Generating Observations of Random Variables in R
Statistics 341

Three steps are required to generate observations from a random variable Y.

1. Find the distribution function $F(y)$ for the random variable Y. To be able to generate observations of the random variable Y, the distribution function $F(y)$ must have a closed form solution. If the distribution function $F(y)$ does not have a closed form solution, an approximation formula can be used, but this topic is outside the scope of this course.

   This step must be done by hand and is not completed in R.

2. Generate observations from the Uniform distribution with $\theta_1 = 0$ and $\theta_2 = 1$. We will call these values u. Each u that is generated is an observed value of the distribution function $F(y)$. So these are the probabilities $P(Y \leq y)$.

   This step is completed in R as the command

   ```
   u <- runif(numobs, 0,1)
   ```

   where **numobs** is the number of observations of the random variable you would like. This step is the same for all data generation problems.

3. For each generated value u, we will set $u = F(y)$ and solve for y. This maps the probabilities $F(y) = P(Y \leq y)$ to a particular value of y. The values of y are the generated observations of the random variable Y.

   The first part of this step must be completed by hand. Once you have a formula for y in terms of u, enter the formula into R as

   ```
   y<- function of u
   ```

   The function of u will change depending on the distribution function of the random variable Y.

Examples.

1. Let Y be a continuous random variable with the following probability density function

   $$f(y) = 2y \qquad 0 \leq y \leq 1$$

   (a) Find $F(y)$. For this distribution, we know that when $y < 0$, $F(y) = 0$ and when $y > 1$, $F(y) = 1$. When $0 \leq y \leq 1$, the value of $F(y)$ will be

   $$F(y) = \int_{-\infty}^{y} f(t)dt = \int_{0}^{y} 2tdt = t^2|_0^y = y^2$$

   (b) Use R to generate 10,000 observations from the uniform distribution with minimum value $\theta_1 = 0$ and maximum value $\theta_2 = 1$. The command is

   ```
   u<- runif(10000,0,1)
   ```

(c) Set these values from the uniform distribution equal to the distribution function $F(y)$ and solve for y. In this case, we have

$$u = y^2 \qquad \text{so therefore} \qquad y = \sqrt{u}$$

Use this equation to calculate the values of y in R. The command is

```
y<- sqrt(u)
```

2. Let Y be a continuous random variable with the following probability density function

$$f(y) = \frac{3}{8}y^2 \qquad 0 \le y \le 2$$

(a) Find $F(y)$. For this distribution, we know that when $y < 0$, $F(y) = 0$ and when $y > 2$, $F(y) = 1$. When $0 \le y \le 2$, the value of $F(y)$ will be

$$F(y) = \int_{-\infty}^{y} f(t)dt = \int_{0}^{y} \frac{3}{8}t^2 dt = \frac{1}{8}t^3|_0^y = \frac{1}{8}y^3$$

(b) Use R to generate 10,000 observations from the uniform distribution with minimum value $\theta_1 = 0$ and maximum value $\theta_2 = 1$. The command is

```
u<- runif(10000,0,1)
```

(c) Set these values from the uniform distribution equal to the distribution function $F(y)$ and solve for y. In this case, we have

$$u = \frac{1}{8}y^3 \qquad \text{so therefore} \qquad y = 2u^{1/3}$$

Use this equation to calculate the values of y in R. The command is

```
y<- 2*u^(1/3)
```