Yuyi Zhang

May 23, 2020

Spring 2020 CPS Quarter Term B

Instructor: Danielle Mauro

Project 1: Chi-squared Goodness of Fit Test Project

ALY 6050\_Introduction to Enterprise Analytics

# **Introduction**

In this report, we are going to illustrate four types of random values by create histogram, select probability distribution, and plot probabilities. After analysis, we will perform the Chi-squared Goodness of Fit test to them, which would help to verify their distribution.

# **Analysis**

## **Problem 1**

### Histogram of X

Below is the histogram of X. This is a right skewed distribution. The peak value is on the left-hand side.

A screenshot of a cell phone

Description automatically generated

*Figure 1*. Histogram of X

### Select distribution

Based on the histogram that generated from the last question, I’ll select exponential distribution for the equation.

θ = mean(X) = 0.967

μ = 0.967

σ = 0.967

### Probability Plot of X

There the plot we’ve got based on the parameters we’ve got from the last question.

A screenshot of a cell phone

Description automatically generated

*Figure 2*. Probability Plot of X

### Chi-squared test

In this phase, we are going to test whether the variables fit on exponential distribution with the level of significance: α = 0.05. State null and alternative hypotheses as below:

Null hypothesis H0: the random variable follows the exponential distribution

Alternative hypothesis H1: the random variable does not follow the exponential distribution

A screenshot of a social media post

Description automatically generated

*Figure 3*. Chi-squared test result

Compare P-value and significance value. In this case, the P-value (1) is larger than the significance level (α = 0.05). So there isn’t sufficient evidence to reject H0. In another word, X fits in exponential distribution.

### Explain what you have learned from this experiment.

Natural logarithm of standard uniform variables belong to the exponential distribution.

## **Problem 2**

### Histogram of X

Below is the histogram of X. This is a right skewed distribution. The peak value is on the left-hand side.

A screenshot of a cell phone

Description automatically generated

*Figure 4*. Histogram of X

### Select distribution

Based on the histogram that generated from the last question, I’ll select gamma distribution for the equation.

α = number of variables = 3

β = rate = 1

μ = α \* β = 3

σ = (α \* β2)0.5

### Probability Plot of X

There the plot we’ve got based on the parameters we’ve got from the last question.

A screenshot of a cell phone

Description automatically generated

*Figure 5*. Probability Plot of X

### Chi-squared test

In this phase, we are going to test whether the variables fit on gamma distribution with the level of significance: α = 0.05. State null and alternative hypotheses as below:

Null hypothesis H0: the random variable follows the gamma distribution

Alternative hypothesis H1: the random variable does not follow the gamma distribution

A screenshot of a social media post

Description automatically generated

*Figure 6*. Chi-squared test result

Compare P-value and significance value. In this case, the P-value (1) is larger than the significance level (α = 0.05). So there isn’t sufficient evidence to reject H0. In another word, X fits in Gamma distribution.

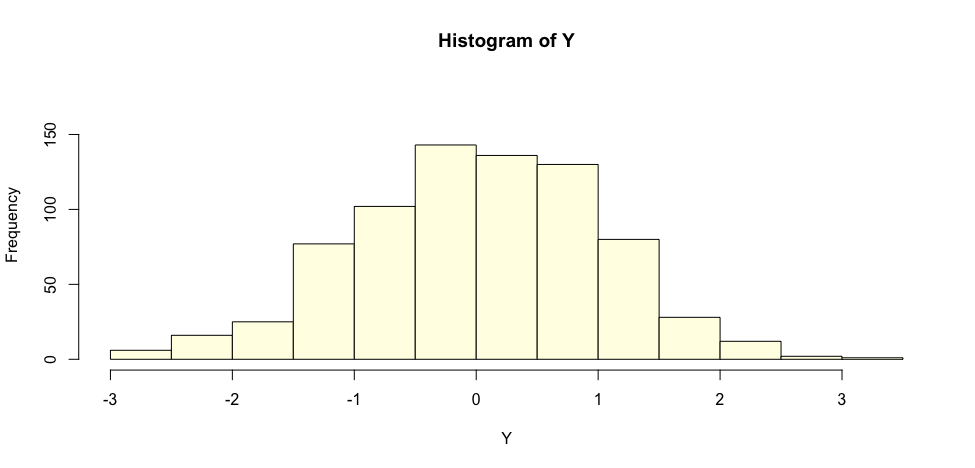
### Explain what you have learned from this experiment.

Three Natural logarithm of standard uniform variables belong to the Gamma distribution.

## **Problem 3**

### Histogram of Y

Below is the histogram of Y. This is a bell shape distribution. The peak value is in the middle of the shape.



*Figure 7*. Histogram of Y

### Select distribution

Based on the histogram that generated from the last question, I’ll select normal distribution for the equation.

μ = E(Y) = 0.0137

σ = 1.0026

### Probability Plot of Y

There the plot we’ve got based on the parameters we’ve got from the last question.

A screenshot of a cell phone

Description automatically generated

*Figure 8*. Probability Plot of Y

### Chi-squared test

In this phase, we are going to test whether the variables fit on normal distribution with the level of significance: α = 0.05. State null and alternative hypotheses as below:

Null hypothesis H0: the random variable follows the normal distribution

Alternative hypothesis H1: the random variable does not follow the normal distribution

A screenshot of a social media post

Description automatically generated

*Figure 9*. Chi-squared test result

In this test, since all entries of ‘x’ must be nonnegative, so 3 was added to the original observed value, for the calculation purposes. Compare P-value and significance value. In this case, the P-value (1) is larger than the significance level (α = 0.05). So there isn’t sufficient evidence to reject H0. In another word, Y fits in normal distribution.

### Explain what you have learned from this experiment.

In this case, two X variables are exponential distributions. Y is for comparing and counting on two logarithm variables. Since there are large number of random values were selected for the experiment, the k value didn’t perform a significant role in this calculation. What we are doing is to compare random number r with 0.5. So the Y is a normal distribution after counting.

## **Problem 4**

### Select expected value and standard deviation of W

μ = E(Y) = 1.3444

σ = 0.0521

### Select distribution

Based on the histogram that generated from the last question, I’ll select uniform distribution for the equation.

μ = E(Y) = 1.3444

σ = 0.0521

A screenshot of a cell phone

Description automatically generated

*Figure 10*. Histogram of W

### Chi-squared test

In this phase, we are going to test whether the variables fit on uniform distribution with the level of significance: α = 0.05. State null and alternative hypotheses as below:

Null hypothesis H0: the random variable follows the uniform distribution

Alternative hypothesis H1: the random variable does not follow the uniform distribution

A screenshot of a social media post

Description automatically generated

*Figure 9*. Chi-squared test result

Compare P-value and significance value. In this case, the P-value (1) is smaller than the significance level (α = 0.05). So there is sufficient evidence to reject H0. In another word, W does not fit in uniform distribution.

### As the number of iterations M becomes larger, the values W will approach a certain limiting value. Investigate this limiting value of W by completing the following table and plotting W versus M. What value do you propose for the limiting value that M approaches to?

Based on the experiment of M in R code, the estimation of W would be 1.3444.

|  |  |
| --- | --- |
| M | W |
| 10 | 1.250 |
| 20 | 1.176 |
| 30 | 1.304 |
| 40 | 1.379 |
| 50 | 1.351 |
| 60 | 1.333 |
| 70 | 1.346 |
| 80 | 1.356 |
| 90 | 1.364 |
| 100 | 1.370 |
| 200 | 1.316 |
| 300 | 1.282 |
| 400 | 1.307 |
| 500 | 1.316 |
| 600 | 1.307 |
| 700 | 1.308 |
| 800 | 1.307 |
| 900 | 1.310 |
| 1000 | 1.312 |

*Table 1*. Compare M and W

### Explain what you have learned from this experiment.

# **Conclusion**

Findings in parts 1 – 4 above by filling the blanks in the sentences below:

1. If R is a standard uniform random variable, then −Ln(R) has the \_\_\_\_\_Exponential\_\_\_\_\_ probability distribution.

2. The sum of three independent and identically distributed \_\_\_Uniform\_\_\_\_ random variables have the \_\_\_\_Gamma\_\_\_ probability distribution.

3. The output of the algorithm of problem 3 has a \_\_\_\_\_Normal\_\_\_\_\_ probability distribution.

4. In step 2 of the algorithm of problem 3, random variables X1 and X2, each of whose probability distribution is \_\_\_\_\_Exponential\_\_\_\_\_ are used to generate a random value Y that has the \_\_\_\_Normal\_\_\_probability distribution.

5. The random value W that was discussed in problem 4, has the \_\_\_\_\_Uniform\_\_\_\_\_\_\_ probability distribution. The expected value of W is: \_\_\_1.34\_\_\_.