Worksheet 14 (Solutions)

1. Consider a simple linear regression where we know that $b_0 = 0$. You can write $b_1 \to b$ to simplify the notation. Write down the likelihood function for the sample. Do not yet simplify.

Solution: The likelihood is given by:

$$\mathcal{L}(b; y_1, \ldots, y_n) = \prod_i \frac{1}{(2\pi\sigma^2)^{1/2}} \times e^{-\frac{1}{2\sigma^2}(y_i - x_i b)^2}.$$

2. Now, (a) compute the log-likelihood function and simplify. (b) Without doing any calculus (that is, just looking at the function), maximizing the log-likelihood with respect to b is equivalent to minimizing what quantity in terms of y_i, x_i, and b? (c) Why might it make sense to minimize this quantity? Note: Ask me about the correct solution before proceeding.

Solution: (a) Taking the logarithm and simplifying yields:

$$\begin{aligned} \updownarrow(b; y_1, \ldots, y_n) &= \sum_i \log\left(\frac{1}{(2\pi\sigma^2)^{1/2}}\right) - \frac{1}{2\sigma^2}(y_i - x_i b)^2 \\ &= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (y_i - x_i b)^2. \end{aligned}$$

(b) Looking at this, we see that to maximize the log-likelihood with b, we need to minimize the quantity $\sum_i (y_i - x_i b)^2$. (c) This is actually a logical thing to do, because these are the squared sums of the residuals, the amount that we are missing the y_i's by our regression line. Making these as small as possible is a reasonable thing; it is also where the term *best-fit line* for the solution comes from.

3. Take the derivative of the quantity that you had in part (b) from the previous question with respect to the parameter b. Set this equal to zero to get the MLE.

Solution: The derivative of the sum of squares is:

$$\frac{\partial}{\partial b}\sum_i (y_i - x_i b)^2 = 2\sum_i x_i \cdot (y_i - x_i b)$$

And solving for zero gives:

$$\begin{aligned} 2\sum_i x_i \cdot (y_i - x_i b) &= 0 \\ \sum_i x_i \cdot y_i - x_i &= \sum_i x_i^2 \widehat{b} \\ \frac{\sum_i x_i \cdot y_i}{\sum_i x_i^2} &= \widehat{b}. \end{aligned}$$

4. What obsevations will have the most influence on the estimate of the slope? Does this make sense?

Solution: Observations farther from the origin will have a higher impact on the output. This makes sense because we are measuring the slope of a line through the origin. Since the variance of Y_i is fixed, we have more signal in points that are farther from the origin.

5. What is the distribution of the MLE of b? Is the estimator unbiased? Under what conditions will it be consistent? Note: This will take several steps.

Solution: We see quickly that $\widehat{b}$ is a sum of independent normals, so it will have a normal distribution. We need only to figure out its mean and variance. These are given by:

$$\begin{aligned} \mathbb{E}\widehat{b} &= \mathbb{E}\left(\frac{\sum_i x_i \cdot y_i}{\sum_i x_i^2}\right) \\ &= \left(\frac{\sum_i x_i \cdot \mathbb{E}y_i}{\sum_i x_i^2}\right) \\ &= \left(\frac{\sum_i x_i \cdot x_i b}{\sum_i x_i^2}\right) \\ &= \left(\frac{\sum_i x_i^2 b}{\sum_i x_i^2}\right) \\ &= b \cdot \left(\frac{\sum_i x_i^2}{\sum_i x_i^2}\right) \\ &= b \end{aligned}$$

So, we see that it is unbiased. The variance is given by:

$$\begin{aligned}
Var(\widehat{b}) &= Var\left(\frac{\sum_i x_i \cdot y_i}{\sum_i x_i^2}\right) \\
&= \left(\frac{\sum_i x_i^2 \cdot Var(y_i)}{\left(\sum_i x_i^2\right)^2}\right) \\
&= \left(\frac{\sum_i x_i^2 \cdot \sigma^2}{\left(\sum_i x_i^2\right)^2}\right) \\
&= \sigma^2\left(\frac{\sum_i x_i^2}{\left(\sum_i x_i^2\right)^2}\right) \\
&= \sigma^2 \cdot \frac{1}{\sum_i x_i^2} = \frac{\sigma^2}{\sum_i x_i^2}.
\end{aligned}$$

The variance will limit to zero as long as $\sum_i x_i^2 \rightarrow \infty$, generally the case as long as we have data points x_i that are not limiting to the origin in some strange way.

6. Go back to the full log-likelihood function. Take the derivative with respect to σ^2 (remember, this is a single parameter, not the square of a parameter). Set this to zero and solve to get the MLE of σ^2. Does this equation make sense to you?

Solution: I will set $v = \sigma^2$ for clarify. Then, we have, at the optimal point of $b = \widehat{b}$, the following:

$$\begin{aligned}
\frac{\partial}{\partial v}(\cdot) &= \frac{-n}{2}\frac{1}{2\pi v} \cdot (2\pi) + \frac{1}{2v^2}\sum_i \widehat{y}_i^2 \\
&= \frac{-n}{2v} + \frac{1}{2v^2}\sum_i \widehat{y}_i^2
\end{aligned}$$

Setting this to zero yields:

$$\begin{aligned}
\frac{n}{2\widehat{v}} &= \frac{1}{2\widehat{v}^2}\sum_i \widehat{y}_i^2 \\
\frac{2\widehat{v}^2}{2\widehat{v}} &= \frac{1}{n}\sum_i \widehat{y}_i^2 \\
\widehat{v} &= \frac{1}{n}\sum_i \widehat{y}_i^2.
\end{aligned}$$

This should make sense because it measures the squared size of the residuals, which we expect to be normally distributed with variance σ^2.

7. The MLE estimator for σ^2 is biased, but we can fix this by dividing by $n-1$ instead of n, just as we did with the one-sample mean.

This unbiased version is independent of $\widehat{b}$. If we take this unbiased estimator and divide by σ^2, we will have a chi-squared distribution with $n-1$ degrees of freedom. Using this, create a pivot statistic that depends only on b and not σ^2.

Solution: This is just a matter of plugging in our answers to the previous questions and using the formula for a T-statistic:

$$
\begin{aligned}
T &= \frac{\frac{\widehat{b}-b}{\sqrt{\sigma^2 / \sum_i x_i^2}}}{\sqrt{\frac{n-2}{n-2} \cdot \frac{1}{\sigma^2} \sum_i \widehat{y}_i^2}} \\
&= \frac{\widehat{b}-b}{\frac{\sum_i \widehat{y}_i^2}{\sum_i x_i^2}}.
\end{aligned}
$$