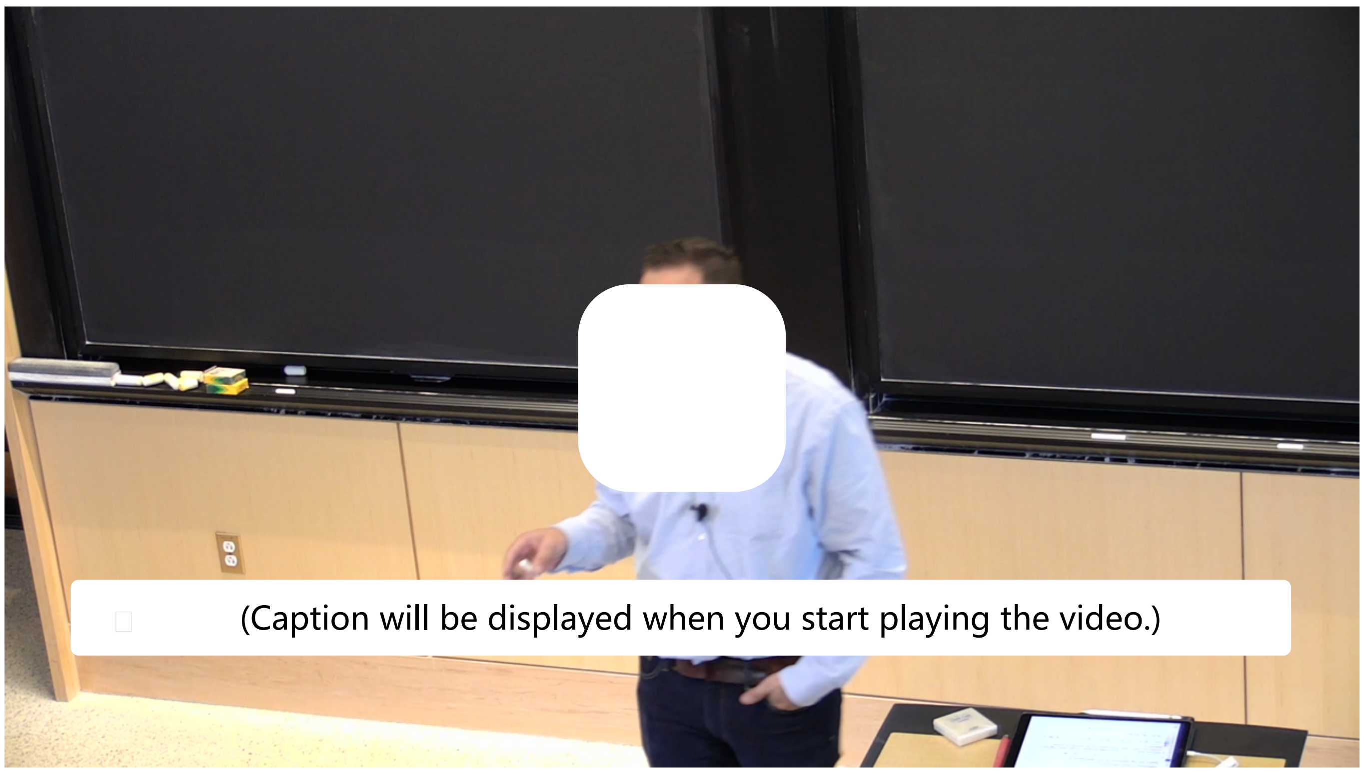
3. Fisher Information

Fisher Information: Definitions



as--

I can think about this covariance, right?

We introduced the covariance matrix of a random vector.

So I have the covariance of gradient L of theta.

And by definition, this was just the expectation of gradient L

theta gradient L theta transpose minus the expectation

of gradient L theta expectation gradient L theta transpose,

right?

So here I'm just writing the definition of what

the covariance actually is.

And this is what you see right here.

So far I'm not doing anything.

I just introduced a quantity.

And for the fun of it, I'm computing its gradient and then

视频
下载视频文件

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

Let $(\mathbb{R}, \{\mathbf{P}_\theta\}_{\theta\in\mathbb{R}})$ denote a continuous statistical model. Let $f_\theta(x)$ denote the pdf (probability density function) of the continuous distribution $\mathbf{P}_\theta$. Assume that $f_\theta(x)$ is twice-differentiable as a function of the parameter θ.

In the next few problems, you will derive the formula

$$\mathcal{I}(\theta) = \int_{-\infty}^{\infty} \frac{\left(\frac{\partial f_\theta(x)}{\partial\theta}\right)^2}{f_\theta(x)}\, dx$$

using the definition $\mathcal{I}(\theta) = \mathsf{Var}(\ell'(\theta))$ and the basic formula $\mathsf{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ for any random variable X.

For computations, it is sometimes convenient to use the above formula for the Fisher information.

Note: The derivation in the next set of problems is presented as a proof in the video that follows, but we encourage you to attempt these problems before watching the video.

Deriving a Useful Formula for the Fisher Information I

2/2 points (graded)

Let $(\mathbb{R}, \{\mathbf{P}_\theta\}_{\theta\in\mathbb{R}})$ denote a statistical model for a continuous distribution $\mathbf{P}_\theta$. Let f_θ denote the pdf (probability density function) of the continuous distribution $\mathbf{P}_\theta$. Recall that

$$\int_{-\infty}^{\infty} f_\theta(x)\, dx = 1$$

for all $\theta \in \mathbb{R}$.

For the next two questions, assume that you are allowed to interchange derivatives and integrals.

What is

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f_\theta(x)\, dx \ ?$$

0

Answer: 0.0

What is

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} f_\theta(x)\, dx \ ?$$

0

Answer: 0.0

STANDARD NOTATION

Solution:

Since we know $\int_{-\infty}^{\infty} f_\theta(x)\, dx = 1$, this implies that

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f_\theta(x)\, dx = \frac{\partial}{\partial\theta} 1 = 0.$$

Since we are allowed to interchange the integral and derivative, this implies that

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f_\theta(x)\, dx = 0.$$

Similarly for the second derivative,

$$\int_{-\infty}^{\infty} \frac{\partial^2}{\partial\theta^2} f_\theta(x)\, dx = \frac{\partial^2}{\partial\theta^2}\int_{-\infty}^{\infty} f_\theta(x)\, dx = \frac{\partial^2}{\partial\theta^2} 1 = 0.$$

Remark: If f is "nice enough," analytically speaking, then we can rigorously justify interchanging the integral and derivative.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Deriving a Useful Formula for the Fisher Information II

1/1 point (graded)

As before, let f_θ denote the pdf (probability density function) of the continuous distribution $\mathbf{P}_\theta$. By definition,

$$\ell(\theta) = \ln L_1(X, \theta) = \ln f_\theta(X)$$

where $X \sim \mathbf{P}_\theta$. Differentiating, we see

$$\ell'(\theta) = \frac{\partial}{\partial\theta} \ln f_\theta(X) = \frac{\frac{\partial}{\partial\theta} f_\theta(X)}{f_\theta(X)}.$$

What is

$$\mathbb{E}[\ell'(\theta)] = \mathbb{E}\left[\frac{\frac{\partial}{\partial\theta} f_\theta(X)}{f_\theta(X)}\right]?$$

0

Answer: 0.0

(Note that $X \sim \mathbf{P}_\theta$.)

STANDARD NOTATION

Solution:

Observe that

$$\mathbb{E}\left[\frac{\frac{\partial}{\partial\theta} f_\theta(X)}{f_\theta(X)}\right] = \int_{-\infty}^{\infty} \left(\frac{\frac{\partial}{\partial\theta} f_\theta(x)}{f_\theta(x)}\right) f_\theta(x)\, dx = \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta} f_\theta(x)\, dx = 0,$$

by the computation in the previous question.

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

Deriving a Useful Formula for the Fisher Information III

1/1 point (graded)

As before, let f_θ denote the pdf (probability density function) of the continuous distribution $\mathbf{P}_\theta$. By definition,

$$\ell(\theta) = \ln L_1(X, \theta) = \ln f_\theta(X)$$

where $X \sim \mathbf{P}_\theta$.

Using the previous question, which of the following are equal to $\mathsf{Var}(\ell'(\theta)) = \mathsf{Var}\left(\frac{\partial}{\partial\theta} \ln f_\theta(X)\right)$? (Choose all that apply.)

- [x] $\mathcal{I}(\theta)$
- [ ] $\mathbb{E}[\ell'(\theta)]$
- [x] $\mathbb{E}\left[(\ell'(\theta))^2\right]$
- [x] $\int_{-\infty}^{\infty} \frac{\left(\frac{\partial}{\partial\theta} f_\theta(x)\right)^2}{f_\theta(x)}\, dx$

Solution:

We consider the choices in order.

- By definition, $\mathcal{I}(\theta) = \textsf{Var}(\ell'(\theta))$, so the first answer choice $\mathcal{I}(\theta)$ is correct.
- By the previous question, $\mathbb{E}[\ell'(\theta)] = 0$, so this answer choice is incorrect.
- By definition of variance,

$$\textsf{Var}(\ell'(\theta)) = \mathbb{E}[\ell'(\theta)^2] - \mathbb{E}[\ell'(\theta)]^2,$$

  and $\mathbb{E}[\ell'(\theta)] = 0$, by the previous question. Hence, $\mathbb{E}[(\ell'(\theta))^2] = \textsf{Var}(\ell'(\theta))$, and so the answer choice $\mathbb{E}[(\ell'(\theta))^2]$ is correct.

- The last choice $\int_{-\infty}^{\infty} \frac{\left(\frac{\partial}{\partial\theta} f_\theta(x)\right)^2}{f_\theta(x)}\, dx$ is correct because, using the previous bullet,

$$\begin{aligned}\textsf{Var}(\ell'(\theta)) &= \mathbb{E}\left[(\ell'(\theta))^2\right] \\ &= \mathbb{E}\left[\left(\frac{\frac{\partial}{\partial\theta} f_\theta(X)}{f_\theta(X)}\right)^2\right] \\ &= \int_{-\infty}^{\infty} \frac{\left(\frac{\partial}{\partial\theta} f_\theta(x)\right)^2}{f_\theta(x)}\, dx.\end{aligned}$$

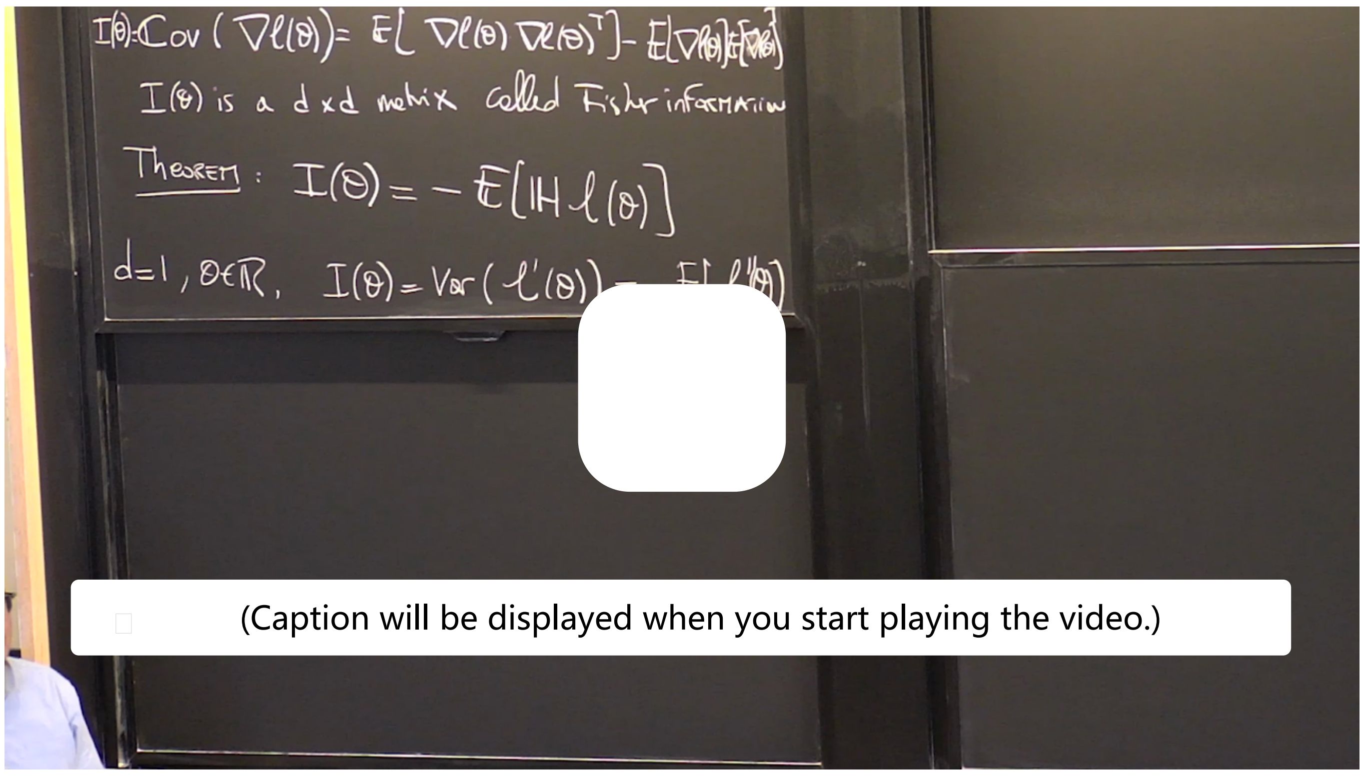
Remark: A convenient way to compute the Fisher information is to use the fourth answer choice, which gives the useful formula

$$\mathcal{I}(\theta) = \int_{-\infty}^{\infty} \frac{\left(\frac{\partial}{\partial\theta} f_\theta(x)\right)^2}{f_\theta(x)}\, dx.$$

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

Proof of Fisher Information Equivalent Formulas for 1 Dimension



Start of transcript. Skip to the end.

So let's just-- again, there's something slightly magical that's happening here.

So let's tell-- let's see for ourselves why this is the case.

And this is the case because--

precisely of the fact that the function we're looking at

is a log-likelihood.

Log of the PDF.

So let's think that x is continuous.

视频
下载视频文件

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

Definition of Fisher Information

Let $\theta \in \Theta \subset \mathbb{R}^d$ and let $\left(E, \{\mathbf{P}_\theta\}_{\theta \in \Theta}\right)$ be a statistical model. Let $f_\theta(\mathbf{x})$ be the pdf of the distribution $\mathbf{P}_\theta$. Then, the Fisher information of the statistical model is

$$\mathcal{I}(\theta) = \mathsf{Cov}\left(\nabla \ell(\theta)\right) = -\mathbb{E}\left[\mathbf{H}\ell(\theta)\right],$$

where $\ell(\theta) = \ln f_\theta(\mathbf{X})$.

The definition when the distribution has a pmf $p_\theta(\mathbf{x})$ is also the same, with the expectation taken with respect to the pmf.

讨论

显示讨论

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