Monte Carlo

Classical Statistics

1. Identify what you want to measure
2. Create Test Statistic and determine its distribution
3. Determine critical value and/or p-value

Issue: Requires large samples (often $n \to \infty$ $\approx n \geq 30$)

Q: Is there a way to estimate the p-value?

- estimating is fine as long as p-value is not "too close" to α

(α = cutoff for significance)

Exact: $p = 0.045$	Exact: $p = 0.045$
Estimate: $p \approx 0.042$	Estimate: $p \approx 0.049$
Cutoff: $\alpha = 0.05$	Cutoff: $\alpha = 0.05$
✓	✗

Monte Carlo utilizes repeated random sampling to predict probabilities of various outcomes

Idea #1: Law of Large Numbers (LLN)

Average converges to Expectation as sample size becomes large

Idea #2: Utilize computer power to generate large samples

Take advantage of technology

Procedure:

B = large integer

For 1, 2, ..., B

Generate sample of size n from some distribution

Compute any interesting values and test statistic

Estimate p-value

$$\text{p-value} \approx \frac{\#\text{ more extreme test statistics} + 1}{B+1}$$

Ex: Goodness-of-Fit test

You want to see if a dice is fair or not. After 60 rolls, you record:

1: 10	2: 3	3: 12
4: 24	5: 6	6: 5

Compute the exact and estimated p-value.

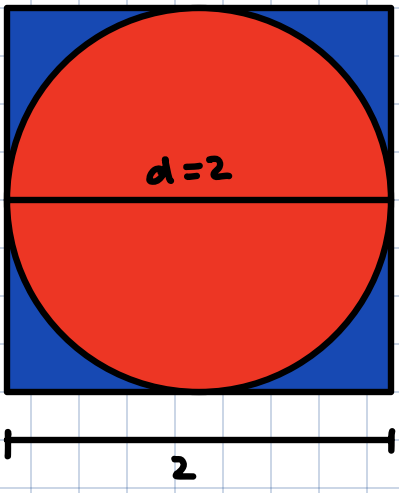
Test: χ^2 goodness of fit test

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 29 \qquad df = k - 1 = 5$$

Calculate p-value: 2.32×10^{-5}

Estimating π

- classic coding problem
- recall dart board problem

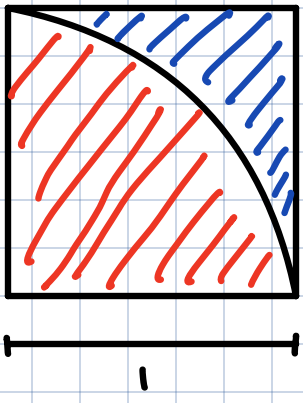


$$\text{Area of Circle} = \pi\left(\frac{d}{2}\right)^2 = \pi$$

$$\text{Total Area} = 4$$

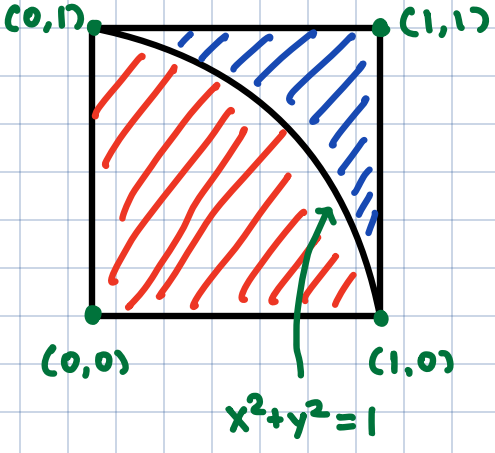
$$\mathbb{P}(\text{circle}) = \frac{\text{Area of Circle}}{\text{Total Area}} = \frac{\pi}{4}$$

Use Quarter Circle



$$\mathbb{P}(\text{red}) = \frac{\text{Red Area}}{\text{Total Area}} = \frac{\frac{1}{4}\pi(1^2)}{1^2} = \frac{\pi}{4}$$

Use whichever scheme is easier to program



Idea:

Generate many random points to see if they fall inside the red or blue area

$$\mathbb{P}(\text{red}) \approx \frac{\#\text{ red points}}{\#\text{ points}}$$

By Law of Large Numbers, as # points grows large, the probability approaches expected value

Simulation Procedure:

1. Generate n points in 1×1 square
2. Determine if point is inside red region

$$\text{distance} = d = \sqrt{x^2 + y^2} \leq 1$$

3. Estimate $\frac{\pi}{4}$ using # points in red / # points
4. Multiply by 4 to estimate π

Experiments:

Class: Try different values of n

Mini-Project: Try half-circle or whole-circle; Apply to other areas