23603-lab-3.R

mattcarr

2023-03-20

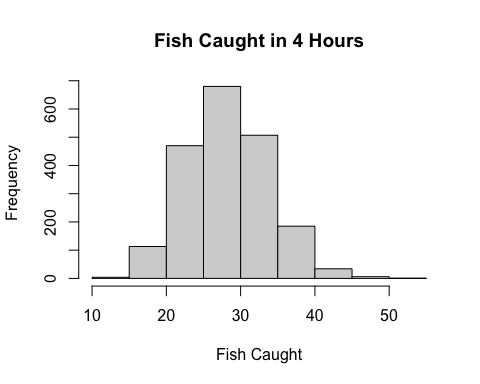
set.seed(2375)  
  
#a  
n=100  
time = 240  
t = rexp(n, 0.12)  
checksum = sum(t)  
print(checksum>time)

## [1] TRUE

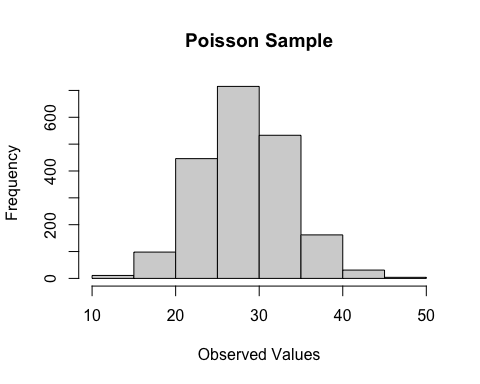
y = rep(0,n)  
  
for(i in 1:n)  
{  
 y[i] = sum(t[1:i])  
}  
x=y[y<=time]  
  
#b  
sample.x <- function()  
{  
 n=100  
 time = 240  
 t = rexp(n, 0.12)  
 y = rep(0,n)  
 for(i in 1:n)  
 {  
 y[i] = sum(t[1:i])  
 }  
 x=y[y<=time]  
}  
  
fish = sample.x()  
  
#c  
m=2000  
Nsample = rep(0,m)  
for (i in 1:m)  
{  
 sample = sample.x()  
 N = length(sample)  
 Nsample[i] = N  
}  
range(Nsample)

## [1] 11 51

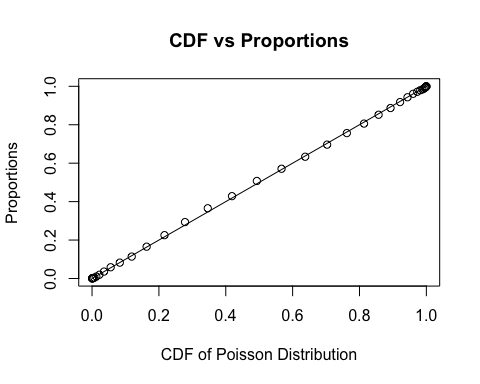
#range(Nsample) produces the minimum and maximum number of fish caught  
#in 2000 samples and gives back a vector including the 2 values  
mean = sum(Nsample) / m  
hist(Nsample, main = 'Fish Caught in 4 Hours', xlab = 'Fish Caught', ylab = 'Frequency')



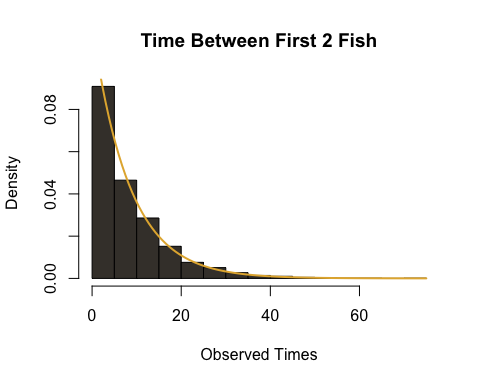
PoisSample = rpois(m, mean)  
hist(PoisSample, main = 'Poisson Sample', xlab = 'Observed Values', ylab = 'Frequency')



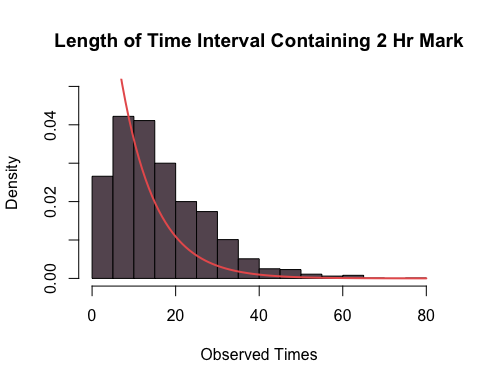
#Based off the histograms, it is plausible that Nsample follows a poisson distribution, as  
#both histograms have similar shape and range.  
  
#To further this investigation, compare cdf of Poisson Distribution with   
#Proportions of sampled values less than the corresponding values.  
  
spread = range(range(PoisSample), range(Nsample))  
numvals = spread[2] - spread[1]  
values = c(spread[1]:spread[2])  
cdf = rep(0, numvals)  
props = rep(0,numvals)  
for (i in 1:numvals)  
{  
 cdf[i] = ppois(values[i], mean)  
 props[i] = (length(Nsample[Nsample <= values[i]])/m)  
}  
plot(cdf, props, xlab = 'CDF of Poisson Distribution', ylab = 'Proportions', main = 'CDF vs Proportions')  
curve(x\*1, add = TRUE)



#Detective Crabbe's claim seems to be accurate. As aforementioned, the histograms of the poisson samples and the  
#actual samples were quite similar, and the same can be seen in a comparison of the cdf of a poisson distribution  
#and the proportion of values less than given values of the support of a poisson distribution.  
#There is a clear linear correlation, and the line y=x fits the plot  
# well, suggesting that the values will be similar  
#for each value. As such, it is likely that Detective Crabbe's claim   
#is correct, and the data follows a poisson distribution with mean equal to the mean  
#of the Observed values  
  
  
#d  
Wsample = rep(0,m)  
for (i in 1:m)  
{  
 sample = sample.x()  
 W = sample[2]-sample[1]  
 Wsample[i] = W  
}  
hist(Wsample, prob = TRUE, main = 'Time Between First 2 Fish', xlab = 'Observed Times', ylab = 'Density', col='#423E37')  
curve(dexp(x, 0.12), add = TRUE, col='#E3B23C', lwd = 2)



#The Distribution of the data seems to follow an Exponential Distribution  
#This fits with my expectation as each value of W is the time between the first  
#and second catch. Since the exponential distribution is memoryless, it follows   
#that the time between the first and second catch should also be an  
#exponential distribution with the same parameter.  
  
#e  
  
Vsample = rep(0,m)  
for (i in 1:m)  
{  
 sample = sample.x()  
 x1 = max(sample[sample<(time/2)])  
 x2 = min(sample[sample>(time/2)])  
 V = x2-x1  
 Vsample[i] = V  
}  
hist(Vsample, prob = TRUE, main = 'Length of Time Interval Containing 2 Hr Mark', xlab = 'Observed Times', ylab = 'Density', col = '#655560', ylim = c(0, 0.05))  
curve(dexp(x, 0.12), add = TRUE, col = '#E85F5C', lwd = 2)



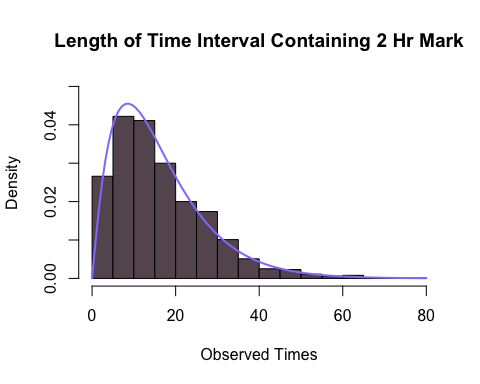
help(hist)  
#I would have expected the distribution to follow an exponential distribution  
#with the same parameter as before using similar reasoning as before  
#that the exponential distribution is memoryless. As such, I would expect  
#that looking for an interval containing a value would not change the distribtion   
#drastically as each interval is an observed value of an exponential random variable  
  
#However, The curve for the pmf of an exponential function  
#is a poor fit for the histogram, so it appears that the observed values do not  
#follow an exponential distribution.   
#Based on the shape of the graph, it appears that Vsample could  
#Follow a gamma distribution.  
#To check, calculate values for parameters k, λ using E(X) and E(X^2)  
  
squared = Vsample^2  
meanVsq = sum(squared) / m  
meanV = sum(Vsample) / m  
meanVsq

## [1] 391.7374

meanV

## [1] 16.28391

#Then k/λ = meanV and k(k+1)/λ^2 = meanVsq  
  
k = 1 / ((meanVsq / meanV^2)-1)  
λ = k / meanV  
#solving gives k = 2.095(3dp) and λ = 0.106(3dp)  
hist(Vsample, prob = TRUE, main = 'Length of Time Interval Containing 2 Hr Mark', xlab = 'Observed Times', ylab = 'Density', col = '#655560', ylim = c(0, 0.05))  
curve(dgamma(x, k, λ), add = TRUE, col = '#9381FF', lwd = 2)



#The curve fits the graph very well, so the Gamma distribution seems to   
#fit the data in Vsample  
  
meanW = sum(Wsample)/m  
meanW

## [1] 8.50276

meanV

## [1] 16.28391

#The mean for Vsample is around double the mean of Wsample for these  
#given trials. The values in Vsample tend to be larger  
#than those in Wsample. This is most likely due to the fact  
#that values in Vsample represent an interval of time containing the   
#half-way mark. Therefore, longer time intervals will be more likely to   
#contain this half-way mark than shorter intervals, which can be seen   
#in the histogram of the data. As such, without this condition,  
#values will follow the exponential distribution which can be seen  
#in Wsample  
  
#f  
  
#The proposed scheme is based on waiting times before an event occurs.   
#This is the same premise as the wait time to catch a fish, the only difference being the   
#event. This is particularly similar to the data in   
#Wsample as Wsample concerned the time between the second and first fish.   
#Since the exponential distribution is memoryless, we can say that Wsample  
#represents the time before catching a fish, which is extremely similar to   
#the proposal, which looks for the time before hip replacement.  
  
#Given the similarities between the values of interest in the fish-catching example   
#and the proposed scheme, we can expect that the data found in this proposed  
#scheme should have similar distributions to those seen in the fish example, but  
#with different parameters. As such, I would tell the investigators that   
#the amount of time before hip replacement surgery should follow an exponential  
#distribution with rate parameter equal to the reciprocal of the mean of the   
#observed values