Generating Random Data*

Alan T. Arnholt
Department of Mathematical Sciences
Appalachian State University
arnholt@math.appstate.edu

Spring 2006 R Notes

Generating Random Samples
Overview of Sampling

Binomial Distribution
Overview of the Binomial Distribution

The R Script

Using `sample()`

- `sample(x, size, replace = FALSE, prob = NULL)` takes a sample of the specified `size` from the elements of `x` by sampling either with or without replacement.

Using `sample()`

- `sample(x, size, replace = FALSE, prob = NULL)` takes a sample of the specified `size` from the elements of `x` by sampling either with or without replacement.
- To sample with replacement, type `replace=TRUE`. To sample without replacement, enter `replace=FALSE`.

Using `sample()`

- `sample(x, size, replace = FALSE, prob = NULL)` takes a sample of the specified `size` from the elements of `x` by sampling either with or without replacement.
- To sample with replacement, type `replace=TRUE`. To sample without replacement, enter `replace=FALSE`.
- The sampling need not be uniform. Probabilities of selecting the values in `x` can be specified with the argument `prob`.

BSDA Example 3.12

Simulate 18,000 rolls of a fair die and determine the frequency of occurrence of each possible outcome.

```
> die <- 1:6
> rolls <- sample(x=die,size=18000,replace=TRUE)
> table(rolls)
rolls
   1    2    3    4    5    6
2969 3063 2994 3042 3021 2911
> round(table(rolls)/length(rolls),3)
rolls
    1     2     3     4     5     6
0.165 0.170 0.166 0.169 0.168 0.162
```

Bernoulli Trial

A Bernoulli trial is a random experiment with only two possible outcomes. The outcomes are mutually exclusive and exhaustive. For example, success or failure, true or false, alive or dead, male or female, etc. A Bernoulli random variable X, can take on two values, where $X(\text{success}) = 1$ and $X(\text{failure}) = 0$. The probability that X is a success is π, and the probability that X is a failure is $\varrho = 1 - \pi$.

Generating Bernoulli Trials with `sample()`

Example 3.15 from BSDA - Suppose a field-goal kicker has an 80% success rate inside the 35 yard line. Simulate eight kicks inside the 35 for ten consecutive seasons.

- To perform the simulation, we will use `sample()`

Generating Bernoulli Trials with `sample()`

Example 3.15 from BSDA - Suppose a field-goal kicker has an 80% success rate inside the 35 yard line. Simulate eight kicks inside the 35 for ten consecutive seasons.

- To perform the simulation, we will use `sample()`
- Eighty Bernoulli trials are generated which can be thought of as ten games with eight field-goal attempts each.

R-Code

```
> set.seed(13)
> fg <- sample(x=c(0,1),size=8*10,replace=TRUE,
+ prob=c(.20,.80))
> fgm <- matrix(fg,nrow=10)
> fgmd <- cbind(fgm,apply(fgm,1,mean)*100)
> fgmd
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]  [,9]
 [1,]    1    1    1    1    1    1    1    1 100.0
 [2,]    1    0    1    1    1    1    1    1  87.5
 [3,]    1    0    1    0    0    1    0    1  50.0
 [4,]    1    1    1    0    1    0    1    1  75.0
 [5,]    0    1    1    1    1    1    1    1  87.5
 [6,]    1    1    1    1    0    1    1    0  75.0
 [7,]    1    1    1    1    1    1    1    1 100.0
 [8,]    1    1    1    1    1    1    1    1 100.0
 [9,]    0    0    1    1    1    1    1    1  75.0
[10,]    1    1    1    1    1    1    1    1 100.0
```

Binomial Distribution

When a sequence of Bernoulli trials conforms to the following list of requirements it is called a **binomial experiment**.

1. The experiment consists of a fixed number (n) of Bernoulli trials.

Binomial Distribution

When a sequence of Bernoulli trials conforms to the following list of requirements it is called a **binomial experiment**.

1. The experiment consists of a fixed number (n) of Bernoulli trials.
2. The probability of success for each trial, denoted by π, is constant from trial to trial. The probability of failure is $\varrho = (1 - \pi)$.

Binomial Distribution

When a sequence of Bernoulli trials conforms to the following list of requirements it is called a **binomial experiment**.

1. The experiment consists of a fixed number (n) of Bernoulli trials.
2. The probability of success for each trial, denoted by π, is constant from trial to trial. The probability of failure is $\varrho = (1 - \pi)$.
3. The trials are independent.

Binomial Distribution

When a sequence of Bernoulli trials conforms to the following list of requirements it is called a **binomial experiment**.

1. The experiment consists of a fixed number (n) of Bernoulli trials.
2. The probability of success for each trial, denoted by π, is constant from trial to trial. The probability of failure is $\varrho = (1 - \pi)$.
3. The trials are independent.
4. The random variable of interest, X, is the number of observed successes during the n trials.

Binomial PDF

The **Binomial** probability distribution function (pdf) is written

$$\mathbb{P}(X = x|n, \pi) = \frac{n!}{(n-x)!x!}\pi^x(1-\pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

If X has a Binomial distribution with parameters n and π, then the mean and standard deviation of X are

- $\mu_X = E[X] = n\pi$

Binomial PDF

The **Binomial** probability distribution function (pdf) is written

$$\mathbb{P}(X = x|n, \pi) = \frac{n!}{(n-x)!x!}\pi^x(1-\pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

If X has a Binomial distribution with parameters n and π, then the mean and standard deviation of X are

- $\mu_X = E[X] = n\pi$
- $\sigma_X = \sqrt{n\pi(1-\pi)}$

Binomial PDF

The **Binomial** probability distribution function (pdf) is written

$$\mathbb{P}(X = x|n, \pi) = \frac{n!}{(n-x)!x!}\pi^x(1-\pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

If X has a Binomial distribution with parameters n and π, then the mean and standard deviation of X are

- $\mu_X = E[X] = n\pi$
- $\sigma_X = \sqrt{n\pi(1-\pi)}$

Using `dbinom()`

- `dbinom(x, size, prob)` computes the probability that $\mathbb{P}(X = x|n, \pi)$ given $X \sim Bin(n = size, \pi = prob)$.

Using `dbinom()`

- `dbinom(x, size, prob)` computes the probability that $\mathbb{P}(X = x|n, \pi)$ given $X \sim Bin(n = size, \pi = prob)$.
- Example: The mortality rate of a certain disease is 34%. Of ten patients who have the disease, what is the probability that more than half will die from the disease?

Using `dbinom()`

- `dbinom(x, size, prob)` computes the probability that $\mathbb{P}(X = x|n, \pi)$ given $X \sim Bin(n = size, \pi = prob)$.
- Example: The mortality rate of a certain disease is 34%. Of ten patients who have the disease, what is the probability that more than half will die from the disease?
- Note: $X \sim Bin(n = 10, \pi = 0.34)$, and we want to find $\mathbb{P}(X > 5) = 1 - \mathbb{P}(X \leq 5) = \mathbb{P}(X = 5) + \mathbb{P}(X = 4) + \mathbb{P}(X = 3) + \mathbb{P}(X = 2) + \mathbb{P}(X = 1) + \mathbb{P}(X = 0)$.

Doing the Math

$$\begin{aligned}\mathbb{P}(X=5) &= \binom{10}{5} \times 0.34^5 \times (1-0.34)^{10-5} \\ &= \frac{10!}{5! \times (10-5)!} \times 0.34^5 \times (1-0.34)^{10-5} \\ &= 0.1433887\end{aligned}$$

```
> choose(10,5)*.34^5*(1-.34)^(10-5)
[1] 0.1433887
> dbinom(5,10,.34)
[1] 0.1433887
```

Link to the R Script

- Go to my web page Script for Generating Random Data
- Homework: problems 3.32 - 3.52
- See me if you need help!