

EPFL

Simple example: maximum likelihood estimation



In this video, we will investigate another estimator of the parameters. Previously, we used the sample average from the contingency table. This is possible because we have a very simple example. But it won't be possible to apply for most complex models that we will see later on. We will actually apply this maximum likelihood estimation. And I just would like to introduce you to: what is maximum likelihood estimation in the very context of this simple example.

Notes

Summary



0m 05s

Maximum likelihood estimation

Likelihood

Probability that the model correctly predicts all the observations

Likelihood function

$$\mathcal{L}^* = \prod_{n=1}^N P(i_n | k_n)$$

For our example

$$\mathcal{L}^* = (\pi_1)^{65} (1 - \pi_1)^{835} (\pi_2)^{55} (1 - \pi_2)^{1045} (\pi_3)^5 (1 - \pi_3)^{495}$$

The likelihood is defined as the probability that the model that we develop correctly predicts all the observations in the sample. To calculate it, we select each individual in the sample (there are N of them) and for each of them we calculate the probability of the explained variable (here the ownership of the electric car) given the explanatory variable (here, the age). In our example, this is relatively simple to calculate because we can calculate the contribution of this likelihood for each cell of the contingency table. So this is the first cell of people who are of age category one who own an electric car. The probability that they do so, that they actually own a car, is given by π_1 by the model and there are 65 observations that are associated with this choice. In the same age category, there are 835 people who do not own an electric car. So the model predicts $1 - \pi_1$ for these people, for this outcome. And then we do the same for the other two age categories. We have 55 people who own an electric car in age category 2. 1045 who don't own an electric car in category 2. And then we have 5 and 495 for Category 3. So this is what is called the likelihood function of the sample for this model.

Notes

Summary



0m 29s

Maximum likelihood estimation

Estimates

- ▶ Values of the parameters that maximize \mathcal{L}^* .
- ▶ In practice, the logarithm is maximized

$$\mathcal{L} = \ln \mathcal{L}^* = \sum_{n=1}^N \ln P(i_n | k_n).$$

As $0 \leq \mathcal{L}^* \leq 1$, we have $\mathcal{L} \leq 0$.

Properties

- ▶ Consistency
- ▶ Asymptotic efficiency

Now we can use this function to estimate the true parameters π_1 , π_2 and π_3 . And the idea is to take the values of the parameter that maximize this likelihood, that maximize the probability that the model correctly predicts the sample. Actually, this function has nasty mathematical properties. It is manipulating very small numbers. These are probabilities with exponents. So in practice it's much easier to deal with the logarithm of the likelihood. It's called the log likelihood. And you can verify that, in this case, the log of this likelihood can be written as the sum of the log of each probability for each individual. The log likelihood is a probability. So it's between 0 and 1. And now, because we take the log of this, we have something which is always negative. Maximum likelihood estimation has some interesting properties in terms of inference of the parameters. The estimates are consistent. It means that they are asymptotically unbiased (asymptotically means when the number of observations tends to infinity) and, also, they are asymptotically efficient. Efficient means that it's not possible to find another estimator with a smaller variance. At least asymptotically, when the number of observations tends to infinity.

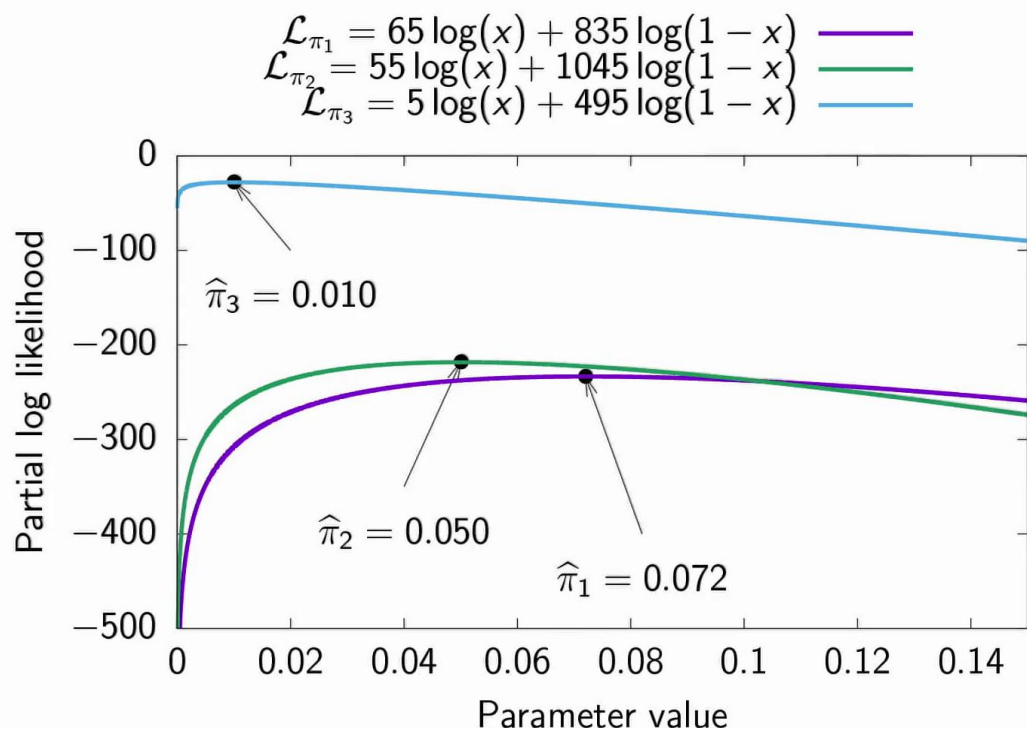
Notes

Summary



2m 03s

Maximum likelihood



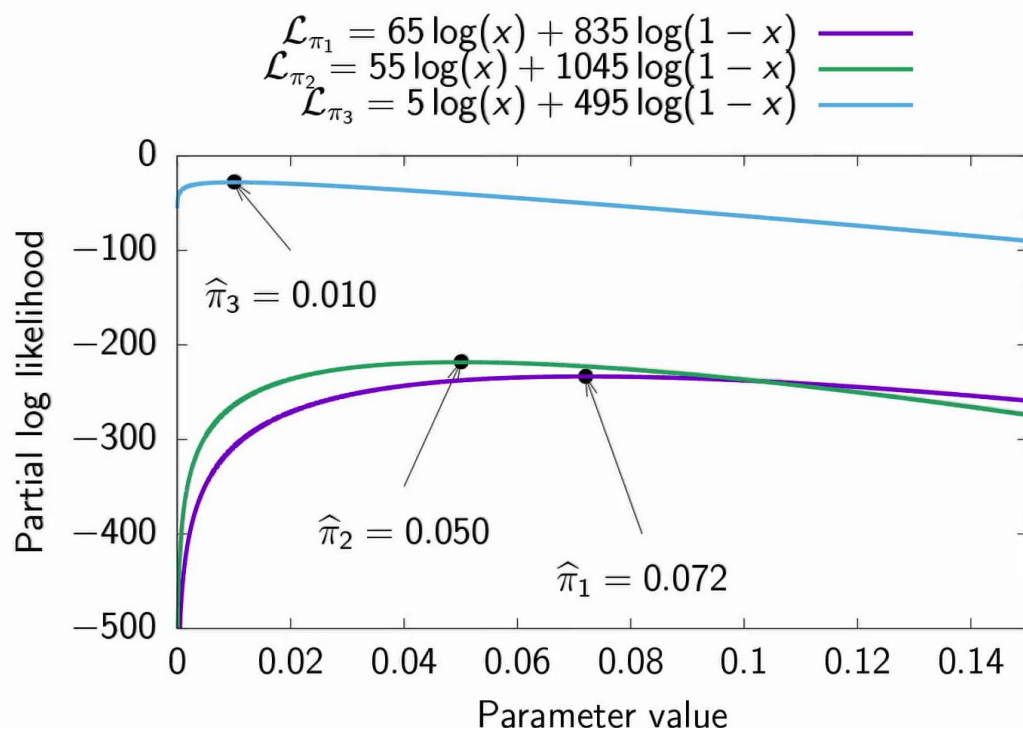
So let's see what they look like. Interestingly, in this example, you can separate the log likelihood function into three parts which are independent. Each part is associated with one age category. So you have here the contribution to the log likelihood associated with the first age category. The second one and the third one. And each of them is associated with a different parameter. I have plotted them on this chart where the x axis corresponds to the parameter value and the y axis corresponds to the partial log likelihood, so the contribution of this partial function to the complete log likelihood. And when you look at it you can find the maximum. So for the blue curve which is corresponding to parameter π_1 , the value of π_1 that maximizes the blue curve is equal to 0.072. This is the estimate of π_1 given by maximum likelihood. For π_2 , this is given by 0.050 which is corresponding to this point here. And for π_3 , this is equal to 0.010 which is the maximum of this third partial log likelihood. As you may notice, interestingly, the estimates that we obtain here using maximum likelihood estimation are exactly the same as we obtained using the sample average from the contingency table.

Notes

Summary



Maximum likelihood



Because the example is simple. Now in practice, for more complex models, the likelihood function is not always separable and will be more complicated to maximize. But, still, the idea is exactly the same. We write the log likelihood function by adding all the contributions of each observation in the sample. And then we find the value of the parameters that maximize this value.

Notes

Summary



Maximum likelihood



Once we have estimated the value of the parameters using maximum likelihood estimation, we are back to the question about what is the validity of these estimates. In this simple example, we will do some hypothesis testing. So I would suggest that you go and read the text on hypothesis testing, that is provided on the web page.

Notes

Summary



5m 15s