

Project 3: Discrete Random Variables

Jose Lazarte

October 12, 2019

Simulation

Utilizing the four random variables studied in class with the parameters of p at 0.4 and a distribution of increasing sizes results as follows:

	Bernoulli		Binomial		Geometric		Poisson	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
10	0.200000	0.4216370	40.10000	3.784471	1.200000	2.201010	25.30000	4.691600
100	0.360000	0.4824182	40.05000	4.965283	1.770000	2.381982	24.35000	5.897055
1000	0.416000	0.4931401	40.06100	4.710782	1.528000	1.977657	25.06000	4.964982
1e+06	0.399661	0.4898289	40.00343	4.893737	1.500093	1.941303	24.99565	5.003777

The following table shows the different means and standard deviations between the four discrete random variables and can be considered an experimental value. The following functions can be seen as the theoretical functions:

Bernoulli

The Bernoulli distribution uses the variable p as the probability of an event occurring, and it is also the expected value.

Therefore, $E[X] = p$

As for the standard deviation the equation is as follows:

$$S = \sqrt{p(1-p)}$$

When utilizing the same parameters as the simulation, where probability is 0.4, our experimental mean should approximate 0.4 and the standard deviation should approximate 0.4899. Directly comparing the

values, we can see that the larger the sample size the more the theoretical and experimental are to one another supporting the simulation's results

Binomial

The Binomial distribution uses the variable p as the probability of an event occurring, and n as the number of times the event occurs.

Therefore, $E[X] = n * p$

As for the standard deviation the equation is as follows:

$$S = \sqrt{np(1-p)}$$

When using the same parameters as the simulation the aforementioned effect of high accuracy is present as the number of trials rises. The theoretical mean and standard deviation using the above functions for the binomial result in a mean of 40 when n is 100 and the probability is 0.4. Furthermore, the standard deviation is a projected 4.8990.

Geometric

The Geometric distribution uses the variable p as the probability of an event occurring.

Therefore, $E[X] = \frac{1-p}{p}$

As for the standard deviation the equation is as follows:

$$S = \frac{\sqrt{(1-p)}}{p}$$

Utilizing the parameter when p is equal to 0.4 the following equations result in a mean and standard deviation of 1.5 and 1.9365, respectively. This theoretical result does align with the simulation results and it also increases in accuracy as the number of trials increases.

Poisson

The Poisson distribution uses the variable λ as the probability of an event occurring.

Therefore, $E[X] = \lambda$

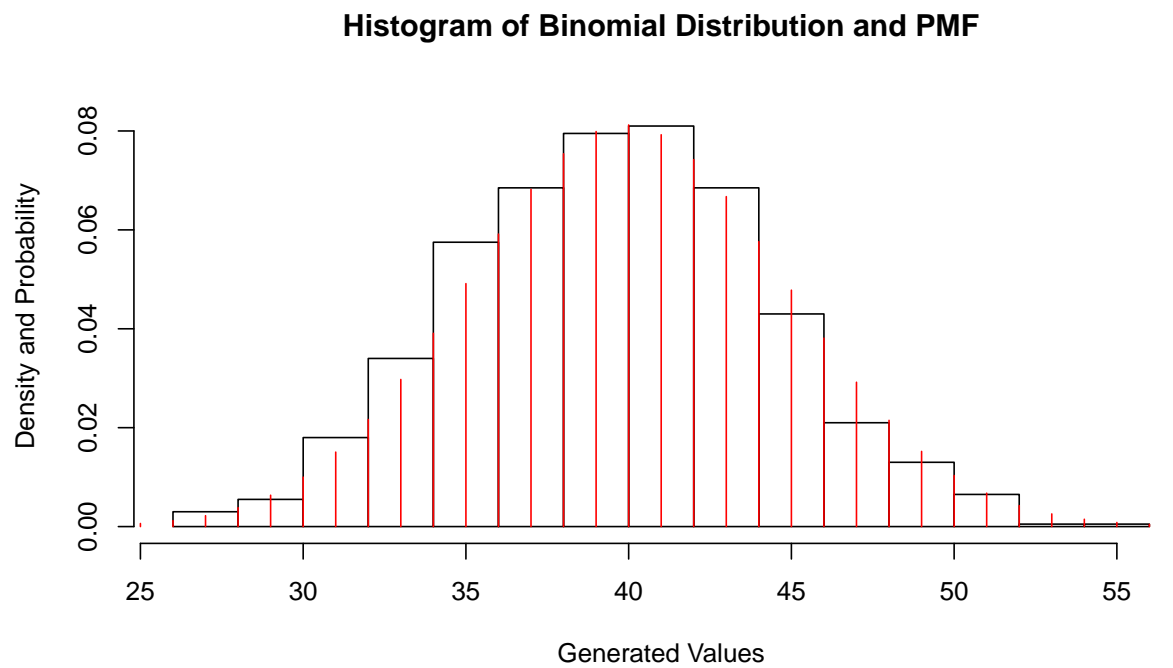
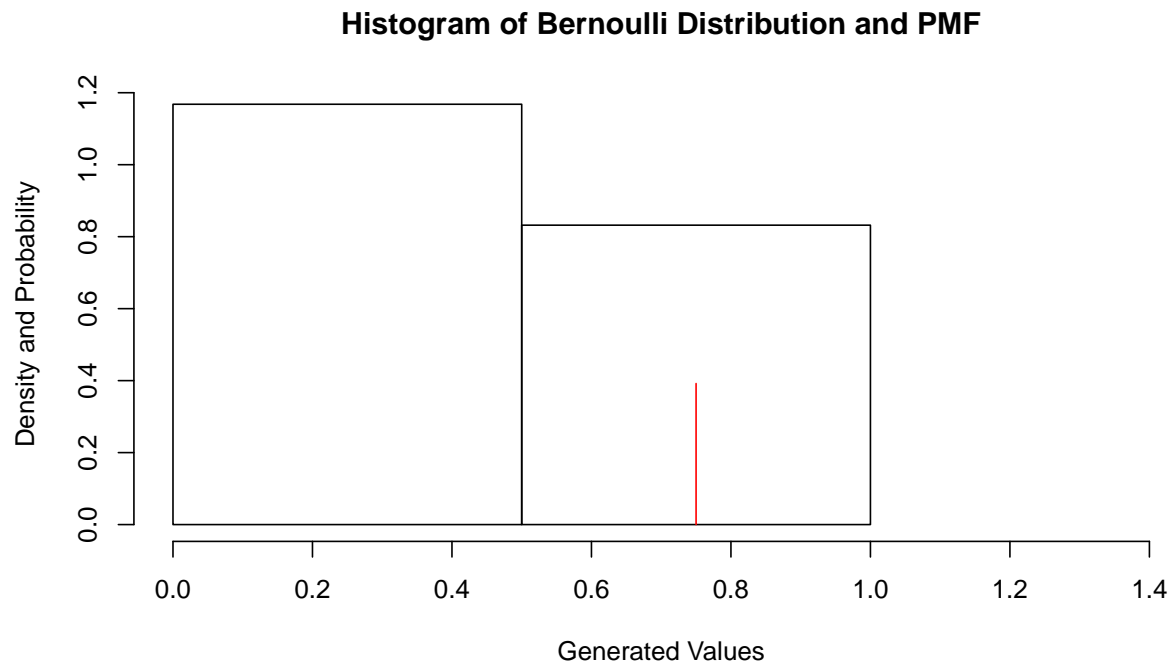
As for the standard deviation the equation is as follows:

$$S = \sqrt{\lambda}$$

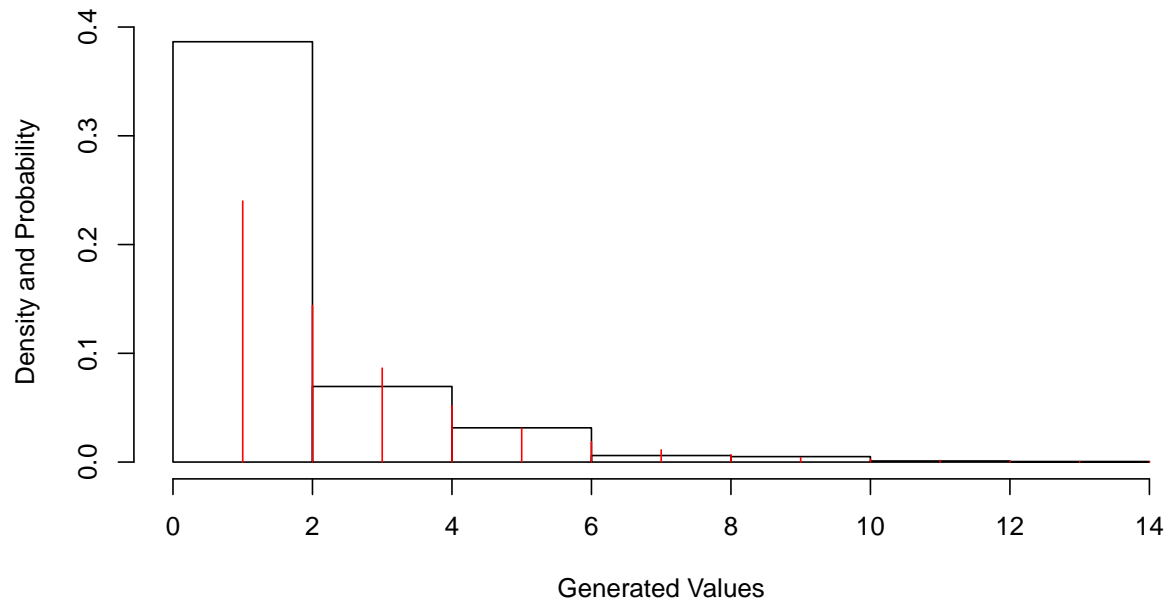
With the value of 25 corresponding to λ , the mean and standard deviation equal to 25 and 5, respectively.

As previous distributions, the accuracy does in fact rise with the rise in test completed.

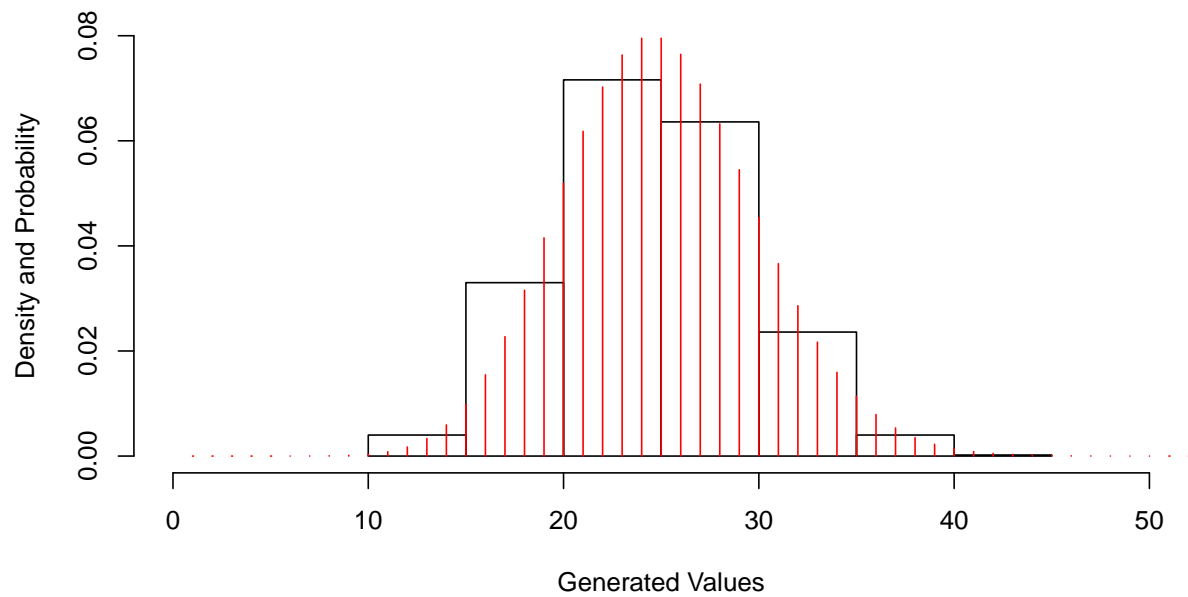
Histogram



Histogram of Geometric Distribution and PMF



Histogram of Poisson Distribution and PMF



Analysis

The first histogram on page 3 represents the simulation of a bernoulli distribution completed 1000 times. Although simple it represents the probability p that was entered. In the case of the histogram it was 0.4. This can be represented as the PMF where under the 1 the redline represents a value of 0.4. Although the line was moved for clarity, the approximated value of the PMF results to be the same as the value p which is also the mean.

The second histogram represents a binomial distribution with the red lines signifying PMF after 1000 trials with numbers ranging from 0 to 100. As seen with the PMF the most likely value to occur would be 40 which is also the mean, $n * p$, as for the histogram, it has a somewhat symmetric shape with a slight right skew since the mean is 40.

The following histogram on page 4 represents the Geometric distribution and PMF of p of 0.4 done 1000 times. With a huge right skew this can be expected as the highest frequencies occur with the first or second occasions. Afterwards the occurrences diminish as the probability of a success after the third or fourth occasions become smaller and smaller. The histogram and PMF both follow this pattern and is a key feature in Geometric distributions.

The final histogram is the Poisson histogram. Although similiar to the Binomial histogram, the biggest difference would be the spread which is much more symmetric compared to the binomial, and the mean which is changed to 25 rather than 40. However, the PMF of the Binomial and Poisson has an identical shape with only small differences in first third and final third of the PMF.

Application

Utilizing a Poisson Distribution the store simulation can be estimated: Where λ is 200 average customers 365 days in a year as the number of total simulations

```
sPois<-rpois(365,200)
```

Using the code above, sPois will hold 365 values that vary in size with the average of 200.

We can now continue to the following problem:

Where 29.95 is the cost of candles we can now multiply sPois array to receive the total money intake in a generated day. By finding the sum and dividing by 365 we are then able to find the average which is around 6000 dollars.

```
sum(sPois*29.95)/365
```

```
[1] 5993.528
```

To the following problem, a good day of 95% percentile can be obtained using the qpois function. Where 0.95 is the wanted percentile and 200 is the λ we receive a value of:

```
qpois(.95,200)
```

```
[1] 224
```

This value is the number of λ that is needed for the store to have a “Good Day” of 95th percentile