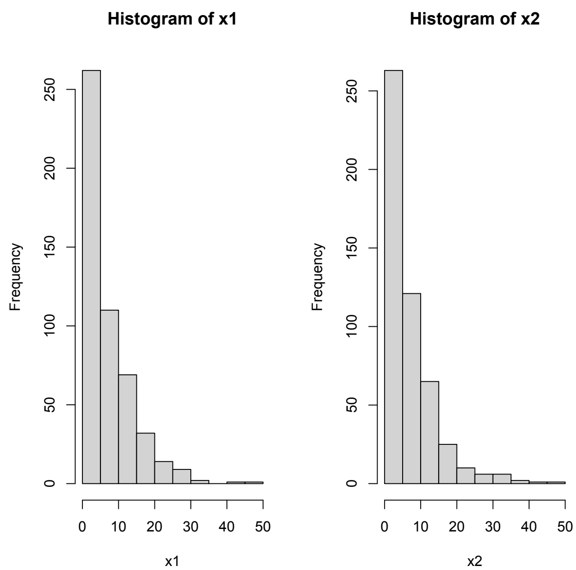
MATH382L R Lab 6. Exponential distribution

Jennifer Minnich
March 10, 2021

1 Problem 1 Exponential and Geometric

- a. Generate samples from the Exponential (mean = 7) distribution in two ways: Directly from Exponential (7) First obtain a sample from Exp(1) random variable, then multiply it by 7. Compare the two samples graphically.

  The two distributions are both right skewed with a very high density of points between 0 to 5. There axis a much greater amount of smaller values than larger values. The values in this exponential distribution range on the x-axis from 0 to 50 and between 0 to 260 on the y-axis. Because both have random generated values, the values are similar (like the graphs), but they are not exactly the same.

- b. Compute the mean of the resulting sample and compare it to the theoretical mean. Do the same for the variance.

```
> mean(x1)
[1] 6.940544
> mean(x2)
[1] 6.761188
> var(x1)
[1] 42.01383
> var(x2)
[1] 48.42749
```

  Because both have random generated values, the central tendency of the data is similar (like the graphs), but they are not exactly the same. The theoretical mean of histogram x1 is 6.94, compared to 6.76, the lesser sample mean of data set 2. We see the variance has more dispersion from the mean in data set x2, the variance being 48.4 versus 42.0 in data set x1.

2 Problem 2

- a. Compute (on paper) the half-life of radioactive material with parameter that is, the time x by which exactly 50% of the material has split. Somewhat surprisingly, it's not the mean of the distribution! It is called the median of the distribution, and it has an obvious relationship to the sample median.

```
> #Problem 2
> #1-e^-x/B=0.5
> #-x/B = ln(0.5)
> #x = B(ln(2))
> 7*log(2)
[1] 4.85203
> #another way to calculate median
> ?qexp
> qexp(.5, rate = 1/7, lower.tail = TRUE, log.p = FALSE)
[1] 4.85203
```

- b. Go back to your Exponential sample from Problem 1 and compare its median to your theoretical result.

```
> #Exponential sample
> median(x1)
[1] 4.761963
```

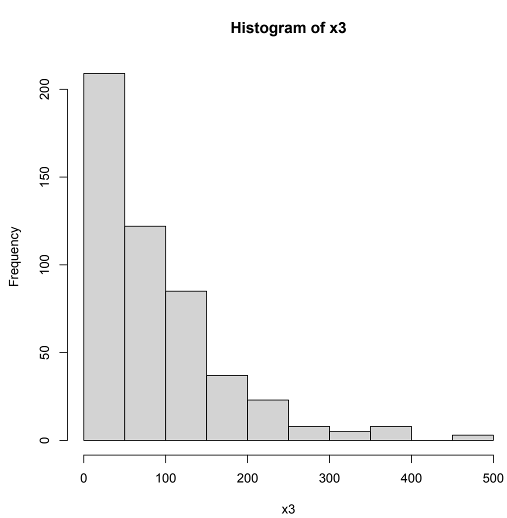
The results of the median are very close in proximity, 4.85 as our calculated median in problem a and 4.76 in the Exponential example from earlier. These are spot on, however, as problem a states, the value of the median is quite different from the mean.

3 Problem 3

- a. Generate 500 Exponential r.v.'s with the mean 88. Now, round the values up (use ceiling function).

```
> #Problem 3
> x3=ceiling(rexp(500,1/88))
> head(x3)
[1]  95  62 162  83  54  22
```

- b. Make a graph and convince yourselves that you are dealing with Geometric distribution.



Our graphs has the attributes of a Geometric distribution. There is a long tail right skew distribution to the right of the mean. All geometric distributions are right skewed with an infinite amount of values that can go to the right. The exponential distribution may be viewed as a continuous counterpart of the geometric distribution.

- c. What is the parameter p of this distribution? [Hint: try to evaluate P(Y=1) using Exponential CDF.]

```
> #3c
> #P(Y=1)
> pexp(1, rate = 1/88, lower.tail = TRUE, log.p = FALSE)
[1] 0.01129931
> ?rgeom
> x3g=rgeom(500, .01129931)
> mean(x3g)
[1] 88.412
```

- d. Verify your guess by generating a sample from Geometric (rgeom function) with that parameter and comparing results.

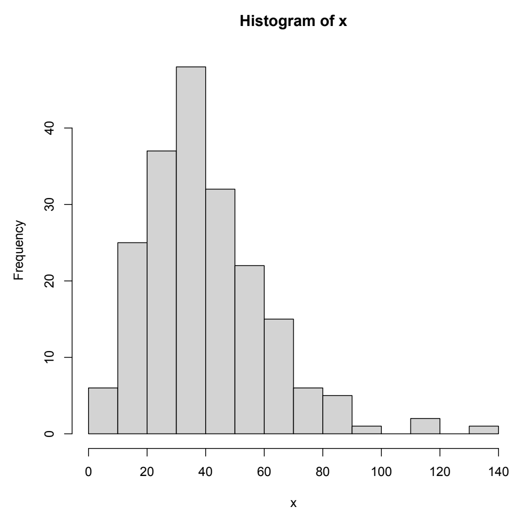
  My results came out well, when I calculated the Exponential CDF with mean = 88, I got a small p .01129931. I thought it was unlikely this was correct but when I calculated the rgeom function using that p, the mean resulting in 88.4 was very close to the original exp mean at 88.

4 Problem 4

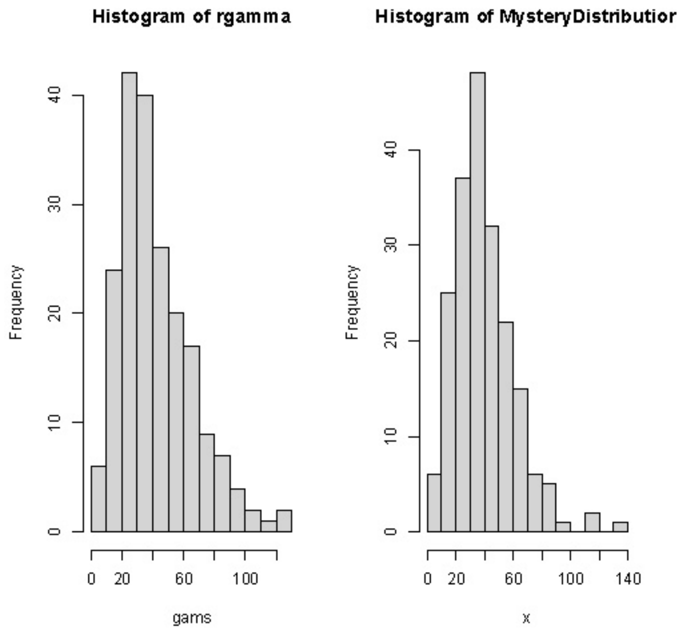
- a. Generate a sample of 200 Gamma (alpha = 3; beta = 0:1) r.v.'s by the summation method. (Do not explicitly use R functions for Gamma.) Obtain the histogram of your sample.

- b. Input the file MysteryDistribution2.csv. What distribution is it? What is (are) the parameter(s)? How can you go about verifying your claim?

```
> #Problem4b
> setwd("/Users/quay17/Desktop/MATH382/StatLab/Lab6/")
> md=read.csv("MysteryDistribution2.csv")
> head(md)
        C1
1 28.45405
2 35.82171
3 16.89504
4 60.02368
5 56.23948
6 21.11792
> x=md[,1]
> head(x)
[1] 28.45405 35.82171 16.89504 60.02368 56.23948 21.11792
> hist(x)
> mean(x)
[1] 39.79028
> var(x)
[1] 433.6886
> std = sqrt(var(x))
> std
[1] 20.82519
```



Our mystery distribution curve looks to match the Gamma distribution. With further calculations of the a(lpha) and B(eta) parameters using the mystery data and running the rgamma function, we can see graphically that the data has an almost exact distribution. Numerically, the mean and std dev of the rgamma data (42 and 23.4 respectively) are also very close to the calculations from the mystery data (39.8 and 20.9).

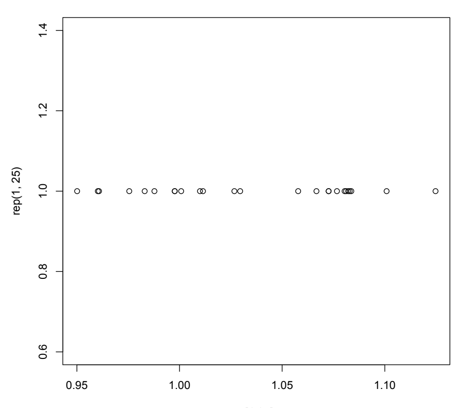



```
> #Gamma distribution parameters
> #alpha = (u/s)^2
> (39.79028/20.82519)^2
[1] 3.650699
> #beta = (s^2/u)4b
> 20.82519^2/39.79028
[1] 10.89936
> gams=rgamma(1:200, 3.6, 1/10.899)
> par(mfrow=c(1,2))
> hist(gams, main="Histogram of rgamma")
> mean(gams)
[1] 40.78468
> var(gams)
[1] 595.6762
> sqrt(var(gams))
[1] 24.40648
```

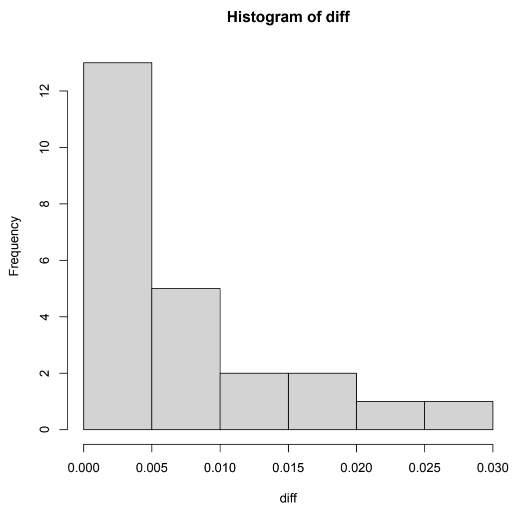
5 Problem 2 Poisson Process

- a. To get an idea of how the event times behave, make a dotplot (kind of one-dimensional scatterplot) of the first 25 event times.

```
> #Problem5
> setwd("/Users/quay17/Desktop/MATH382/StatLab/Lab6/")
> x=read.csv("eq-time.csv")
> dim(x)
[1] 766   7
> x1 = x[766:742,]
> dim(x1)
[1] 25  7
> var1=x1[,6]
> plot(var1[1:25],rep(1,25))
```



- b. Compute the differences between event times to obtain interarrival times for earthquakes. Make a histogram of the results. Do they appear to have Exponential distribution?



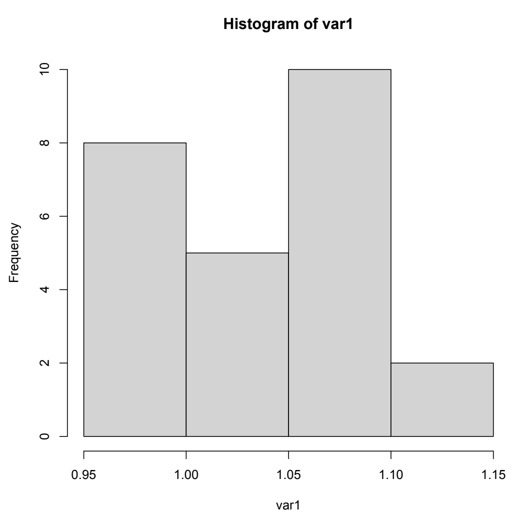

```
diff=c(0.023831,0.017199,0.0007755,0.0007523,0.0010185,0.00070
61,0.0037384,0.0040509,1.16e-05,0.0059259,0.0089005,0.0282523,
0.0028241,0.0153009,0.0014583,0.0091551,0.0031019,8.1e-05,0.00
98495,0.0047454,0.0075116,0.0147338,0.0005902,0.0100695)
```

This is definitely an exponential distribution, the graph is a declining curve, this is the probability distribution of the time between events in a Poisson point process.

- c. Estimate the mean time between earthquakes. How much is it in minutes?

```
> mean(diff)
[1] 0.007274304
```
=10 mins, 29 seconds

- d. Make a histogram of the event times. Do they appear to spread uniformly over time? If they do not, then we have a Poisson process of non-constant intensity!



Poisson process of non-constant intensity.