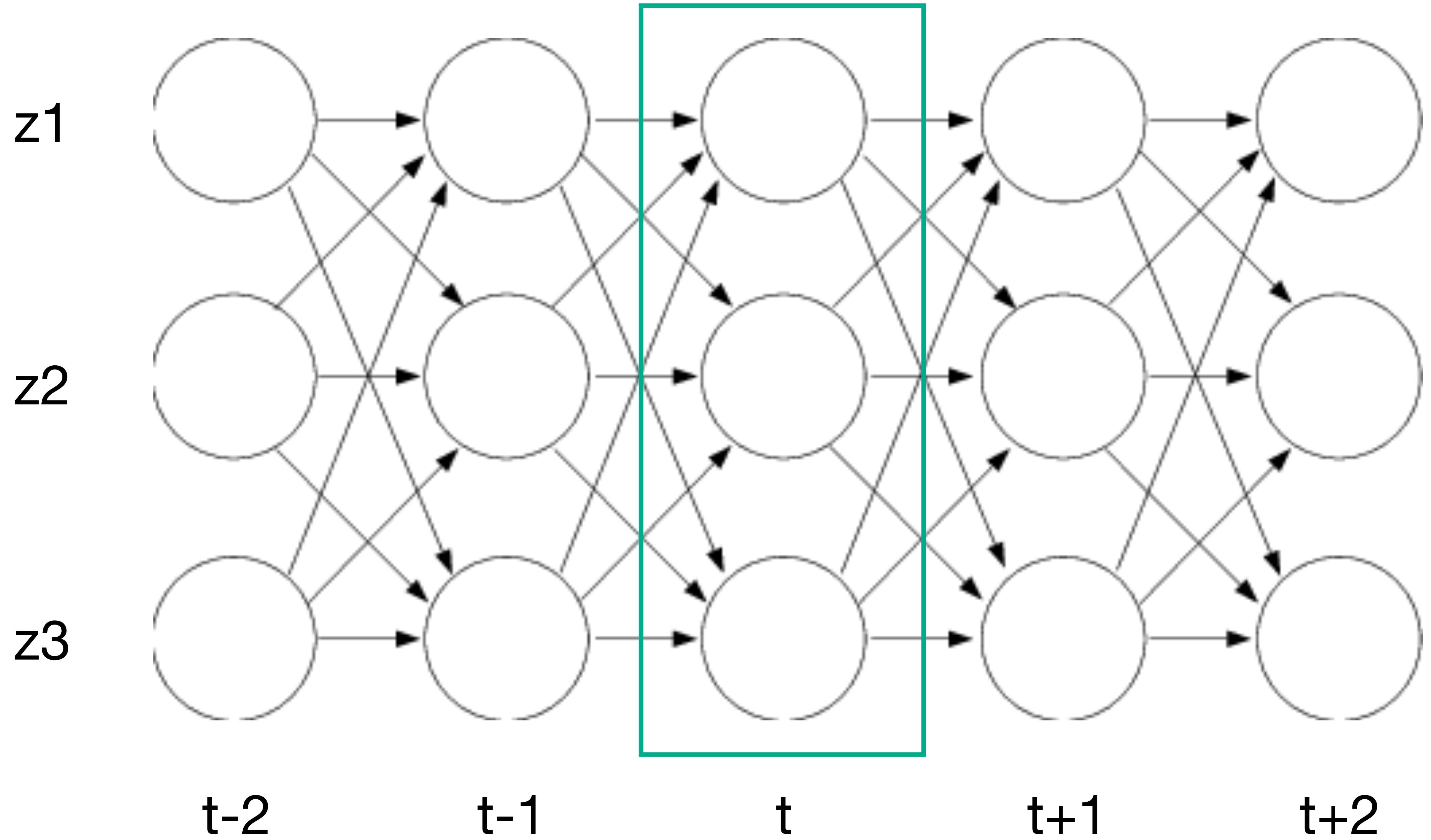
		d6	d20
d6	[0.917	0.083	
d20	0.025	0.975	

$$a_{i,j} = P(z_{t+1} = j | z_t = i)$$

- Observations: $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T \in \mathbb{R}^d$
- Latent states: $z_1, z_2, \dots, z_T \in \{1, 2, \dots, k\}$
- Emission distribution: $P(\mathbf{x}|z; \theta)$
- Initial probabilities: $\pi = [\pi_1, \pi_2, \dots, \pi_k]$
- Transition matrix: $\mathbf{A} \in \mathbb{R}^{k \times k}$

- Probability of an observation or sequence
 - given (X, π, A, θ) , estimate $P(X)$
- Likelihood of parameters
 - given (X, π, A, θ) , estimate $L(\pi, A, \theta)$
- Parameter fitting
 - given (maybe lots of) X , estimate (π, A, θ)
- Infer hidden state sequence
 - given (X, π, A, θ) , estimate (at least some of) Z

$$P(\mathbf{x}_{1:t}, z_t) = P(\mathbf{x}_t | z_t) \sum_{z_{t-1}} P(z_t | z_{t-1}) P(\mathbf{x}_{1:t-1}, z_{t-1})$$



Forward

$$\alpha_t(j) = P(\mathbf{x}_{1:t}, z_t = j) = \sum_i^k \alpha_{t-1}(i) a_{i,j} P(\mathbf{x}_t; \theta_j)$$

Forward-Backward

$$\beta_t(j) = P(\mathbf{x}_{t:T}, z_t = j) = \sum_i^k \beta_{t+1}(i) a_{i,j} P(\mathbf{x}_{t+1}; \theta_j)$$

Viterbi

$$\alpha_t(j) = \max_i \alpha_{t-1}(i) a_{i,j} P(\mathbf{x}_t; \theta_j)$$

$$\zeta_t(j) = \operatorname{argmax}_i \alpha_{t-1}(i) a_{i,j} P(\mathbf{x}_t; \theta_j)$$

Baum-Welch



Questions?

Next: Deep Learning Applications

