MATH 205: Statistical methods

Lab 5: Random continuous data

Goals: Generate random data

- simulate uniform distribution
- simulate continuous 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 continuous random variables

Theorem

Let X be a continuous random variable with probability distribution function F. Then F(X) is a uniform random variable over (0,1).

Proof.

Let $Y = F(X)$, then $Y \in [0,1]$ and for all $y \in (0,1)$

$$F_Y(y) = P[Y \le y] = P[F(X) \le y] = P[X \le F^{-1}(y)] = F(F^{-1}(y)) = y$$

thus

$$f_Y(y) = \begin{cases} 1 & \text{if } y \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

□

Simulate continuous random variables

- Question: How to simulate samples from the following distribution

$$f(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- The distribution function of X is

$$F(x) = \begin{cases} 1 - e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Step 1: generate u from the uniform distribution on $(0, 1)$
- Step 2: Solve equation $F(x) = u$
- Set x as the solution

$$x = -\frac{1}{2} \ln(1 - u)$$

Practice problem 1

Let X be a continuous random variable with the following
probability density function

$$f(x) = \begin{cases} 2x, & \text{for } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

- Simulate a sample of 1000 random draws from the distribution described above
- Compute the mean and produce a histogram of the sample
- Compare the mean of the dataset and the expected value of X

- 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