semester="3T2019"info_sidebar_name="RelatedLinks"start="2019-09-03T23:
59"end="2019-06-06T23:59"display_coursenumber="18.6501x"course_organization="MITx"

18.6501x

December 22, 2019

edX Course: 18.6501x

Contents

Midterm Exam 2

Midterm Exam 2

edXvertical: 1. Rules
edXtext: Rules

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 **STRICT closing time, November 27 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. While we expect you to be able to do all problems yourselves, you are allowed to use (online) calculators to computer integrals, derivatives, or do algebraic manipulations. However, you must abide by the honor code, and **must not ask for answers directly from any aide** . This means you may not refer to discussion forums from previous runs of this course, or ask for help from others.

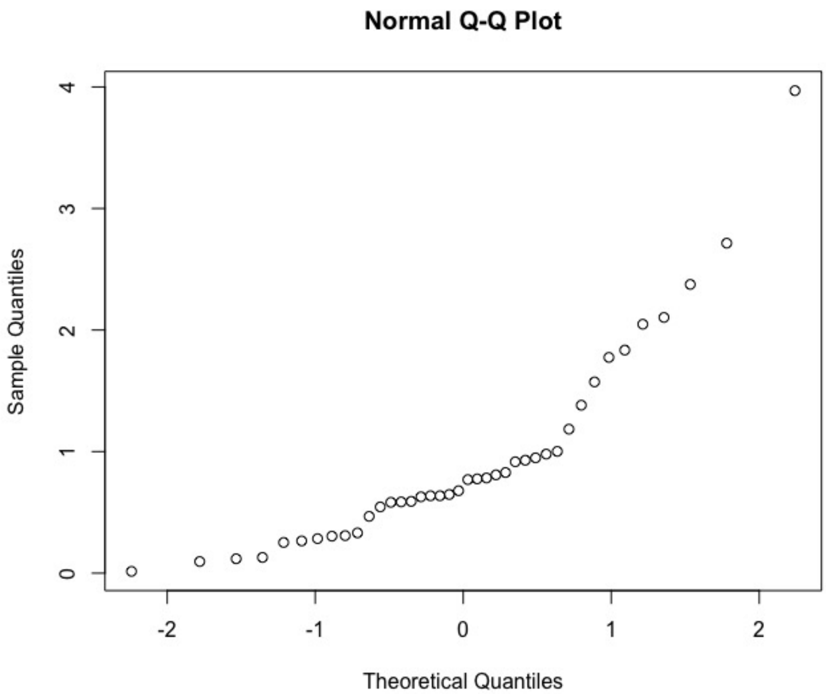
3. You **must not share the exam content** with anyone in any way, including posting anywhere on the internet.

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

6. **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. Your post must not shed any light on the contents or concepts in the exam. **Violators will receive a failing grade or grade reduction in this exam** .

7. **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.

edXvertical: 2. Multiple Choice
edXproblem: QQ-Plot

Consider an i.i.d. sample $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} P$ that has been reordered as $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. In the image below, we compare the empirical quantiles (from the data) to the standard Gaussian quantiles using a QQ-plot.

Does the distribution underlying the data have a heavier or lighter **right** tail than a Gaussian distribution?

Answer choices:
heavier
lighter

Which of the following best represents the support of the distribution underlying the data?

Answer choices:
$(0,1)$
$(-2,2)$
$(1,4)$
$[0,\infty)$
$\mathbb{R}$

Solution:

- Since the empirical quantiles on the right are larger than the theoretical quantiles, the right tail of the distribution underlying the data must be **heavier** than that of the standard Gaussian.
- The plotted data points take values between 0 and 4. Also from the plot, the empirical quantiles $q_{sample}(1-\alpha)$ (recall $P(X \le q_{sample}(1-\alpha) = (1-\alpha))$ for $1-\alpha \sim 0$ are close to 0. Hence, the most reasonable guess for the range of data would be $[0,\infty)$.

(We could guess that data follows an exponential distribution, since it satisfies both conditions above.)

edXproblem: Multivariate Gaussian

Let X and Y be two independent random variables with distribution $\mathcal{N}(0,1)$. Which of the following statements are correct? (Choose all that apply.)

Answer Choices:
(X,Y) is a Gaussian vector, i.e. a multivariate Gaussian variable
$(X+Y)$ and $(X-Y)$ are independent
$\mathsf{Cov}(2X+Y,3X)=0$
$(X,Y)\sim\mathcal{N}_2(0,I_2)$

Grading Note: Partial credit is given.

Solution:

Let us start with the correct choices:

- Since X and Y are independent, (X,Y) is a 2-dimensional Gaussian vector.
- Moreover, since $\mathsf{Cov}(X,Y)=I_2$, we have $(X,Y)\sim\mathcal{N}_2(0,I_2)$.

- To see that $X+Y, X-Y$ are independent, we change variables using $Z_+ = X+Y$ and $Z_- = X-Y$, and see that $f_{(Z_+,Z_-)}(z_+,z_-) = f_{Z_+}(z_+)f_{Z_-}(z_-)$. Recall the (joint) pdf $f_{(X,Y)}(x,y)$

$$f_{(X,Y)}(X,Y) = \qquad \frac{1}{\sqrt{(2\pi}} \exp\left(-\frac{1}{2}\begin{pmatrix} x & y \end{pmatrix} I_2 \begin{pmatrix} x \\ y \end{pmatrix}\right)$$

where the covariance matrix (and its inverse) has been simplied to the identity matrix I_2 using $\sigma_x^2 = \sigma_y^2 = 1$. We write the change of variables in vector form:

$$\begin{pmatrix} z_+ \\ z_- \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \Longleftrightarrow \qquad \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} z_+ \\ z_- \end{pmatrix}.$$

This gives the joint pdf $f_{(Z_+,Z_-)}(z_+,z_-)$ to be

$$\begin{aligned} f_{(Z_+,Z_-)}(z_+,z_-) = &\quad C\exp\left[-\frac{1}{2}\begin{pmatrix} z_+ & z_- \end{pmatrix}\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} I_2 \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} z_+ \\ z_- \end{pmatrix}\right] \\ = &\quad C\exp\left[-\frac{1}{2}\begin{pmatrix} z_+ & z_- \end{pmatrix}\frac{1}{2} I_2 \begin{pmatrix} z_+ \\ z_- \end{pmatrix}\right] \\ = &\quad C\exp\left[-\frac{z_+}{4}\right]\exp\left[-\frac{z_-}{4}\right] \\ = &\quad f_{Z_+}(z_+)f_{Z_-}(z_-) \end{aligned}$$

where C is an overall normalization constant that does not depend on the variables z_+ and z_-(since the change of variable is linear). Hence $Z_+ = X+Y$ and $Z_- = X-Y$ are 2 independent 1-d Gaussian variables.

<div>
</div>Let us examine the wrong choice: Since $\mathsf{Cov}(X,Y)=0$, we have $\mathsf{Cov}(2X+Y,3X) = \mathsf{Cov}(2X,3X) = 6\mathsf{Var}(X) \neq 0$.

edXproblem: Median

Let X be a random variable with pdf $f(x) = \dfrac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$. The median of X minimizes which of the following functions?

(The notation $m \mapsto m^2$ denotes the function that takes input m and gives output m^2, and similarly for other output functions.)

Answer Choices:
$m \mapsto \mathbb{E}(X-m)^2$
$m \mapsto \mathbb{E}|X-m|$

$m \mapsto m^4$
$m \mapsto \mathsf{Var}(X - m)$

Grading Note: All answers are graded as correct.

Solution:

We will go through the choices in order.

- The notation $\mathbb{E}(X-m)^2$ was not clear and could be interpreted as $\mathbb{E}\left[(X-m)^2\right]$ or $(\mathbb{E}[X-m])^2$. We will go through both.
  1. $\mathbb{E}\left[(X-m)^2\right]$: For general continuous r.v. X, the minimizer of $m \to \mathbb{E}\left[(X-m)^2\right]$ is $m = \mathbb{E}[X]$. In this problem, $\mathbb{E}[X] = \text{median}(X) = 0$, $\text{median}(X) = 0$ also minimizes $m \to \mathbb{E}\left[(X-m)^2\right]$.
  2. $(\mathbb{E}[X-m])^2$: The minimizer of $m \to (\mathbb{E}[X-m])^2$ is again $m = \mathbb{E}[X]$. As above, $\mathbb{E}[X] = \text{median}(X) = 0$ implies $\text{median}(X) = 0$ also minimizes $m \to (\mathbb{E}[X-m])^2$.
- **For general continuous random variable X, the median is minimizes $m \mapsto \mathbb{E}|X-m|$.**
- The minimizer of $m \to m^4$ is $m = 0$. Since $\text{median}(X) = 0$ for the given X, this is also correct.
- $\mathsf{Var}(X-m) = \mathsf{Var}(X)$ is constant for all m. Hence any value of m, in particular $m = \text{median}(X) = 0$, is a minimizer of $m \to \mathsf{Var}(X-m)$.

edXvertical: 3.
edXproblem: Two sample t-statistic

Let

- $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu_0, \sigma_0^2)$
- $Y_1, \ldots, Y_m \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$
- the X_i's are independent of the Y_i's,

where $\mu_0, \mu_1, \sigma_0^2, \sigma_1^2$ are **unknown** parameters.

What is the two-sample t-statistic $T_{n,m}$?

Answer Choices:

$\left(\overline{X}_n - \overline{Y}_m\right)\sqrt{(\widehat{\sigma_1}^2 + \widehat{\sigma_0}^2)/(m+n)}$
$\dfrac{\overline{X}_n - \overline{Y}_m}{\sqrt{\widehat{\sigma_1^2}/m + \widehat{\sigma_0^2}/n}}$
$\dfrac{\overline{X}_n - \overline{Y}_m}{\sqrt{\sigma_1^2/m + \sigma_0^2/n}}$
$\dfrac{\overline{X}_n - \overline{Y}_m}{\sqrt{\sigma_1^2 + \sigma_0^2/(m+n)}}$
The t-statistic cannot be defined for two samples.

edXproblem: Distribution of t-statistic with Small Sample Size

Continuing with the setup as above, suppose the data you collected are summarized as follows:

- Sample sizes: $n = m = 5$;
- Sample means: $\overline{X}_n = 1.2$, $\overline{Y}_m = 1.0$;
- Sample variances: $\widehat{\sigma_1^2} = \widehat{\sigma_2^2} = 0.5$.

Which of the following distributions **best approximates** the distribution of the t-statistic $T_{n=5,m=5}$ under the null hypothesis $H_0 : \mu_0 = \mu_1$?

Answer Choices:

Normal distribution with mean $\mu \neq 0$
Standard normal distribution
t-distribution
F-distribution

If applicable, specify the degrees of freedom of the distribution above (i.e. the distribution that **best approximates** the distribution of the t-statistic). If not applicable, enter -1.

Degrees of freedom: **Assessment content removed**

Solution:

Since the sample sizes are small, the two-sample T-statistic $T_{5,5} = \dfrac{\overline{X}_n - \overline{Y}_m}{\sqrt{\widehat{\sigma_1^2}/m + \widehat{\sigma_0^2}/n}}$ is approximated by a t-distribution with degree of freedom given by the Welsh-Satterthwaite formula.

A sanity check is that the degree of freedom should be at least $\min m, n = 5$.

edXvertical: 4.
edXtext: Problem setup

Setup:

Suppose you have observations X_1, X_2, X_3, X_4, X_5 which are i.i.d. draws from a Gaussian distribution with unknown mean μ and unknown variance σ^2.

Given Facts:

You are given the following:

$$\frac{1}{5}\sum_{i=1}^{5} X_i = 0.90, \qquad \frac{1}{5}\sum_{i=1}^{5} X_i^2 = 1.31$$

edXproblem: Choose a test

To test the null hypothesis $H_0 : \mu = 0$ versus the alternative hypothesis $H_1 : \mu \neq 0$ using the data above, which of the following test(s) is appropriate?

(Choose all that apply.)

Answer Choices:
t-test
Z-test: i.e. the test based on the central limit theorem
Wald's test

Grading Note: Partial credit is given.

Solution:

The rule of thumb in this course is to require 30 or more observations in order to obtain accurate asymptotic approximations. Since the sample size is small, asymptotic tests may not give accurate results, hence t-test is the only appropriate test.

edXproblem: Unbiased Sample Variance

Compute the **unbiased sample variance** S.

(Enter a numerical answer accurate to at least 2 decimal places.)

$S =$ **Assessment content removed**

Solution:

Recall the unbiased sample variance is

$$\begin{aligned} S &= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) \\ &= \frac{n}{n-1}\left(\overline{X_n^2}-\overline{X_n}^2\right) \\ &= \frac{5}{4}(1.31-0.9^2) \;=\; 0.625. \end{aligned}$$

(Without the 5/4 factor, the biased variance is ~ 0.5.)

edXproblem: MLE

Give an estimate the mean μ and variance σ^2. Use the maximum likelihood estimator.

$\hat{\mu} =$ **Assessment content removed**

$\hat{\sigma}^2 =$ **Assessment content removed**

Solution:

The MLE in the Gaussian model is the sample mean and sample variance:

$$\begin{aligned} \overline{X_5} &= 0.9 \\ S_5 &= \overline{X_5^2}-(\overline{X_5})^2 \;=\; 1.31-0.9^2 = 0.5. \end{aligned}$$

edXproblem: T test

Find the value of the t-statistic for testing the hypotheses above:

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu \neq 0$$

given this set of data.

(Enter a numerical answer accurate to at least 2 decimal places.)

t$-$ statistic: **Assessment content removed**

If we allow 5% of samples to wrongly reject H_0 when H_0 is in fact true, what can we conclude from the t-test?

Answer Choices:
reject H_0
accept H_0
fail to reject H_0

Solution:

Recall that the t-distribution with d degrees of freedom is the law of a ratio

$$\frac{Z}{\sqrt{V/d}} \quad \text{where} \quad Z \sim \mathcal{N}(0,1) \text{ indep } V \sim \chi_d^2.$$

The t-statistic is the sample mean scaled by the square root of the sample size, devided by the sample variance, where the variance must be the unbiased estimator with $\frac{1}{n-1}$ normalization, and the t statistics has $d = n-1$ degrees of freedom. Hence, in this problem, the t-statistic is

$$T_n = \frac{\sqrt{n}\overline{X}_n}{\sqrt{S_n^{\text{unbiased}}}} = \frac{\sqrt{5}0.9}{\sqrt{0.625}} = 2.5456.$$

The rejection region for this two-sided test with confidence level $\alpha = 0.05$ is

$$\psi_\alpha = \mathbf{1}\left(|T_5| > q_{\alpha/2}\right) = \mathbf{1}\left(|T_5| > 2.78\right)$$

where $q_{\alpha/2} = 2.78$ is the $(1-\alpha/2)$ quantile of t_4 distribution. We fail to reject the null hypothesis in this sample.

edXproblem: Confidence interval

Provide a non-asymptotic confidence interval $[A, B]$ for μ that is symmetric around the sample mean and covers the true mean in 95% of samples.

(To avoid double jeopardy, enter your answer in terms of the (unbiased) sample variance S from above, or directly enter numerical answers accurate to at least 2 decimal places.)

(Be careful that $A < B$.)

Lower bound $A =$ **Assessment content removed**

Upper bound $B =$ **Assessment content removed**

Solution:

The corresponding confidence interval for μ with coverage probability 0.95 and level $\alpha = 1 - 0.95$ is

$$\left[\overline{X_5} - q_{\alpha/2}\sqrt{\frac{S_5^{\text{unbiased}}}{5}}, \overline{X_5} + q_{\alpha/2}\sqrt{\frac{S_5^{\text{unbiased}}}{5}}\right] = [-0.0816, 1.8816].$$

edXvertical: 5.
edXtext: Setup

Wald's Test and Likelihood Ratio test

Consider a sample of i.i.d. random variables $X_1, \ldots, X_n$ and assume their common density is given by

$$f_\theta(x) = \frac{x}{\theta}\exp\left(-\frac{x^2}{2\theta}\right)\mathbf{1}(x \geq 0),$$

where $\theta > 0$ is an unknown parameter

Consider the following set of hypotheses:

$$H_0 : \theta = 1 \quad \text{and} \quad H_1 : \theta \neq 1.$$

You will perform Wald's test and the likelihood ratio test at significance level 7% for these hypotheses.

edXproblem: Maximum Likelihood Estimator

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

$\hat{\theta} =$ **Assessment content removed**

edXinclude: ../xml/standardnotation.xml

Solution:

The given pdf f_X is not one of the familiar families

$$\begin{aligned} l_n(X_1, \ldots, X_n; \theta) &= \sum_i \ln(X_i) - n\ln\theta - \frac{1}{2\theta}\sum_i X_i^2 \qquad (x \geq 0) \\ \frac{dl_n(\theta)}{d\theta} &= -\frac{n}{\theta} + \frac{1}{2\theta^2}\sum_i X_i^2 \end{aligned}$$

Hence, the maximum likelihood estimator $\hat{\theta}$ is

$$\hat{\theta} = \frac{1}{2n}\sum_i X_i^2 = \frac{\overline{X_n}}{2}.$$

edXproblem: Fisher Information

Compute the Fisher information $I(\theta)$, as a function of θ.

Useful fact: $\int_0^\infty u^3 e^{-u^2/2}du = 2.$

$I(\theta) =$ **Assessment content removed** edXinclude: ../xml/standardnotation.xml

Solution:

Recall $I(\theta) = -\mathbb{E}[l''(\theta)]$.

$$\begin{aligned}
l(\theta) &= \ln(X) - \ln\theta - \frac{X^2}{2\theta} \qquad (X,\theta \geq 0) \\
l'(\theta) &= -\frac{1}{\theta} + \frac{X^2}{2\theta^2} \\
l''(\theta) &= +\frac{1}{\theta^2} - \frac{X^2}{\theta^3} \\
-\mathbb{E}[l''(\theta)] &= -\frac{1}{\theta^2} + \frac{1}{\theta^3}\mathbb{E}[X^2] \\
\mathbb{E}[X] &= \frac{1}{\theta}\int_0^\infty x^3 \exp\left(-\frac{x^2}{2\theta}\right) dx \\
&= \theta\int_0^\infty u^3 \exp\left(-\frac{u^2}{2}\right) du \qquad u = \frac{x}{\sqrt{\theta}} = 2 \\
I(\theta) \;=\; -\mathbb{E}[l''(\theta)] &= -\frac{1}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}.
\end{aligned}$$

edXproblem: Wald's Test

Write down the test statistic T_n^{Wald} for Wald's test in terms of the maximum likelihood estimator $\hat{\theta}$ of θ, the Fisher information I, and the sample size n. Use the Fisher information $I(\hat{\theta})$ evaluated at $\hat{\theta}$ in the definition of Wald's test.

(To avoid double jeopardy, enter **I** for the $I = I(\hat{\theta})$; or enter your answer directly in terms of $\hat{\theta}$ only. Enter **hattheta** for $\hat{\theta}$.)

$T_n^{\text{Wald}} =$ **Assessment content removed**

When do we reject the null hypothesis in Wald's test

Answer Choices:

When $T_n^{\text{Wald}} > C$ for some $C > 0$

When $T_n^{\text{Wald}} < C$ for some $C>0$

Find C such that the Wald's test has asymptotic level 7%.

(Enter a numerical value accurate to at least 2 decimal places.)

$C =$ **Assessment content removed**

Solution:

Wald's test is defined as

$$\Psi = \mathbf{1}(T_n^{\text{Wald}} > C) \qquad \text{where } T_n^{\text{Wald}} = nI(\hat{\theta})(\hat{\theta}-1)^2 = n\frac{(\hat{\theta}-1)^2}{\hat{\theta}^2}.$$

To obtain asymptotic level 7%, $C = q_{\chi^2}(0.07) = 3.28302$ where $q_{\chi^2}(0.07)$ is the 0.93-quantile of the χ^2 distribution with 1 degree of freedom.

edXproblem: P-value

Assume that the sample size is $n = 500$. You observe that

- The sample mean is $\frac{1}{n}\sum_{i=1}^{n} X_i = 0.86$;
- The (biased) sample variance is $\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 = 1.09$.

Enter a numerical value for $\hat{\theta}$. (Enter a value acccurate to at least 2 decimal places.)

$\hat{\theta} =$ **Assessment content removed**

Compute the p-value of the Wald's test above. (As above, the Fisher information $I(\hat{\theta})$ evaluated at the MLE $\hat{\theta}$ in the definition of Wald's test.)

(Enter a numerical value accurate to at least 3 decimal places.) p-value: **Assessment content removed**

Does your test reject H_0 at the following asymptotic levels? (Choose all that apply.)

Answer Choices:
reject H_0 at asymptotic level 1%
reject H_0at asymptotic level 7%
reject H_0 at asymptotic level 10%
Fail to reject at asymptotic level 10%.

Solution:

THe MLE is

$$\hat{\theta} = \frac{1}{2}\left(1.09 + (0.86)^2\right) = 0.9148.$$

The p-value is

$$p = 1 - \Phi_{\chi^2}^{-1}\left(\frac{n}{\hat{\theta}^2}(\hat{\theta}-1)^2\right) = 1 - \Phi_{\chi^2}^{-1}(4.337) = 0.037.$$

Since $1\% < p < 7\% < 10\%$, we reject H_0 at 7% and 10%.

edXproblem: Likelihood Ratio Test

You perform the Likelihood Ratio test for the same set of hypotheses:

$$H_0 : \theta = 1 \quad \text{and} \quad H_1 : \theta \neq 1.$$

Write down the test statistic T_n^{LR} for the likelihood ratio test in terms of $\hat{\theta}$ and n.

(Enter **hattheta** for $\hat{\theta}^{\text{MLE}}$).

$T_n^{\text{LR}} =$ **Assessment content removed**

When do we reject the Null hypothesis using the likelihood ratio test?

Answer Choices:

When $T_n^{\text{LR}} > C$ for some $C > 0$

When $T_n^{\text{LR}} > C$ for some $C<0$

Find C such that the Likelihood ratio test has asymptotic level 7%.

(Enter a numerical value accurate to at least 2 decimal places.)

$C =$ **Assessment content removed**

Compute the p-value of the likelihood ratio test using the same sample as in the previous problem.

(Enter a numerical value accurate to at least 3 decimal places.) p-value: **Assessment content removed**

Does your test reject H_0 at the following asymptotic levels? (Choose all that apply.)

Answer Choices:

reject H_0 at asymptotic level 1%

reject H_0 at asymptotic level 7%

reject H_0 at asymptotic level 10%

Fail to reject at asymptotic level 10%.

Solution:

The likelihood ratio test statistic is

$$\begin{aligned} T_n^{\text{LR}} &= 2(l_n(\hat{\theta}) - l_n(1)) \\ &= n(\hat{\theta} - \ln(\hat{\theta}) - 1). \end{aligned}$$

The likelihood ratio test with level 7% is

$$\psi^{\mathrm{LR}} = \qquad \mathbf{1}(T_n^{\mathrm{LR}} > C) \qquad \text{where } C = q_{0.07,\chi^2} = 3.28302.$$

Since $T_n^{\mathrm{LR}} = 1000(0.9148 - \ln(0.9148) - 1) = 3.85$, the p-value is 0.04975. Hence we reject H_0 at levels 7% and 10%.

edXvertical: 6.
edXtext: setup

Setup: All problems on this page use the following setup.

Let $\mathcal{X} = \{1, \ldots, k\}$ be a sample space of k possible outcomes of an experiment, and $(\mathcal{X}, \{P_\theta\}_{\theta\in\Theta})$ be a statistical model for this experiment.

Suppose that this model has d parameters, so that $\Theta \subset \mathbb{R}^d$.

Suppose that the Fisher information matrix $I(\theta)$ of this statistical model is nonsingular (i.e. invertible) and that the MLE $\hat{\theta}_n$ has the Gaussian asymptotic distribution

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{\text{in } (d) \text{ under } P_\theta} \mathcal{N}\left(0, I^{-1}(\theta)\right)$$

for every $\theta \in \Theta$.

Let $X_1, \ldots, X_n$ be i.i.d. observations from repeated trials of the experiment. Let

$$N_j = \sum_{i=1}^{n} \mathbf{1}(X_i = j)$$

be the counts of the number each outcome is observed in the data.

edXproblem: Dimension of the Parameter

What must be true about d, the dimension of the model?

Answer Choices:
$d = 1$
$d = k - 1$
$d \leq k - 1$
$d \geq k - 1$

Solution:

Since the asymptotic distribution of the MLE is non-degenerate, we must have $d \leq k - 1$.

edXproblem: Degrees of Freedom

Consider the following χ^2 test statistics:

$$T_n = \qquad n\sum_{i=1}^{k}\frac{\left(N_i/n - P_\theta(i)\right)^2}{P_\theta(i)}.$$

Under the null hypothesis

$$H_0 : \theta = \theta^0$$

what is the number of degrees of freedom of the asymptotic distribution, which is a χ^2 distribution, of T_n?

Degrees of freedom: **Assessment content removed**

edXproblem: Alternate Chi Square Test

Now, consider the alternate χ^2 test statistic below:

$$\widetilde{T_n} = \qquad n\sum_{i=1}^{k}\frac{\left(N_i/n - P_{\hat{\theta}}(i)\right)^2}{P_{\hat{\theta}}(i)}.$$

(The difference between T_n and $\widetilde{T_n}$ is that θ is changed to $\hat{\theta}$ as the parameter for P.)

Suppose that q_α is chosen such that

$$\lim_{n\to\infty} P_{\theta^0}\left(T_n > q_\alpha\right) = \alpha$$

where $T_n = n\sum_{i=1}^{k}\frac{\left(N_i/n - P_\theta(i)\right)^2}{P_\theta(i)}$ was defined as in the previous problem.

Which of the following must be true?

Answer Choices:

$\lim_{n\to\infty} P_{\theta^0}\left(\widetilde{T_n} > q_\alpha\right) = \alpha$

$\lim_{n\to\infty} P_{\theta^0}\left(\widetilde{T_n} > q_\alpha\right) < \alpha$

$\lim_{n\to\infty} P_{\theta^0}\left(\widetilde{T_n} > q_\alpha\right) > \alpha$

None of the above

Solution:

- Under the null hypothesis, $T_n^{(j=0)} \xrightarrow[n\to\infty]{(d)} \chi^2_{k-1}$ and coincides with the Wald's test of the multinomial statistical model. This follows from the discussion of Lecture 15.

- By estimating the parameter θ with MLE in the test statistic, we obtain an asymptotic χ^2 distribution with **fewer** degrees of freedom:

$$T_n^{(j)} \xrightarrow[n\to\infty]{(d)} \chi^2_{k-1-j}.$$

Therefore, the threshold $q_\alpha^{(j)}$ of the test $\psi_j\left(q_\alpha^{(j)}\right) = \mathbf{1}\left(T_n^{(j)} > q_\alpha^{(j)}\right)$ with asymptotic level α **decreases** with the number j of parameters that are estimated from data. Consequently, $\psi_{j+1}(q)$ is more **conservative** (rejects less often) than $\psi_j(q)$.

edXvertical: 7.
edXtext: Problem setup

Setup:

Suppose you observe an i.i.d. sample $X_1,\ldots,X_n$ of a Bernoulli random variable X with unknown parameter $\theta_0 = \mathbb{P}(X=1)$.

We use the Bayesian approach to statistical inference and model the unknown parameter θ_0 as a random variable Θ (defined jointly with X). That is, let the conditional distribution of X given Θ be

$$X|\Theta \sim \mathsf{Ber}(\Theta)$$

and let

$$\pi_\Theta(\theta) = 3\theta^2\mathbf{1}(\theta\in[0,1])$$

be the prior distribution of Θ.

Let P denote the probabilities implied by the Bayesian model.

Useful Facts:

The $\mathsf{Beta}(\alpha,\beta)$ distribution has pdf

$$f(t) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)} \quad \text{where} \quad B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt,$$

and

$$\begin{aligned}
\text{Mean} &= \frac{\alpha}{\alpha+\beta}\\
\text{Variance} &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\\
\text{Mode} = \text{argmax} f &= \frac{\alpha-1}{\alpha+\beta-2} \text{ for } \alpha,\beta>1.
\end{aligned}$$

The **mode** is the value of the random variable at which the pdf attains its **maximum**.

edXproblem: Posterior Distribution

The posterior distribution $\pi_{\Theta|X_1,\ldots,X_n}$ of Θ given the sample $X_1,\ldots,X_n$ is a Beta distribution $\mathsf{Beta}(a,b)$. Specify the parameters a and b below.

(Enter **barX_n** for $\overline{X}_n$.)

$a =$ **Assessment content removed**

$b =$ **Assessment content removed**

Solution:

$$\pi_{\Theta|\vec{X}_n}(\theta) = \frac{\theta^{\sum_{i=1}^n X_i+2}(1-\theta)^{n-\sum_{i=1}^n X_i}}{B(3+\sum_{i=1}^n X_i, n+1-\sum_{i=1}^n X_i)} = \mathrm{Beta}\left(3+\sum_{i=1}^n X_i, n+1-\sum_{i=1}^n X_i\right).$$

edXproblem: Bayes Estimator and its Limit

What is the mean $\hat{\theta}_n^{\text{Bayes}}$ of the posterior distribution in terms of $\overline{X}_n$?

(Enter **barX_n** for $\overline{X}_n$).

$\hat{\theta}^{\text{Bayes}} = \mathbb{E}[\Theta|\vec{X}_n] =$ **Assessment content removed**

The estimator $\hat{\theta}_n^{(\pi)}$ converges in probability to a constant as $n \to \infty$? What is this limiting constant?

(Enter in terms of the true parameter θ_0.) $\hat{\theta}_n^{(\pi)} \xrightarrow{\text{in } \mathbb{P}}$ **Assessment content removed**

edXinclude: ../xml/standardnotation.xml

Note: The two notation $\hat{\theta}^{\text{Bayes}}$ and $\hat{\theta}_n^{(\pi)}$ referred to the same estimator in this problem.

Solution:

$$\begin{aligned}\hat{\theta}_n^{(\pi)} &= \mathbb{E}[\Theta|\vec{X}_n] \\ &= \frac{\sum_{i=1}^n X_i + 3}{n+4} \\ &\overset{\text{in } P^*}{\to} \theta_0 \quad \text{as } n \to \infty.\end{aligned}$$

edXproblem: MAP Estimator

What is the maximum a posteriori (MAP) estimator $\hat{\theta}_n^{\text{MAP}}$ of θ_0 in terms of $\overline{X}_n$? The maximum a posteriori (MAP) estimator is the value of θ where the posterior distribution is maximum, i.e. $\hat{\theta}^{\text{MAP}} = \arg\max_\theta \pi_{\Theta|\vec{X}_n}(\theta)$.

(Enter **barX_n** for $\overline{X}_n$).

$\hat{\theta}^{\text{MAP}} = \arg\max_\theta \pi_{\Theta|\vec{X}_n}(\theta) =$ **Assessment content removed**

What is the limit in probability of $\hat{\theta}_n^{\text{MAP}}$ as $n \to \infty$? (There is no answer box for this question.)

edXinclude: ../xml/standardnotation.xml

Solution:

$$\begin{aligned}\hat{\theta}_n^{\text{MAP}} &= \text{argmax}_\theta \pi_{\Theta|\vec{X}_n}(\theta)\\ &= \frac{\sum_{i=1}^n X_i + 2}{n+2}\\ &\overset{\text{in } P^*}{\to} \theta_0 \quad \text{as } n \to \infty.\end{aligned}$$

edXproblem: MLE

What is the MLE $\hat{\theta}^{\text{MLE}}$ of θ in terms of $\overline{X}_n$?

(Enter **barX_n** for the mean $\overline{X}_n$.)

$\hat{\theta}^{\text{MLE}} =$ **Assessment content removed**

The MLE $\hat{\theta}_n^{\text{MLE}}$ converges in probability as $n \to \infty$ to what constant? $\hat{\theta}_n^{\text{MLE}} \xrightarrow{\text{in } \mathbb{P}}$ **Assessment content removed**

edXinclude: ../xml/standardnotation.xml

Solution:

$$\begin{aligned}\hat{\theta}_n^{\text{MLE}} &= \text{argmax}_\theta P(\vec{X}_n|\Theta = \theta)\\ &= \bar{X}_n \overset{\text{in } P^*}{\to} \theta_0 \quad \text{as } n \to \infty.\end{aligned}$$

edXtext: Food for thought

Ungraded Questions for Thought

The following questions do not have answer boxes and are for you to explore.

1. What is the Fisher information $\mathcal{I}(\theta)$?
2. What is the variance $\mathsf{Var}\left(\Theta|\vec{X}_n\right)$ of the posterior distribution?
3. What is the limit of $n\mathsf{Var}\left(\Theta|\vec{X}_n\right)$ in terms of the Fisher information $\mathcal{I}(\theta)$?
4. *added (December):* Assume that for large sample sizes n the posterior distribution $\pi_{\Theta|\vec{X}_n}$ can be approximated by the normal distribution with the same mean $\hat{\theta}_n^{(\pi)}$ and variance $\mathsf{Var}\left(\Theta|\vec{X}_n\right)$ or $\frac{1}{n}\mathcal{I}(\theta)^{-1}$. (This result is called the **Bernstein-von Mises Theorem**). Recall that for Gaussian random variables
$$N(a, V/n) \quad \text{has the same law as} \quad a + \frac{1}{\sqrt{n}}N(0, V).$$
Does the 0.95-quantile of the centered posterior distribution $\pi_{\Theta|\vec{X}_n} - \hat{\theta}_n^{(\pi)}$ decrease/increase or has constant magnitude with the sample size n? What is the rate of this change with n?

Solution

1.
$$\begin{aligned} I(\theta) &= -\mathbb{E}[\frac{d^2}{d\theta^2}\ln P(\vec{X}_n|\Theta=\theta)] \\ &= -\mathbb{E}\Big[\frac{X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\Big] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta(1-\theta)} \end{aligned}$$
2.
$$V[\Theta|\vec{X}_n] = \frac{n^2(\bar{X}_n + \frac{3}{n})(1+\frac{1}{n} - \bar{X}_n)}{(n+4)^2(n+5)} = \frac{1}{n}\bar{X}_n(1-\bar{X}_n) + o_P(n)$$
3. Therefore,
$$nV[\Theta|\vec{X}_n] = \bar{X}_n(1-\bar{X}_n) + o_P(1) \overset{\text{in } P^*}{\to} \theta^*(1-\theta^*) = I^{-1}(\theta^*)$$
4. By Bernstein-von Mises theorem, the posterior distribution can be approximated (in probability in total variation norm) by the normal distribution
$$\pi_{\Theta|\vec{X}_n} \approx \hat{\theta}_n^{(\pi)} + \tfrac{1}{\sqrt{n}}N\big(0, I^{-1}(\theta)\big)$$
so the quantiles of the centered posterior distribution $\pi_{\Theta|\vec{X}_n}\left(\theta - \hat{\theta}_n^{(\pi)}\right)$ shrink to zero at the rate $O_P(\sqrt{n})$. This means that asymptotically Bayesian inference is the same as the Maximum Likelihood inference.