12. Examples of Maximum Likelihood Estimators

Note: The following problem will be presented in lecture (at the bottom of this page), but we encourage you to attempt it first.

Maximum Likelihood Estimator of a Bernoulli Statistical Model I

3/3 points (graded)

In the next two problems, you will compute the MLE (maximum likelihood estimator) associated to a Bernoulli statistical model.

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}\,(p^*)$ for some unknown $p^* \in (0,1)$. You construct the associated statistical model $\left(\{0,1\}, \{\mathrm{Ber}\,(p)\}_{p\in(0,1)}\right)$. Let L_n denote the likelihood of this statistical model. Recall that in the fourth problem "Likelihood of a Bernoulli Statistical Model" from two slides ago that you derived the formula

$$L_n\left(x_1, \ldots, x_n, p\right) = \prod_{i=1}^{n} p^{x_i}\left(1-p\right)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}\left(1-p\right)^{n-\sum_{i=1}^{n} x_i}.$$

Oftentimes for computating the MLE it is more convenient to work with and optimize the **log-likelihood** $\ell\,(p) := \ln L_n\left(x_1, \ldots, x_n, p\right)$.

The derivative of the log-likelihood can be written

$$\frac{\partial}{\partial p}\ln L_n\left(x_1, \ldots, x_n, p\right) = A/p - \left(n-A\right)/B$$

where A can be expressed in terms of $\sum_{k=1}^{n} x_i$ and B can be expressed in terms of p. Fill in the blanks with the appropriate values for A and B

(Enter **Sigma_n** for entire sum $\sum_{k=1}^{n} x_i$).

$A =$ Sigma_n ☐ **Answer:** Sigma_n

Σ_n

$B =$ 1-p ☐ **Answer:** 1-p

$1 - p$

For which p does $\frac{\partial}{\partial p}\ln L_n\left(x_1, \ldots, x_n, p\right) = 0$? Denote this critical point by $\hat{p}$.

- ○ $\hat{p} = 0$
- ○ $\hat{p} = 1$
- ○ $\hat{p} = \sum_{k=1}^{n} x_i$
- ● $\hat{p} = \frac{1}{n}\sum_{k=1}^{n} x_i$ ☐

STANDARD NOTATION

Solution:

Observe that

$$\begin{aligned}\ln L_n\left(x_1,\ldots,x_n,p\right) &= \ln\left(p^{\sum_{i=1} x_i}\left(1-p\right)^{n-\sum_{i=1} x_i}\right)\\ &= \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n-\sum_{i=1}^{n} x_i\right)\ln\left(1-p\right).\end{aligned}$$

Taking the derivative with respect to p,

$$\frac{\partial}{\partial p}\ln L_n\left(x_1,\ldots,x_n,p\right) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n-\sum_{i=1}^{n} x_i}{1-p}.$$

We set this to be 0 and solve for p:

$$\begin{aligned}\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n-\sum_{i=1}^{n} x_i}{1-p} &= 0 \Leftrightarrow\\ \frac{\left(1-p\right)\sum_{i=1}^{n} x_i - p\left(n-\sum_{i=1}^{n} x_i\right)}{p\left(1-p\right)} &= 0 \Leftrightarrow\\ \frac{\sum_{i=1}^{n} x_i - np}{p\left(1-p\right)} &= 0.\end{aligned}$$

Since the derivative blows up at $p=0,1$, we can assume $0<p<1$ and ignore the denominator for the purpose of solving for p. Hence $\hat{p}=\frac{1}{n}\sum_{i=1}^{n} x_i$ is the unique critical point of the log-likelihood.

提交 你已经尝试了1次（总共可以尝试2次）

☐ Answers are displayed within the problem

Maximum Likelihood Estimator of a Bernoulli Statistical Model: Second Derivative Test

5/5 points (graded)

Setup:

As above, let $X_1,\ldots,X_n \overset{iid}{\sim} \mathrm{Ber}\left(p^*\right)$ for some unknown $p^* \in (0,1)$. You construct the associated statistical model $\left(\{0,1\},\{\mathrm{Ber}\left(p\right)\}_{p\in(0,1)}\right)$. Let L_n denote the likelihood of this statistical model. Recall from a previous problem that

$$L_n\left(x_1,\ldots,x_n,p\right) = \prod_{i=1}^{n} p^{x_i}\left(1-p\right)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}\left(1-p\right)^{n-\sum_{i=1}^{n} x_i}.$$

As stated, it will be more convenient to work with the **log-likelihood** $\ell\left(\theta\right) = \ln L_n\left(x_1,\ldots,x_n,p\right)$.

Question:

Next we will do the second derivative test to see if the critical point $\hat{p}$ obtained from the previous question is a local maximum. The second derivative of the log-likelihood can be written

$$\frac{\partial^2}{\partial p^2}\ln L_n\left(x_1,\ldots,x_n,p\right) = -\frac{C}{p^2} - \frac{n-C}{D}$$

where C depends on $\sum_{i=1}^{n} x_i$ and D depends on p. Fill in the blanks with the correct values of C and D.

(Type **Sigma_n** for the entire sum $\sum_{k=1}^{n} x_i$)

$C =$ Sigma_n ☐ **Answer:** Sigma_n

Σ_n

$D =$ (1-p)^2 ☐ **Answer:** (1-p)^2

$(1-p)^2$

Next we will test the endpoints of our optimization problem. Fill in the blanks with the correct values:
(Note that here we are working with the **likelihood**, *not* the **log-likelihood**)

$L_n(x_1, \ldots, x_n, 0) =$ 0 ☐ **Answer:** 0.0

$L_n(x_1, \ldots, x_n, 1) =$ 0 ☐ **Answer:** 0.0

What is the maximum likelihood estimator (MLE) $\hat{p}_n^{MLE}$ for the true parameter p^*?

○ 0

○ 1

○ $\sum_{i=1}^n X_i$

● $\frac{1}{n}\sum_{i=1}^n X_i$ ☐

Solution:

The second derivative is

$$\frac{\partial}{\partial\theta}\left(\frac{\sum_{i=1}^n X_i}{p} - \frac{n-\sum_{i=1}^n X_i}{1-p}\right) = -\frac{\sum_{i=1}^n x_i}{p^2} - \frac{n-\sum_{i=1}^n x_i}{(1-p)^2}.$$

Since this expression is always negative, this implies that the critical point $\hat{p}$ is a **local maximum**.

Testing the endpoints we see

$$\begin{aligned} L_n(x_1, \ldots, x_n, 0) &= 0^{\sum_{i=1}^n x_i}(1)^{n-\sum_{i=1}^n x_i} = 0 \\ L_n(x_1, \ldots, x_n, 1) &= 1^{\sum_{i=1}^n x_i}(0)^{n-\sum_{i=1}^n x_i} = 0 \end{aligned}$$

Since the likelihood is non-negative, the endpoints are actually **global minima**.

Hence, the global maximum is achieved at $\hat{p} = \frac{1}{n}\sum_{i=1}^n x_i$. Plugging in the random variables $X_1, \ldots, X_n$, we derive the MLE

$$\hat{p}_n^{MLE} = \frac{1}{n}\sum_{i=1}^n X_i$$

which is precisely the **sample mean**.

Remark 1: This problem illustrates the conceptually nice fact that the **maximum likelihood estimator** for a Bernoulli statistical model is the **sample mean**.

Remark 2: Note that for this problem, we derived the maximum likelihood estimator by optimizing $\ln L_n$ treating $x_1, \ldots, x_n$ as abstract variables. At the end, we plugged in our random samples $X_1, \ldots, X_n$. In practice, we would have access to observations $X_1 = x_1, \ldots, X_n = x_n$, and we can simply plug in $x_1, \ldots, x_n$ for the values of $X_1, \ldots, X_n$ in the expression for the MLE to get our estimate of the true parameter.

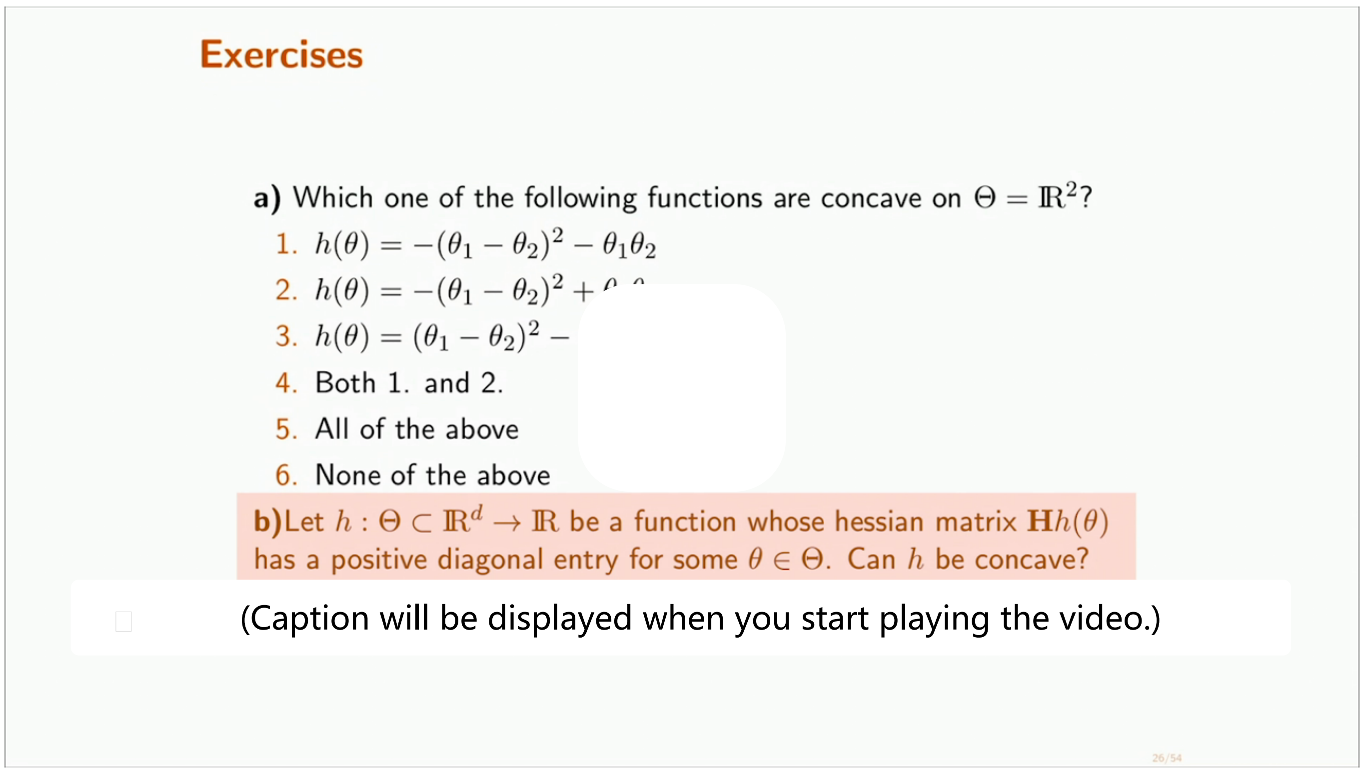
Remark 3: Alternatively, to get the estimate for p^*, we can first plug in the observations $X_1 = x_1, \ldots, X_n = x_n$ and then optimize the log-likelihood $\ln L_n\left(x_1, \ldots, x_n, p\right)$ as a function of p. You will get the same answer either way.

提交 你已经尝试了1次（总共可以尝试3次）

☐ Answers are displayed within the problem

Maximum Likelihood Estimator of Bernoulli Trials

Start of transcript. Skip to the end.



Exercises

a) Which one of the following functions are concave on $\Theta = \mathbb{R}^2$?

1. $h(\theta) = -(\theta_1 - \theta_2)^2 - \theta_1\theta_2$
2. $h(\theta) = -(\theta_1 - \theta_2)^2 +$
3. $h(\theta) = (\theta_1 - \theta_2)^2 -$
4. Both 1. and 2.
5. All of the above
6. None of the above

b) Let $h : \Theta \subset \mathbb{R}^d \to \mathbb{R}$ be a function whose hessian matrix $\mathbf{H}h(\theta)$ has a positive diagonal entry for some $\theta \in \Theta$. Can h be concave?

(Caption will be displayed when you start playing the video.)

So check one thing.

Actually, let's do the b together.

Let h be a function whose hessian matrix has

a positive diagonal entry.

So it's any matrix.

So I don't tell you anything except for the fact

that there is.

视频
下载视频文件

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

讨论

显示讨论

主题： Unit 3 Methods of Estimation:Lecture 9: Introduction to Maximum Likelihood Estimation / 12. Examples of Maximum Likelihood Estimators