MATH 205: Statistical methods

Lab 5: Random continuous data

Goals: Generate random data

- simulate uniform distribution [lab4]
- simulate continuous distribution
- the law of large numbers

Distribution function

For continuous random variable:

$$F(t) = P(X \leq t) = \int_{(-\infty,t]} f(x)dx$$
$$= \int_{-\infty}^{t} f(x)dx$$

Distribution function

For continuous random variable:

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx = F(b) - F(a)$$

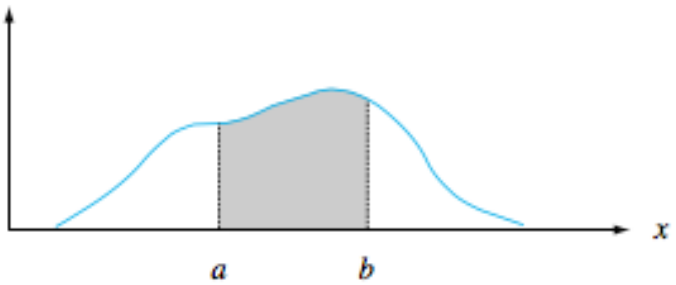

Figure 4.2 $P(a \leq X \leq b)$ = the area under the density curve between a and b

Moreover:

$$f(x) = F'(x)$$

Simulate continuous random variables: principles

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 $y \in (0,1)$

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

thus

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

□

Simulate continuous random variables

To simulate a continuous random variable with density function f:

- Preparation:
  - Step 1: Compute (by hand) the distribution function F(x)
  - Step 2: Solve (by hand) the equation $F(x) = u$ for general $u \in (0, 1)$ to obtain $x = g(u)$
- Simulation:
  - Step 1: Generate u from the uniform distribution in $[0, 1]$
  - Step 2: Set $x = g(u)$ as the sample

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

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

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

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