AMS 597: Statistical Computing

Pei Fen Kuan
Generating random variables

- **Aim:** Simulate random variables from specified probability distributions
- A suitable generator of uniform pseudo random numbers is essential. Methods for generating r.v. from other probability distributions all depend on the uniform random number generator.
- In this course, we assume that a suitable uniform pseudo random number generator is available.
Generating random variables

- We have previously learnt to generate random variables using built-in functions in R, e.g. rnorm, rbeta, etc.
- We will now learn a few algorithms for generating random variables.
Theorem (Probability Integral Transformation)
If $X$ is a continuous random variable with cdf $F_X(x)$, then $U = F_X(X) \sim U(0, 1)$. 
The method can be applied for generating continuous or discrete random variables.

1. Derive the inverse function \( F^{-1}_X(u) \).
2. Write a command or function to compute \( F^{-1}_X(u) \).
3. Generate a random \( u \) from Uniform(0,1).
4. Define \( x = F^{-1}_X(u) \).
Exercise: Simulate from 10000 X’s from Exp(2), i.e., E(X)=1/2 using the inverse transform method.
Example: Generate 1000 random numbers whose probability density function is \( f(x) = x \) for \((0, c)\) and 0 otherwise. (Figure out the appropriate \( c \))
Inverse Transform Method

- Theoretically you can generate any random variables from the uniform distribution but the disadvantage of the integral transform method is that it is practical only if the c.d.f. are explicitly available. For instance the c.d.f. of a normal r.v. cannot be expressed explicitly.
In Inverse Transform Method

- If $X \sim N(0, 1)$, the c.d.f. of $X$ is given by

  \[ \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp \left( \frac{z^2}{2} \right) dz. \]

- It can be shown that

  \[ \Phi^{-1}(x) \approx t - \frac{a_0 + a_1 t}{1 + b_1 t + b_2 t^2} \]

  for some constants $a_i$ and $b_i$ ($t^2 = -2\log x$), where $a_0 = 2.30753$, $a_1 = 0.27061$, $b_1 = 0.99229$, $b_2 = 0.04481$. 
Exercise: Simulate 10000 $N(0,1)$ using the inverse transform method.
Suppose that $X$ and $Y$ are random variables with pdf $f$ and $g$ respectively, and there exists a constant $c$ such that

$$\frac{f(t)}{g(t)} \leq c$$

for all $t$ such that $f(t) > 0$, and that we know how to generate $Y$. Then the acceptance-rejection method can be used to generate $X$. 
Acceptance-Rejection Method

1. Generate Y from g(y)
2. Generate U from U(0,1)
3. If $U \leq \frac{f(Y)}{c g(Y)}$, then set $X = Y$ (accept); otherwise go back to 1 (reject).
Exercise: Simulate 10000 r.v’s that follow probability density $f(y)=2y$, $0<y<1$. 
Exercise: Simulate 10000 r.v’s from Beta(2,2) using AR method. On average, how many iterations needed?
Acceptance-Rejection Method

- AR method may not be an efficient algorithm.
- To use the accept-reject method, the distributions $f$ and $g$ should be somewhat similar to have a sufficiently good algorithm.
Transformation Methods

- If $X \sim \text{Ga}(a, \lambda)$ and $Y \sim \text{Ga}(b, \lambda)$, $X$ and $Y$ are independent, then $X/(X+Y) \sim \text{Beta}(a, b)$
Exercise: Simulate 10000 r.v’s from Beta(2,2) using transformation method.
Box-Muller Method

If $U_1$ and $U_2$ are independent $U(0,1)$ random variables, then

$$X_1 = \sqrt{-2\ln U_1} \cos(2\pi U_2)$$
$$X_2 = \sqrt{-2\ln U_1} \sin(2\pi U_2)$$

are independent $N(0,1)$
Exercise: Simulate 10000 $N(0,1)$ using the Box Muller method.