MLE for Gaussian population

In the videos, you got an intuition of what the Maximum Likelihood Estimation (MLE) should look like for the mean and variance of a Gaussian population.

In this reading item, you will learn the derivation of both results.

Mathematical formulation

Suppose you have n samples $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ from a Gaussian distribution with mean μ and variance σ^2. This means that $X_i \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma)$.

If you want the MLE for μ and σ the first step is to define the likelihood. If both μ and σ are unknown, then the likelihood will be a function of these two parameters.

$$\begin{aligned} L(\mu, \sigma; \boldsymbol{x}) &= \prod_{i=1}^{n} f_{X_i}(x_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n} e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}} \end{aligned}$$

Now all you have to do is find the values of μ and σ that maximize the likelihood $L(\mu, \sigma; \boldsymbol{x})$.

You might remember from the calculus course that one way to do this analytically is by taking the derivative of the Likelihood function and equating it to 0. The values of μ and σ that make the derivative zero, are the extreme points. In particular, for this case, they will be maximums.

Taking the derivative of the likelihood is a cumbersome procedure, because of all the products involved. However, there is a nice trick you can use to simplify things. Note that the logarithm function is always increasing, so the values that maximize $L(\mu, \sigma; \boldsymbol{x})$ will also maximize its logarithm. This is the **log-likelihood**, and it is defined as

$$\ell(\mu, \sigma) = \log(L(\mu, \sigma; \boldsymbol{x}))$$

The logarithm has the property of turning a product into a sum, this means that $\log(a \cdot b) = \log(a)\log(b)$. This makes taking the derivative of the log-likelihood very straight forward. To get the simplest expression for the log-likelihood for a Gaussian population, you will also need the following properties of the logarithm:

$$\log(1/a) = -\log(a)$$

and

$$\log(a^k) = k\log(a).$$

Putting it all together you get:

$$\begin{aligned} \ell(\mu, \sigma) &= \log\left(\frac{1}{\left(\sqrt{2\pi}\right)^n \sigma^n} e^{-\frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2}}\right) \\ &= -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2}\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^2} \end{aligned}$$

Now to find the MLE for μ and σ, all there is left to do is take the partial derivatives of the log-likelihood, and equate them to zero.

For the partial derivative with respect to μ note that the first two terms do not involve μ, so you get:

$$\begin{aligned} \frac{\partial}{\partial \mu}\ell(\mu, \sigma) &= -\frac{1}{2}\frac{\sum_{i=1}^{n} 2(x_i-\mu)}{\sigma^2}(-1) \\ &= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu\right) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right) \end{aligned}$$

Now, for the partial derivative with respect to σ you get that

$$\frac{\partial}{\partial \sigma}\ell(\mu, \sigma) = -\frac{n}{\sigma} - \frac{1}{2}\left(\sum_{i=1}^{n}(x_i-\mu)^2\right)(-2)\frac{1}{\sigma^3} = -\frac{n}{\sigma} + \left(\sum_{i=1}^{n}(x_i-\mu)^2\right)\frac{1}{\sigma^3}$$

The next step is equating this to 0 to find the estimates for μ and σ. Let's begin with the partial derivative with respect to μ:

$$\frac{\partial}{\partial \mu}\ell(\mu, \sigma) = \frac{1}{\sigma}\left(\sum_{i=1}^{n} x_i - n\mu\right) = 0$$

First, observe that since $\sigma > 0$, the only option is that $\sum_{i=1}^{n} x_i - n\mu = 0$. Simple algebraic manipulations show that the MLE for μ has to be

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x},$$

which is the sample mean.

Next, find the value of σ that achieves $\frac{\partial}{\partial \sigma}\ell(\mu, \sigma) = 0$:

$$\frac{\partial}{\partial \sigma}\ell(\mu, \sigma) = -\frac{n}{\sigma} + \left(\sum_{i=1}^{n}(x_i-\mu)^2\right)\frac{1}{\sigma^3} = 0$$

In this case, first note that since $\sigma > 0$ you can simplify the expression to

$$\frac{\partial}{\partial \sigma}\ell(\mu, \sigma) = -n + \left(\sum_{i=1}^{n}(x_i-\mu)^2\right)\frac{1}{\sigma^2} = 0$$

Also, you can replace μ by its estimate $\hat{\mu} = \overline{x}$, because yuo want both partial derivatives to be 0 at the same time. You get

$$\frac{\partial}{\partial \sigma}\ell(\mu, \sigma) = -n + \left(\sum_{i=1}^{n}(x_i-\overline{x})^2\right)\frac{1}{\sigma^2} = 0$$

This gives you

$$\sigma^2 = \frac{\sum(x_i-\overline{x})^2}{n},$$

so the MLE for the standard deviation is

$$\hat{\sigma} = \sqrt{\frac{\sum(x_i-\overline{x})^2}{n}}$$

This expression tells you that the MLE for the standard deviation of a Gaussian population is the square root of the average squared difference between each sample and the sample mean. This expression is very similar to the one you learnt in Week 2 for the sample standard deviation. The only difference is the normalizing constant: for the MLE you have $1/n$ while for the sample standard deviation you use $1/(n-1)$.

A final comment: formally, what you just did was the derivation of the critical point. To make it all complete, you would need to show that these are the coordinates of a maximum point (and not a minimum or saddle point). However, this proof would require a little bit more complicated math and we will skip it here.

A simple example

Now, let's see how this looks like with an example. Suppose you are interested on distribution of heights od 18 year olds in the US. You have the following 10 measurements:

$$\begin{array}{ccccc} 66.75 & 70.24 & 67.19 & 67.09 & 63.65 \\ 64.64 & 69.81 & 69.79 & 73.52 & 71.74 \end{array}$$

Each measurement is supposed to come from a Gaussian distribution with unknown parameters μ and σ. The MLE estimation for the parameters with this samples are

$$\hat{\mu} = \frac{66.75 + 70.24 + 67.19 + 67.09 + 63.65 + 64.64 + 69.81 + 69.79 + 73.52 + 71.74}{10} = 68.442$$

and

$$\begin{aligned} \hat{\sigma} &= \sqrt{\frac{1}{10}\left(\begin{array}{l} (66.75-68.442)^2 + (70.24-68.442)^2 + (67.19-68.442)^2 + (67.09-68.442)^2 + \\ (63.65-68.442)^2 + (64.64-68.442)^2 + (69.81-68.442)^2 + (69.79-68.442)^2 + \\ (73.52-68.442)^2 + (71.74-68.442)^2 \end{array}\right)} \\ &= 2.954 \end{aligned}$$

Completed **Go to next item**

Like **Dislike** **Report an issue**