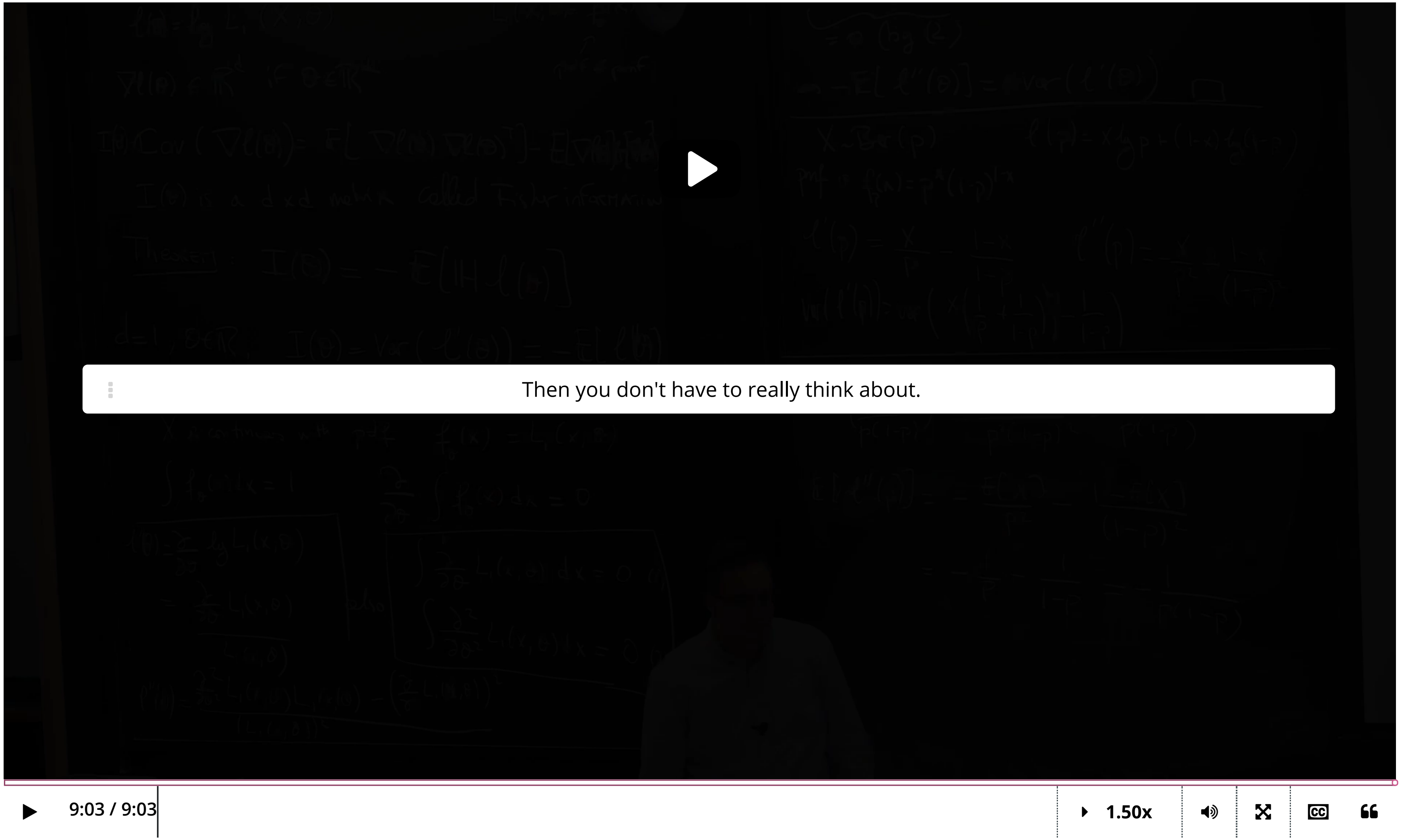
4. Examples of Fisher Information Computation

Fisher Information of the Bernoulli Random Variable



Video

[Download video file](#)

Transcripts

[**Download SubRip (.srt) file**](#)

[**Download Text (.txt) file**](#)

Fisher Information of the Binomial Random Variable

0/1 point (graded)

Let X be distributed according to the binomial distribution of n trials and parameter $p \in (0,1)$. Compute the Fisher information $\mathcal{I}(p)$.

Hint: Follow the methodology presented for the Bernoulli random variable in the above video.

$\mathcal{I}(p)$: (n+p-2*n*p)/p/(1-p)^2 ✖ **Answer:** n/(p*(1-p))

$\frac{n+p-2\cdot n\cdot p}{p\cdot(1-p)^2}$

STANDARD NOTATION

Solution:

The logarithm of the pmf of a binomial random variable X, treated as a random function, can be written as

$$\ell(p) \triangleq \ln\binom{n}{X} + X\ln p + (n-X)\ln(1-p), \quad X \in \{0,1,\ldots,n\}.$$

The derivative of $\ell(p)$ with respect to p is

$$\ell'(p) = \frac{X}{p} - \frac{n-X}{1-p},$$

which means the second derivative is

$$\ell''(p) = -\frac{X}{p^2} - \frac{n-X}{(1-p)^2}.$$

The Fisher information $\mathcal{I}(p)$, therefore, is

$$\mathcal{I}(p) = -\mathbb{E}[\ell''(p)] \;= \mathbb{E}\left[\frac{X}{p^2} + \frac{n-X}{(1-p)^2}\right]$$

$$= \frac{np}{p^2} + \frac{n-np}{(1-p)^2}$$
$$= \frac{n}{p(1-p)}.$$

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Fisher Information of a Bernoulli-Like Random Variable

1/1 point (graded)

Consider the following experiment: You take a coin that lands a head (H) with probability $0 < p < 1$ and you toss it twice. Define X as the following random variable:

$$X = \begin{cases} 1 & \text{if outcome is HH} \\ 0 & \text{otherwise} \end{cases}$$

Compute the Fisher information $\mathcal{I}(p)$.

$\mathcal{I}(p)$: 4/(1-p^2) ✔ **Answer:** 4/(1-p^2)

$\frac{4}{1-p^2}$

STANDARD NOTATION

Solution:

Following the Bernoulli and binomial examples,

$$\ell(p) \triangleq 2X\ln p + (1-X)\ln\left(1-p^2\right), \quad X \in \{0,1\}.$$

The derivative of $\ell(p)$ with respect to p is

$$\ell'(p) = \frac{2X}{p} - 2p \cdot \frac{1-X}{1-p^2},$$

which means the second derivative is

$$\ell''(p) = -\frac{2X}{p^2} - 2 \cdot \frac{(1-X)}{1-p^2} - 4p^2 \cdot \frac{1-X}{(1-p^2)^2}.$$

The Fisher information $\mathcal{I}(p)$, therefore, is

$$\begin{aligned}\mathcal{I}(p) = -\mathbb{E}[\ell''(p)] &= \mathbb{E}\left[\frac{2X}{p^2} + 2 \cdot \frac{(1-X)}{1-p^2} + 4p^2 \cdot \frac{1-X}{(1-p^2)^2}\right] \\ &= \frac{2p^2}{p^2} + \frac{2(1-p^2)}{(1-p^2)} + 4p^2 \cdot \frac{1-p^2}{(1-p^2)^2} \\ &= 4 + \frac{4p^2}{1-p^2} \\ &= \frac{4}{1-p^2}\end{aligned}$$

Submit

You have used 1 of 3 attempts

Answers are displayed within the problem

Fisher Information of a Modified Gaussian Random Vector

4/4 points (graded)

Let $\mathbf{X}$ be a gaussian random vector with **independent** components $X^{(i)} \sim \mathcal{N}(\alpha + \beta t_i, 1)$ for $i = 1, \ldots, d$, where t_i are known constants and α and β are unknown parameters.

Compute the Fisher information matrix $\mathcal{I}(\theta)$ using the formula $\mathcal{I}(\theta) = -\mathbb{E}[\mathbf{H}\ell(\theta)]$.

Use **S_1** for $\sum_{i=1}^{d} t_i$ and **S_2** for $\sum_{i=1}^{d} t_i^2$.

$\mathcal{I}(\theta)_{1,1} =$ d

d

✔ **Answer:** d + 0*S_1 + 0*S_2

$\mathcal{I}(\theta)_{1,2} =$ S_1

S_1

✔ **Answer:** 0*d + S_1 + 0*S_2

$\mathcal{I}(\theta)_{2,1} =$ S_1

S_1

✔ **Answer:** 0*d + S_1 + 0*S_2

$\mathcal{I}(\theta)_{2,2} =$ S_2

S_2

✔ **Answer:** 0*d + 0*S_1 + S_2

Hint: Let $\theta = [\alpha \;\; \beta]^T$ denote the paramaters of the statistical model. $\ell(\theta)$ is a real-valued function of θ as given by the joint pdf at any fixed $\mathbf{x}$.

Solution:

Let $\theta = [\alpha \;\; \beta]^T$ denote the paramaters of the statistical model. The Gaussian random vector $\mathbf{X}$ has the pdf

$$f_\theta(\mathbf{x}) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}\sum_{i=1}^{d}\left(x^{(i)}-\alpha-\beta t_i\right)^2}, \quad \mathbf{x} = \left[x^{(1)} \;\; x^{(2)} \;\; \cdots \;\; x^{(d)}\right]^T \in \mathbb{R}^d,$$

as the variance of each individual component is equal to 1 and the components are independent.

Taking $\ln$ of the pdf yields (written as a random function)

$$\ell(\theta) = -\frac{d}{2}\ln(2\pi) - \frac{1}{2}\left[\sum_{i=1}^{d}\left(\left(X^{(i)} - \beta t_i\right)^2 - 2\alpha\left(X^{(i)} - \beta t_i\right) + \alpha^2\right)\right]$$

Therefore,

$$\nabla\ell(\theta) = \begin{bmatrix} \sum_{i=1}^{d}\left(X^{(i)} - \beta t_i - \alpha\right) \\ \sum_{i=1}^{d}\left(t_i X^{(i)} - \beta t_i^2 - \alpha t_i\right) \end{bmatrix},$$

from which we can obtain the hessian

$$\mathbf{H}\ell(\theta) = \begin{bmatrix} \sum_{i=1}^{d}(-1) & \sum_{i=1}^{d}(-t_i) \\ \sum_{i=1}^{d}(-t_i) & \sum_{i=1}^{d}(-t_i^2) \end{bmatrix}.$$

Therefore,

$$\mathcal{I}(\theta) = -\mathbb{E}\left[\mathbf{H}\ell(\theta)\right] = \begin{bmatrix} d & \sum_{i=1}^{d} t_i \\ \sum_{i=1}^{d} t_i & \sum_{i=1}^{d} t_i^2 \end{bmatrix},$$

where the expectation is taken with respect to the pdf of the random vector $\mathbf{X}$. Since none of the entries of the hessian contained any $X^{(i)}$, the expectation was simply the hessian matrix itself.

Submit You have used 1 of 4 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 3 Methods of Estimation:Lecture 11: Fisher Information, Asymptotic Normality of MLE; Method of Moments / 4. Examples of Fisher Information Computation

Add a Post

All Posts

[STAFF] Was able to get Modified Gaussian Random Vector on first attempt but would appreciate more explanation in solution.

question posted 2 days ago by DriftingWoods

Could you add a few more lines to the beginning of the solution including starting with the generic multivariate Gaussian formula and what happens to each term in there? Even though linear algebra is a recommended prerequisite for the course not everyone has that much exposure with it and knowing explicitly what happened to the determinant of the co-variance matrix and the inverse of the co-variance matrix in the formula could be very helpful to some students.

This post is visible to everyone.

synnfusion
a day ago - marked as answer a day ago by **sudarsanvsr_mit** (Staff)

I didn't use any of that stuff. The components of the random vector are independent so you can get the pdf right away as a product of the pdf's of the components.

This is the correct answer. The covariance matrix is just an identity matrix in this case.

posted a day ago by **sudarsanvsr_mit** (Staff)

I figured out it was the identity matrix and the determinant was 1 so it reduces to the case of the pdfs being multiplied. Just wanted that explicitly stated in the solution so students could see how the general formula simplifies based on the info given.

posted a day ago by **DriftingWoods**

Ok will add this to the solution. Thank you!

posted about 19 hours ago by **sudarsanvsr_mit** (Staff)

Add a comment

Preview

Submit

Learn About Verified Certificates