Live Session 7

Jennifer Shin

February 20, 2017

Maximum Likelihood

Question: What is likelihood? How is it different from the probability density functions we've been working with?

The probability function returns probabilities of the data, given the sample size and the parameters, while the likelihood function gives the relative likelihoods for different values of the parameter, given the sample size and the data.

Question: Why do data scientists need to understand likelihood?

R: Optimize Function

In this lesson, we'll practice maximizing likelihood functions. You can often maximize a function using calculus, but if you can write an R function to compute the likelihood or log-likelihood, then you can also maximize it numerically. One way to do this in R is to use the optimize() function.

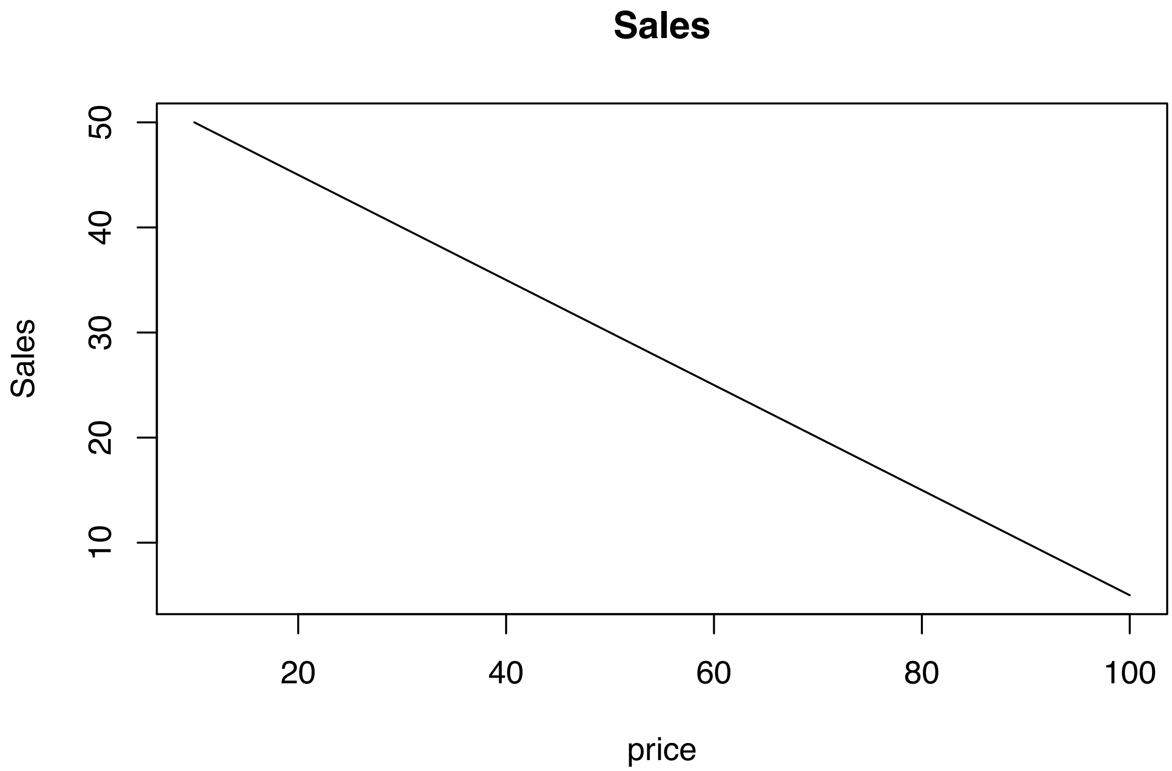
Example 1:

Suppose you are interested in selling a product that is priced between $10 and $100 and you are interested in finding the price that maximizes your revenue. We assume that sales and revenue are defined as follows:

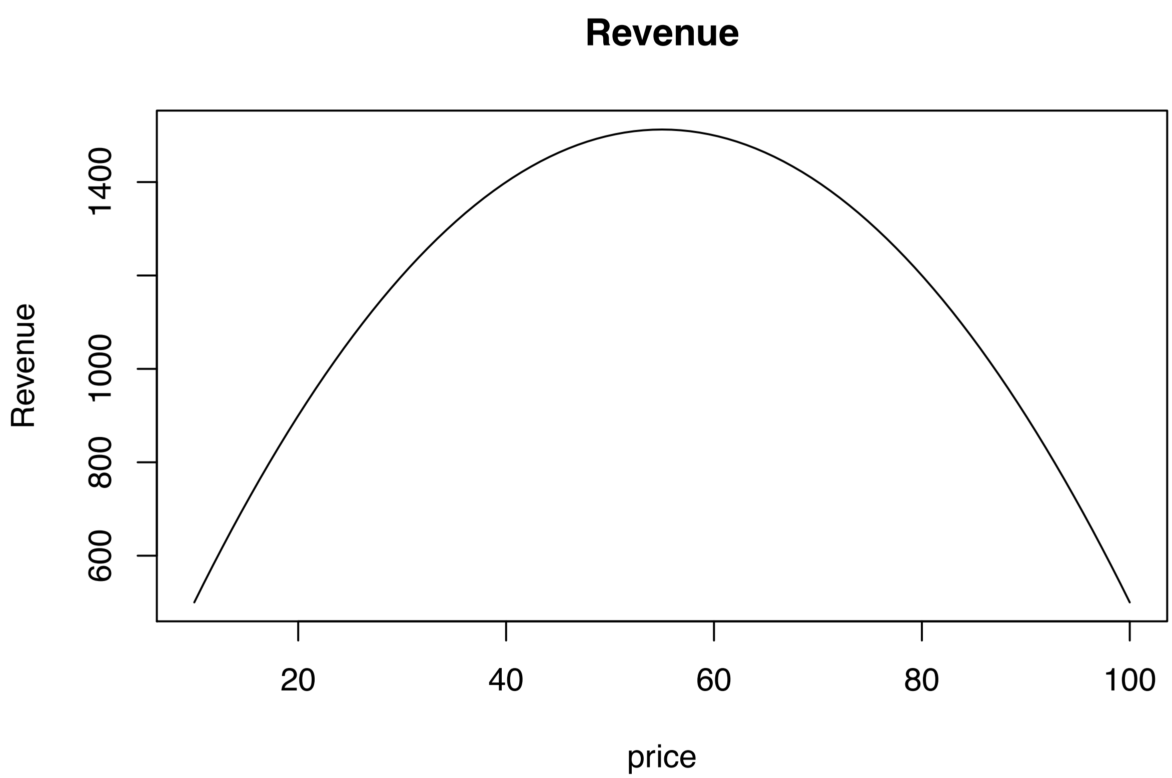
```r
sales <- function(price) { 55 - 0.5 * price }
revenue <- function(price) { price * sales(price) }
```

By plotting both functions in R, we can visually identify the optimal price:

```r
curve(sales, from=10, to=100, xname="price", ylab="Sales", main="Sales")
```




```r
curve(revenue, from=10, to=100, xname="price", ylab="Revenue", main="Revenue")
```



Question: How would you find the price that maximizes revenue using calculus?

Question: Does the first-order condition always hold at the maximum value of a function?

Another way to find the optimal price – or at least get close to it - is to use the optimize() function in R. This function uses iterative numerical methods to compute an approximation to the maximizing price.

```
optimize(revenue, interval=c(10, 100), maximum=TRUE)
```

```
## $maximum
## [1] 55
##
## $objective
## [1] 1512.5
```

Hence, By charging $55, you will receive a revenue of $1512.5.

Maximum Likelihood Estimation of a Bernoulli Variable

Suppose that you've got a sequence of values from an unknown Bernoulli variable like so:

```
p=0.5
S <- rbinom(10, 1, p)
S
```

```
##  [1] 0 1 0 0 1 0 1 1 0 1
```

Assume we are given this sequence of data and p is unknown to us. For example, we could be testing an unfair coin to see what the probability of heads is. Or we could be testing a new phone to see what the probability of a phone explosion is.

Question: What does the likelihood function, L(p), represent in this case?

Question: Does the integral of the likelihood function equal 1?

Question: Write down an expression for the likelihood function, L(p).

SOLUTION:

f(p) = p(1-p) for $0 \leq p \leq 1$.

$X\ Bin(n, p)$

$L(p; x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)}$

Using the maximum likelihood approach, we want to select the parameter with the highest likelihood. In this case, for a Bernoulli variable we can search through the space of values for p (i.e [0, 1]) that makes the data most probable to have been observed.

In the async, Paul demonstrated how to find the maximizing value of p using calculus.

To find this parameter numerically, we need to define a function that specifies the probability of our data set. You may do this using code like the following.

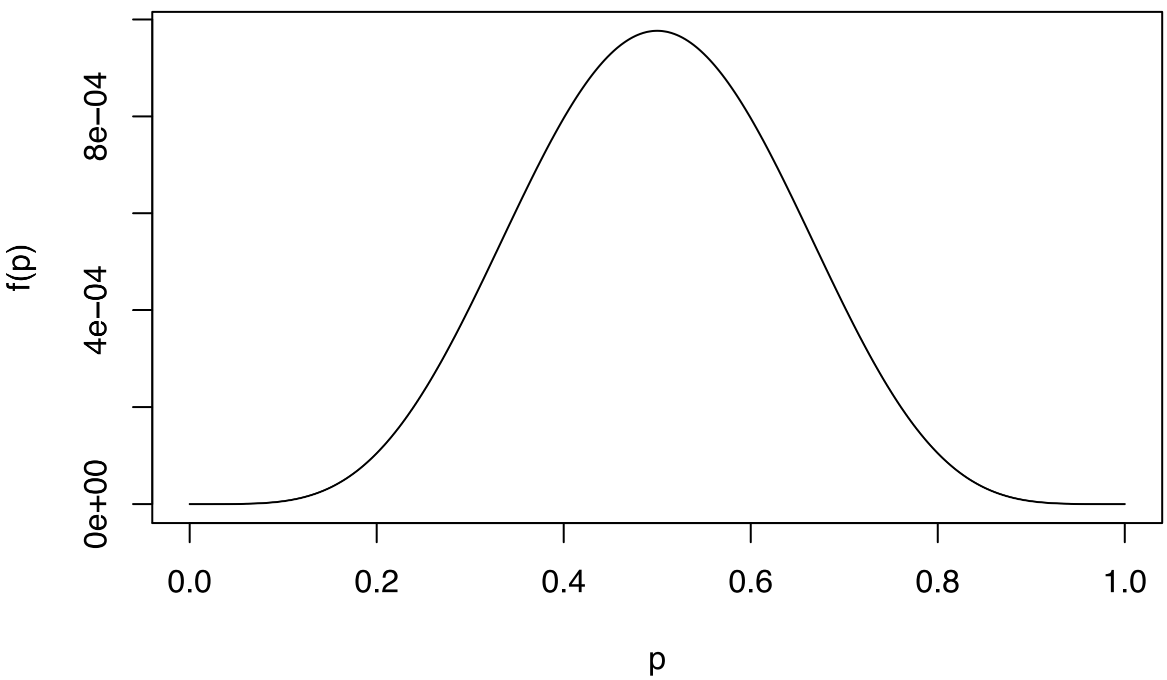
Let's define the function:

```
likelihood <- function(s, p)
{
  likelihood <- 1

  for (i in 1:length(s))
  {
    if (s[i] == 1)
    {
      likelihood <- likelihood * p
    }
    else
    {
      likelihood <- likelihood * (1 - p)
    }
  }
  return(likelihood)
}
```

We can graph the likelihood

```
P <- seq(0, 1, by = 0.001)
plot(P,  likelihood(S, P),type="l",xlab="p",ylab="f(p)")
```



Optional graph (only for student's who already have ggplot2):

```
P <- seq(0, 1, by = 0.001)
jpeg('Likelihood.jpg')
library('ggplot2')
qplot(P,
      sapply(P, function (p) {likelihood(S, p)}),
      geom = 'line',
```

```
      main = 'Likelihood as a Function of P',
      xlab = 'P',
      ylab = 'Likelihood')
dev.off()
```

```
## pdf
##   2
```

Question: use R's optimize function to find the maximum likelihood estimate of p.

Question: Compare the MLE estimate of p to the sample mean of the Bernoulli trials. If you can, explain whether you expected this result.

MLE for Poisson Distribution

A Poisson process is a simple model that statisticians use to describe how events occur over time. Imagine that time stretches out on the x-axis, and each event is a single point on this axis.

The key feature of a Poisson process is that it is *memoryless*. Loosely speaking, the probability that an event occurs in any (differentially small) instant of time is a constant. It doesn't depend on how long ago the previous event was, nor does it depend on when future events occur. Statisticians might use a Poisson process (or more complex variations) to represent: - The scoring of goals in a world cup match - The arrival of packets to an internet router - The arrival of customers to a website - The failure of servers in a cluster - The time between large meteors hitting the Earth

To understand a Poisson process, imagine an experiment in which you observe the arrival of cars at an intersection. Assume that the probability density that a car arrives in a differentially small interval of time is just a constant. The intersection is no more busy during the day than during the night.

Moreover, the probability density that a car arrives at a particular instant does not depend on when the previous cars arrived, not when future cars are going to arrive. Each moment of time is independent. This is an example of what we call a memory-less process.

Next, suppose we use a camera to record the intersection for a particular length of time, and we write down the number of cars that arrive in that interval. This is what we call a Poisson random variable. It has a well-known probability mass function, given by, f(x|λ) $= \frac{\lambda^x e^{-\lambda}}{x!}$

Here, λ is a parameter, which represents the mean number of cars in an interval. (You may take the expectation to check this). The following graph from Wikipedia shows the probability mass function for different values of λ.

Suppose we take a random sample of 300 1 minute intervals from our intersection. The following frequency table shows the data we get.

We want to find the maximum likelihood estimate for λ.

First, we can define a dataset, X, and generate the data using the table above:

```
X <- c(rep(0,14), rep(1,30), rep(2,36), rep(3,68), rep(4, 43),
       rep(5,43), rep(6, 30), rep(7,14), rep(8,10), rep(9, 6),
       rep(10,4), rep(11,1), rep(12,1))
```

NOTE: rep(x, n) means repeat the number x for n times, and c( ) is the operation to combine all the numbers into a long vector.

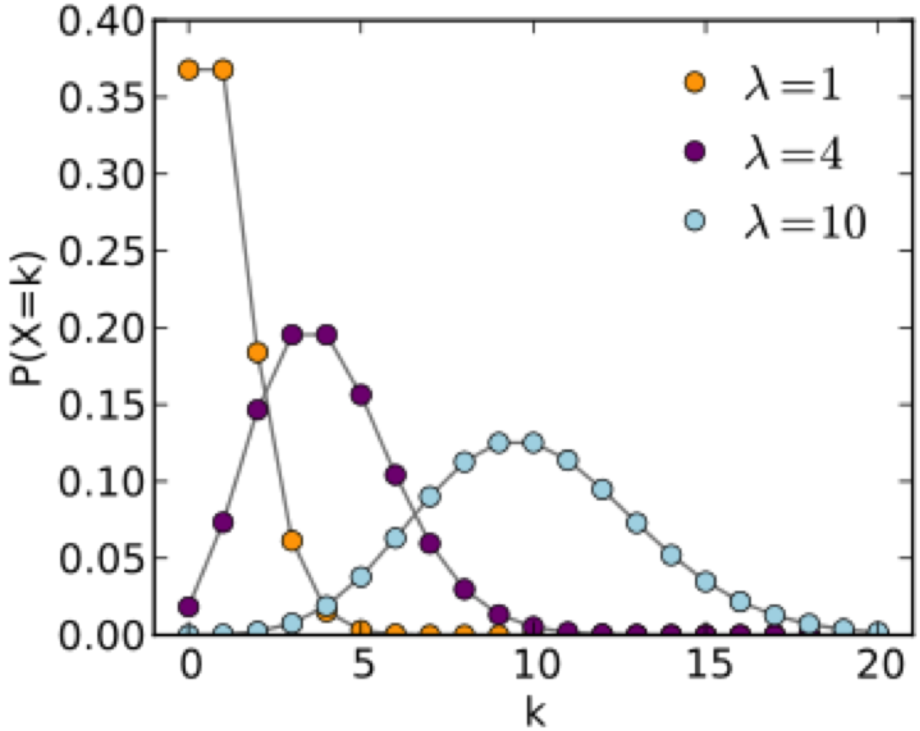


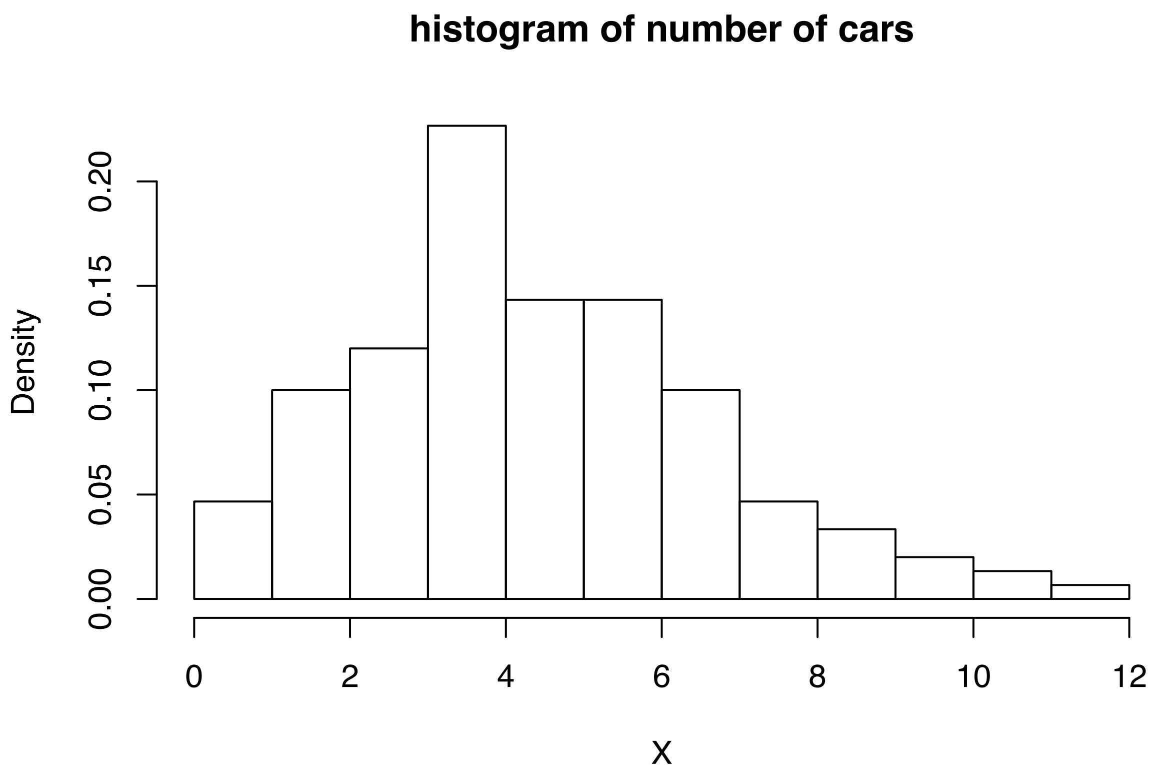
Figure 1:

n	Frequency
0	14
1	30
2	36
3	68
4	43
5	43
6	30
7	14
8	10
9	6
10	4
11	1
12	1

Figure 2:

To see whether the Poisson model is a good distribution, we can plot the histogram of the data and compare to the point mass function of Poisson.

```r
hist(X, main="histogram of number of cars", right=FALSE, prob=TRUE)
```



Question: Write down an expression for the likelihood function, L(λ).

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

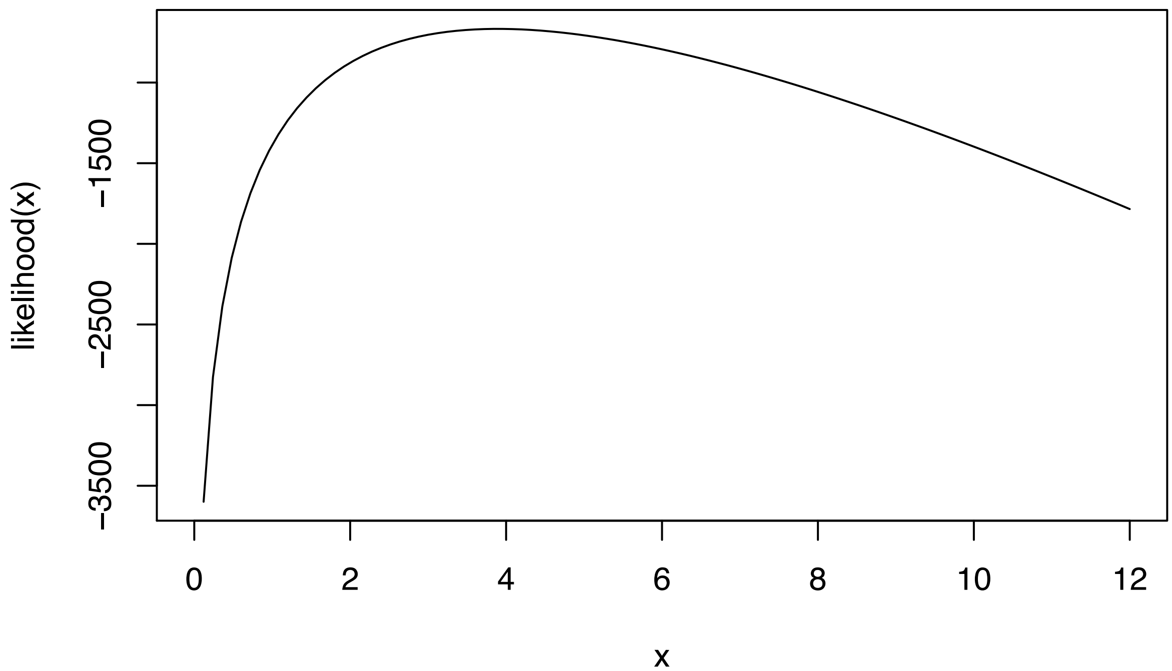
$$f(x_1, x_2, ..., x_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Question: To make things easier, take the log of your likelihood function and simplify it.

$$\log(L(\lambda)) = \sum_{i=1}^{n} (x_i \log(\lambda) - \lambda - \log(x_i!))$$

Question: Define a function in R to compute the likelihood function, then plot it in the interval from 0 to 12.

```r
likelihood = function(lambda){
  log(lambda)*sum(X)-length(X)*lambda-sum(log(factorial(X)))
}
curve(likelihood, from=0, to=12)
```



We can either maximize the likelihood function using R or using calculus.

Question: Use optimize to find the maximum likelihood estimate for λ.

```r
optimize(likelihood, interval=c(0, 12), maximum=TRUE)
```

```
## $maximum
## [1] 3.893332
##
## $objective
## [1] -667.183
```

Question: maximize the likelihood function using calculus

1. take the 1st derivative of likelihood with respect to λ and set it equal to 0. This is the first order condition.
2. Solve for λ

Write the likelihood function of the sample:

$$f(x_1, x_2, ..., x_n|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Take the Log of the Likelihood

$$\sum_{i=1}^{n} (x_i \log(\lambda) - \lambda - \log(x_i!))$$

Take the 1st derivative with respect to λ and set equal to zero

$$\frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

Solve for λ

$$\hat{\lambda}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}$$