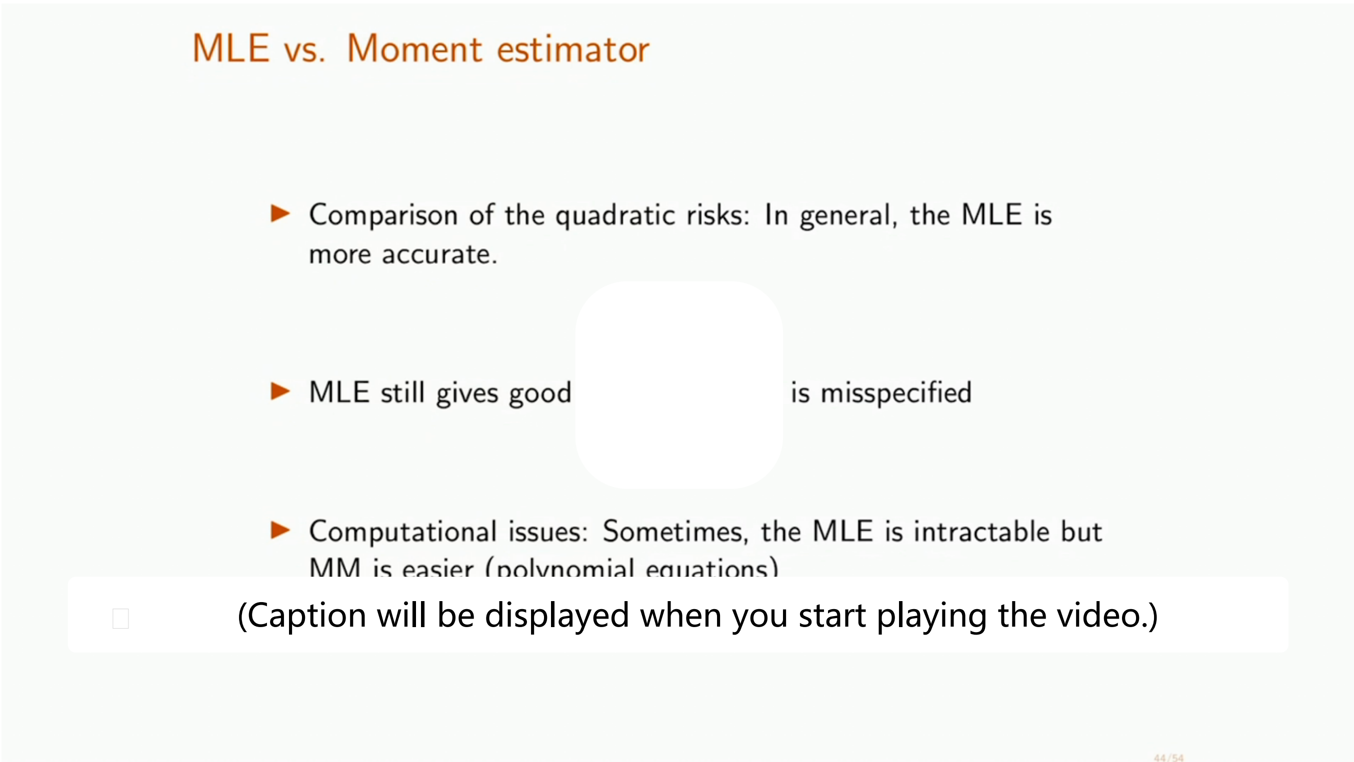
11. MLE versus Method of Moments

MLE versus Method of Moments



Start of transcript. Skip to the end.

So how do you compare them?

Which one should you pick?

The maximum likelihood estimator or the moment estimator?

Well there's we have ways to actually compare estimators,

right?

So for example, we could look at the quadratic risk,

which is a combination of bias and variance.

视频
下载视频文件

字幕
下载 SubRip (.srt) file
下载 Text (.txt) file

MLE vs. Method of Moments

1/1 point (graded)
Which of the following are advantages of using the MLE over the method of moments estimator? (Choose all that apply.)

Remark: All of the choices below are true statements; your task is to figure out which of these choices are indeed advantages.

- [x] In general, the MLE provides a more accurate estimator than the method of moments estimator.
- [ ] If the likelihood has several local maxima, then we may not be able to compute the MLE efficiently
- [x] The method of moments requires you to find d so that the first d moments uniquely determine the distribution of interest. To compute the MLE, this step is not necessary.

Solution:

We examine the choices in order.

- As stated in the slides, if we compare the quadratic risks of the method of moments estimator and the MLE, then the MLE has better performance in general. Hence "In general, the MLE provides a more accurate estimator than the method of moments estimator." is correct.

- Since the MLE is not always computationally tractable, this is a disadvantage. Optimizing the likelihood function can be very inefficient if the likelihood function is complicated and has several local maxima which require testing. Hence "If the likelihood has several local maxima, then we may not be able to compute the MLE efficiently" is an incorrect response.

- "The method of moments requires you to find d so that the first d moments uniquely determine the distribution of interest. To compute the MLE, this step is not necessary." is correct. The expression of the moments map ψ in terms of the parameter θ can be quite complicated, so it may be difficult to deduce how many moments (or *degrees of freedom*) are needed to uniquely recover the true distribution from moments. It is not necessary to make assumptions on or work with the moments map to use the MLE, so this is another advantage.

提交 你已经尝试了2次（总共可以尝试3次）

Answers are displayed within the problem

讨论

隐藏讨论

主题： Unit 3 Methods of Estimation:Lecture 11: Fisher Information, Asymptotic Normality of MLE; Method of Moments / 11. MLE versus Method of Moments

Add a Post

All Posts

May I get some examples of "MLE still gives good results if model is misspecified"?

question posted 2 days ago by butterandfly

Thanks!

此帖对所有人可见。

添加回复

3 responses

dfannius (Community TA)
2 days ago

This isn't a very concrete answer, but we know that a Gaussian distribution is a pretty decent approximation for a Poisson distribution as λ gets larger. So if your data was actually Poisson, but you did your MLE assuming it was Gaussian, your estimated Gaussian distribution might not be very far off (by some measure like KL divergence) from the true Poisson distribution, despite being from the wrong family.

It's a nice one! Thanks!

butterandfly 在a day ago前发表

添加评论

sudarsanvsr_mit (Staff)
2 days ago

When we derived MLE we started with the definition of KL divergence.

The MLE is the minimizer of $\mathrm{KL}\left(\mathbf{P}_{\theta^*}, \mathbf{P}_{\theta}\right)$. Say the data was generated from $\mathbf{P}_{\theta^*}$ but you think the family of distributions is $\mathbf{Q}_{\theta}$. Then, MLE is the minimizer of $\mathrm{KL}\left(\mathbf{P}_{\theta^*}, \mathbf{Q}_{\theta}\right)$. That is, the ML estimate is still the closest in the family of distributions that we considered to be the family that generated the data to the actual true distribution in the true family.

Right! Now I get it! Thanks!

butterandfly 在a day ago前发表

Very nicely explained, thanks 😁

katicicar 在about 8 hours ago前发表

添加评论

JenniferVoitle
a day ago

I'm reading this now. it gives some good examples such as incorrectly assuming that the mean of a population is zero when estimating the variance, and so on. https://www.jstor.org/stable/1912526

Wow, thanks for sharing!

butterandfly 在a day ago前发表

添加评论

显示所有的回复

添加一条回复：

Preview

提交