PairProgramming_Module4_Poisson

HD Sheets

2024-08-27

Poisson Distributions, Pair Programming exercise, DSE5001

HD Sheets, August 27, 2024 checked 01/03/2025

Questions taken from Open Intro to Statistics, Diez et al

Name: Ryan Waterman Partner Name: Date: 2/6/25

Question 4.33

4.33 How many cars show up? For Monday through Thursday when there isn't a holiday, the average number of vehicles that visit a particular retailer between 2pm and 3pm each afternoon is 6.5, and the number of cars that show up on any given day follows a Poisson distribution.

a. What is the probability that exactly 5 cars will show up next Monday?
b. What is the probability that 0, 1, or 2 cars will show up next Monday between 2pm and 3pm?
c. There is an average of 11.7 people who visit during those same hours from vehicles. Is it likely that the number of people visiting by car during this hour is also Poisson? Explain.

Solution 4.31

Okay, by now, I hope you realize that using distributions to solve problems is about

-figuring out which distribution will work

-{with real world applications we would then validate the choice of the distribution using *Goodness of Fit* testing}

-determine the parameter values to use- either from textbook values or by estimating them from data

-determine whether we need to us the pdf, cdf, or quantiles

-it never hurts to create some graphs

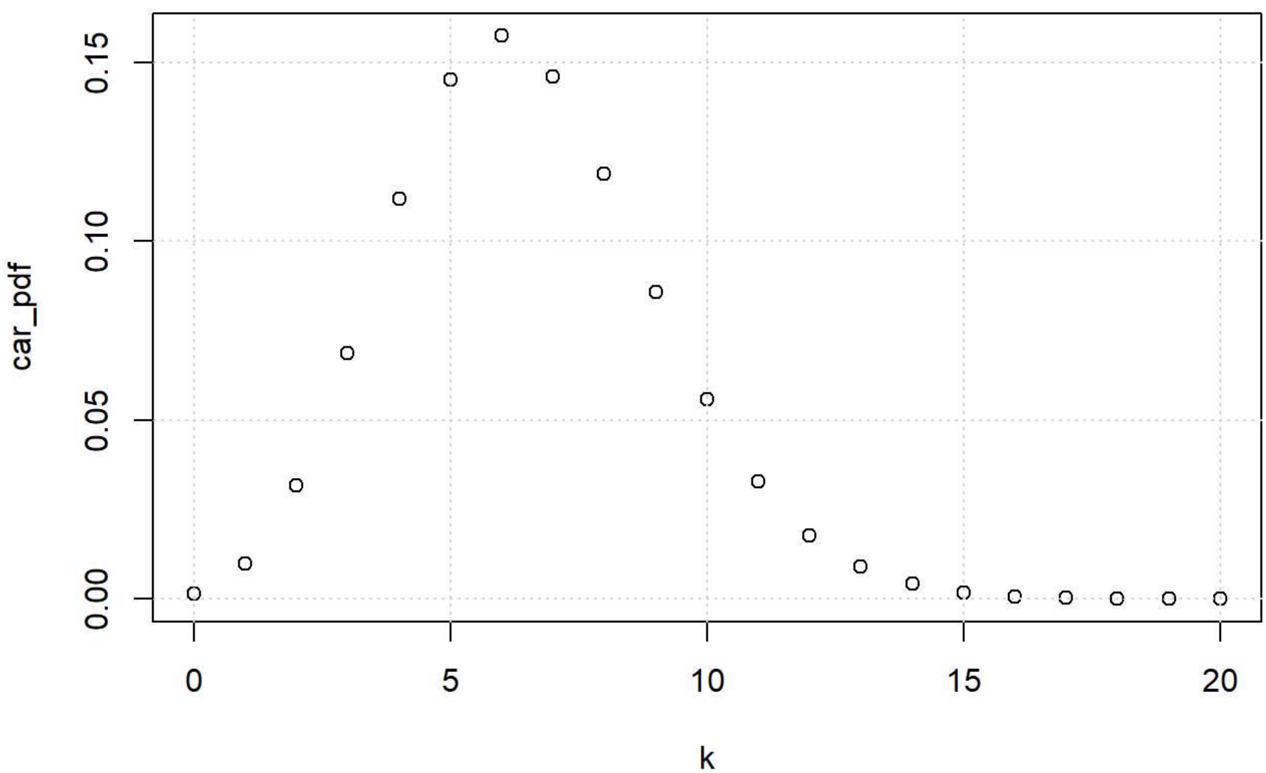
For this problem, we know it is the Poisson, and that the mean is 6.5

The poisson is a discrete distribution, which makes sense since cars are discrete, countable things

Looking at the poisson, the only parameter is lambda, and the mean equals lamba

Let's plot the distribution, we have dpois(), ppois(), qpois() and rpois() available

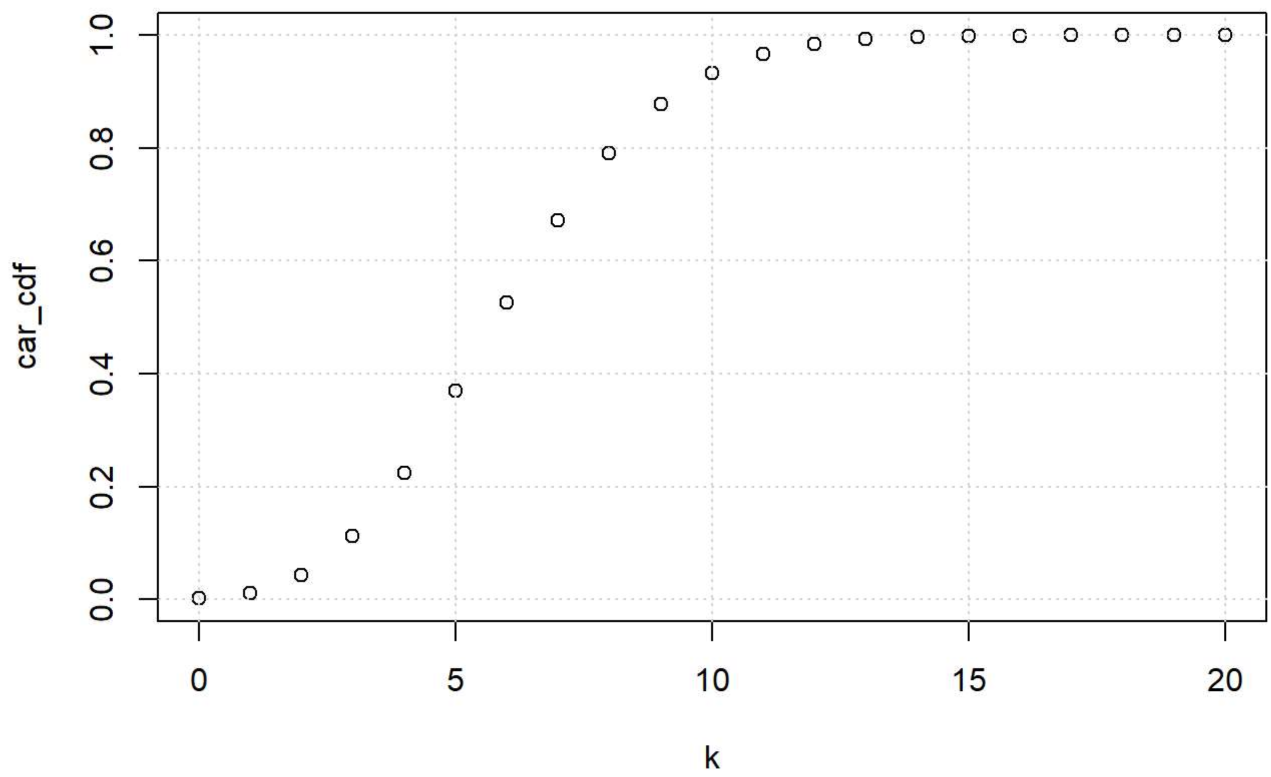
```
#set up k occurrences, for k=0 to 20, so we can graph the pdf and cdf

k=0:20

car_pdf=dpois(k,lambda=6.5)

plot(k,car_pdf)
grid()
```



It strikes me that this is pretty broad, the dispersion is high

Let’s plot the cdf

```
car_cdf=ppois(k,lambda=6.5)
plot(k,car_cdf)
grid()
```



Okay, let's find some answers

a. What is the probability that exactly 5 cars will show up next Monday?

Looking at this, we ask, is the the pdf or cdf

It is probability of 1 specific value, so this is a pdf

```
# get the cdf of a poison, with lambda 6.5 at k=5

dpois(5,lambda=6.5)
```

```
## [1] 0.1453689
```

b. What is the probability that 0, 1, or 2 cars will show up next Monday between 2pm and 3pm?

We could do this several ways

it could be the sum of p(k=0)+p(k=1)+p(k=2) using the pdf

or p(k<=2) using the cdf

I'll use the cdf

```
ppois(2,lambda=6.5)
```

```
## [1] 0.04303595
```

c. There is an average of 11.7 people who visit during those same hours from vehicles. Is it likely that the number of people visiting by car during this hour is also Poisson? Explain.

It does seem likely this is also Poisson. The number of people per car is always at least 1 though, there is probably an upper bound, but it might be a pretty large upper bound. It might be possible to treat this as a poisson.

Question/Action, complete this question as a Pair

-figure out what distribution you are dealing with -find the parameters -plot the pdf and the cdf for a reasonable range of outcome values -for each portion of the question, figure out if you need the pdf or the cdf -locate the R functions you will need

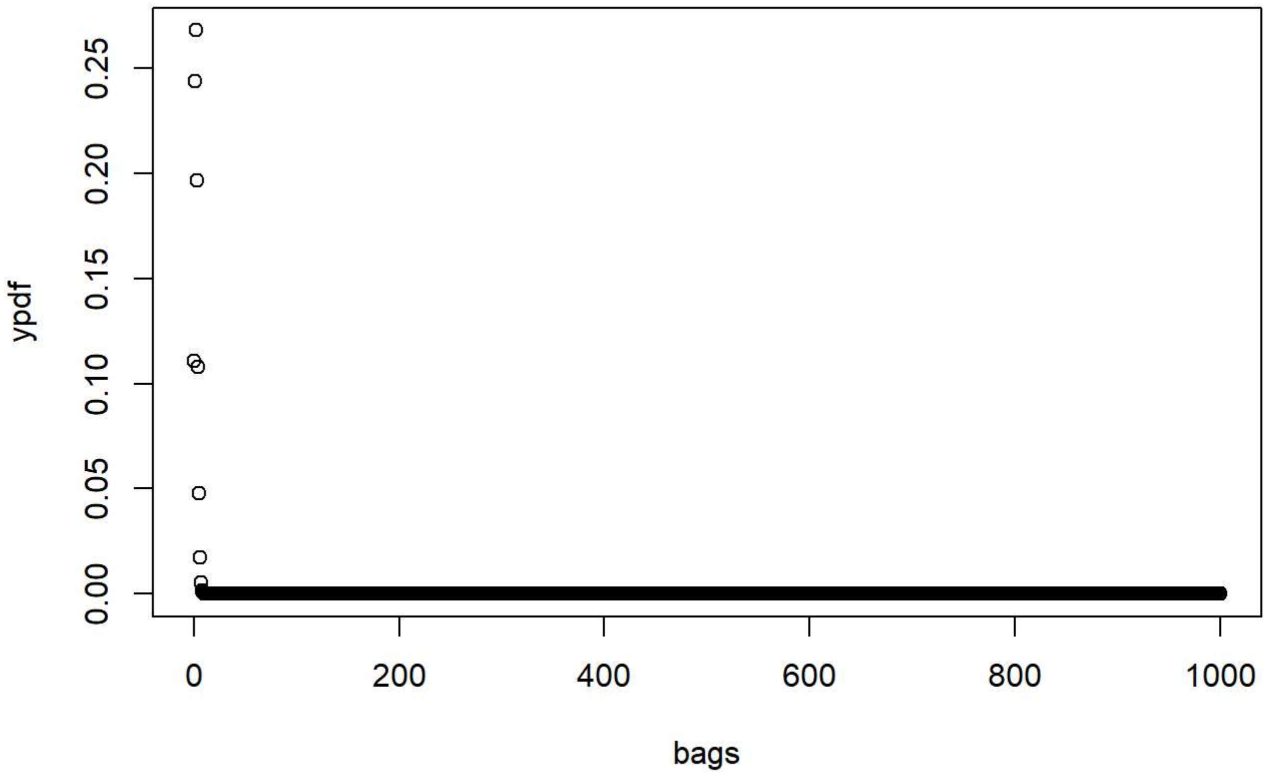
Question 4.34

4.34 Lost baggage. Occasionally an airline will lose a bag. Suppose a small airline has found it can reasonably model the number of bags lost each weekday using a Poisson model with a mean of 2.2 bags.

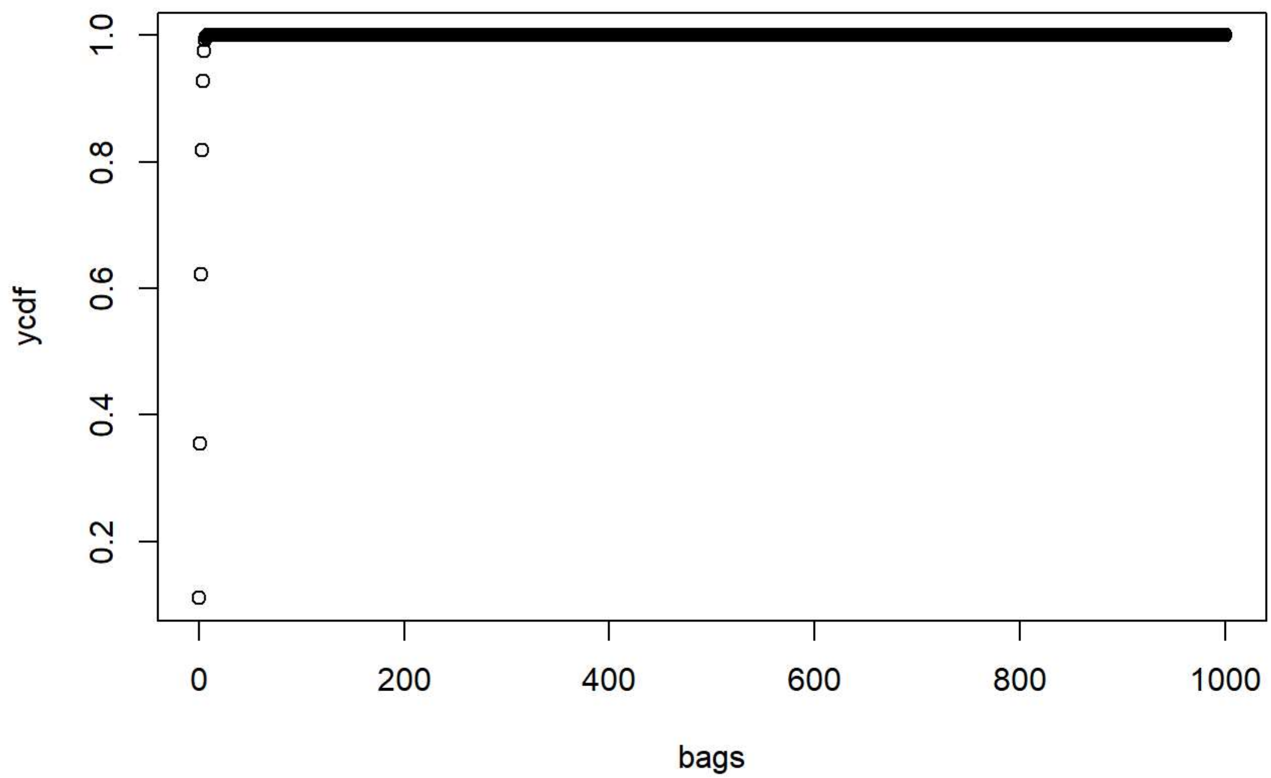
This is a poisson distribution because the data is discrete and the variable is time dependent.

a. What is the probability that the airline will lose no bags next Monday?

```
#Plotting pdf and cdf
bags = 0:1000

ypdf = dpois(bags, 2.2)
plot(bags, ypdf)
```




```
ycdf = ppois(bags, 2.2)
plot(bags, ycdf)
```




```
ppois(0, 2.2)
```

```
## [1] 0.1108032
```

About 11%

b. What is the probability that the airline will lose 0, 1, or 2 bags on next Monday?

```
ppois(2,2.2)
```

```
## [1] 0.6227137
```

Using the cdf capture x<=2, resulting in a probability of 62.2%

c. Suppose the airline expands over the course of the next 3 years, doubling the number of flights it makes,and the CEO asks you if it's reasonable for them to continue using the Poisson model with a mean of 2.2. What is an appropriate recommendation? Explain.

The Poisson model with a mean of 2.2 would be outdated. The number of bags lost per unit time is directly proportional to the number of bags flown per unit time. A better assumption would be that the mean bags lost per week scales linearly with the number of bags flown per week, assuming staffing has also grown proportionally.