semester="1T2019"info_sidebar_name="RelatedLinks"start="2019-02-12T23:59"end="2019-06-06T23:59"display_coursenumber="18.6501x"course_organization="MITx"

18.6501x

May 27, 2019

edX Course: 18.6501x

Contents

Exam 1

Exam 1

Exam Rules

1. You have opened a timed exam with a **48 hours** time limit. Please use the timer to see the time remaining. If you had opened this exam too close to the exam **closing time, April 9 23:59UTC** , you will not have the full 48 hours, and the exam will close at the closing time.

2. This is an **open book exam** and you are allowed to refer back to all course material and use (online) calculators. However, you must abide by the honor code, and **must not ask for answers directly from any aide** .

3. As part of the honor code, you **must not share the exam content** with anyone in any way, i.e. **no posting of exam content anywhere on the internet** . Violators will be removed from the course.

4. You will be given **no feedback** during the exam. This means that unlike in the problem sets, you will not be shown whether any of your answers are correct or not. This is to test your understanding, to prevent cheating, and to encourage you to try your very best before submitting. Solutions will be available after the exam closes.

5. You will be given **3 attempts** for each (multipart) problem. Since you will be given no feedback, the extra attempts will be useful only in case you hit the "submit" button in a haste and wish to reconsider. **With no exception, your last submission will be the one that counts** . **DO NOT FORGET TO SUBMIT** your answers to each question. The "end your exam" button will not submit answers for you.

6. The exam will only be **graded 1 day after the due date** , and the **Progress Page will show fake scores while the exam is open** .

7. **Error and bug reports:** While the exam is open, you are **not allowed to post on the discussion forum on anything related to the exam, except to report bugs/platform difficulties** . If you think you have found a bug, please **state on the forum only what needs to be checked on the forum** . You can still post questions relating to course material, but **the post must not comment on the exam** , and in particular **must not shed any light on the contents or concepts in the exam** . **Violators will receive a failing grade or a grade reduction in this exam** .

8. **Clarification:** If you need clarification on a problem, please first **check the discussion forum** , where staff may have posted notes. After that, if you still need clarification that will **strictly not lead to hints of the solution** , you can email staff at 186501exam@mit.edu. If we see that the issue is indeed not addressed already on the forum, we will respond within 28 hours and post a note on the forum; otherwise– if the issue has been addressed on the forum, we will **not** respond and assume your responsibility to check the forum for answers.

Problem 1: Multiple Choice

Instructions :

Be very careful with the multiple choice questions below. Some are choose all that apply and some test your knowledge on the definitions of terms.

As in the rest of this exam, only your last submission will count.

(a)

All maximum likelihood estimators are asymptotically normal.

Solution:
The maximum likelihood estimator is asymptotically normal only if certain technical conditions are satisfied. One technical condition that is often violated is that the support of the distributions depend on the parameters being estimated; e.g. estimating θ when $X_1, \ldots, X_n$ come from a $\text{Unif}[0, \theta]$.

(b)

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables with some unknown parameter $p \in (0, 1)$. Then which of the following is/are valid confidence interval(s) for p with **nonasymptotic** confidence level 95%?
(Choose all that apply.)

"$[\bar{X}_n - 1.96\sqrt{\frac{p(1-p)}{n}}, \bar{X}_n + 1.96\sqrt{\frac{p(1-p)}{n}}]$"
"$(0, 1)$"
"$[0, \bar{X}_n + \frac{.83}{n}]$"
"None of the Above"

Correction Note: An earlier version of this problem contained an ambiguity and asked for confidence intervals for p with nonasymptotic level 5%, without stating whether the level refers to a signifance level or confidence level.

Solution:
We consider each of the possible choices in turn.

- The first choice is a "confidence interval" for p that depends on p and thus is clearly invalid.
- The second choice is a 100% confidence interval as p is always between 0 and 1 and is therefore is also a 95% confidence interval. (Note that a constant interval could also be viewed as a random interval, just as a constant can be viewed as a random variable.)
- Finally, the third choice is invalid. For small n (e.g. $n = 1$), it has (maximal) level p, which makes this a 95% nonasymptotic interval if $p > 0.95$ (p should be in the CI with probability 0.95 for all p). It is not even a valid 95% nonasymptotic CI, since it has the wrong dependence on n, i.e. $\frac{.83}{n}$ rather than $\frac{.83}{\sqrt{n}}$. This makes the interval too narrow to be a valid confidence interval for p with asymptotic or nonasymptotic level 95%.

(c)

Which of the following is/are valid statistical model(s)?
(Choose all that apply.)

"$\left(\mathbb{R}, \{\mathcal{N}(\theta,1)\}_{\theta>10}\right)$"
"$\left([\theta,\infty), \{t \mapsto e^{\theta-t}\mathbf{1}(t>\theta)\}_{\theta>0}\right)$"
"$\left([0,\infty), \{\mathcal{N}(\mu,\sigma^2)\}_{\mu>\sigma^2}\right)$","$\left([0,\infty), \{x \mapsto e^{\theta x}\mathbf{1}(x>0)\}_{\theta>0}\right)$

Solution:
We consider each of the possible choices in turn.

- The first choice is a valid statistical model.
- The second choice is invalid as the sample space of the random variables depends on the parameter of interest θ.
- The third choice is invalid as a normal distribution has support over $\mathbb{R}$ and not the subset $[0,\infty)$. (Do not confuse restricting the parameter space Θ with incorrectly restricting the support, or sample space, E of the family of distributions.)
- Finally, the fourth choice is invalid as $x \mapsto e^{\theta x}\mathbf{1}(x>0)\}_{\theta>0}$ does not define a probability distribution.

Problem 2

Setup :

Let $X_1, \ldots, X_n$ be i.i.d. random variable with pdf f_θ defined as follows:

$$f_\theta(x) = \theta x^{\theta-1}\,\mathbf{1}(0 \le x \le 1)$$

where θ is some positive number.

(a)

Is the parameter θ identifiable?
"Yes"
"No"

Solution:
Yes it is identifiable. If $\theta_1 \neq \theta_2$, then the pdfs $f_{\theta_1}(x) \neq f_{\theta_2}(x)$.

(b)

Compute the maximum likelihood estimator $\widehat{\theta}$ of θ.

Maximum likelihood estimator $\widehat{\theta} =$

Solution:
The likelihood of $X_1, \ldots, X_n$ given a parameter θ is

$$L(X_1, \ldots, X_n; \theta). = \theta^n \prod_{i=1}^{n} X_i^{\theta-1}.$$

Taking the logarithm we find that log-likelihood

$$\ell_n(\theta) = n\ln(\theta) + (\theta - 1)\sum_{i=1}^{n} \ln(X_i).$$

Setting $\ell'(\theta) = 0$ we find that

$$\widehat{\theta} = \frac{-n}{\sum_{i=1}^{n} \ln X_i}.$$

This is the unique maximum as

$$\ell_n''(\theta) = \frac{-n}{\theta^2} < 0.$$

(c)

Compute the Fisher information.

$I(\theta) =$

Solution:
By definition, the Fisher information is defined as

$$I(\theta) = -\mathbb{E}[\ell''(X;\theta)]$$

where $\ell(\theta) = \ln L(X;\theta)$ is the log-likelihood defined using a sample of size 1.
The likelihood of X given a parameter θ is

$$L(X;\theta) = \theta(X^{\theta-1}).$$

Taking the logarithm we find that log-likelihood

$$\ell(\theta) = \ln(\theta) + (\theta-1)\ln(X_i).$$

Taking the second derivative we find that

$$\ell''(\theta) = \frac{-1}{\theta^2}$$

and therefore we have that

$$I(\theta) = -\mathbb{E}[\ell''(X;\theta)] = \frac{1}{\theta^2}.$$

(d)

What kind of distribution does the distribution of $\sqrt{n}\widehat{\theta}$ approach as n grows large? "Bernoulli"
"Poisson"
"Normal"
"Exponential"

Solution:
The theorem for MLE applies in this example as the following conditions hold:

- θ is identifiable
- $I(\theta)$ is invertible
- Support of f_θ does not depend on θ

Hence $\widehat{\theta}$ is asymptotically normal:

$$\sqrt{n}\left(\widehat{\theta}-\theta\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, I^{-1}(\theta)).$$

This means that as n grows large,

$$\widehat{\theta} \overset{\text{approx}}{\sim} \mathcal{N}\left(\theta, \frac{I^{-1}(\theta)}{n}\right)$$

and hence $\sqrt{n}\widehat{\theta}$ is also approximately normal.

(e)

What is the asymptotic variance $V(\widehat{\theta})$ of $\widehat{\theta}$?

To avoid double jeopardy, you may use I for the Fisher information $I(\theta)$ evaluated at $\theta,$ or you may enter your answer without using I.

$V(\widehat{\theta}) =$

Solution:
By the theorem for the MLE the asymptotic variance of the estimator is $I(\theta)^{-1} = \frac{1}{\theta^2}$.

(f)

Using the MLE $\widehat{\theta}$, find the shortest confidence interval for θ with asymptotic level 85% using the plug-in method.

$\mathcal{I}_{\text{plug-in}} = [A, B]$ where $A =$

$B =$

Solution:
Using the previous question on the asymptotic normality of the MLE it follows that

$$\lim_{n\to\infty} \mathbb{P}[\widehat{\theta} \in [\theta - 1.44\frac{\theta}{\sqrt{n}}, \theta + 1.44\frac{\theta}{\sqrt{n}}]] = .85.$$

Therefore it follows that

$$\lim_{n\to\infty} \mathbb{P}[\theta \in [\widehat{\theta} - 1.44\frac{\theta}{\sqrt{n}}, \widehat{\theta} + 1.44\frac{\theta}{\sqrt{n}}]] = .85$$

and since $\widehat{\theta}$ approaches θ almost surely we get the confidence interval

$$[\widehat{\theta} - 1.44\frac{\widehat{\theta}}{\sqrt{n}}, \widehat{\theta} + 1.44\frac{\widehat{\theta}}{\sqrt{n}}]$$

via the plug-in method.

Problem 3

Setup:

As on the previous page, let $X_1, \ldots, X_n$ be i.i.d. with pdf

$$f_\theta(x) = \theta x^{\theta-1}\mathbf{1}(0 \le x \le 1)$$

where $\theta > 0$.

(a)

Assume we do not actually get to observe $X_1, \ldots, X_n$. Instead let $Y_1, \ldots, Y_n$ be our observations where $Y_i = \mathbf{1}(X_i \le 0.5)$. Our goal is to estimate θ based on this new data.

What distribution does Y_i follow?

First, choose the type of the distribution: "Bernoulli"
"Poisson"
"Normal"
"Exponential"

Second, enter the parameter of this distribution in terms of θ. Denote this parameter by m_θ. (If the distribution is normal, enter only 1 parameter, the mean).

$m_\theta =$

Solution:
Note that Y is distributed over only two values and therefore is a distributed as a Bernoulli random variable. The parameter of the Bernoulli is

$$\mathbb{P}[Y = 1] = \int_0^{1/2} \theta x^{\theta-1} dx = \frac{1}{2^\theta}$$

(b)

Write down a statistical model associated to this experiment. Is the parameter θ identifiable?

"Yes"
"No"

Solution:
Yes it is identifiable since $2^{-\theta}$, which is the parameter of the Bernoulli, is an injective function of θ.

(c)

Compute the Fisher information $I(\theta)$.
(To answer this question correctly, your answer to part (a) needs to be correct.)

$I(\theta) =$

Solution:
The log likelihood of an observation Y is

$$\begin{aligned}\ell(Y;\theta) &= \mathbf{1}_{Y=0}\ln(1-2^{-\theta}) + \mathbf{1}_{Y=1}\ln(2^{-\theta})\\ &= \mathbf{1}_{Y=0}\ln(1-2^{-\theta}) - \theta\mathbf{1}_{Y=1}\ln(2).\end{aligned}$$

Taking the second derivative one finds that

$$\ell''(Y;\theta) = \mathbf{1}_{Y=0}\frac{2^{\theta}(\ln(2))^2}{(2^{\theta}-1)^2}$$

and therefore the Fisher Information is

$$I(\theta) = \mathbb{E}[-\ell''(Y;\theta)] = \frac{(\ln(2))^2}{2^{\theta}-1}.$$

(Note that by definition, the Fisher Information does not depend on n.)

(d)

Compute the maximum likelihood estimator $\widehat{\theta}$ for θ in terms of $\overline{Y_n}$.

$\widehat{\theta} =$

Solution:
Let n_0 and n_1 denote the number of 0's and 1's among $Y_1, \ldots, Y_n$. The log-likelihood of this observation is then

$$\ell(\theta) = n_0\ln(1-2^{-\theta}) - n_1\theta\ln(2).$$

The MLE $\widehat{\theta}$ satisfies

$$\ell'(\widehat{\theta}) = 0$$

which is equivalent to

$$\ell'(\widehat{\theta}) = \frac{n_0\ln(2)2^{-\theta}}{1-2^{-\theta}} - n_1\ln(2).$$

Rearranging and solving for $\widehat{\theta}$ it follows that

$$\widehat{\theta} = \frac{-\ln(\frac{n_1}{n_0+n_1})}{\ln(2)} = \frac{-\ln\bar{Y}_n}{\ln(2)}.$$

Note that $\ell''(\theta) < 0$ so this is the unique maximum.

(e)

Compute the method of moments estimator $\widetilde{\theta}$ for θ.

$\widetilde{\theta} =$

Solution:
Note trivially that

$$\mathbb{E}[Y_i] = 2^{-\theta}$$

and therefore $2^{-\widetilde{\theta}} = \overline{Y_n}$. Thus

$$\widetilde{\theta} = \frac{-\ln(\overline{Y_n})}{\ln 2}.$$

(f)

What is the asymptotic variance $V(\tilde{\theta})$ of the method of moments estimator $\tilde{\theta}$?

$V(\widetilde{\theta}) =$

Solution:
Note the the method of moments estimator and the MLE estimator are the same! Thus we can use the Theorem on MLE to determine that the asymptotic variance is

$$I(\theta)^{-1} = \frac{2^{\theta} - 1}{(\ln(2))^2}.$$

(g)

Give a **formula** for the p-value for the test of

$$H_0 : \theta \leq 1 \quad \text{vs.} \quad H_1 : \theta > 1$$

based on the asymptotic distribution of $\widehat{\theta}$.

To avoid double jeopardy, you may use V for the asymptotic variance $V(\theta_0)$, I for the Fisher information $I(\theta_0)$, **hattheta** for $\widehat{\theta}$, or enter your answer directy without using V or I or **hattheta**.

p-value:

Assume $n = 50$, and $\overline{Y_n} = 0.46$. Will you reject the null hypothesis at level $\alpha = 5\%$?

"Yes, reject the null hypothesis at level $\alpha = 5\%$."
"No, cannot reject the null hypothesis at level $\alpha = 5\%$."

Correction Note: In an earlier version of this problem, the input instruction was: " To avoid double jeopardy, you may use V for the appropriate estimator of the asymptotic variance $V(\widehat{\theta})$ of the MLE $\hat{\hat{\theta}}$, I for the Fisher information $I(\widehat{\theta})$ evaluated at $\hat{\theta}$, **hattheta** for $\widehat{\theta}$, or enter your answer directy without using V or I or **hattheta**."

Solution:
Define the test statistic for this one-sided test as

$$T_n = \sqrt{\frac{n}{V}}\left(\hat{\theta}-1\right)$$

where $V = V(1)$ is the asymptotic variance evaluated at the boundary of the null hypothesis. Recall the p-value is the smallest level at which this test will reject H_0. Hence

$$p = 1 - \Phi\left(\sqrt{\frac{n}{V(1)}}\left(\hat{\theta}-1\right)\right)$$

If $\bar{Y_n} = .46$ the MLE is

$$\widehat{\theta} = \frac{-\ln(.46)}{\ln(2)} = 1.12.$$

The asymptotic variance of $\bar{\theta}$ given $\theta = 1$ is

$$V(1) = \frac{2^{\theta}-1}{(\ln(2))^2} = 2.08.$$

Therefore the desired p value is

$$\begin{aligned} p &= \mathbb{P}(Z \geq \sqrt{n}(2.08)^{\frac{-1}{2}}(\bar{\theta}-1)) \\ &= 1 - \mathbb{P}(Z \geq .59) \\ &= .2776 \end{aligned}$$

where Z is distributed as an $\mathcal{N}(0,1)$. Since $p > 0.05$, we fail to reject the null hypothesis at level $\alpha = 5\%$.

Problem 4

Let $X_1, \ldots, X_n$ be i.i.d. normal variable following the distribution $\mathcal{N}(\mu, \tau)$, where μ is the mean and τ is the variance.

Denote by $\widehat{\mu}$ and $\widehat{\tau}$ the maximum likelihood estimators of μ and τ respectively based on the i.i.d. observations $X_1, \ldots, X_n$.

(In our usual notation, $\tau = \sigma^2$. We use τ in this problem to make clear that the parameter being estimated is σ^2 not σ.)

(a)

Is the estimator $2(\widehat{\mu})^2 + \widehat{\tau}$ of $2\mu^2 + \tau$ asymptotically normal?

"yes"
"no"
"not enough information to determine"

Solution:
Let

$$g(\widehat{\mu}, \widehat{\tau}) = 2(\widehat{\mu})^2 + \widehat{\tau}$$

Since g is continuously differentiable, the delta method gives

$$\sqrt{n}\,(g(\widehat{\mu}, \widehat{\tau}) - g(\mu, \tau)) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \nabla g(\mu, \tau)^T (\mathbf{I}(\mu, \tau))^{-1} \nabla g(\mu, \tau)^T.$$

where $\mathbf{I}(\mu, \tau)$ is the Fisher Information matrix of any of the Gaussian variables X_i.

(b)

Let

$$g(\mu, \tau) = 2\mu^2 + \tau.$$

and let $\mathbf{I}$ be the Fisher information matrix of $X_i \sim \mathcal{N}(\mu, \tau)$.

The asymptotic variance of $2(\widehat{\mu})^2 + \widehat{\tau}$ is...

"$\nabla g(\mu, \tau)^T \mathbf{I}(\mu, \tau) \nabla g(\mu, \tau)$"
"$\nabla g(\mu, \tau)^T \left(\mathbf{I}(\mu, \tau)\right)^{-1} \nabla g(\mu, \tau)$"
"$\nabla g(\mu, \tau)^T \mathbf{I}(\mu, \tau)$"
"$\nabla g(\mu, \tau)^T \left(\mathbf{I}(\mu, \tau)\right)^{-1}$"

Solution:
Refer to the solution to the previous problem.

(c)

Using the results from above and referring back to homeowork solutions if necessary, compute the asymptotic variance $V(2(\widehat{\mu})^2 + \widehat{\tau})$ of the estimator $2(\widehat{\mu})^2 + \widehat{\tau}$.

Hint: The inverse of a diagonal matrix $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ where $a, b \neq 0$ is the diagonal matrix $\begin{pmatrix} 1/a & 0 \\ 0 & 1/b \end{pmatrix}$.

$V(2(\widehat{\mu})^2 + \widehat{\tau})\ =$

Solution:
Recall from homework 6 problem 2 that the Fisher Information of a Gaussian distribution $\mathcal{N}(\mu, \tau)$ where the μ and $\tau = \sigma^2$ are the parameters to be estimated is

$$I(\mu, \tau) = \begin{pmatrix} \frac{1}{\tau} & 0 \\ 0 & \frac{1}{2\tau^2} \end{pmatrix}.$$

Using this and the results from the previous parts, we obtain the asymptotic variance $V(2(\widehat{\mu})^2 + \widehat{\tau})$ as

$$\begin{aligned} V(2(\widehat{\mu})^2 + \widehat{\tau}) &= \nabla g(\mu, \tau)^T \left(\mathbf{I}(\mu, \tau)\right)^{-1} \nabla g(\mu, \tau) \\ &= \begin{pmatrix} 4\mu & 1 \end{pmatrix} \begin{pmatrix} \tau & 0 \\ 0 & 2\tau^2 \end{pmatrix} \begin{pmatrix} 4\mu \\ 1 \end{pmatrix} \\ &= 16\mu^2\tau + 2\tau^2. \end{aligned}$$