Simulating Random Variable Using R

Andy Kirtland

Using a Seed

- Setting the seed in R allows you to duplicate results.

  set.seed(###) to turn on

  set.seed() to turn off

Distributions Available in R

- Uniform → `x <- runif(n, min=0, max=1)`
- Normal → `x <- rnorm(n, mean=0, sd=1)`
- Lognormal → `x <- rlnorm(n, meanlog =0, sdlog = 1)`
- Poisson → `x <- rpois(n, lambda)`
- Binomial → `x <- rbinom(n, size, prob)`
- Geometric → `x <- rgeom(n, prob)`
- T-Distribution → `x <- rt(n, df, ncp)`
- F-Distribution → `x <- rf(n, df1, df2, ncp)`
- Chi-squared → `x <- rchisq(n, df, ncp=0)`

Generating Variables from Other Distributions

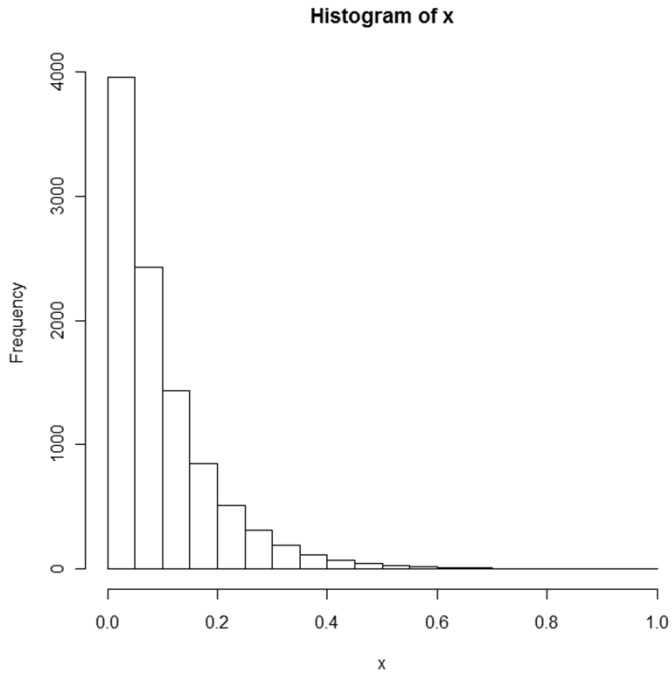
- First check to see if the distribution isn't already available in R!
- If not,
  - Determine desired cumulative distribution function, F(x)
  - Set F(x) = unif(0,1)
  - Solve for x

Exponential Example

$$F(x) = 1 - e^{-\lambda \cdot x}$$

$$unif(0,1) = 1 - e^{-\lambda \cdot x}$$

$$\rightarrow x = \frac{-\log\{1 - unif(0,1)\}}{\lambda}$$




```
set.seed(333)
n <- 10000
lambda <- 10
x <- -log( runif(n,0,1) ) / lambda
hist(x)
```

Sampling From a Data Set

- Useful when bootstrapping

```
sample(x, size, replace = false, prob = null)
```

- Example

```
set.seed(333)
n <- 10000
s_size <- round(.2*n,0)
sample(x=1:n,size=s_size)
psuedoData <- Data[sample(x=1:n,size=s_size)]
```