2019

Validation of a Lottery

Xiaona Zhou
CUNY New York City College of Technology

Validation of a Lottery
Xiaona Zhou

Honors Program
Mentor: Prof. Jonathan Natov, Mathematics Department

ABSTRACT

The NY Pick 4 lottery consists of four "randomly" chosen digits from 0 to 9. For this to be fair, each digit should be equally likely to occur. To determine whether this is the case, a Chi-squared goodness of fit test will be applied to historical data. This provides a quantitative way of measuring how well the observed frequency of digits matches our expectations of a fair lottery. We also explore the "Lucky Sum", which is also a part of the Pick 4. We determine which sum is most likely to occur, and what the odds of winning are if you play the sum most likely to occur. We also developed a MATLAB program, which allows us to find Lucky Sum of any pick. Furthermore, we found the equation that fits the curve approximately.

INTRODUCTION

The fairness of Pick 4 lottery can be determined by using the Chi-squared test. The Q value and a Chi-squared goodness of fit test are determined in R (which is a programming language). To find which Lucky Sum is most likely to occur, similar codes are used. After that, a fair model of Lucky Sum is generated using parallelogram number planes, and this fair model is used to find how well the frequencies of the Lucky Sum match the model. Lastly, the MATLAB codes for finding the Lucky Sum of pick n are demonstrated, which returns the distribution of the Lucky Sum, the most common Lucky Sum, and the chance of get that Lucky Sum.

DETERMINE THE FAIRNESS OF PICK 4

To determine the fairness of Pick 4, a Chi-squared test is used on data over a period of thirty days. The sample R-codes used in doing a Chi-squared test are illustrated below:

> raw<-c(winning numbers)
> length(raw)
> count<-0
> for(k in 0:9){count[k+1]<-sum(k==raw)}
> Q<-sum((count-24)^2/24)
> pchisq(Q,9)

The 95th percentile of the Chi-squared distribution with 9 degrees of freedom is 16.91898, and we failed to reject any outcomes. Since one of the Q values is equal to 1.33333 and seems suspicious, I computed another 25 groups of data, and the Q values are ranged from 1.333333 to 14.58333. A histogram of Q values is displayed below.
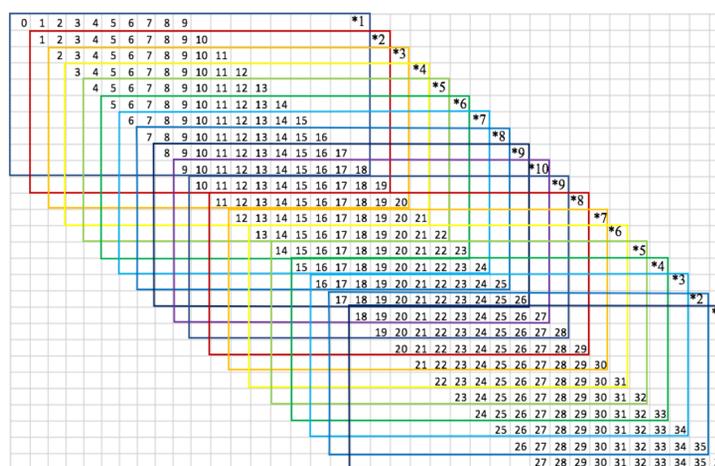
Histogram of Q_Values
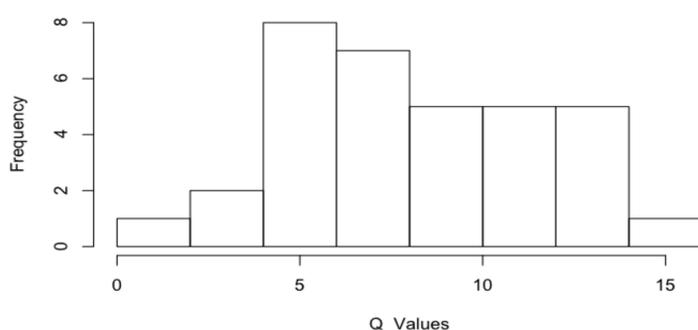
GENERATE A FAIR MODEL FOR LUCKY SUM

Lucky Sum is the sum of the four winning numbers. You can win a prize if the sum of your numbers matches Lucky Sum. In graph below, it contains the Lucky Sum of all the possible combinations in Pick 4.

*(number) represent the number of times this plane has been generated

An example of generating sum 18:

1*1+2*2+3*3+4*4+5*5+6*6+7*7+8*8+9*9+10*10+9*9+8*8+7*7+6*6+5*5+4*4+3*3+2*2+1*1=670

(# of sum 18 in Plane 16) (# of Plane 16)

After doing the same computation for all the Lucky Sums, we found that Lucky Sum 18 is most likely to occur, and according to historical data, Lucky Sum 18 occurred 824 times and Lucky Sum 19 occurred 822 times(the total is 12330 times) from 12/02/2001 to 09/23/2018.

HOW WELL THE FREQUENCIES MATCH

To find how well the frequencies match, we applied Chi-squared test on historical data and fair model. This can be done in R use codes below:

> model<-c(numbers of each sums in fair model)
> observed<-c(numbers of each sums in historical data from 12/02/2001 to 09/23/2018)
> for(k in 0:36){observed[k+1]<-sum(observed==k)}
> Q<-sum((model-observed)^2/model)

The Q value is 24.7085. Since the 95th percentile of the Chi-squared distribution with 36 degrees of freedom is 50.99846, it can be said that the frequencies of the "Lucky Sum" match the model quite well.

FIND PICK N USING MATLAB

Since it is unrealistic to find the Lucky Sum of pick n using parallelogram number planes, I wrote a program, which will find the Lucky Sum of pick n and the most common Lucky Sum, display the graph of the Lucky Sum, and calculate the chance of getting that Lucky Sum. The result returned from MATLAB is exactly what we got using parallelogram number planes. A part of MATLAB codes, and the graph of Lucky Sum distribution returned from the program are illustrated below. By using the curve fitting tool in MATLAB, I found the following equation that fits the curve with R-squared =0.999.

$$f(x) = a * e^{-\left(\frac{x-b}{c}\right)^2}$$

where $a \approx \max(y)$; $b =$ most common Lucky Sum; $c \approx$ standard deviation of all the Lucky Sums $* \sqrt{2}$.

```
1 -    prompt = 'Please enter your number from 2 to n: ';
2 -    ci= input(prompt);
3 -    pick=0:9;%x contains pick 1: numbers from 0-9.
4 -    pickk=[];
5 -    for n=1:ci-1%repeat this loop ci time(s)
6 -        for z=0:9
7 -            x=pick+z;%add numbers from 0-9 to each number in pick n-1 to get pick n.
8 -            pickk=[pickk,x]; %contain all the lucky sum can be generated in pick n
9 -        end
10 -       pick=pickk;
11 -       pickk=[];%make pickk an empty set.
12 -   end
13 -   A=pick;
14 -   ss=[];
15 -   for i=0:9*ci
16 -       s=sum(A==i);%count each lucky sum
17 -       ss=[s,ss];%contains the total number of each lucky sum in pick n
18 -   end
```

Command Window
```
>> picknnnforproject
Please enter your number from 2 to n: 4
lucky Sum 18 is most common one
The number of ways of getting this lucky sum is 670
The probability of getting this sum is 6.7%
```
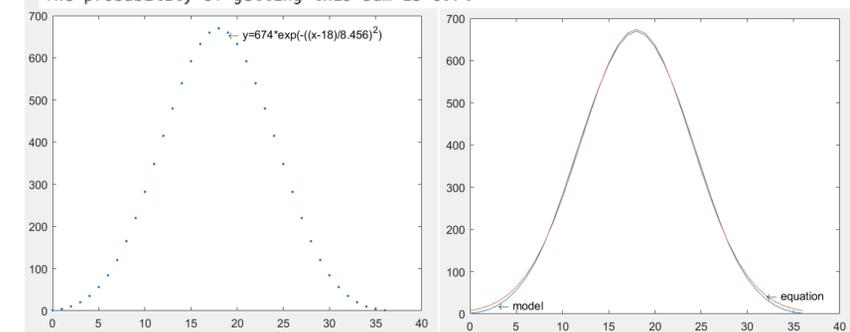

CONCLUSION

In this project, we determined that the NY Pick 4 lottery is a fair game. Both in the fair model and in the historical Lucky Sum, we found that the sum 18 is most likely to occur. The sum 18 occurred 819.410 times in fair model and 824 times in NY Pick 4 lottery. Therefore, the chance of winning when you have lucky sum 18 is about 6.7%. After applying Chi-square test to Lucky Sum, we found that the Q value was 24.7085. The Q value is not large enough to reject, so we can't conclude that Lucky Sum is unfair. And the MATLAB program makes the process of finding Lucky Sum easier. We were also able to find an equation that fits the curve approximately.

SELECTED REFERENCES

Lottery Daily Numbers/Win-4 Winning Numbers: Beginning 1980.
https://data.ny.gov/Government-Finance/Pick4/hj4u-8nyt

This histogram shows the shape of the distribution of Q values. The shape is close to a belt shape, which is what Chi-square distribution looks like. Therefore, we believe that having a Q value equal to 1.33333 is a chance event. In conclusion, it can be said that Pick 4 lottery is a fair game.