Problem 4

Setup:

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

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

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

(a)

1/1 point (graded)

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.

- [x] $f_\theta(x) \ge 0$ for all $x \in \mathbb{R}$ ✔
- [ ] $f_\theta(x) \le 1$ for all $x \in \mathbb{R}$
- [x] $\int_{\mathbb{R}} f_\theta(x)\,dx = 1$ ✔
- [ ] $f_\theta(x) = 0$ 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.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

(b)

2.0/2 points (graded)

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** ($X_i \ge 0$). (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.

$\hat{\theta}^{\text{MLE}} =$ sqrt(a/(a+b)) ✔ **Answer:** sqrt(a/n)

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

- (●) yes ✔
- ( ) no
- ( ) not enough information to determine

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** ($X_i \geq 0$)".

STANDARD NOTATION

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.)

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(c)

2.0/2.0 points (graded)

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

$V(\theta) =$ (1-theta^2)/4 ✔ **Answer:** (1-theta^2)/4

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

STANDARD NOTATION

Solution:

Again, note that random variable $Y_i = \mathbf{1}_{X_i<0}$ is a $\text{Ber}(\theta^2)$. Therefore the variance $\text{Var}(Y_i) = \theta^2(1-\theta^2)$. 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 $\hat{\theta}^{\text{MLE}}$ is not the above but is instead

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

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

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

(d)

5.0/5 points (graded)

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:

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

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.)

$T_n^{\text{Wald}} =$ (n*(hattheta - sqrt(1/2))^2)/V ✔ **Answer:** n/V*(hattheta-1/sqrt(2))^2

What is the form of this Wald's test?

- (selected) $\mathbf{1}(T_n^{\text{Wald}} > C)$ for some $C>0$ ✔
- $\mathbf{1}(T_n^{\text{Wald}} > C)$ for some $C<0$
- $\mathbf{1}(T_n^{\text{Wald}} < C)$ for some $C>0$
- $\mathbf{1}(T_n^{\text{Wald}} < C)$ for some $C<0$

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

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

$C =$ 3.841 ✔ **Answer:** 3.84

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%?

- (selected) We reject H_0. ✔
- We fail to reject H_0 ✔
- Cannot be determined without more information.

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: 0.03473 ✔ **Answer:** 0.035

STANDARD NOTATION

Solution:

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

$$\begin{aligned} H_0 &= \theta = \frac{1}{\sqrt{2}} \\ H_1 &= \theta \neq \frac{1}{\sqrt{2}} \end{aligned}$$

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

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

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

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

hence we reject H_0. The p-value is

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

Remark Because the null hypothesis consists of only 1 value of $\theta,$ 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 $V\left(\hat{\theta}^{\text{MLE}}\right)$ in T_n, then

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

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

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

(e)

5.0/5 points (graded)
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:

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

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

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

$T_n^{\text{LR}} =$ 2*(2*a*ln(hattheta)+b*ln(1 - hattheta^2) - 2*a*ln(sqrt(1/2)) - b*ln(1/2))

Answer: 2*(a*ln(hattheta^2)+b*ln(1-hattheta^2)-n*ln(1/2))

What is the form of the likelihood ratio test?

- ◉ $\mathbf{1}\left(T_n^{\mathrm{LR}} > C\right)$ for some $C>0$ ✔
- ○ $\mathbf{1}\left(T_n^{\mathrm{LR}} > C\right)$ for some $C<0$
- ○ $\mathbf{1}\left(T_n^{\mathrm{LR}} < C\right)$ for some $C>0$
- ○ $\mathbf{1}\left(T_n^{\mathrm{LR}} < C\right)$ for some $C<0$

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

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

$C =$ 3.841 ✔ **Answer:** 3.84

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%?

- ◉ We reject H_0. ✔
- ○ We fail to reject H_0
- ○ Cannot be determined without more information.

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

p-value: 0.04477 ✔ **Answer:** 0.045

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** ($X_i > 0$)".

STANDARD NOTATION

Solution:

We are again testing the hypotheses

$$\begin{aligned} H_0 &= \theta = \frac{1}{\sqrt{2}} \\ H_1 &= \theta \neq \frac{1}{\sqrt{2}} \end{aligned}$$

The test statistics for the likelihood ratio test is

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

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

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

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

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

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

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

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Show Discussion

Topic: Midterm Exam 2:Midterm Exam 2 / Problem 4