edX Course: 18.6501x

Contents

Exam 2

Midterm Exam 2

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, Tuesday May 7, 2019 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.

True or False

Instructions :

Be very careful with the multiple choice questions below. Some are "choose all that apply," and many tests your knowledge of when particular statements apply.

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

(a)

The likelihood ratio test is used to obtain a test with non-asymptotic level α.

"True"
"False"

Solution:

The likelihood ratio test relies on Wilks Theroemis and (under subject to technical conditions), and is used to obtain test with specified asymptotic levels.

(b)

The sample mean and the sample variance of i.i.d. random variables are always independent.

"True"
"False"

Solution:

This is a consequence of Cochran's theorem when the random variables are Gaussian random variables. However Cochran's theorem does not hold for general i.i.d. random variables.

(c)

Let U be a standard Gaussian random variable and V be a χ^2 random variable with d degrees of freedom. (No other assumptions are made about U and V.) Then, $\sqrt{d}\frac{U}{\sqrt{V}}$ is a Student t random variable with d degrees of freedom.

"True"

"False"

Solution:

Note that $\sqrt{\frac{U}{\sqrt{V}}}$ is a Student random variable with d degrees of freedom only if U and V are **independent**. If U and V are not independent then $\sqrt{d}\frac{U}{\sqrt{V}}$ does not have the distribution of a Student random variable with d degrees of freedom.

(d)

Student's t test can be run even with a small sample size, as long as the data are i.i.d. Gaussian.

"True"
"False"

Solution:

The Student's test applies for all sample sizes, because the given data is Gaussian, and the Central Limit Theorem is not needed as a source of normality.

(e)

If $X_1, \ldots, X_n$ are i.i.d. random variables on a finite discrete sample space E, a χ^2-test can be run in order to test if the unknown distribution of X_1 is uniform on E.

Note: (added May 4) You may assume n is sufficiently large.

"True"
"False"

Solution:

A χ^2 test can be used to check whether a random variable has any distribution over a finite space E. In particular it can be used to test whether the samples are uniform for E. (As seen in lectures and homework, the χ^2 test can also be adapted to test for continuous distributions or even families of distributions.)

Multiple Choice Questions (cont.)

(f)

If $X_1, \ldots, X_n$ are i.i.d. random variables uniformly distributed on $[0, \theta]$ for some unknown $\theta > 0$, Wald's test can be used to test whether $\theta = 1$.

"True"
"False"

Solution:

Wald's test applies only when the asymptotic normality of the MLE estimate applies. In the case of $\mathsf{Unif}[0, \theta]$, the technical conditions for the MLE do not apply as the support this distribution depends on the parameter θ that is being estimated.

(g)

Let $X_1, \ldots, X_n \sim X$ be n i.i.d random variables. To test if they follow a Gaussian distribution, (i.e. if their distribution belongs to the Gaussian family), you can use...
(Check all that apply.)

Note: Here, we are not testing whether X follows $\mathcal{N}(\mu, \sigma^2)$ for specific μ and σ^2. We are testing whether there exists μ and σ^2 such that X follows $\mathcal{N}(\mu, \sigma^2)$.

Note (added May 1): The test is not required to be rigorous.

"The χ^2 goodness of fit test"
"The Kolmogorov-Smirnov test"
"The Kolmogorov-Lilliefors test"
"A normal QQ-test"

Grading note: *Since there are a lot of ambiguity in the problem statement and correction notes, you will receive full credit as long as you have selected the Kolmogorov-Lilliefors test as one of your answers. This will take some time.*

Solution:

We consider each of the choices in turn.

- To use the χ^2 test to test if the data follow a Gaussian distribution, first use a consistent estimator to estimate μ and σ^2 (e.g. the MLE), then normalize the data appropriately, bucket various data points, and then perform the test. (For details, see Homework 8.) **Remarks:** The χ^2 test is an asymptotic test, hence n needs to be large.

- The Kolmogorov-Smirnov test is used to test whether the distribution underlying the data is a particular distribution with fixed parameters, e.g. $\mathcal{N}(\mu, \sigma^2)$ where μ and σ^2 are fixed, and not whether or not it belongs to the Gaussian family. **Remark:** The KS test can be used for small n. The KS test statistic is defined using specific parameters. If we instead plug in estimators of μ and σ^2 into the KS test statistic, the KS test may lead us to fail to reject while the more accurate Kolmogorov-Lilliefors test will lead us to reject the null hypothesis that the underlying distribution is Gaussian.

- A Kolmogorov-Lilliefors tests whether the data is Gaussian by testing whether or not the distribution fits the Gaussian of the sample mean and sample variance well and therefore is appropriate.

- Finally, a normal QQ-plot directly checks whether a distribution is Gaussian; if the plot is not nearly linear, then this is an indication that the distribution is not Gaussian. Gaussians with different mean and variances will correspond to linears with different slopes and intercepts.

(h)

Consider the model $\{\mathbb{R}, \{\mathcal{N}(\theta, 1)\}_{\theta\in\mathbb{R}}$. Then the prior $\pi(\theta) \propto 1$ for all $\theta \in \mathbb{R}$ is called... (check all that apply).

"Improper"
"A Jeffreys prior"
"None of the above"

Grading Note: Partial credit are given.

Solution:

The prior distribution $\pi(\theta)$ is improper since the function $\pi(\theta)$ is not integrable over $\mathbb{R}$, i.e. $\int_{-\infty}^{\infty} \pi(\theta) d\theta$ is not finite. Finally note that the likelihood of $L(X;\theta) = \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\theta)^2}{2}}$ and therefore the log-likelihood is

$$\ell(x;\theta) = \frac{-(x-\theta)^2}{2} - \frac{1}{2}\log(2\pi).$$

Taking the second derivative with respect to θ it follows that

$$\ell''(x;\theta) = -1$$

and therefore the Jeffreys prior is proportional to $\sqrt{I(\theta)} = \sqrt{1} = 1$. Therefore the uniform prior is also the Jeffreys prior.

(i)

The maximum likelihood estimator is always unbiased.

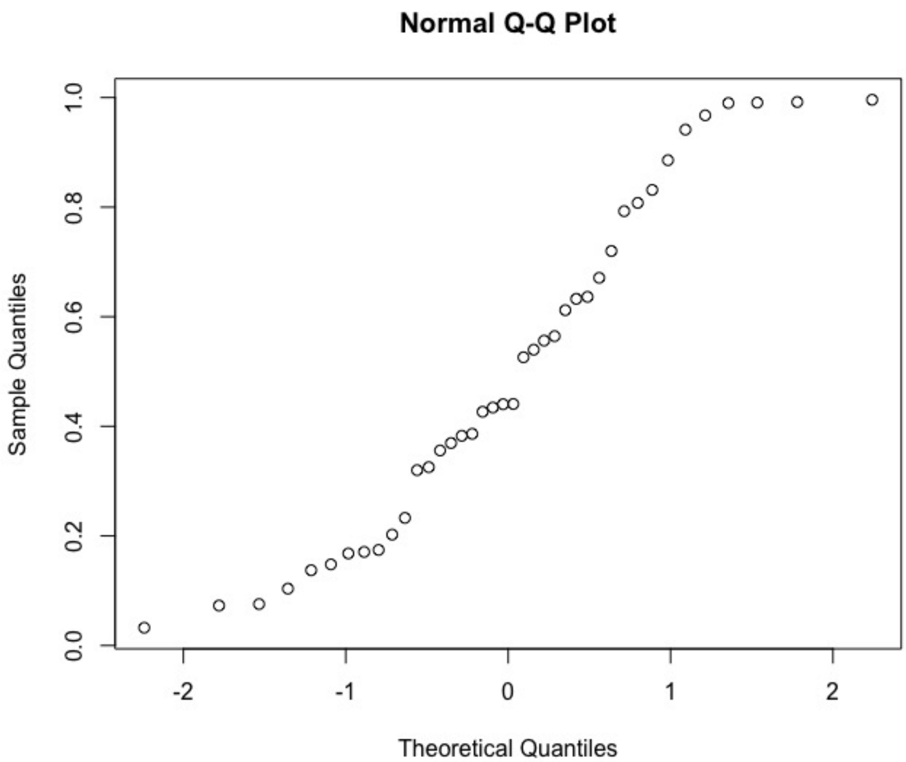
"True"
"False"

Solution:

The maximum likelihood estimator can be biased. For example, consider $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{Unif}[0, \theta]$. The MLE $\theta^{\text{MLE}} = \max(X_1, \ldots, X_n)$ is biased because it always less than or equal to the true θ.
Another example of biased MLE is the MLE $\widehat{\sigma^2}^{\text{MLE}}$ for a Gaussian statistical model.

Problem 2

(a)



Refer to the QQ-plot above. Which of the following best represent the support of the distribution underlying the data?

"$(0, 1)$"
"$(-2.3, 2.3)$"
"$(-2, 1)$"
"$\mathbb{R}$"

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

Solution:

- Recall that the points on a normal QQ-plot are $(x, y) = (\Phi^{-1}(i/n), X_{(i)})$ where Φ is the cdf of the standard normal distribution and $X_{(i)}$ is i^{th} largest data point in the sample. Hence, the range of the y-values on the QQ-plot gives a visual estimate of the support of distribution underlying the data, which is [0.1] in this case.

- The QQ-plot "flattens" on its right side; this means the quantiles of the empirical quantiles are smaller than the normal distribution at the right tail of the distributions. This is the case only if the distribution underlying the data has a lighter right tail than that of the standard normal distribution.

Problem 3

Setup: Let $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} \mathsf{Unif}([\theta, \theta^2])$ for some unknown $\theta > 1$. That is, the pdf of X_i is

$$f_\theta(x) = C\mathbf{1}_{[\theta,\theta^2]}(x).$$

where $C = \frac{1}{\theta^2 - \theta}$.

(a)

Is the parameter θ identifiable?

"True"
"False"

Solution:

Yes, θ is identifiable as the minimum of the support of the $f_\theta(x)$.

(b)

Which of the following statements are true regarding the samples? Note that $X_{(i)}$ denote the order statistics, i.e. $X_{(i)}$ represents the $i^{(th)}$ smallest value of the sample. For example, $X_{(1)}$ is the smallest and $X_{(n)}$ is the greatest of a sample of size n.
(Check all that apply.)

"$X_{(2)} \geq X_{(1)}$"
"$X_{(2)} \geq X_{(3)}$"
"$X_{(1)}^2 \geq X_{(n)}$"
"$X_{(2)} \geq X_{(3)}^2$"

Grading note: Partial credits are given.

Solution:

Since $X_{(i)}$ are the order statistics of the data $X_1, \ldots, X_n$, $X_{(i)}$ are the rearrangement of $X_1, \ldots, X_n$ such that $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$. Hence, the first choice is true while the second is not.
For the third choice, the key point is that since $\theta > 1$, we have $\theta^2 > \theta$, and therefore $X_{(1)} \geq \theta$ and $X_{(n)} \leq \theta^2$. Therefore $X_{(1)}^2 \geq X_{(n)}$.
The fourth choice is false, since $X_3 > \theta > 1$, $X_{(3)}^2 > X_{(3)} > X_{(2)}$.

(c)

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

(If applicable, enter **m** for the minimum $\min_i(X_i)$ of the X_i, **M** for the maximum $\max_i(X_i)$ of the X_i, and **barX_n** for $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$.)

$\hat{\theta}^{\mathrm{MLE}} =$

Solution:

Note that the likelihood of $X_1, \ldots, X_n$ is

$$L(X_1, \ldots, X_n; \theta) = \left(\frac{1}{\theta(\theta-1)}\right)^n \mathbf{1}(\theta \leq X_{(1)} < X_{(n)} \leq \theta^2)$$

In the region $\theta > 1$, the function $\frac{1}{(\theta(\theta-1))^n}$ is strictly decreasing, so $\hat{\theta}^{\mathrm{MLE}}$ is the smallest possible θ that can lead to the given data, i.e. a greatest lower bound on θ. Since $X_{(n)} < \theta^2$ can be arbitrarily close θ^2, we have $\sqrt{X_{(n)}} < \theta$. Therefore, the MLE is so $\hat{\theta}^{\mathrm{MLE}} = \sqrt{X_{(n)}}$.

(d)

Suppose that $X_{(1)} = 2.01$ and $n = 18$. Design a test for the hypotheses

$$H_0 : \theta = 2 \text{ vs. } H_1 : \theta < 2.$$

that uses $T_n = X_{(1)}$ as the test statistic.

Compute the p-value of this test.

(Enter a numerical value accurate to at least 3 decimal places.)

p-value:

Do we reject the null hypothesis at the level 5%?

"Yes"
"No"

Correction Note (added May 5):. An earlier version of the problem statement was "Design a test for the hypotheses

$$H_0 : \theta = 2 \text{ vs. } H_1 : \theta < 2.$$

that uses $X_{(1)}$ as the test statistic."

Solution:

The test for the given hypotheses

$$H_0 : \theta = 2 \text{ vs.} H_1 : \theta < 2.$$

and the estimator $X_{(1)} = \min_i\{X_i\}$, is

$$\psi = \mathbf{1}\left(X_{(1)} < C\right)$$

for some threshold C. Hence, the p-value of this test is

$$\begin{aligned} p - \text{value} = &\quad \mathbb{P}_{\theta=2}\left(X_{(1)} < 2.01\right) \\ = &\quad 1 - \mathbb{P}_{\theta=2}\left(X_{(1)} \geq 2.01\right) \\ = &\quad 1 - \left(\prod_{i=1}^{18} \mathbb{P}_{\theta=2}(X_i \geq 2.01)\right) \\ = &\quad 1 - \left(1 - \frac{2.01-2}{2}\right)^{18} \approx .086. \end{aligned}$$

Since the p-value is greater than 0.05, we fail to reject the null hypothesis.

(e)

Again, $X_{(1)} = 2.01$ and $n = 18$, and consider the same hypotheses as above:

$$H_0 : \theta = 2 \text{ vs. } H_1 : \theta < 2$$

Now, design a test using the test statistic $T_n = X_{(n)}$. What is the largest value of $X_{(n)}$ that would lead to a rejection of H_0 at level 5%?

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

To reject H_0 at level 5%,

$X_{(n)} \leq$

Solution:

The test for the given hypotheses

$$H_0 : \theta = 2 \text{ vs.} H_1 : \theta < 2.$$

using the given test statistic $T_n = X_{(n)} = \max_i\{X_i\}$, is

$$\psi = \mathbf{1}\left(X_{(n)} < C\right)$$

for some threshold C. We want C such that

$$\begin{aligned}
&\mathbb{P}_{\theta=2}(X_{(n)} < C) &&= 0.05\\
\iff &\prod_{i=1}^{n} \mathbb{P}_{\theta=2}(X_i < C) &&= 0.05\\
\iff &\left(\frac{C-2}{4-2}\right)^{18} &&= 0.05\\
\iff &C &&= 2 + 2(0.05)^{\frac{1}{18}} \approx 3.69.
\end{aligned}$$

This C is the largest value of $X_{(n)}$ so that H_0 will be rejected at 5% by this test.

Problem 4

Setup:

For $x \in \mathbb{R}$ and $\theta \in (0,1)$, define

$$f_\theta(x) = \begin{cases} \theta^2 & \text{if } -1 \leq x < 0 \\ 1-\theta^2 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $X_1, \ldots, X_n$ be i.i.d. random variables with density f_θ, for some unknown $\theta \in (0,1)$.

(a)

To prepare, sketch the pdf $f_\theta(x)$ for different values of $\theta \in (0,1)$.

Which of the following properties of $f_\theta(x)$ guarantee that it is a probability density? (Check all that apply)

Note (added May 3): Note that you are **not** asked which of the following are properties of $f_\theta(x)$, but rather, which properties ensure that $f_\theta(x)$ is a density.

Note (added May 4): To be precise, select the smallest subset of choices below that would guarantee that $f_\theta(x)$ is a probability density.
"$f_\theta(x) \geq 0$ for all $x \in \mathbb{R}$"
"$f_\theta(x) \leq 1$ for all $x \in \mathbb{R}$"
"$\int_{\mathbb{R}} f_\theta(x)dx = 1$"
"$f_\theta(x) = 0$ for $|x| > 1$"
Grading note: Partial credit are given.

Solution:

In order for $f_\theta(x)$ to be a probability density we need the function to be non-negative and the function to integrate to 1. Therefore, the first and third choices are the correct choices. The remaining choices are true properties of f_θ that do not guarantee f_θ to be a density.

(b)

Let a be the number of X_i which are **negative** ($X_i < 0$) and b be the number of X_i which are **non-negative** ($X_i \geq 0$). (Note that the total number of samples is $n = a + b$ and be careful not to mix up the roles of a and b.)

What is the maximum likelihood estimator $\hat{\theta}^{\text{MLE}}$ of θ?

Note (added May 3): Different correct forms of the answer will be graded as correct.

$\hat{\theta}^{\text{MLE}} =$

Is $\hat{\theta}^{\text{MLE}}$ asymptotically normal (in this example)?

Correction Note: An earlier version of the problem statement contains minor errors and was "Let a be the number of X_i which are **negative** ($X_i < 0$) and b be the number of X_i which are **non-negative** ($X_i \geq 0$)".

Solution:

Observe that $Y_i = \mathbf{1}(X_i < 0)$ are i.i.d. $\mathsf{Ber}(\theta^2)$, so we can write down the likelihood for Bernoulli random variables and obtain:

$$L(X_1, \ldots, X_n; \theta) = (\theta^2)^a (1 - \theta^2)^b.$$

(Alternatively, work out the likelihood directly from definition.)
Therefore the log-likelihood is

$$\ell(\theta) = 2a \log(\theta) + b \log(1 - \theta^2).$$

Taking the derivative:

$$\ell'(\theta) = +\frac{2a}{\theta} + \frac{-2b\theta}{1 - \theta^2},$$

and setting $\ell'(\theta) = 0$ gives

$$\hat{\theta}^{\text{MLE}} = \left(\frac{a}{a+b}\right)^{\frac{1}{2}}.$$

Note that

$$\ell''(\theta) = \frac{-2a}{\theta^2} - \frac{2b(\theta^2 + 1)}{(1 - \theta^2)^2} < 0$$

and therefore this maximum is unique.
This MLE $\hat{\theta}^{\text{MLE}}$ is asymptotically normal because the conditions 1-4 on the slide on this page holds:

1. θ is identifiable

2. For all $\theta \in (0, 1)$, the support of f_θ does not depend on θ

3. θ^* is not on the boundary of $(0,1)$, i.e. $\theta^* \notin \{0,1\}$;
4. $I(\theta)$ is invertible

(In this course, we generally do not need to worry about the few more technical conditions listed on this slide.)

(c)

What is the asymptotic variance $V(\theta)$ for $\hat{\theta}^{\mathrm{MLE}}$?

$V(\theta) =$

Solution:

Again, note that random variable $Y_i = \mathbf{1}_{X_i<0}$ is a $\mathsf{Ber}(\theta^2)$. Therefore the variance $\mathsf{Var}(Y_i) = \theta^2(1-\theta^2)$. Thus by the Central Limit Theorem it follows that

$$\sqrt{n}(\bar{Y}_n - \theta^2) \xrightarrow[n\to\infty]{(d)} N(0, \theta^2(1-\theta^2)).$$

However, our estimator $\hat{\theta}^{\mathrm{MLE}}$ is not the above but is instead

$$\hat{\theta}^{\mathrm{MLE}} = \sqrt{\bar{Y}_n} = \sqrt{\frac{a}{n}}.$$

Therefore we need the delta method to derive its asymptotic variance: Let $g(x) = x^{1/2}$, so $g'(x) = \frac{1}{2x^{1/2}}$. The delta method gives

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \theta^2(1-\theta^2)g'(\theta^2)^2) = \mathcal{N}(0, \frac{(1-\theta^2)}{4})$$

Therefore, the asymptotic variance for $\hat{\theta}^{\mathrm{MLE}}$ is $\frac{(1-\theta^2)}{4}$. **Alternatively,** use $V(\theta) = I(\theta)^{-1}$ to obtain the same answer.

(d)

Recall from the setup that $X_1, \ldots, X_n \sim X$ are i.i.d. random variables with density f_θ, for some unknown $\theta \in (0,1)$:

$$f_\theta(x) = \begin{cases} \theta^2 & \text{if } -1 \le x < 0 \\ 1-\theta^2 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Consider the following hypotheses:

$$H_0 : X \sim \mathsf{Unif}(-1, 1)$$
$$H_1 : X \text{ not distributed as } \mathsf{Unif}(-1, 1).$$

Write down the test statistic T_n^{Wald} for Wald's test for the above hypothesis. **Use the value of θ that defines H_0 as the argument of the asymptotic variance** $V(\theta)$.

Hint: Rewrite the hypothesis in terms of the parameter θ.

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

(To avoid double jeopardy, you may use **V** for the asymptotic variance $V(\theta)$ under H_0.)

$T_n^{\text{Wald}} =$

What is the form of this Wald's test?

Find C such that Wald's test has asymptotic level 5%.

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

$C =$

You obtain a sample of size $n = 100$, of which 40 of the X_i are **negative** ($X_i < 0$) and 60 of the X_i are **non-negative** ($X_i \geq 0$).

Do we reject H_0 at asymptotic level 5%?

What is the p-value for this test? (Again, use the value of θ that defines H_0 as the argument of the asymptotic variance $V(\theta)$.)

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

p-value:

Solution:

First, rewrite the hypothesis in terms of the parameter θ: X_1 is $\mathsf{Unif}(-1, 1)$ is equivalent to $\theta^2 = \frac{1}{2}$. Hence the null and alternative hypotheses are

$$H_0 = \qquad \theta = \frac{1}{\sqrt{2}}$$
$$H_1 = \qquad \theta \neq \frac{1}{\sqrt{2}}$$

Then, Wald's Theorem gives, under the null hypothesis:

$$T_n = nI(\theta_0)\left(\hat{\theta} - \frac{1}{\sqrt{2}}\right)^2 \xrightarrow[n\to\infty]{(d)} \chi_1^2.$$

where the Fisher information, or equivalently inverse asymptotic variance, is $I(\theta_0) = \frac{4}{1-\theta_0^2})$. Thus our Wald's test with asymptotic level 5% is

$$\psi = \mathbf{1}_{T_n > q_{0.05}} \qquad \text{where } q_{0.05} = q_{0.05}(\chi_1^2) \approx 3.84.$$

For $n = 100$, $a = 40$, $b = 60$,

$$T_{100}^{\text{Wald}} = \qquad 800\left(\sqrt{0.4} - \sqrt{0.5}\right)^2 \approx 4.46 > 3.84;$$

hence we reject H_0. The p-value is

$$\begin{aligned} p - \text{value} &= \qquad \mathbb{P}_{\chi_1^2}\left(y > \frac{4n}{1-\theta_0^2}\left(\hat{\theta}^{\text{MLE}} - \sqrt{\theta_0}\right)^2\right) \\ &= \qquad \mathbb{P}_{\chi_1^2}\left(y > 800\left(\sqrt{0.4} - \sqrt{0.5}\right)^2\right) \approx 0.035. \end{aligned}$$

Remark Because the null hypothesis consists of only 1 value of θ, we had chosen to implement Wald's test with θ_0 (as opposed to $\hat{\theta}^{\text{MLE}}$) as the argument of the asymptotic variance. If we have used $V(\hat{\theta}^{\text{MLE}})$ in T_n, then

$$T_{100}^{\text{Wald}} = \qquad 100\left(\frac{4}{60/100}\right)\left(\sqrt{0.4} - \sqrt{0.5}\right)^2 \approx 3.715 < 3.84.$$

This would lead to a p-value of 0.054 which is larger than 0.05and a failure to reject H_0.

(e)

As above, $X_1, \ldots, X_n \sim X$ are i.i.d. random variables with density f_θ, for some unknown $\theta \in (0, 1)$:

$$f_\theta(x) = \begin{cases} \theta^2 & \text{if } -1 \le x < 0 \\ 1-\theta^2 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

As in part (b), let a denote the number of X_i which are **negative** ($X_i < 0$) and b be the number of X_i which are **non-negative** ($X_i \ge 0$).

(Be careful not to mix up the roles of a and b).

Again, we consider the hypotheses:

$$H_0 : X \sim \mathsf{Unif}(-1, 1)$$
$$H_1 : X \text{ not distributed as } \mathsf{Unif}(-1, 1).$$

Write down the test statistic T_n^{LR} for the likelihood ratio test.

(To avoid double jeopardy, you may use **hattheta** for $\hat{\theta}^{\mathrm{MLE}}$, or directly enter your answer for $\hat{\theta}^{\mathrm{MLE}}$).

$T_n^{\mathrm{LR}} =$

What is the form of the likelihood ratio test?

Find C such that the Likelihood ratio test has asymptotic level 5%.

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

$C =$

For the same sample as in the previous part, i.e. a sample of size $n = 100$, of which 40 of the X_i are **negative** ($X_i < 0$) and 60 of the X_i are **non-negative** ($X_i \geq 0$).

Do we reject H_0 at asymptotic level 5%?

What is the p-value for the likelihood ratio test? (Enter a numerical value accurate to at least 3 decimal places.)

p-value:

Correction Note: An earlier version (before April 27 9pm EST) of the problem statement contains an error and was "As in part (b), let a denote the number of X_i which are **non-negative** ($X_i \leq 0$) and b be the number of X_i which are **negative** ($X_i > 0$)".

Solution:

We are again testing the hypotheses

$$H_0 = \qquad \theta = \frac{1}{\sqrt{2}}$$
$$H_1 = \qquad \theta \neq \frac{1}{\sqrt{2}}$$

The test statistics for the likelihood ratio test is

$$T_n^{\mathrm{LR}} = \quad 2\left(a\ln((\hat{\theta}^{\mathrm{MLE}})^2) + b\ln(1-(\hat{\theta}^{\mathrm{MLE}})^2) - n\ln(0.5)\right)$$

Since $T_n^{\mathrm{LR}} \xrightarrow[n\to\infty]{(d)} \chi_1^2,$ the likelihood ratio test of asymptotic level 5% is

$$\psi_n = \mathbf{1}(T_n^{\mathrm{LR}} > q_{0.05}(\chi_1^2)).$$

where $q_{0.05}(\chi_1^2) \approx 3.84$ is the quantile of the χ^2 distribution with degrees of freedom 1. For $n = 100$, $a/n = 0.4$, $b/n = 0.6$,

$$T_{100}^{\mathrm{LR}} = \quad 2\left(40\ln(0.4) + 60\ln(0.6) - 100\ln(0.5)\right) \approx 4.04 > q_{0.05}(\chi_1^2)$$

Hence, we can reject H_0. The p-value for this test is

$$p - \text{value} = \quad \mathbb{P}_{\chi_1^2}\left(y > T_{100}^{\mathrm{LR}}\right) = \mathbb{P}_{\chi_1^2}\left(y > 4.04\right) \approx 0.045.$$

Problem 5

(a)

You have five coins in your pocket. You know a priori that one coin gives heads with probability 0.4, and the other four coins give heads with probability 0.7.

You pull out one of the five coins at random from your pocket (each coin has probability $\frac{1}{5}$ of being pulled out), and you want to find out which of the two types of coin it is. To that end, you flip the coin 6 times and record the results $X_1, \ldots, X_6$ of each coin flip where $X_i = 1$ if "heads" and $X_i = 0$ if "tails".

Let $p = P(X_1 = 1)$. Which of the following is the space of all possible values of the parameter p? In other words, what is the smallest sample space of p?

Note (Added May 6): This question is asking for the sample space (domain) of the prior distribution of p.

"$[0.4, 0.7]$"
"$(0, 1)$"
"$\{0.2, 0.8\}$"
"$\{0.4, 0.7\}$"

Correction Note (added May 4): An earlier version of the statement did not include the second sentence "In other words, what is the smallest sample space of p".

Solution:

Note that by the problem statement that the possible values of p are .4 and .7 and the result follows.

(b)

Find the pmf π that quantifies my prior knowledge of p.

Then, enter the value of the prior evaluated at $p = 0.7$ i.e. $\pi(p = 0.7)$, below.

$\pi(p = 0.7) =$

Correction note (added May 5): An earlier version of the problem statement was "Enter $\pi(p = 0.7)$, below." for the second sentence.

Solution:

Since four of the five coins have probability of heads 0.7, we have $\pi(p = 0.7) = 4/5$; the final coin with probability of heads 0.4 gives $\pi(p = 0.4) = 1/5$.

(c)

Suppose that $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 3$. Find the maximum likelihood estimate $\hat{p}^{\text{MLE}}$ for p using the given data. (Enter a numerical value accurate to at least 3 decimal places.)

$\hat{p}^{\text{MLE}} =$

Grading note: Though not the intention of the question, there is a possibility that you interpreted this question as asking for the constrained MLE,i.e. the value of p within $\{0.4, 0.7\}$ that maximizes the likelihood. Because this was not clearly specified, you wil also receive full credit if you computed the correct constrained MLE.

Solution:

Since $X_i \sim \mathsf{Ber}(p)$, we have $\hat{p}^{\text{MLE}} = \bar{X}_n = 3/6 = 0.5$. Note that the maximum likelihood estimator does not take into account the prior distribution.

Constrained MLE:

Since the likelihood of obtaining 3 heads and 3 tails of a coin with probability of heads p is $\binom{6}{3}p^3(1-p)^3$, which is $\binom{6}{3}(0.013824)$ for $p = 0.4$ and $\binom{6}{3}(0.00926)$ for $p = 0.7$, we have $\hat{p}^{\text{MLE}}_{\text{constrained}} = 0.4$. Again, the MLE does not take into account the prior distribution of p.

(d)

Find the Bayes estimate $\hat{p}^{\text{Bayes}}$ of p based on $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 3$.

(Recall the Bayes estimator is the mean of the posterior distribution.)

(Give your answer to 3 decimal places.)

$\hat{p}^{\text{Bayes}} =$

Solution:

Given a coin of probability of heads p, the probability of flipping 3 heads and 3 tails is $\binom{6}{3}p^3(1-p)^3$.

The posterior distribution is

$$\pi\left(p \,\middle|\, \sum_{i=1}^{6} X_i = 3\right) \propto L\left(\sum_{i=1}^{6} X_i = 3 \,\middle|\, p\right)\pi(p) \propto \begin{cases} 0.2(0.4)^3(0.6)^3 = 0.0027648 & \text{for } p = 0.4 \\ 0.8(0.7)^3(0.3)^3 = 0.0074088 & \text{for } p = 0.7 \end{cases}.$$

The Bayes estimator is the expectation of p under the posterior distribution:

$$\begin{aligned} p^{\text{Bayes}} &= \frac{0.4\,\pi\left(0.4|\sum_{i=1}^{6} X_i = 3\right) + 0.7\,\pi\left(0.7|\sum_{i=1}^{6} X_i = 3|\sum_{i=1}^{6} X_i = 3\right)}{\pi\left(0.4|\sum_{i=1}^{6} X_i = 3\right) + \pi\left(0.7|\sum_{i=1}^{6} X_i = 3\right)} \\ &= 0.61847. \end{aligned}$$

(e)

Find the maximum a posteriori estimate $\hat{p}^{\text{MAP}}$ of p, i.e. the value of p at which the posterior distribution is maximum, based on $X_1 + X_2 + X_3 + X_4 + X_5 + X_6 = 3$.

$\hat{p}^{\text{MAP}} =$

Solution:

The value of p at which the posterior distribution is maximum is $p = 0.7$. Hence $\hat{p}^{\text{MAP}} = 0.7$.