Lecture 18 Worksheet

Chrysafis Vogiatzis

Every worksheet will work as follows.

1. You will be entered into a Zoom breakout session with other students in the class.
2. Read through the worksheet, discussing any questions with the other participants in your breakout session.
   - You can call me using the "Ask for help" button.
   - Keep in mind that I will be going through all rooms during the session so it might take me a while to get to you.
3. Answer each question (preferably in the order provided) to the best of your knowledge.
4. While collaboration between students in a breakout session is highly encouraged and expected, each student has to submit their own version.
5. You will have 24 hours (see Compass) to submit your work.

Worksheet 1: Our first MLE

As a reminder, to get the maximum likelihood estimators, we:

1. build the likelihood function $L(\theta)$ by multiplying the probability mass function (for discrete) or the probability density function (for continuous) for each of the sample values.
2. get the derivative(s) $\frac{\partial L(\theta)}{\partial \theta}$ for the unknown parameter(s).
3. equate them to zero and solve a system of equations.

Maximum likelihood estimators can get pretty tough to calculate as we have more and more samples, so let us start easy.

Problem 1: Building a likelihood function

Assume we have been told that a continuous population X is distributed with pdf $f(x) = \frac{1}{2}(1+\theta x), -1 \leq x \leq 1$. We have also collected a sample of size $n = 3$: $X_1 = 0.75, X_2 = 0.5, X_3 = 0.80$. What is the likelihood function $L(\theta)$?

Answer to Problem 1.

Problem 2: MLE for a sample of $n = 3$ observations

Based on your likelihood function from Problem 1, what is the maximum likelihood estimator for θ? [1]

Answer to Problem 2.

[1] If you are solving a quadratic equation, you are bound to get more than one solutions. That is because setting the derivative equal to 0 gives both maxima and minima: we only want the maximum here! From the two solutions, then, pick the one that gives maximum likelihood.

Problem 3: Success probabilities

In the notes we saw how to calculate the likelihood function and the maximum likelihood estimator for a Bernoulli random variable. Let us revisit this here and see some other discrete random variables, too.

Assume you are observing an experiment that is successful with probability $0 < p < 1$ and unsuccessful with probability $1 - p$. [2] You observe 8 experiments and you note $n_1 = 6$ successes and $n_2 = 2$ failures. What is the maximum likelihood estimator for p?

[2] For a Bernoulli random variable we have that its pmf is p for a success and $1 - p$ for a failure.

Answer to Problem 3.

Check your answer: is it the same as the estimator we would have gotten from the method of moments? Do you think this is true every time?

Worksheet 2: Streaming services and their data (reloaded)

Let us revisit the example from last time. Remember that TV streaming giant with the data they had collected? Well, they are back and they would like to see what else they can do to come up with estimators.

As a reminder, they have provided you with data on the time people spend during a session (continuous); as well as the number of episodes people watch during a session (discrete number).

	Session #									
	1	2	3	4	5	6	7	8	9	10
Time (in hours)	0.75	1.20	1.33	0.97	0.80	1.43	0.87	1.41	1.09	1.05
# of episodes	1	3	3	3	2	5	1	5	3	1

Let us help them find good estimators using the maximum likelihood method this time around!

Problem 4: Poisson rates in the general case

So far we have been providing you with the actual observations in the sample. What if we have not collected a sample yet? Can you still calculate the maximum likelihood estimator as a function of the sample, whatever it may be? Let's put this to practice for the rate of a Poisson distributed population. What is λ as a function of the sample collected $X_1, X_2, \ldots, X_n$?

Answer to Problem 4.

Problem 5: Poisson rates in the general case (reloaded)

While the previous result is not terribly difficult to derive, it is still confusing sometimes to take the derivative of a product of functions, as is the case with general likelihood.

Following our logic from earlier, the likelihood function for a general sample of size n (let it be $X_1, X_2, \ldots, X_n$) would be:

$$L(\lambda) = e^{-\lambda} \cdot \frac{\lambda^{X_1}}{X_1!} \cdot e^{-\lambda} \cdot \frac{\lambda^{X_2}}{X_2!} \cdot \ldots e^{-\lambda} \cdot \frac{\lambda^{X_n}}{X_n!}.$$

We would need the derivative of this [3], and this is not always easy. In the notes, we mentioned something called the log-likelihood. For the log-likelihood, you should still calculate the likelihood function but then take its logarithm to obtain $\ln(L(\theta))$. Then, you can take its derivative and equate it to 0. [4] Using the logarithm properties we may calculate

$$\ln(L(\lambda)) = \ln\left(e^{-\lambda} \cdot \frac{\lambda^{X_1}}{X_1!}\right) + \ln\left(e^{-\lambda} \cdot \frac{\lambda^{X_2}}{X_2!}\right) + \ldots + \ln\left(e^{-\lambda} \cdot \frac{\lambda^{X_n}}{X_n!}\right),$$

which is **so much easier to differentiate**! What would be the maximum likelihood estimator?

Answer to Problem 5.

It is of course the same either way! Using the likelihood or the log-likelihood to differentiate and equate to 0 will always give the same result.

[3] In terms of λ, recall that the observations X_i in the sample are supposed to be known values!

[4] Some properties logarithms have:

$$\ln(a \cdot b) = \ln a + \ln b$$
$$\ln\left(a^b\right) = b \cdot \ln a$$

Problem 6: Solving the example

Based on your answer in Problem 4 or 5, use the data provided for the number of episodes watched per session and calculate the maximum likelihood estimator for the rate λ. Is it the same as the rate you got when you used the method of moments during the previous lecture?

Answer to Problem 6.

Worksheet 3: Extra details

In this worksheet we see a couple of special cases and answer the following questions:

1. Are the method of moments and MLE always giving us the same solutions?

2. How can we estimate two or more parameters using MLE?

Problem 7: Not always the same

Consider the following probability density function [5] $f(x) = \theta x^{\theta-1}$ defined over $0 \leq x \leq 1$. Assume we have been provided a sample of size n: $X_1, X_2, \ldots, X_n$. We can apply the method of moments to get:

$$E[X] = \sum_{i=1}^{n} X_i \implies \int_0^1 x f(x) dx = \overline{X} \implies \int_0^1 \theta x^{\theta} dx = \overline{X} \implies$$

$$\implies \theta \frac{x^{\theta+1}}{\theta+1}\bigg|_0^1 = \overline{X} \implies \frac{\theta}{\theta+1} = \overline{X} \implies \hat{\theta} = \frac{\overline{X}}{1-\overline{X}}.$$

How about the maximum likelihood estimator? What would it be? [6]

Answer to Problem 7.

So, as we just saw the method of moments and the maximum likelihood estimation method match often; but not always.

[5] Taken from last year's quiz! So... good practice!

[6] You will need to take the derivative of a^x as far as x is concerned. We have:

$$\frac{\partial a^x}{\partial x} = a^x \cdot \ln a.$$

Problem 8: The normal distribution

Assume that you have collected $n = 5$ observations from a normally distributed random variable with unknown mean μ and unknown variance σ^2. What are the maximum likelihood estimators for μ and σ^2? You may use the following data for your sample: $X_1 = 113, X_2 = 118, X_3 = 116, X_4 = 127, X_5 = 116$. [7]

Answer to Problem 8.

[7] We need the pdf of the normal distribution to answer this question. You may use that $f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.