Problem Set 5

John Hancock

July 15, 2017

Contents

1 References and License

We are answering questions in the material from MIT OpenCourseWare course 18.05, Introduction to Probabil-

ity 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 [2]. 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 [3], section 4.
Then the likelihood function f_i for one point (x_i, y_i) is:

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

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

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

We can rewrite the product above as:

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

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:

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

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

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

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

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

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

2.2.1 General formulation

We gave the general formulation for the likelihood function above:

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

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.

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

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

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

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

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

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

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.

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

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. *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).

[3] 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).