Problem Set 5

John Hancock

July 12, 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 [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:

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

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:

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

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:

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

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

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

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

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

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).