5. Fisher Information and the Asymptotic Normality of the ML Estimator

A Geometric View on the Fisher Information

1/1 point (graded)

Let $(\mathbb{R}, \{\mathbf{P}_\theta\}_{\theta\in\mathbb{R}})$ denote a statistical model. Recall that the MLE (maximum likelihood estimator) for one observation maximizes the log-likelihood for one observation, which is the random variable $\ell(\theta) = \ln L_1(X,\theta)$ where $X \sim \mathbf{P}_\theta$. Suppose we observe $X_1 = x_1$, and now consider the graph of the function $\theta \mapsto \ln L_1(x_1,\theta)$.

What does the Fisher information $\mathcal{I}(\theta)$ represent?

Hint: Use the definition $\mathcal{I}(\theta) = -\mathbb{E}[\ell''(\theta)]$.

- It gives you an approximation for the true parameter θ^*.
- It tells you the average slope of the function $\theta \mapsto \ln L(x_1,\theta)$ is.
- It tells you, on average, how curved the function $\theta \mapsto \ln L(x_1,\theta)$ is.

Solution:

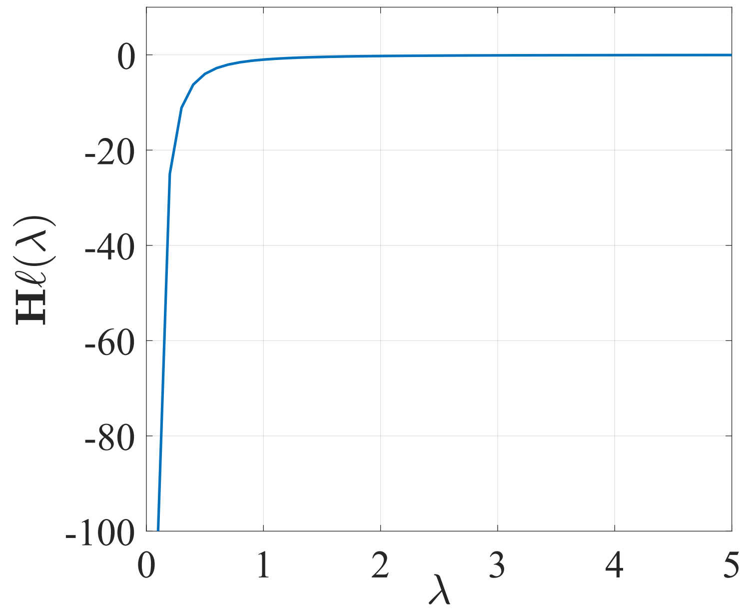
The third choice "It tells you, on average, how curved the function $\theta \mapsto \ln L(x_1,\theta)$ is." is correct. Recall that the Fisher information is also equal to the expected second-derivative of the log-likelihood: $\mathcal{I}(\theta) = -\mathbb{E}[\ell''(\theta)]$. Since the second-derivative measures concavity/convexity (how curved a function is at a particular point), $\mathcal{I}(\theta)$ measures the *average* curvature of the function $\theta \mapsto \ell(\theta) = \ln L_1(x_1,\theta)$.

Remark: It turns out that the Fisher information tells how curved (on average) the log-likehood $\ln L_n(x_1,\ldots,x_n,\theta)$ for several samples $X_1 = x_1,\ldots,X_n = x_n$ is. In particular, $\mathcal{I}(\theta^*)$ tells how curved (on average) the log-likelihood is near the true parameter. As a rule of thumb, if the Fisher information $\mathcal{I}(\theta^*)$ is large, then we expect the MLE to give a good estimate for θ^*.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Consider the exponential statistical model with $f_\lambda(x) = \lambda e^{-\lambda x}, x \geq 0$ and $\lambda \in (0,\infty)$. The second derivative of $\ell(\lambda)$, which is $\mathbf{H}\ell(\lambda) = \ell''(\lambda)$ (you will compute later in a homework exercise), does not depend upon x and is shown in the following figure.



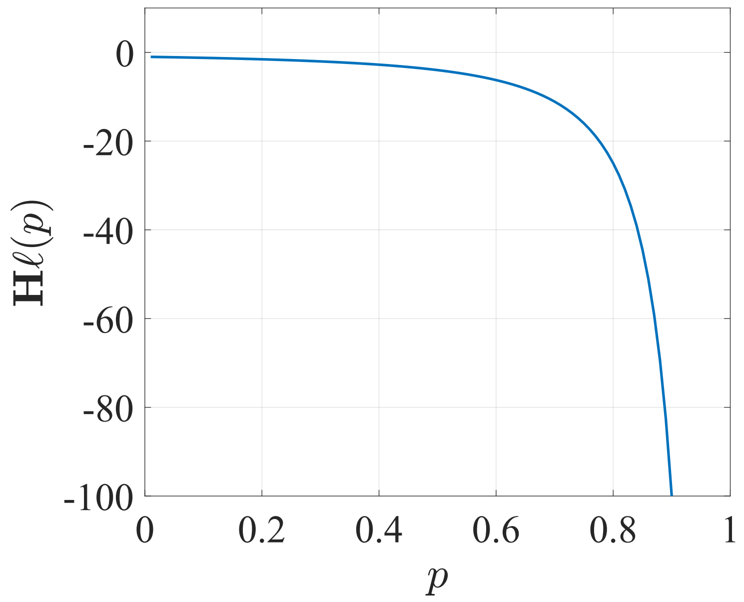
$\ell''(\lambda)$ for the exponential statiscal model with parameter λ for all X.

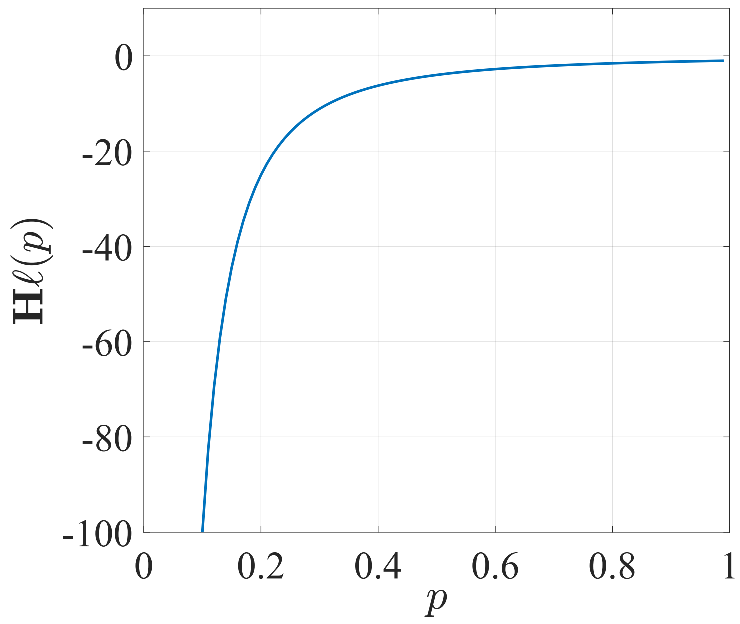
The Fisher information, $\mathcal{I}(\theta) = -\mathbb{E}[\mathbf{H}\ell(\theta)]$, captures the negative of the expected curvature of $\ell(\theta)$. For example, for the exponential statistical model, the expected curvature of $\ell(\lambda)$ is $\ell''(\lambda)$ itself and this is shown in the figure above. The Fisher information in this case is always positive. The fact that $\ell''(\lambda)$ is negative for all x also means that the log-likelihood function $\ell(\lambda)$ is a concave function of λ for all x.

If we consider the Bernoulli statistical model with parameter $p \in (0,1)$, we derived in a lecture video (also can be seen in the slides) that

$$\ell''(p) = -\frac{X}{p^2} - \frac{1-X}{(1-p)^2}, \quad X \in \{0,1\}.$$

Here, we see that $\ell''(p)$ is a random function that depends upon X. The following two figures show $\ell''(p)$ for $X=0$ and $X=1$.





$\ell''(p)$ for the Bernoulli statiscal model with parameter $p \in (0,1)$ for $X=0$ (left) and $X=1$ (right).

The **asymptotic normality of the ML estimator** , which will be discussed in the upcoming video, depends upon the Fisher information. For a one-parameter model (like the exponential and Bernoulli), the asymptotic normality result will say something along the lines of following: that the asymptotic variance of the ML estimator is inversely proportional to the value of Fisher information at the true parameter θ^* of the statistical model. This means that if the value of Fisher information at θ^* is high, then the asymptotic variance of the ML estimator for the statistical model will be low.

Asymptotic Normality of the Maximum Likelihood Estimator

the smaller the asymptotic variance to have.

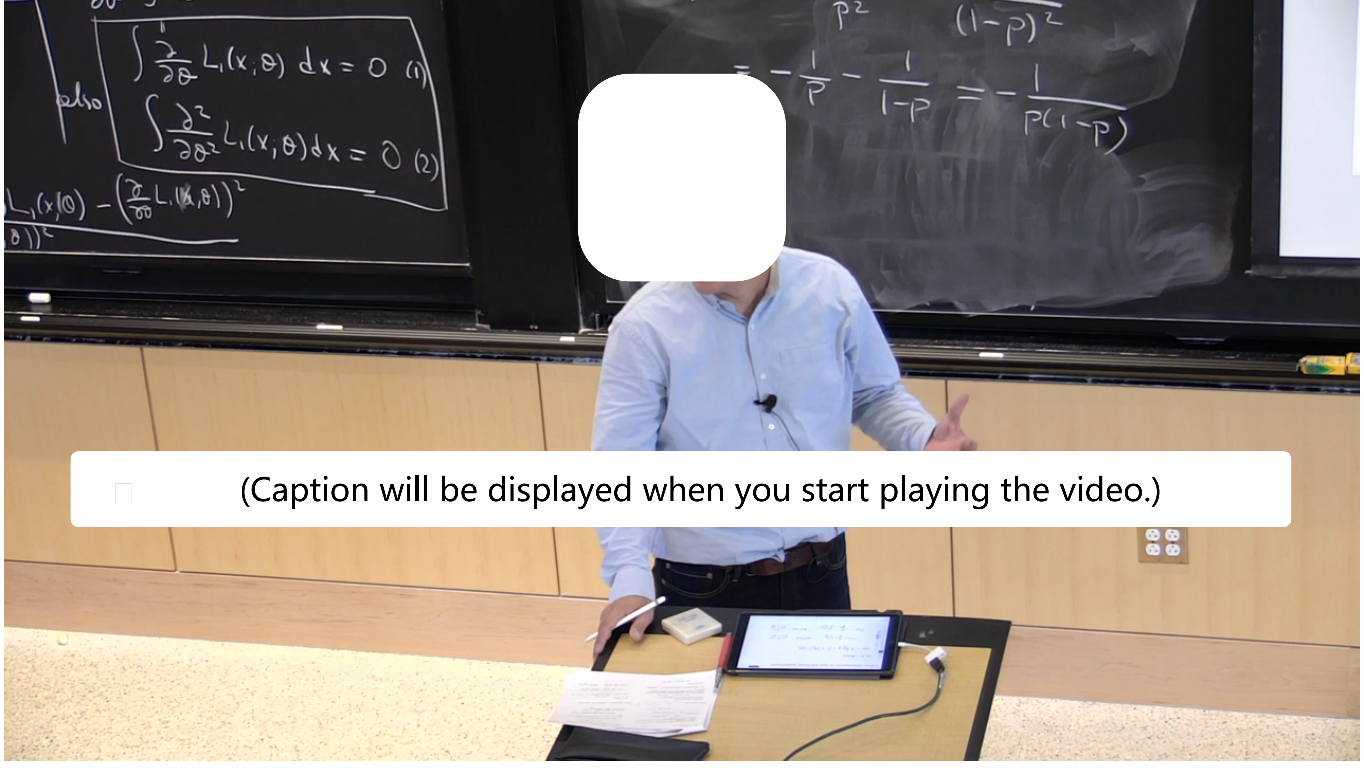
I reduce my uncertainty by having more information.

So this inverse, at least from semantic perspective,

this sort of makes sense.

Now I need to tell you why this is what it is.

Of course, if D is equal to 1 this is just 1 over I of theta.



End of transcript. Skip to the start.

视频
下载视频文件

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

Weak Consistency of the MLE

1/1 point (graded)

Let $(\mathbb{R}, \{\mathbf{P}_\theta\}_{\theta\in\mathbb{R}})$ denote a statistical model associated to a statistical experiment $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathbf{P}_{\theta^*}$ for some true parameter θ^* that we would like to estimate. You construct the maximum likelihood estimator $\hat{\theta}_n^{\text{MLE}}$ for θ^*. Which of the following conditions is **not** necessary for the MLE $\hat{\theta}_n^{\text{MLE}}$ to converge to θ^* in probability ?

- The model $(\mathbb{R}, \{\mathbf{P}_\theta\}_{\theta\in\mathbb{R}})$ is identified. (Recall that the parameter θ is identified if the map $\theta \mapsto \mathbf{P}_\theta$ is injective.)
- For all $\theta \in \Theta$, the support of $\mathbf{P}_\theta$ does not depend on θ.
- The MLE $\hat{\theta}_n^{\text{MLE}}$ is given by the sample average.
- The Fisher information $I(\theta)$ is non-zero in an interval containing true parameter θ^*. (Note that this is what it means for a 1×1 matrix, a scalar, to be invertible.)

Solution:

The power of the theorem for the convergence of the MLE is that it applies even in situations where the MLE is *not* the sample average. Hence, the third choice, "The MLE $\hat{\theta}_n^{\text{MLE}}$ is given by the sample average.", is correct, as it is not an assumption needed for the theorem statement. On the other hand, the first, second, and fourth choices are all hypotheses in the theorem statement regarding the convergence of the MLE. Hence, "The model $(\mathbb{R}, \{\mathbf{P}_\theta\}_{\theta\in\mathbb{R}})$ is identified.", "For all $\theta \in \Theta$, the support of $\mathbf{P}_\theta$ does not depend on θ.", and "The Fisher information $I(\theta)$ is invertible in an interval containing the true parameter θ^*" are all incorrect responses.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Informal Idea of Proof of Asymptotic Normality of the MLE

Start of transcript. Skip to the end.