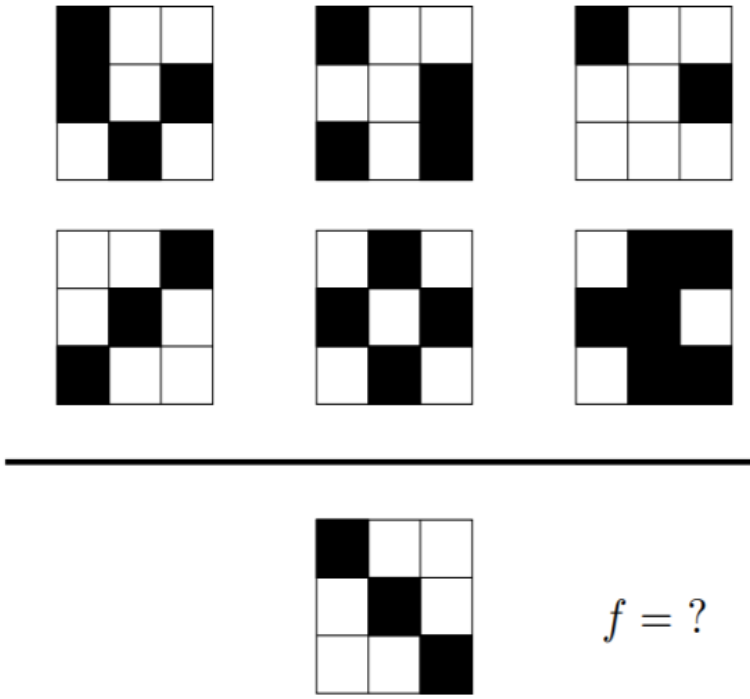
Machine Learning from Data

Lecture 3: Spring 2021

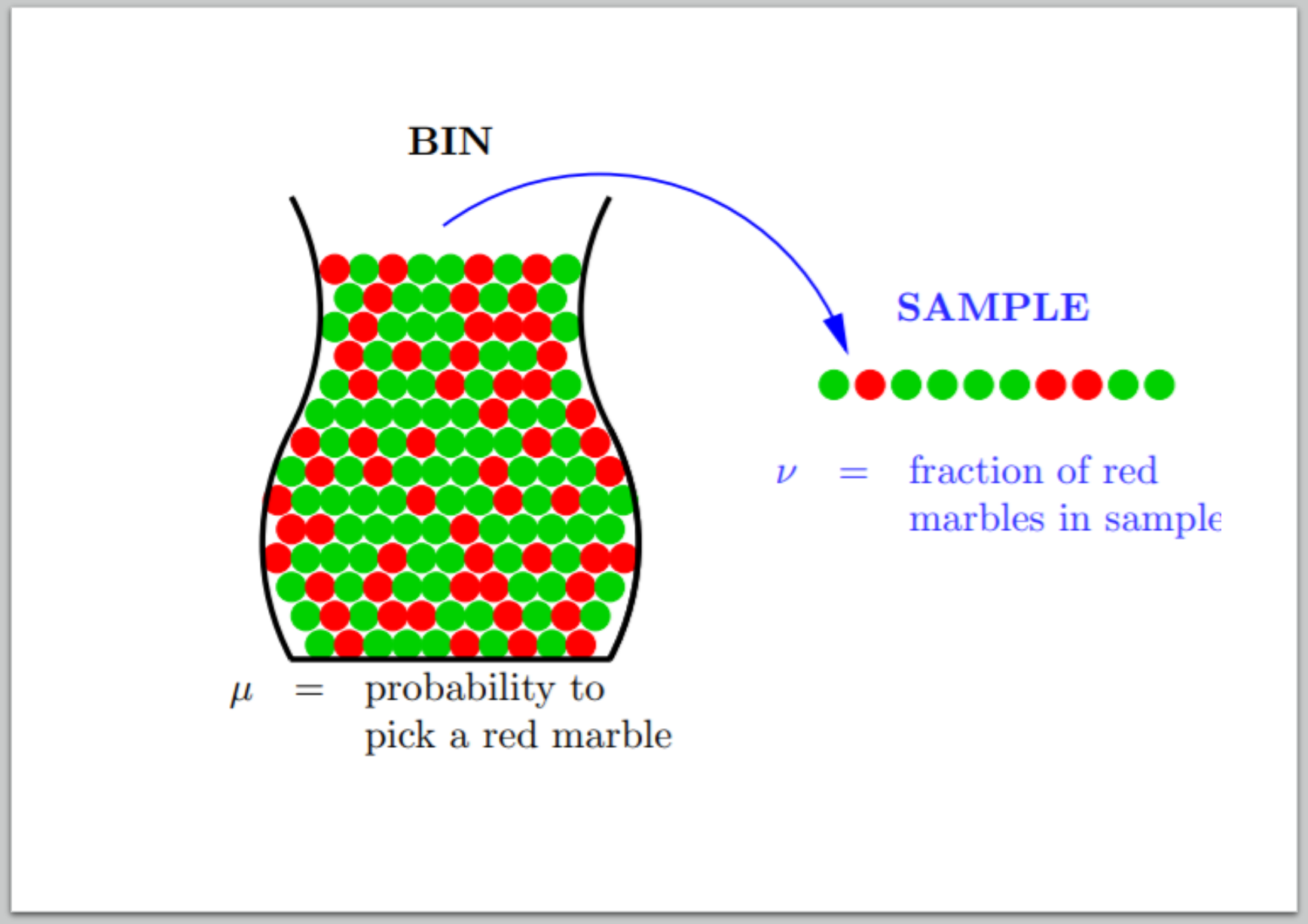
Today's Lecture

- Is Learning Feasible?
- Infer outside the dataset D
- Hoeffding's Inequality

Outside Data



Population Mean from the Sample Mean



Probability to the Rescue

-

Hoeffding’s Inequality

•

Hoeffding’s Inequality continued…

•

Example

-

Relationship with Learning

-

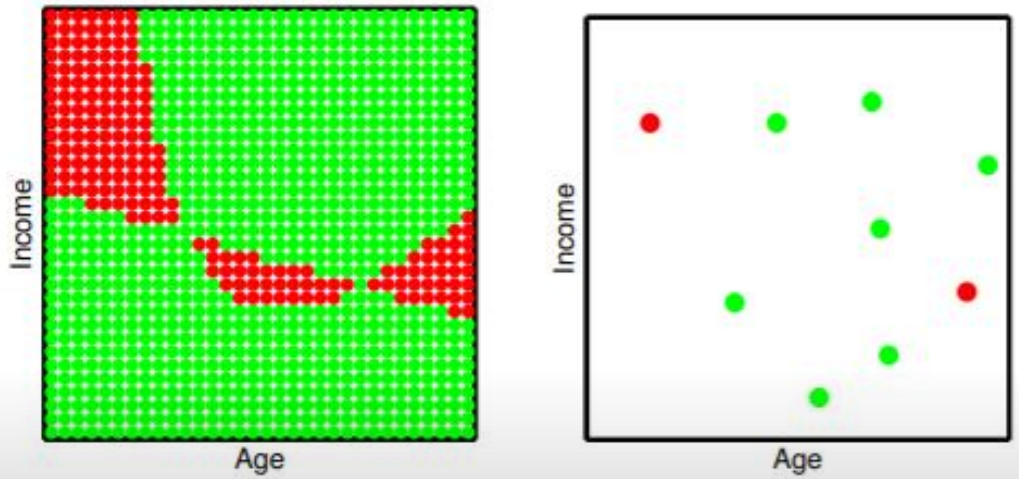
Relationship with Learning

-

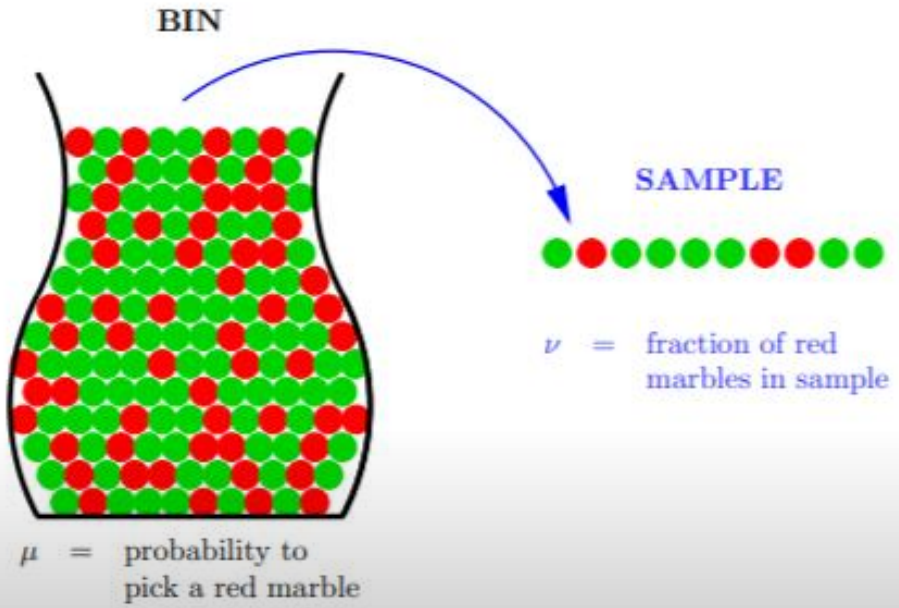
Hoeffding's Summarized

$$\mathbb{P}\left[\left|E_{\text{in}}(h) - E_{\text{out}}(h)\right| > \epsilon\right] \leq 2e^{-2\epsilon^2 N}, \qquad \text{for any } \epsilon > 0.$$

$$\mathbb{P}\left[\left|E_{\text{in}}(h) - E_{\text{out}}(h)\right| \leq \epsilon\right] \geq 1 - 2e^{-2\epsilon^2 N}, \qquad \text{for any } \epsilon > 0.$$



Unknown f and $P(\mathbf{x})$, **fixed** h



Function Vs Bin

Learning

input space $\mathcal{X}$

$\mathbf{x}$ for which $h(\mathbf{x}) = f(\mathbf{x})$

$\mathbf{x}$ for which $h(\mathbf{x}) \neq f(\mathbf{x})$

$P(\mathbf{x})$

data set $\mathcal{D}$

Out-of-sample Error: $E_{\text{out}}(h) = \mathbb{P}_{\mathbf{x}}[h(\mathbf{x}) \neq f(\mathbf{x})]$

In-sample Error: $E_{\text{in}}(h) = \frac{1}{N} \sum_{n=1}^{N} [\![h(\mathbf{x}) \neq f(\mathbf{x})]\!]$

Bin Model

Bin

green marble

red marble

randomly picking a marble

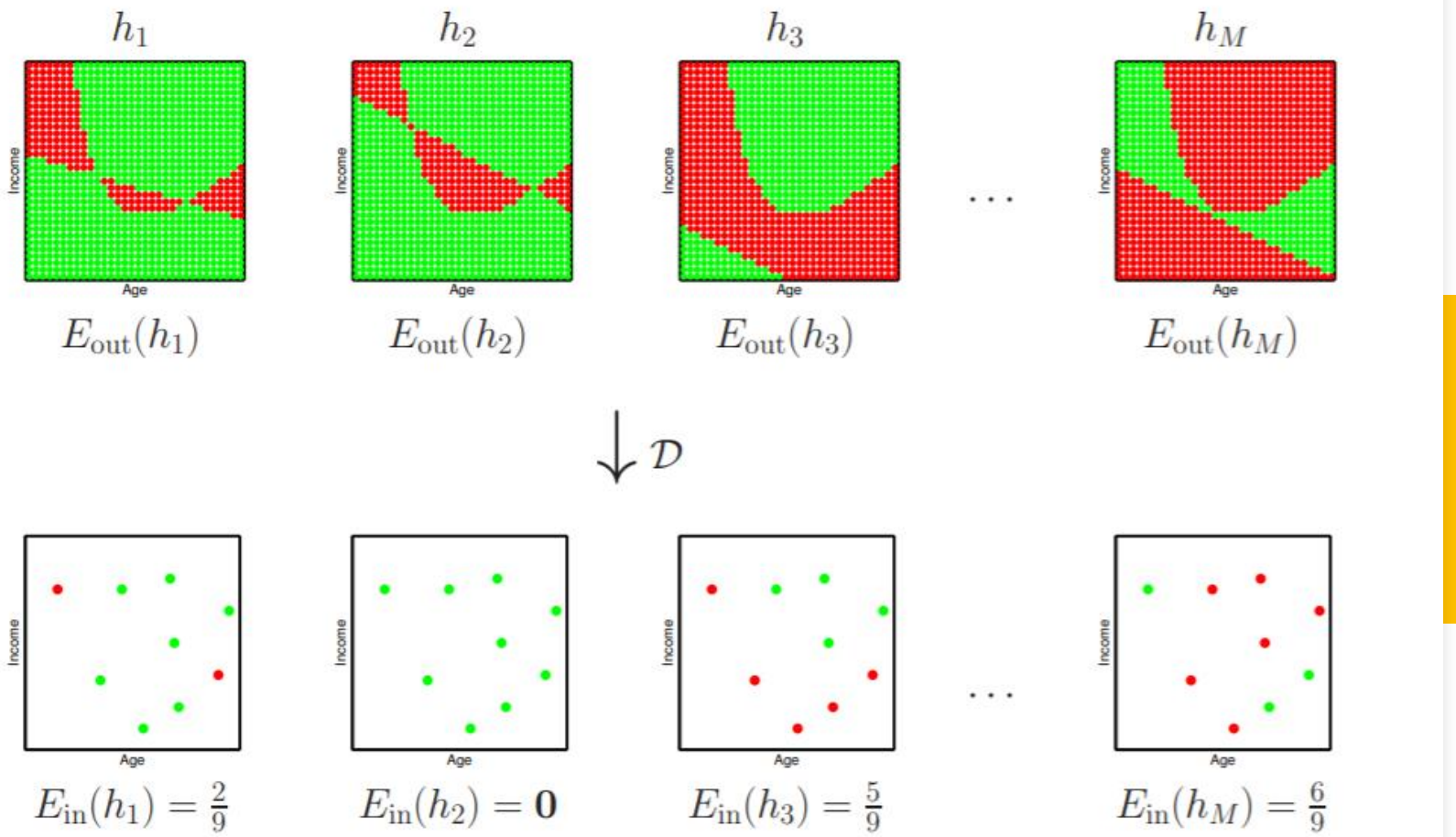
sample of N marbles

μ = probability of picking a red marble

ν = fraction of red marbles in the sample

Verification Vs. Real Learning

•



Finite Learning Model