Test for Uniformity

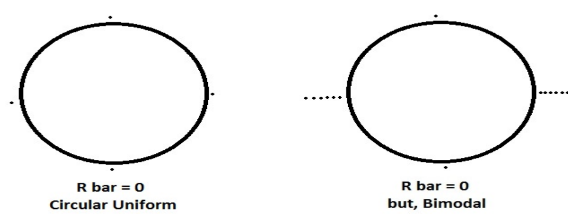
Directional Data Analysis Project

Abhishek Ghosh MS1509
Pratyay Sengupta MS1505
Samiran Kundu MS1513
Soumajit Biswas MS1514

Supervisor: Prof. Ashis Sengupta, ASU

1 Introduction

Uniformity refers to the situation in which all values around the circle are equally likely. Occasionally, it is useful to perform a statistical test whether a set of data follows the uniform distribution. Several tests of uniformity have been developed. Note that when any of these existing tests is rejected, we can conclude that the data were not uniform. However, when the test is not rejected, we cannot conclude that the data follows the uniform distribution. Rather, we do not have enough evidence to reject the null hypothesis of uniformity. The above fact can be easily verified from the following figures of a bi-modal distribution (considering Rayleigh Test).



2 Objective

Our objective is to suggest some suitable tests which overcome the drawbacks of the existing tests. To summarize,we shall perform Raleigh's test first.If the null hypothesis is rejected we conclude that the population is not uniform.But if we do not have enough evidence to reject the null hypothesis,we perform any of the following suggested tests to conclude.

3 Suggested methods

In this project we are suggesting three test procedures which are described in the subsection below:

3.1 Test based on Mobius transformation

3.1.1 Intuition

The test is based on the fact that Mobius transformation (one-to-one transformation) on circular uniform distribution gives rise to linear Cauchy distribution. Therefore the problem of testing uniformity on the circular distribution boils down to the problem of testing whether the corresponding transformed linear distribution follows standard Cauchy distribution.

3.1.2 Method

Let, $\theta_1, \theta_2, ..., \theta_n$ be i.i.d. observations from F. Our objective is to test:

$$H_0\text{: F is circular uniform distribution. ag } H_1 : H_0 \text{ is not true.}$$

Using Mobius transformation, we get on $X_i = tan(\frac{\theta_i}{2})$ for all $i = 1, 2, ..., n$. Let, $X_1, X_2, ..., X_n$ follows F'. Thus the problem boils down to testing:

$$H_0'\text{: F' is Standard Cauchy distribution. ag } H_1' : H_0' \text{ is not true.}$$

We test the above using the Kolmogorov–Smirnov test.

3.1.3 Simulation

We simulate observations from Circular Uniform and Bimodal Circular Uniform (modes at 0 *and pi*) distributions and perform the Mobius transformation for both set of observations. After that we perform KS test to test for uniformity in both cases. The results are as follows:

```
> t <- runif(100,0,2*pi)     #generating circular uniform variables
> x <- tan(t/2)    #Mobius tranformation
> y <- rcauchy(100)     #generating Cauchy(0,1) variables
> ks.test(x,y)

        Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.12, p-value = 0.4676
alternative hypothesis: two-sided
```

As p-value > 0.05, we accept the test for uniformity in case of Circular Uniform distribution, which is desired.

```
> t <- bimodal(100)     #generating circular uniform variables
> x <- tan(t/2)    #Mobius tranformation
> y <- rcauchy(100)     #generating Cauchy(0,1) variables
> ks.test(x,y)

        Two-sample Kolmogorov-Smirnov test

data:  x and y
D = 0.24, p-value = 0.006302
alternative hypothesis: two-sided
```

As p-value < 0.05, we accept the test for uniformity in case of Bimodal Circular Uniform distribution, which is desired.

3.2 Shift Method/Shift and Goodness of fit method

3.2.1 Intuition

This method relies on the fact that Circular Uniform distribution remains the same under translation or shift by any angle whereas, in the other cases, a shift changes the whole distribution. The shift angle is chosen to be very small because, even if the underlying distribution is not uniform, it may be invariant of the shift by a suitable angle.

3.2.2 Method

Let, $\theta_1, \theta_2, ..., \theta_n$ be i.i.d. observations from F. We shift the whole data by δ and obtain a new set of independent observations $\theta'_1, \theta'_2, ..., \theta'_n$ where

$$\theta'_i = (\theta_i + \delta) mod(2\pi)$$

for all $i = 1, 2, ..., n$. Then, $\theta'_1, \theta'_2, ..., \theta'_n$ are i.i.d. observations from F'. Our aim is to check the difference of these two distributions. Let, F_n and F'_n, be the empirical distributions of the original sample and the sample obtained after shift respectively. We check the difference of these two by GOF test. We propose the following test statistic to test:

H_0: F is Circular Uniform distribution. ag $H_1 : H_0$ is not true.

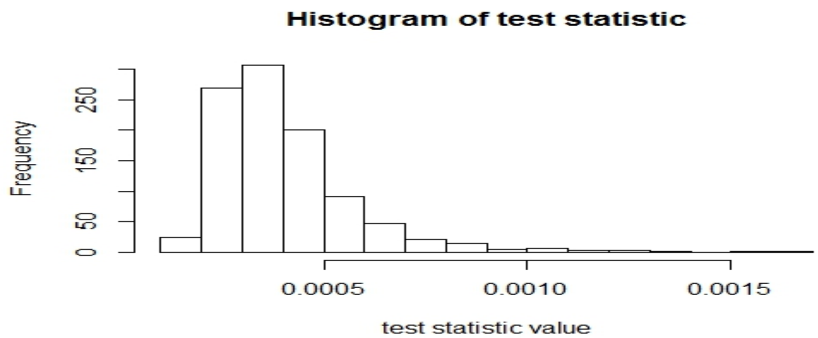
Here our test statistics is

$$T_n(\delta) = \int \Big(F_n(x) - F'_n(x) - \int \left(F_n(x) - F'_n(x) \right) \Big)^2 dF_n$$

Now if $T_n(\delta)$ is large we have enough reasons to suspect the uniformity of the distribution. Now, in order to test the uniformity using this statistics we need to find the cut-off points. It is clear that we will reject H_0 if $T_n(\delta) > c$. Now in order to find c such that $P(T_n(\delta) > c) = \alpha$ under H_0, we can generate 1000 random samples each of size 100 and get $T_n(\delta)$ for each of the random samples are get the sample $(1-\alpha)^{th}$ quantile to get the required cut-off point for the test.

3.2.3 Simulation

Firsly, the cut off value is computed from the empirical distribution of the test statistic, which is represented by the histogram of the test statistic values after repeated simulations.



The 95^{th} percentile is coming out to be 0.0007201.

We simulate observations from Circular Uniform and Bimodal Circular Uniform distributions and perform the Shift method to test for uniformity in both cases. The results are as follows:

```
> theta <- runif(100,0,2*pi)
> t = 10*2*pi/360
> ifelse(w2(theta,t) > cut, "reject","accept")
     95%
"accept"
```

As the value of the test statistic < the cut-off value(95^{th} percentile), we accept the test for uniformity in case of Circular Uniform distribution, which is desired.

```
> theta <- bimodal(100)
> t = 10*2*pi/360
> ifelse(w2(theta,t) > cut, "reject","accept")
     95%
"reject"
```

As the value of the test statistic > the cut-off value(95^{th} percentile), we accept the test for uniformity in case of Bimodal Circular Uniform distribution, which is desired.

3.3 Supremum Method

3.3.1 Intuition

A slight modification to the previous method gives rise to this method. Here, in spite of taking a particular angle of shift, we take shifts by all possible angles and take the supremum of the differences corresponding to each shift over the angle of shifts.

3.3.2 Method

Let, $\theta_1, \theta_2, ..., \theta_n$ be i.i.d. observations from F. We shift the whole data by an angle δ and obtain a new set of independent observations $\theta'_1, \theta'_2, ..., \theta'_n$ where

$$\theta'_i = (\theta_i + \delta) mod(2\pi)$$

for all $i = 1, 2, ..., n$. Then, $\theta'_1, \theta'_2, ..., \theta'_n$ are i.i.d. observations from F'..Our aim is to check the difference of these two distributions. Let, F_n and F'_n, be the empirical distributions of the original sample and the sample obtained after shift respectively. We check the difference of these two by GOF test. But this depends on the shift δ. Hence, we consider different values of δ and calculate the supremum of the differences of the original distribution and the shifted distribution by GOF test. We propose the following test statistic to test:

$$H_0\text{: F is circular uniform distribution. ag } H_1 : H_0 \text{ is not true.}$$

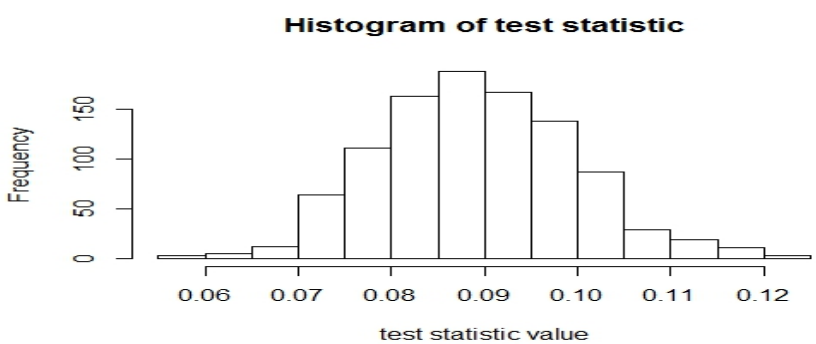
Here our test statistics is

$$T_n = \sup_{\delta} T_n(\delta) = \sup_{\delta} \left\{ \int \Big(F_n(x) - F'_n(x) - \int (F_n(x) - F'_n(x)) \Big)^2 dF_n \right\}$$

Now if T_n is large we have enough reasons to suspect the uniformity of the distribution. Now in order to test the uniformity using this statistics we need to find the cut-off points. It is clear that we will reject H_0 if $T_n > c$. Now in order to find c such that $P(T_n > c) = \alpha$ under H_0, we can generate 1000 random samples each of size 100 and get T_n for each of the random samples are get the sample $(1 - \alpha)^{th}$ quantile to get the required cut-off point for the test.

3.3.3 Simulation

Firsly, the cut off value is computed from the empirical distribution of the test statistic, which is represented by the histogram of the test statistic values after repeated simulations.



The 95^{th} percentile is coming out to be 0.10240.

We simulate observations from Circular Uniform and Bimodal Circular Uniform (modes at 0 *and pi*) distributions and perform the Supremum method to test for uniformity in both cases. The results are as follows:

```
> ##unimodal circular uniform
> theta <- runif(100,0,2*pi)
> t2 <- (seq(0,90))*2*pi/360
> ifelse(w2t(theta,t2) > cut2, "reject","accept")
     95%
"accept"
```

As the value of the test statistic < the cut-off value(95^{th} percentile), we accept the test for uniformity in case of Circular Uniform distribution, which is desired.

```
> ##bimodal circular uniform
> theta <- bimodal(100)
> t2 <- (seq(0,90))*2*pi/360
> ifelse(w2t(theta,t2) > cut2, "reject","accept")
    95%
"reject"
```

As the value of the test statistic > the cut-off value(95^{th} percentile), we accept the test for uniformity in case of Bimodal Circular Uniform distribution, which is desired.

4 Conclusion

As we can see from the above simulations, all of our suggestions successfully detected the multi-modal cases when the Rayleigh Test is accepted.

5 Further Scopes

- Instead of finding the cut off values from simulating several samples and evaluating the value of the test statistic,we can try to find out the exact theoretical distribution of the test statistic.
- We can also think about the situations where each of the above mentioned tests will perform better that the other two.
- In spite of using the GOF test one can use a measure of difference of these two distributions which is independent of change of origin.