MATH 205: Statistical methods

Lab 4: Random data

Goals: Generate random data

- simulate uniform distribution
- simulate discrete distribution
- the law of large numbers

Continuous uniform distribution

- the distribution is often abbreviated $U(a, b)$, where U stands for uniform distribution
- distributes probability for all points in $[a, b]$ equally
- all intervals of the same length on $[a, b]$ are equally probable

Simulate uniform distribution

- the uniform distribution on (a, b) has density

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{elsewhere} \end{cases}$$

- To generate the uniform distribution on $(0, 1)$, use the function *runif*

$$b = runif(200)$$

Simulate a biased coin

- Assume that we want to simulate a Bernoulli random variable

$$p(x) = \begin{cases} 0.6 & \text{if } x = 0 \\ 0.4 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

- Step 1: generate u from the uniform distribution on $(0,1)$
- Step 2: If $u < 0.4$, then set $x = 1$; otherwise set $x = 0$

Simulate discrete random variables

- Question: How to simulate samples from the following distribution

$$p(x) = \begin{cases} 0.2 & \text{if } x = 3 \\ 0.3 & \text{if } x = 5 \\ 0.5 & \text{if } x = 7 \\ 0 & \text{otherwise} \end{cases}$$

Law of Large Number: discrete

- Generate n = 2000 samples from the following distribution

$$p(x) = \begin{cases} 0.2 & \text{if } x = 3 \\ 0.3 & \text{if } x = 5 \\ 0.5 & \text{if } x = 7 \\ 0 & \text{otherwise} \end{cases}$$

- Let n_3, n_5, n_7 be the number of 3, 5, 7 appear in the sample, then

$$\frac{n_3}{n} \approx 0.2, \qquad \frac{n_5}{n} \approx 0.3, \qquad \frac{n_7}{n} \approx 0.5$$

- So naturally,

$$\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{3n_3 + 5n_5 + 7n_7}{n} \approx \sum_{x=3,5,7} xp(x)$$

Expected value: discrete variables

Definition

Given a discrete random variable X which takes values in the set $\mathcal{D}$ and which has probability distribution P, we define the expected value of X as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{D}} xP(X = x)$$

This is sometimes written $\mathbb{E}_P[X]$, to clarify which distribution one has in mind.

Law of Large Number

Theorem

Let $X_1, X_2, \ldots, X_n, \ldots$ be a sequence of copies of a random variable X, then

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \longrightarrow E[X]$$

as n approaches infinity.

Practice problem 1

Let X be a discrete random variable with the following probability mass function

$$p(x) = \begin{cases} \frac{x}{10}, & \text{for} \quad x = 1, 2, 3, 4 \\ 0 & \text{otherwise} \end{cases}$$

- Simulate a sample of 500 random draws from the distribution described above
- Compute the mean and produce a bar plot of the sample
- Compare the mean of the dataset and the expected value of X

Practice problem 2

- Repeat "Practice Problem 1" 2000 times, each time record the mean of a dataset as an element in a vector v (of length 2000)
- Produce a histogram of v