Final Exam

Thursday, April 21, 2022 5:25 PM

MTH 375
Winter 2022
Final Exam

Cason Konzer :)

Please take no more than about 2.5 hours to complete this exam. It is due at midnight tonight (Thursday, 4/21). If you find an error, please let me know right away.

You may use R, your textbook, and any notes or documents you have from class. You may use the result of any theorem or computation from any of those sources, provided you cite it. You may not use any website, nor any person other than me as a source. I will provide hints if you request them by email. As usual, I grade solutions, not just answers.

1. Suppose the random variables $X_1, \ldots, X_n$ are iid, with common density

$$f_X(x) = \begin{cases} \theta(\theta+1)x(1-x)^{\theta-1} & \text{if } 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where θ is an unknown parameter.

Find a sufficient statistic for θ.

Step 1: determine the likelihood function

$$L(\vartheta) = \vartheta(\vartheta+1)x_1(1-x_1)^{\vartheta-1} \cdots \vartheta(\vartheta+1)x_n(1-x_n)^{\vartheta-1}$$

$$= \vartheta^n(\vartheta+1)^n \prod_{i=1}^{n} x_i(1-x_i)^{\vartheta-1}$$

By the Fisher-Neyman Factorization Lemma we may inspect $g(S,\vartheta)$ in $L(\vartheta) = g(S,\vartheta)\cdot h(\vec{x})$ in order to find our sufficient statistic.

Notice $L(\vartheta) = \underbrace{\left(\vartheta^n(\vartheta+1)^n \prod_{i=1}^{n}(1-x_i)^{\vartheta-1}\right)}_{g(S,\vartheta)} \cdot \underbrace{\left(\prod_{i=1}^{n} x_i\right)}_{h(\vec{x})}$

Thus from inspection we can notice that

$\boxed{S = \prod_{i=1}^{n}(1-x_i) \text{ is a sufficient statistic for } \vartheta}$

16/16

2. Let X be geometric with parameter p. You are testing the hypotheses

$$H_0 : p = .8 \text{ vs. } H_a : p < .8.$$

and decide that the test will be "Reject H_0 if $X > 2$."

(a) What is the level of significance of this test?

(b) Find a simple formula for the power function.

Hints: (i) A geometric RV is discrete with pmf $p_X(n) = p(1-p)^{n-1}$ for $n = 1, 2, 3, \ldots$.

(ii) The sum of a geometric series is $a + ar + ar^2 + ar^3 + \cdots = \frac{a}{1-r}$ for $-1 < r < 1$.

A H_0 is true iff $p = 0.8$ & we have only one sample

— H_0 is true iff $p = 0.8$ & we have only one sample

> pgeom(q = 1, prob = 0.8, lower.tail = FALSE)
[1] 0.04

in R: $n = 0, 1, 2, 3 \ldots$ thus use 1 not 2

Utilizing R we can see that our level of significance, α, is 0.004

[B]

$$\text{Power}(p) = P(\text{reject } H_0 \mid p) = P(X > 2 \mid p)$$

$$= \sum_{n=1}^{\infty} p(1-p)^{n-1} - p(1-p)^{1-1} - p(1-p)^{2-1}$$

$$= \frac{p}{1-(1-p)} - p - p(1-p) = 1 - 2p + p^2$$

$$\boxed{\text{power}(p) = (1-p)^2}$$

16/16

3. Suppose $Y_1, \ldots, Y_{10}$ are iid with common pdf

$$f(y) = \begin{cases} 2y/\theta^2 & \text{if } 0 < y < \theta \\ 0 & \text{otherwise} \end{cases}$$

for some $\theta > 0$.

(a) Show that $Q = \dfrac{\max\{Y_1, \cdots, Y_{10}\}}{\theta}$ is a pivotal quantity,

(b) Find a 90% confidence interval for θ based on Q.

[A]

A pivotal quantities distribution does not depend on ϑ

$$L(\vartheta) = \left(2^{10} \prod_{i=1}^{10} y_i\right) \cdot \left(\frac{1}{\vartheta^{2n}} \text{ for } 0 \le Y_{(n)} \le \vartheta\right), \text{ thus}$$

$Y_{(n)}$ is a sufficient statistic for ϑ.

$$F_{Y_i} = \int_0^{y_i} \frac{2y}{\vartheta^2}\, dy = \frac{y^2}{\vartheta^2} \Big|_0^{y_i} = \frac{y_i^2}{\vartheta^2}$$

$$F_{Y_{(10)}} = P(Y_{(10)} \le q) = P(Y_1 \le q \;\&\; \cdots \;\&\; Y_{10} \le q) = (F_{Y_i})^{10}$$

$$P\left(\frac{Y_{(10)}}{\vartheta} \le q\right) = P(Y_{(10)} \le \vartheta q) = [P(Y_i \le \vartheta q)]^{10} = \left(\frac{(\vartheta q)^2}{\vartheta^2}\right)^{10} = q^{20}$$

which does not depend on ϑ

16/16

[B]

$$P(a \le Q \le b) = P\left(a \le \frac{Y_{(10)}}{\vartheta} \le b\right) = P\left(\frac{a}{Y_{(10)}} \le \frac{1}{\vartheta} \le \frac{b}{Y_{(10)}}\right)$$

$$= P\left(\frac{Y_{(10)}}{a} \ge \vartheta \ge \frac{Y_{(10)}}{b}\right) = 0.90$$

Thus we consider $P(0.05 \le Q \le 0.95)$

Thus we consider $P(0.05 \le Q \le 0.95)$

$$P(Q \le 0.95) = 0.95^{20}$$

$$P(Q \ge 0.05) = 1 - P(Q \le 0.05) = 1 - 0.05^{20}$$

Last our 90% Confidence interval is

$$\frac{Y_{(10)}}{(1-0.05^{20})} \le \theta \le \frac{Y_{(10)}}{0.95^{20}}$$

4. A coin with probability p of landing 'heads' is tossed 5 times. A Bayesian statistician believes that p may be 1/4, 1/2 or 3/4, and assigns prior probabilities of $\pi(1/4) = 1/6$, $\pi(1/2) = 1/3$ and $\pi(3/4) = 1/2$.

The coin is tossed 5 times, and the result is 3 heads and 2 tails.

(a) What is the posterior distribution of p?

(b) Assuming a quadratic loss function (i.e., $L(\hat{p}, p) = (\hat{p} - p)^2$), what is the Bayesian estimate of p?

$\boxed{A}$

$$\text{prior} = \pi(p) = \begin{cases} 1/6 & \text{if } p = 1/4 \\ 1/3 & \text{if } p = 1/2 \\ 1/2 & \text{if } p = 3/4 \end{cases}$$

X_i's are iid $\sim$ Binomial$(1, p)$ & $\Sigma x = 3$, Thus

$$f_X(x; p) = p^x (1-p)^{1-x} \Rightarrow L(p) = p^{\Sigma x}(1-p)^{5-\Sigma x}$$

$$f(p \mid \vec{x}) = \frac{\pi(p)\, p^3 (1-p)^{5-3}}{\sum_p \pi(p)\, p^3 (1-p)^{5-3}}$$

$$\sum_p \pi(p)\, p^3 (1-p)^2 = \left(\frac{1}{6}\cdot\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 + \frac{1}{3}\left(\frac{1}{2}\right)^5 + \frac{1}{2}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2\right)$$

Thus our posterior is

$$f(p \mid \vec{x}) = \frac{\pi(p)\, p^3 (1-p)^2}{\left(\frac{1}{6}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^2 + \frac{1}{3}\left(\frac{1}{2}\right)^5 + \frac{1}{2}\left(\frac{3}{4}\right)^3\left(\frac{1}{4}\right)^2\right)}$$

$\boxed{B}$

$$f(p = 1/4 \mid \vec{x}) \simeq 0.05844$$
$$f(p = 1/2 \mid \vec{x}) \simeq 0.41558$$
$$f(p = 3/4 \mid \vec{x}) \simeq 0.35065$$

RIght ideas, but for some reason these 3 values don't sum to 1.

15/16

$$\hat{p} = E(p \mid \vec{x}) = \tfrac{1}{4}(0.05844) + \tfrac{1}{2}(0.41558) + \tfrac{3}{4}(0.35065)$$

$$= 0.48539$$

5. Suppose the pairs $(X_1, Y_1), \ldots, (X_{10}, Y_{10})$ satisfy the usual hypotheses for linear regression, i.e., $Y = \beta_0 + \beta_1 X + \varepsilon$ and so on. Ten such points turn out to be

X_i	.01	.03	.03	.19	.24	.25	.66	.77	.86	.94
Y_i	.87	.31	.60	.24	.67	.16	.80	.95	.69	.64

(a) Test at level of significance .05 the hypotheses

$$\mathrm{H}_0 : \beta_1 = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \beta_1 \neq 0.$$

(b) Plot the points and the regression line, and explain why the result of (a) is reasonable.

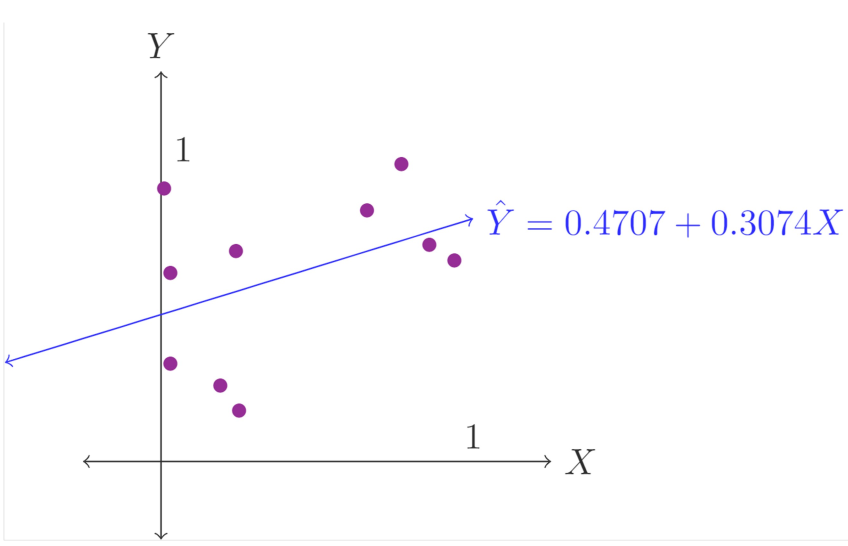
A

This question will be solved via R

```
> c(0.01,0.03,0.03,0.19,0.24,0.25,0.66,0.77,0.86,0.94) -> X
> c(0.87,0.31,0.60,0.24,0.67,0.16,0.80,0.95,0.69,0.64) -> Y
> lm(Y~X) -> reg
> summary(reg)

Call:
lm(formula = Y ~ X)

Residuals:
     Min       1Q   Median       3Q      Max
-0.38751 -0.15731  0.03755  0.12624  0.39626

Coefficients:
            Estimate  Std. Error t value Pr(>|t|)
(Intercept)  0.4707     0.1245    3.780  0.00539 **
X            0.3074     0.2349    1.309  0.22701
```

We can see the p-value to test against the null hypothesis is 0.227, Thus for any $\alpha < 0.227$ we fail to reject the null $\beta_1 \neq 0$.

B



16/16

It is reasonable to fail to reject H_0 as the given (x, y) coordinates seem to have no relevant correlation.

6. Let $X_1, \cdots, X_n$ be iid normal with unknown μ and $\sigma = 1$.
(a) Find an unbiased estimator for e^{μ}.
(b) Is your estimator an UMVUE for e^{μ}? Why or why not?
Hint: It may be useful to look up a relevant moment-generating function.

A We know $\bar{X}$ is UMVUE for μ

A We know $\bar{X}$ is UMVUE for μ.

Similarly we know $\bar{X} \sim N(\mu, \sigma^2/n) = N(\mu, 1/n)$

The mgf for normal distributions is $e^{\mu t + \sigma^2 t^2/2}$

Thus $M_{\bar{X}}(t) = \mathbb{E}[e^{t\bar{X}}] = e^{\mu t + \frac{t^2}{2n}}$

setting $t=1$; $M_{\bar{X}}(1) = \mathbb{E}[e^{\bar{X}}] = e^{\mu + \frac{1}{2n}}$

Now $e^{\mu} = e^{\mu + \frac{1}{2n} - \frac{1}{2n}} = \dfrac{\mathbb{E}[e^{\bar{X}}]}{\mathbb{E}[e^{\frac{1}{2n}}]} = \mathbb{E}[e^{\bar{X} - \frac{1}{2n}}]$

As $\mathbb{E}[e^{\bar{X} - \frac{1}{2n}}] - e^{\mu} = 0$ we have our unbiased statistic $e^{\bar{X} - \frac{1}{2n}}$

B As $\bar{X}$ is a complete ~~&~~ sufficient statistic for μ, we can utilize the Lehmann Scheffe theroms as the function $h(S) = e^{S - \frac{1}{2n}}$ is a function of S along as $S = \dfrac{\sum_{i=1}^{n} x_i}{n}$ & n is known.

Thus $\boxed{e^{\bar{X} - \frac{1}{2n}} \text{ is UMVUE for } e^{\mu}}$

16/16