11/27 Kengo Kato

STSCI 5080 Practice Final Exam

Instructions:

- The actual exam is 2 and half hours long. In the exam, you can bring notes and textbooks, but can not bring any electronic device.
- When you are asked to find a confidence interval with (asymptotic) level $1-\alpha$, you need to find a confidence interval whose coverage probability is (approaching) $1-\alpha$, where $0 < \alpha < 1$ (e.g., $\alpha = 0.05$). Likewise, when you are asked to find a test with asymptotic level α, you need to find a test whose probability of the type I error is (approaching) α.
- You may use 1.96 for the 97.5%-quantile of the $N(0,1)$ distribution.

Problem 1. Circle the correct choice in each of the following questions.

(1) Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ for some $-\infty < \mu < \infty$ and $\sigma^2 > 0$, where $n \geq 2$. The sample mean and variance are defined by $\overline{X} = n^{-1} \sum_{i=1}^n X_i$ and $S^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \overline{X})^2$, respectively. What is the distribution of $\overline{X}$?

a. $N(\mu, \sigma^2)$ b. $N(\mu, \sigma^2/(n-1))$ $\boxed{\text{c.}}$ $N(\mu, \sigma^2/n)$ d. $N(0,1)$

(2) In the previous question, what is the distribution of $(n-1)S^2/\sigma^2$?

$\boxed{\text{a.}}$ $\chi^2(n-1)$ b. $\chi^2(n)$ c. $\chi^2(n+1)$ d. $N(\mu, \sigma^2/n)$

(3) Let $X_1, \ldots, X_n \sim Ex(\lambda)$ i.i.d. for some $\lambda > 0$. What is the likelihood function for λ?

a. $\lambda e^{-\lambda \sum_{i=1}^n X_i}$ $\boxed{\text{b.}}$ $\lambda^n e^{-\lambda \sum_{i=1}^n X_i}$ c. $e^{-\lambda \sum_{i=1}^n X_i}$ d. $\lambda^n e^{-\lambda^n \sum_{i=1}^n X_i}$

(4) Suppose that you are given a statistical model $\{f_\theta \mid -\infty < \theta < \infty\}$, and you have an estimator $\widehat{\theta}_n$ such that $\widehat{\theta}_n \xrightarrow{P} \theta/2$ as $n \to \infty$ for any value of θ. Find a consistent estimator for θ from the following choices. Only one of them is consistent.

a. $\widehat{\theta}_n$ b. $10\sqrt{\widehat{\theta}_n}$ $\boxed{\text{c.}}$ $2\widehat{\theta}_n$ d. $4\widehat{\theta}_n$

(5) You observe the numbers of failures occurring in machines in a factory in a year. Denote by X_i the number of failures of the i-th machine, and suppose that

$$X_1, \ldots, X_n \sim Po(\lambda) \text{ i.i.d.}$$

where n is the number of machines in the factory. If the MLE is $\widehat{\lambda} = 0.4$, what is the MLE of $P_\lambda(X_1 = 0)$?

$\boxed{\text{a.}}$ $e^{-0.4}$ b. $0.4e^{-0.4}$ c. $e^{0.4}$ d. $0.4e^{0.4}$

(6) Suppose that the heights of women in a certain population follow the normal distribution with mean μ and variance 9. If the sample size $n = 36$ and the sample mean $\overline{X}$ are given, then what is a confidence interval for μ whose coverage probability is 95%?

a. $[\overline{X}-0.49, \overline{X}+0.49]$ $\boxed{\text{b.}}$ $[\overline{X}-0.98, \overline{X}+0.98]$ c. $[\overline{X}-1.96, \overline{X}+1.96]$ d. $[\overline{X}-2.45, \overline{X}+2.45]$

Problem 2. Let f_θ be a pmf of the form

$$f_\theta(x) = \begin{cases} \frac{2}{3}\theta & \text{if } x = 0 \\ \frac{1}{3}\theta & \text{if } x = 1 \\ \frac{2}{3}(1-\theta) & \text{if } x = 2 \\ \frac{1}{3}(1-\theta) & \text{if } x = 3 \\ 0 & \text{elsewhere} \end{cases},$$

where $0 < \theta < 1$. We observe a random sample

$$(X_1, X_2, \ldots, X_{10}) = (3, 0, 2, 1, 3, 2, 1, 0, 2, 1) \tag{*}$$

from such a pmf.

(a) Find the joint pmf of $(X_1, X_2, \ldots, X_{10})$ at (*). You need to derive an explicit form of the joint pmf.

(b) Find the MLE of θ at (*).

(a) The joint pmf at (*) is

$$f_\theta(3) f_\theta(0) \cdots f_\theta(1) = \{f_\theta(0)\}^2 \{f_\theta(1)\}^3 \{f_\theta(2)\}^3 \{f_\theta(3)\}^2 = \frac{2^5}{3^{10}} \theta^5 (1-\theta)^5.$$

(Comment). If you only write

$$f_\theta(3) f_\theta(0) \cdots f_\theta(1)$$

or

$$f_\theta(X_1) f_\theta(X_2) \cdots f_\theta(X_{10}),$$

then you still get partial credit but not full credit. I will ignore small mistakes. For example, the constant $\frac{2^5}{3^{10}}$ is not important for the joint pmf as a function of θ, and so even if the constant is not correct, you will get full credit if the other parts are correct.

(b) The log likelihood function is

$$\ell(\theta) = \log \frac{2^5}{3^{10}} \theta^5 (1-\theta)^5 = \log \frac{2^5}{3^{10}} + 5\{\log \theta + \log(1-\theta)\}.$$

The FOC is

$$\ell'(\theta) = 5\left(\frac{1}{\theta} - \frac{1}{1-\theta}\right) = 0.$$

Solving the FOC, we obtain the MLE (maximum likelihood estimate)

$$\widehat{\theta} = \frac{1}{2}.$$

(Comment). The solution consists of two steps. The first step is to derive the FOC. So if you write something like $\ell'(\theta) = 0$, you will get partial credit. The second step is to find the MLE. If your FOC is explicit, then you can skip a derivation of the MLE (like the above solution). But if you only say that "Solving $\ell'(\theta) = 0$, we get $\widehat{\theta} = \frac{1}{2}$", then you will not get full credit.

Problem 3. Let

$$X_1, \dots, X_n \sim N(0, \theta) \text{ i.i.d.}$$

where $\theta > 0$ is unknown.

(a) Find the log likelihood function for θ. You need to derive an explicit form of the log likelihood function.

(b) Verify that the MLE of θ is $\widehat{\theta} = n^{-1} \sum_{i=1}^n X_i^2$.

(a) The pdf of $N(0, \theta)$ is

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/(2\theta)},$$

and so the joint pdf is

$$\prod_{i=1}^n f_\theta(x_i) = \frac{1}{(2\pi\theta)^{n/2}} e^{-\sum_{i=1}^n x_i^2/(2\theta)}.$$

The likelihood function is

$$L_n(\theta) = \frac{1}{(2\pi\theta)^{n/2}} e^{-\sum_{i=1}^n X_i^2/(2\theta)}.$$

The log likelihood function is

$$\ell_n(\theta) = \log L_n(\theta) = -\frac{n}{2} \log(2\pi\theta) - \frac{1}{2\theta} \sum_{i=1}^n X_i^2.$$

(Comment). The same rule as in Part (a) of Problem 2 applies.

(b) We note that

$$\ell_n'(\theta) = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n X_i^2.$$

So the FOC is

$$-\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^n X_i^2 = 0.$$

Solving the FOC w.r.t. θ, we have $\widehat{\theta} = n^{-1} \sum_{i=1}^n X_i^2$.

(Comment). The same rule as in Part (b) of Problem 2 applies.

Problem 3 (continued).

(c) Find the limiting distribution of $\sqrt{n}(\widehat{\theta}-\theta)$ as $n \to \infty$. You may use the fact that $E_\theta(X_1^2) = \theta$ and $E_\theta(X_1^4) = 3\theta^2$ without derivations.

(d) Find a confidence interval for θ with asymptotic level 95% by estimating the asymptotic variance given in Part (c). You don't need any derivation in this question and so only have to state your confidence interval.

(c) Since X_1^2 has mean θ and variance $\mathrm{Var}_\theta(X_1^2) = E_\theta(X_1^4) - \{E_\theta(X_1^2)\}^2 = 2\theta^2$, we have

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, 2\theta^2)$$

by CLT.

(Comment). In this example, $\widehat{\theta}$ is the sample mean of X_i^2's, and so it is clear that we can apply CLT. So you will get full credit if the limiting distribution is correct.

(d)

$$\left[\widehat{\theta} - \frac{1.96\widehat{\theta}\sqrt{2}}{\sqrt{n}}, \widehat{\theta} + \frac{1.96\widehat{\theta}\sqrt{2}}{\sqrt{n}}\right]$$

is a desired confidence interval.

Problem 3 (continued).

(e) Show that $\sqrt{n}\left(\frac{1}{\sqrt{2}}\log\widehat{\theta} - \frac{1}{\sqrt{2}}\log\theta\right) \xrightarrow{d} N(0,1)$ as $n \to \infty$.

(f) Find, with a brief justification, a confidence interval for θ with asymptotic level 95% using the result of Part (e).

(e) Let $g(\theta) = \frac{1}{\sqrt{2}}\log\theta$. Since $g'(\theta) = \frac{1}{\sqrt{2}\theta}$, we have

$$\sqrt{n}\left(\frac{1}{\sqrt{2}}\log\widehat{\theta} - \frac{1}{\sqrt{2}}\log\theta\right) = \sqrt{n}\{g(\widehat{\theta}) - g(\theta)\} \xrightarrow{d} N(0, \{g'(\theta)\}^2 2\theta^2) = N(0,1)$$

by the delta method.

(Comment). You need to show how you apply the delta method. For example, you will not be given full credit if you only write

$$\sqrt{n}\left(\frac{1}{\sqrt{2}}\log\widehat{\theta} - \frac{1}{\sqrt{2}}\log\theta\right) \xrightarrow{d} N(0, \{g'(\theta)\}^2 2\theta^2) = N(0,1),$$

without specifying what is $g(\theta)$ (though you will get partial credit). But if you correctly specify $g(\theta)$, then you will get full credit.

(f)

$$\left[\frac{1}{\sqrt{2}}\log\widehat{\theta} - \frac{1.96}{\sqrt{n}}, \frac{1}{\sqrt{2}}\log\widehat{\theta} + \frac{1.96}{\sqrt{n}}\right]$$

is a CI for $\frac{1}{\sqrt{2}}\log\theta$ with asymptotic level 95%. Since

$$\begin{aligned} &\frac{1}{\sqrt{2}}\log\theta \in \left[\frac{1}{\sqrt{2}}\log\widehat{\theta} - \frac{1.96}{\sqrt{n}}, \frac{1}{\sqrt{2}}\log\widehat{\theta} + \frac{1.96}{\sqrt{n}}\right] \\ &\Leftrightarrow \theta \in \left[\exp(\log\widehat{\theta} - 1.96\sqrt{2}/\sqrt{n}), \exp(\log\widehat{\theta} + 1.96\sqrt{2}/\sqrt{n})\right], \end{aligned}$$

a desired confidence interval is

$$\left[\exp(\log\widehat{\theta} - 1.96\sqrt{2}/\sqrt{n}), \exp(\log\widehat{\theta} + 1.96\sqrt{2}/\sqrt{n})\right].$$

(Comment). The derivation can be brief and so the above solution is enough to get full credit. But you need to derive an explicit CI without using $g(\theta)$.

Problem 4. Let $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d. where $\lambda > 0$ is unknown. You know that the MLE of λ is $\widehat{\lambda} = \overline{X} = n^{-1}\sum_{i=1}^n X_i$ and $\sqrt{n}(\widehat{\lambda} - \lambda) \xrightarrow{d} N(0, \lambda)$ as $n \to \infty$. Consider the following testing problem:

$$H_0 : \lambda = 2 \quad \text{vs.} \quad H_1 : \lambda \neq 2.$$

Find a test based on the MLE that has asymptotic level 5%. Justify that your test has asymptotic level 5%.

A desired test is given by

$$\left| \frac{\sqrt{n}(\widehat{\lambda} - 2)}{\sqrt{2}} \right| > 1.96 \Rightarrow \text{reject } H_0.$$

If $\lambda = 2$, we have

$$\frac{\sqrt{n}(\widehat{\lambda} - 2)}{\sqrt{2}} \xrightarrow{d} Z \sim N(0, 1),$$

and so

$$P_{\lambda=2}\left\{ \left| \frac{\sqrt{n}(\widehat{\lambda} - 2)}{\sqrt{2}} \right| > 1.96 \right\} \approx P(|Z| > 1.96) = 0.05.$$

This implies that the test has asymptotic level 5%.

(Comment). The derivation can be brief and so the above solution is enough to get full credit. But you need to derive an explicit test.

Problem 5. Suppose that you are a market researcher interested in the proportion p of a population that will buy a product, and ask n potential customers about their willingness to buy the product. Discuss how to determine the sample size n of the survey using a confidence interval for p. You need to derive an explicit way to determine the sample size.

(Comment). Summarize Lecture 22 slides pages 13–15. I will ask the same question in the actual exam.