Problem Set 5

John Hancock

July 16, 2017

Contents

1 References and License

We are answering questions in the material from MIT OpenCourseWare course 18.05, Introduction to Probability and Statistics.
In this document we are answering questions Orloff and Bloom ask in [1].
Please see the references section for detailed citation information.
The material for the course is licensed under the terms at `http://ocw.mit.edu/terms`.
We use documentation in to write the LaTeXsource code for this document.

2 Fit line to data

In this section we answer questions about a random variable Y drawn from the random variable $Y_i \sim ax_i + b + \epsilon_i$, where $epslion_i$ is a random variable with mean 0 and variance σ^2.
Orloff and Bloom grant us that the ϵ_i are independent.

2.1 Likelihood function

We derive the likelihood function $f\left(y_i \mid a, b, x_i, \sigma\right)$.
To derive f we assume x_i, y_i, and σ are known values.
It is of paramount importance to note:

$$\epsilon_i \sim N\left(0, \sigma\right). \tag{1}$$

We then look at the random variable:

$$Y_i = ax_i + b + \epsilon_i \tag{2}$$

ϵ_i is a random variable that follows a normal distribution. In the context of this discussion, it is not a fixed value, its value depends on what we choose for a, and b. Keep in mind that we are trying to find values for a, and b that maximize the likelihood of the linear relationship between X and Y.
So, if $\epsilon_i \sim N\left(0, \sigma^2\right)$, then

$$ax_i + b + \epsilon_i \sim N\left(ax_i + b, \sigma^2\right). \tag{3}$$

That is, since ϵ_i is a random variable with mean 0, then the random variable $ax_i + b + \epsilon_i$ will have mean $ax_i + b$. Orloff and Bloom show this in [3]. In this case we are treating $ax_i + b$ as constants. This is really confusing, because we are trying to find values for a and b that maximize a probability. So we are considering varying values of a

and b so that we find the best values for them. However, assuming we choose values for a and b, then $ax_i + b + \epsilon_i$ will have mean $ax_i + b$.
In order to make the leap to a probability density function that we are going to maximize, we cite the reasoning Orloff and Bloom give in [4], section 4.
Then the likelihood function f_i for one point (x_i, y_i) is:

$$f_i\left(y_i \mid x_i, a, b, \sigma\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-(ax_i+b))^2}{2\sigma^2}}. \tag{4}$$

The likelihood function f of all points is the product of the function above for all values of x_i, and y_i:

$$f = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i-(ax_i+b))^2}{2\sigma^2}}. \tag{5}$$

We can rewrite the product above as:

$$f\left(y_i \mid x_i, a, b, \sigma\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\sum_{i=1}^{n}(y_i-ax_i+b))^2}{2\sigma^2}}. \tag{6}$$

The right hand side of the equation above is the likelihood function.

2.2 Likelihood and log-likelihood functions for particular values

We suppose we have the following data:

$(1,8),(3,2),(5,1)$.
We write down the liklihood and log likelihood functions for these data:

$$f(y_i \mid x_i, a, b, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(8-(a+b))^2+(2-(3a+b))^2+(1-(5a+b))^2}{2\sigma^2}}. \tag{7}$$

$$ln(f(y_i \mid x_i, a, b, \sigma)) = ln\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(8-(a+b))^2+(2-(3a+b))^2+(1-(5a+b))^2}{2\sigma^2}}\right). \tag{8}$$

We simplify the right hand side of the equation above in several steps:

$$ln(f(y_i \mid x_i, a, b, \sigma)) = ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + ln\left(e^{-\frac{(8-(a+b))^2+(2-(3a+b))^2+(1-(5a+b))^2}{2\sigma^2}}\right) \tag{9}$$

$$ln(f(y_i \mid x_i, a, b, \sigma)) = ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \frac{-\left((8-(a+b))^2+(2-(3a+b))^2+(1-(5a+b))^2\right)}{2\sigma^2} ln(e). \tag{10}$$

$$ln\left(f\left(y_i \mid x_i, a, b, \sigma\right)\right) = ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \frac{\left(-\left(8-(a+b)\right)^2 + \left(2-(3a+b)\right)^2 \cdots\right.}{2\sigma^2} \tag{11}$$

2.2.1 General formulation

We gave the general formulation for the likelihood function above:

$$f\left(y_i, a, b, \sigma\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\sum_{i=1}^{n}(y_i - ax_i + b)\right)^2}{2\sigma^2}}. \tag{12}$$

Note we have removed the x_i from the left hand side of the equation as a function parameter because the x_i are constants.

We obtain the log likelihood function applying the natrual logarithm function to both sides of the equation above, and then simplifying using the laws of logarithms.

$$ln\left(f\left(y_i, a, b, \sigma\right)\right) = ln(left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\left(\sum_{i=1}^{n}(y_i - ax_i + b)\right)^2}{2\sigma^2}}. \tag{13}$$

$$ln\left(f\left(y_i,a,b,\sigma\right)\right)=ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)+lnleft(e^{-\frac{\left(\sum_{i=1}^{n}(y_i-ax_i+b)\right)^2}{2\sigma^2}}. \tag{14}$$

$$ln\left(f\left(y_i,a,b,\sigma\right)\right)=ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)+ln\left(e^{-\frac{\left(\sum_{i=1}^{n}(y_i-ax_i+b)\right)^2}{2\sigma^2}}\right). \tag{15}$$

$$ln\left(f\left(y_i,a,b,\sigma\right)\right)=ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)-\frac{\left(\sum_{i=1}^{n}\left(y_i-ax_i+b\right)\right)^2}{2\sigma^2}. \tag{16}$$

2.3 Maximum likelihood estimates for a, and b

For this problem, Orloff and Bloom allow us to assume that σ is a constant, known value. They ask us to find the maximum likelihood estimates for a, and b, under these circumstances.
In this case, we will be working with partial derivatives of

$$ln\left(f\left(y_i\mid,a,b,\sigma\right)\right)=ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)+\frac{\left(-\left(8-(a+b)\right)^2+\left(2-(3a+b)\right)^2+(\right.}{2\sigma^2} \tag{17}$$

At this point we have an exercise in calculus and linear algebra to obtain two equations in two unknowns by setting the parital derivatives of the function f above with respoect to a, and b to zero, and then solving the system for a and b.

$$\frac{\delta}{\delta a} ln\left(f\left(y_i\mid, a, b, \sigma\right)\right) = \frac{-2\left(8-(a+b)\right)-2\left(6\right)\left(2-(3a+b)\right)-2\left(5\right)\left(1-\right.}{2\sigma^2} \tag{18}$$

We obtain a similiar partial derivative with respect to b, then solve the resulting system of equations for a, and b.

3 Estimating uniform parameters

subsectionEstimate with set of specific numbers
Orloff and Bloom give us a dataset S:

$$S = \{1.2, 2.1, 1.3, 10.5, 5\} \tag{19}$$

In [4] Orloff and Bloom state that the likelihood function for data that follows a uniform distribution over the interval $[a, b]$ is maximized when the uniform distribution has parameters $\hat{a} = \min(S)$, and $\hat{b} = \max(S)$.
Therefore in this example the maximum likelihood estimate for the distribution that the data in S follows is a uniform distribution on the interval $[1.2, 10.5]$.

subsectionEstimate with general set of numbers
Orloff and Bloom change the question above. They ask us what would be the maximum likelihood estimate for parameters of a uniform distribution where we have a data set S

$$S = \{x_1, x_2, \ldots, x_n\} . \tag{20}$$

We just re-defined S, here, but the result is the same, the maximum likelood estimate of the parameters a and b for the uniform distribution that the data set is drwan from is $\hat{a} = \min(S)$, and $\hat{b} = \max(S)$ for a, and b respectively.

4 Monty Hall sober and drunk

4.1 Bayes' table

Hypothesis A is that the car is behind door A, and similarly for hypthotheses B, and C.
The data, D, is that Monty opens door B,and reveals a goat.
In order to procede we need to make some assumptions, but we can make these assumptions without losing generality.
We are going to assume we chose door A.
There are two cases:

- the car is behind door A, and
- the car is not behind door A.

If the car is behind door A, then there is a 0.5 probabilty that hypothesis A is correct because Monty opened door B. Since we chose a door with a car behind it Monty chose one of the doors with a goat behind it with 0.5 probabilty. We capture this case with the first row of the Bayes table below.
If the car is not behind door A, then we selected a door with a goat behind it, so Monty is forced to show us what is behind the only other door with a goat behind it, whiich in this case is door B. Therefore, in this case, the likelihood of the hypothesis that the car is behind door C has probabilty 1. We capture this in the third row of the Bayes table below.

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(D \mid \mathcal{H})$	$P(D \mid \mathcal{H})$	$P(\mathcal{H} \mid D)$
A	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{3}$
B	$\frac{1}{3}$	0	0	
C	$\frac{1}{3}$	1	$\frac{1}{3}$	$\frac{2}{3}$
			$\frac{1}{2}$	

5 Dice

We are continuing the dice problem from [2].
The hypotheses A - E, are:

- we chose the 4 sided die
- we chose the 6 sided die
- we chose the 8 sided die
- we chose the 12 sided die
- we chose the 20 sided die

5.1 Rolling 7's

If we rolled n 7's in a row, we can calculate the posterior probability of that event in a Bayes table:

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(D \mid \mathcal{H})$	$P(D \mid \mathcal{H})$	$P(\mathcal{H} \mid D)$
A	$\frac{1}{5}$	0	$\frac{1}{6}$	$\frac{1}{3}$
B	$\frac{1}{5}$	0	0	
C	$\frac{1}{5}$	$\frac{1}{8^n}$	$\frac{1}{5}\left(\frac{1}{8^n}\right)$	$\frac{\frac{1}{8^n}}{\frac{1}{8^n}+\frac{1}{12^n}+\frac{1}{20^n}}$
D	$\frac{1}{5}$	$\frac{1}{12^n}$	$\frac{1}{5}\left(\frac{1}{12^n}\right)$	$\frac{\frac{1}{12^n}}{\frac{1}{8^n}+\frac{1}{12^n}+\frac{1}{20^n}}$
E	$\frac{1}{5}$	$\frac{1}{20^n}$	$\frac{1}{5}\left(\frac{1}{20^n}\right)$	$\frac{\frac{1}{20^n}}{\frac{1}{8^n}+\frac{1}{12^n}+\frac{1}{20^n}}$
			$\frac{1}{5}\left(\frac{1}{8^n}+\frac{1}{12^n}+\frac{1}{20^n}\right)$	

References

[1] Jeremy Orloff and Jonathan Bloom. *18.05 Problem Set 5, Spring 2014*. Available at `https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/assignments/MIT18_05S14_ps5.pdf` (Spring 2014).

[2] Jeremy Orloff and Jonathan Bloom. *Bayesian Updating: Odds Class 12, 18.05 Jeremy Orloff and Jonathan Bloom*. Available at `https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading12b.pdf` (Spring 2014).

[3] Jeremy Orloff and Jonathan Bloom. *Expectation, Variance and Standard Deviation for Continuous Random Variables Class 6, 18.05, Spring 2014 Jeremy Orloff and Jonathan Bloom*. Available at https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading6a.pdf (Spring 2014).

[4] Jeremy Orloff and Jonathan Bloom. *Maximum Likelihood Estimates Class 10, 18.05 Jeremy Orloff and Jonathan Bloom*. Available at https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading10b.pdf (Spring 2014).