A Very Basic Introduction to R – Part IV

Simulation of Binomial Random Variables

The `rbinom()` function can be used to simulate N independent binomial random variables.

For example, we can generate 15 binomial random numbers with parameters $n = 4$ and $p = .7$ as follows:

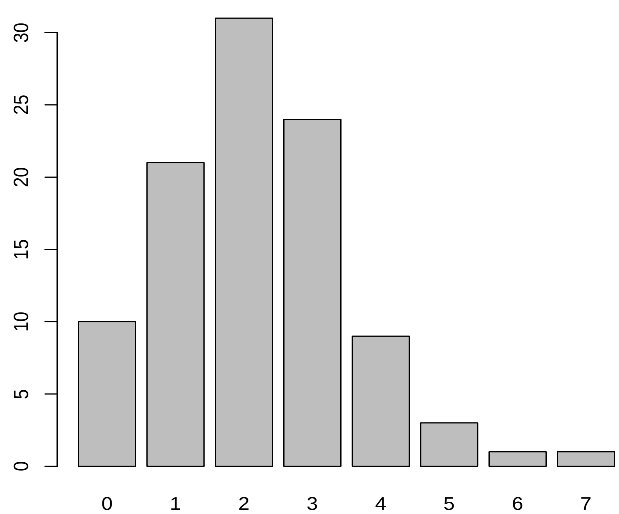
```
> rbinom(15, 4, .7)

 [1] 3 3 3 2 4 4 2 3 2 2 4 3 3 3 4
```

In the next example, we will assign 100 independent binomial numbers with parameters $n = 7$ and $p = .3$ to a vector object called `V`.

```
> V <- rbinom(100, 7, .3)
```

We can plot a bar chart of the numbers:

```
> barplot(table(V))
```



The bar plot is an estimate of the probability distribution $P(V = v)$. It is more appropriate than a histogram, because the data are discrete, not continuous.

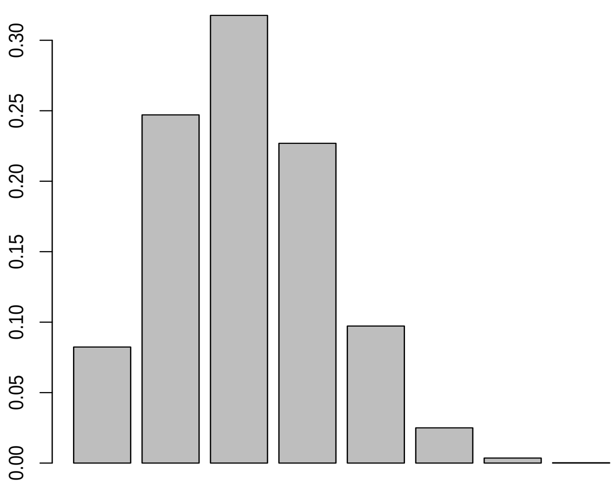
The theoretical distribution is

$$P(V = v) = \binom{n}{v)} p^v (1-p)^{n-v}$$

for $v = 0, 1, 2, \ldots, n$, and 0, otherwise. It can be calculated using the `dbinom()` function.

For example,

```
> barplot(dbinom(0:7, 7, .3))
```



We can calculate the mean, standard deviation and variance of the sample in `V`:

```
> mean(V)

[1] 2.19

> sd(V)

[1] 1.368439

> var(V)

[1] 1.872626
```

The sample mean is an estimate of the population mean (or expected value). The expected value of a binomial random variable is given by

$$E[V] = np$$

The variance of a random variable is calculated from the formula:

$$\mathrm{Var}(X) = np(1-p)$$

The standard deviation is $\sqrt{np(1-p)}$.

For the binomial random variable with $n = 7$ and $p = .3$, we can calculate these quantities by hand or by using R:

```
> n <- 7
> p <- .3
> n*p # Expected Value (or Mean)

[1] 2.1

> n*p*(1-p) # Variance

[1] 1.47

> sqrt(n*p*(1-p)) # Standard Deviation

[1] 1.212436
```

We can compare these with what we would obtain from a simulated sample of 10000 binomial random variables:

```
> B <- rbinom(10000, 7, .3)
> mean(B) # sample mean

[1] 2.1094

> var(B) # sample variance

[1] 1.451977

> sd(B) # sample standard deviation

[1] 1.20498
```

Exercises

1. Generate 20 binomial random numbers with $n = 17$ and $p = .45$, and plot a histogram of the resulting sample.

2. Calculate the expected value and variance of the above numbers.

3. Assign 2000 binomial random numbers with parameters $n = 40$ and $p = .3$ to a vector called `U`.

4. Calculate the mean value of `U`. Compare with the theoretical value.

5. Calculate the variance of `U`. Compare with the theoretical value.

6. Plot the histogram of the sample in `U`. Compare with a plot of the theoretical probability distribution.

7. Use simulation to estimate the expected value of U^3 where U is binomial with parameters $n = 10$ and $p = .55$.

8. Suppose V is a binomial random variable with mean 10 and variance 5. Find n and p. Then simulate 10000 values of V and compute their mean and variance, comparing with the theoretical values.

9. Simulate 200 binomial numbers with parameters $n = 17$ and $p = .3$. Find their sample mean and variance. Compare with the theoretical values.