STAT420 Homework 7

James Zhang*

December 6, 2023

*Email: jzhang72@terpmail.umd.com

12/07: HW 7 (due before 11:00 p.m.) & REVIEW (Final)

8.1(1,5,7); 8.2(5,11); 9.6(9,14); 9.8(1); 9.9(2)

8.1.1. In Example 8.1.2 of this section, let the simple hypotheses read $H_0 : \theta = \theta' = 0$ and $H_1 : \theta = \theta'' = -1$. Show that the best test of H_0 against H_1 may be carried out by use of the statistic $\overline{X}$, and that if $n = 25$ and $\alpha = 0.05$, the power of the test is 0.9996 when H_1 is true.

Solution.
By the Neyman-Pearson Theorem, we seek to show that $\frac{L(0;x)}{L(-1;x)} \leq k, k \in \mathbb{R}^+$. The likelihood function is

$$L(\theta; x) = (\frac{1}{\sqrt{2\pi}})^n \exp(-\sum_{i=1}^{n}(x_i - \theta)^2/2)$$

Testing $H_0 : \theta = \theta' = 0$ against $H_0 : \theta = \theta'' = -1$, we get

$$\frac{L(\theta'; x)}{L(\theta''; x)} = \frac{(1/\sqrt{2\pi})^n \exp(-\sum_{i=1}^{n} x_i^2/2)}{(1/\sqrt{2\pi})^n \exp(-\sum_{i=1}^{n}(x_i + 1)^2/2)}$$

$$= \exp(-\sum_{i=1}^{n} x_i^2/2 + \sum_{i=1}^{n}(x_i + 1)^2/2) = \exp(-\sum_{i=1}^{n} \frac{x_i^2}{2} + \sum_{i=1}^{n} \frac{x_i^2}{2} + x_i + \frac{1}{2})$$

$$= \exp(\frac{n}{2} + \sum_{i=1}^{n} x_i) \leq k$$

$$\frac{n}{2} + \sum_{1}^{n} x_i \leq \log k$$

$$\sum_{i}^{n} x_i \leq \log k - \frac{n}{2} = c$$

Note that this event is equivalent to the event

$$\bar{X} \leq \frac{c}{n} = c_1$$

and so we have showed that this test can be carried out by use of the statistic $\bar{X}$. Now, if $n = 25, \alpha = 0.05$, then to find the power we must first find the c corresponding to this value of $\alpha = 0.05$. Under H_0, $\bar{X} \sim N(0, \frac{1}{25})$. Thus, $\alpha =$P(Type I Error), which is when we reject H_0 when it is true.

$$P_{\theta=0}(\bar{X} \leq c_1) = 0.05 \implies c_1 = -0.32897$$

Therefore, we compute the power to be

$$1 - P_{\theta=-1}(\bar{X} \geq -0.32897) = 1 - 0.0004 = 0.9996$$

as desired.

□

8.1.5. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution having pdf of the form $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, show that a best critical region for testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$ is $C = \{(x_1, x_2, \ldots, x_n) : c \leq \prod_{i=1}^{n} x_i\}$.

Solution.
By the Neymean-Pearson Theorem, C is the best critical region of size α if $\alpha = P_{H_0}(X \in C)$ and

$$\text{(a) } \frac{L(1)}{L(2)} \leq k \ \forall \ x \in C$$

$$\text{(b) } \frac{L(1)}{L(2)} \geq k \ \forall \ x \in C^c$$

Let us first find the value of k that satisfies condition (a) and then we will confirm this is the best critical region by showing condition (b). The likelihood function is

$$L(\theta; x) = \prod_{i}^{n} \theta x_1^{\theta-1} = \theta^n (\prod_{1}^{n} x_i)^{\theta-1}$$

Thus, we test

$$\frac{L(1)}{L(2)} = \frac{1}{2^n(\prod_1^n x_i)} \leq k, k \in \mathbb{R}^+$$

Rewriting the inequality yields

$$\frac{1}{k2^n} \leq \prod_{1}^{n} x_i$$

and if we set $c = \frac{1}{k2^n}$ then

$$c \leq \prod_{1}^{n} x_i$$

as desired. We confirm that this is the best critical region by showing that $\frac{L(1)}{L(2)} \geq k \ \forall \ x \in C^c$. Thus,

$$c > \prod_{1}^{n} x_i \implies \frac{1}{k2^n} > \prod_{1}^{n} x_i \implies \frac{1}{k \prod_1^n x_i} > k \implies \frac{L(1)}{L(2)} > k$$

which shows condition (b) and proves that C is a best critical region for this specific test.

□

8.1.7. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution $N(\theta, 100)$. Show that $C = \{(x_1, x_2, \ldots, x_n) : c \leq \overline{x} = \sum_1^n x_i/n\}$ is a best critical region for testing $H_0 : \theta = 75$ against $H_1 : \theta = 78$. Find n and c so that

$$P_{H_0}[(X_1, X_2, \ldots, X_n) \in C] = P_{H_0}(\overline{X} \geq c) = 0.05$$

and

$$P_{H_1}[(X_1, X_2, \ldots, X_n) \in C] = P_{H_1}(\overline{X} \geq c) = 0.90,$$

approximately.

Solution.
First, let us show that $C = \{(x_1, \cdots, x_n) : c \le \bar{x}\}$ is in fact a best critical region for testing. Note that the pdf and likelihood functions are

$$f(x;\theta) = \frac{1}{10\sqrt{2\pi}} \exp(-\frac{(x-\theta)^2}{200})$$

$$L(\theta; x) = (\frac{1}{10\sqrt{2\pi}})^n \exp(\sum_1^n -(x_i - \theta)^2/200)$$

It is desired to test the hypothesis $H_0 : \theta = 75$ against the alternative hypothesis $H_1 : \theta = 78$ such that by the Neyman-Pearson Theorem,

$$\frac{L(75)}{L(78)} = \frac{(\frac{1}{10\sqrt{2\pi}})^n \exp(\sum_1^n -(x_i - 75)^2/200)}{(\frac{1}{10\sqrt{2\pi}})^n \exp(\sum_1^n -(x_i - 78)^2/200)}$$

$$= \exp(-\frac{1}{200}\sum_1^n (x_i - 75)^2 + \frac{1}{200}\sum_1^n (x_i - 78)^2) \le k$$

$$-\frac{1}{200}\sum_1^n (x_i - 75)^2 + \frac{1}{200}\sum_1^n (x_i - 78)^2 \le \log k$$

$$\sum_1^n (x_i - 78)^2 - (x_i - 75)^2 \le 200 \log k$$

$$\sum_1^n -156x_i + 78^2 + 150x_i - 75^2 \le 200 \log k$$

$$459n - 6\sum_1^n x_i \le 200 \log k$$

$$\sum_1^n x_i \ge \frac{200 \log k - 459n}{6}$$

$$c = \frac{200 \log k - 459n}{6n} \le \frac{1}{n}\sum_1^n x_i$$

which is of the given best critical region format, as desired. Now we will find c and n that satisfy the given probability computations. Recall that by normal sampling theory, $\bar{X} \sim N(\theta, \frac{100}{n})$. By properties of the standard normal distribution,

$$\bar{X} \sim N(\theta, \frac{100}{n}) \implies \frac{\sqrt{n}(\bar{X} - \theta)}{10} \sim N(0, 1)$$

The first condition we are given is

$$P_{H_0}(\bar{X} \ge c) = 0.05$$

$$P(\frac{\sqrt{n}(\bar{X} - 75)}{10} \ge \frac{\sqrt{n}(c - 75)}{10}) = 0.05$$

$$\frac{\sqrt{n}(c-75)}{10} = 1.64485$$

Now we apply a similar idea for the second condition

$$P_{H_1}(\bar{X} \geq c) = 0.90$$

$$P(\frac{\sqrt{n}(\bar{X}-78)}{10} \geq \frac{\sqrt{n}(c-78)}{10}) = 0.90$$

$$\frac{\sqrt{n}(c-78)}{10} = -1.28155$$

Now we have 2 equations and 2 unknowns so we can solve a system of equations

$$\frac{\frac{\sqrt{n}(c-75)}{10}}{\frac{\sqrt{n}(c-78)}{10}} = \frac{1.64485}{-1.28155} \approx -1.28348 = \frac{c-75}{c-78}$$

$$-1.28348(c-78) = c-75$$

$$c + 1.28348c = 1.28348(78) + 75$$

$$c \approx 76.686216$$

Setting this known value into one of the equations in and solving for n, we get

$$n = \frac{10(1.64485)}{76.686216-75} \approx 95.15384 \implies 96$$

□

8.2.5. Consider Example 8.2.2. Show that $L(\theta)$ has a monotone likelihood ratio in the statistic $\sum_{i=1}^{n} X_i^2$. Use this to determine the UMP test for $H_0 : \theta = \theta'$, where θ' is a fixed positive number, versus $H_1 : \theta < \theta'$.

Solution.
Example 8.2.2 defines $X_1, \cdots, X_n$ as a random sample from $N(0,\theta), \theta \in \mathbb{R}^+$. The likelihood function is

$$L(\theta; x) = (\frac{1}{2\pi\theta})^{n/2} \exp(-\frac{1}{2\theta}\sum_{1}^{n} x_i^2)$$

Now let us consider $\frac{L(\theta';x)}{L(\theta'';x)}$ if $\theta'' > \theta'$, as described in the given simple and alternative composite hypotheses of Example 8.2.2.

$$\frac{L(\theta';x)}{L(\theta'';x)} = \frac{(\frac{1}{2\pi\theta'})^{n/2}\exp(-\frac{1}{2\theta'}\sum_1^n x_i^2)}{(\frac{1}{2\pi\theta''})^{n/2}\exp(-\frac{1}{2\theta''}\sum_1^n x_i^2)} = (\frac{\theta''}{\theta})^{n/2}\exp(-(\frac{\theta''-\theta'}{2\theta'\theta''})\sum_1^n x_i^2)$$

$$= (\frac{\theta''}{\theta'})^{n/2}\exp((\frac{\theta'-\theta''}{2\theta'\theta''})\sum_1^n x_i^2)$$

Note that $\frac{\theta'-\theta''}{2\theta'\theta''} < 0$, given that $\theta' \in \mathbb{R}^+$ and $\theta'' > \theta'$. Thus, since $e^{-ax}, a < 0$ is a monotone decreasing function, the ratio is a decreasing function of $y = \sum x_i^2$. Thus, we have a monotone likelihood ratio in the statistic $Y = \sum X_i^2$

Alternatively, now let the alternative composite hypothesis be $H_1 : \theta < \theta'$ as given in Exercise 8.2.5. We wish to find the UMP test for $H_0 : \theta = \theta'$ against $H_1 : \theta < \theta'$. We revisit our ratio of likelihood functions, now considering that $\theta' > \theta''$ and not vice versa.

$$\frac{L(\theta'; x)}{L(\theta''; x)} = (\frac{\theta''}{\theta'})^{n/2} \exp((\frac{\theta' - \theta''}{2\theta'\theta''}) \sum_1^n x_i^2) \le k$$

Note that now, the ratio is a monotonically increasing a function of $Y = \sum X_i^2$ since the ratio $\frac{\theta'-\theta''}{2\theta'\theta''} > 0$. Taking the logs of both sides,

$$\frac{n}{2} \log(\frac{\theta''}{\theta'})(\frac{\theta' - \theta''}{2\theta'\theta''}) \sum_1^n x_i^2 \le \log k$$

$$\sum_1^n x_i^2 \le \frac{2\theta'}{\theta''}(\log k - \frac{n}{2}\log(\frac{\theta''}{\theta'}))$$

which is essentially the result of Example 8.2.2 with the inequality reversed. This is expected. Let α denote the significance level. By the proven claim following the definition of monotone likeliood ratio (mlr), the UMP level α decision rule for testing H_0 against H_1 is given by

$$\text{Reject } H_0 \text{ if } Y = \sum_1^n X_i^2 \le c$$

where c is derived from $\alpha = P_{\theta'}(Y = \sum_1^n X_i^2 \le c)$.

$\square$

8.2.11. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pdf $f(x;\theta) = \theta x^{\theta-1}$, $0 < x < 1$, zero elsewhere, where $\theta > 0$. Show the likelihood has mlr in the statistic $\prod_{i=1}^n X_i$. Use this to determine the UMP test for $H_0 : \theta = \theta'$ against $H_1 : \theta < \theta'$, for fixed $\theta' > 0$.

Solution.
The likelihood function is

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \theta x_i^{\theta-1} = \theta^n \prod_{i=1}^n x_i^{\theta-1}$$

Now consider

$$\frac{L(\theta')}{L(\theta'')} = \frac{\theta'^n \prod_{i=1}^n x_i^{\theta'-1}}{\theta''^n \prod_{i=1}^n x_i^{\theta''-1}} = (\frac{\theta'}{\theta''})^n \prod_{i=1}^n (x_i)^{\theta'-\theta''}$$

On the interval $0 < x < 1$, if $\theta' > \theta''$, then the ratio is monotone increasing. Similarly, on the interval, if $\theta' < \theta''$, then the ratio is monotone decreasing. Either way, we have a monotone likelihood ratio (mlr) in the statistic $Y = \prod_1^n X_i$. To find the UMP test, we proceed with the Neyman-Pearson theorem,

$$(\frac{\theta'}{\theta''})^n \prod_{i=1}^n (x_i)^{\theta'-\theta''} \le k$$

$$(\frac{\theta'}{\theta''})^n(\prod_{i=1}^n x_i)^{\theta'-\theta''} \leq k$$

Note that we are given $\theta'' < \theta' \implies$ the ratio is monotone increasing. Taking the logs of both sides,

$$n\log(\frac{\theta'}{\theta''}) + (\theta' - \theta'')\sum_{i=1}^n x_i \leq \log k$$

$$\sum_{i=1}^n x_i \leq \frac{\log k - n\log(\frac{\theta'}{\theta''})}{\theta' - \theta''} = c$$

Let the significance level be α. Therefore, the UMP test is

$$\text{Reject } H_0 \text{ if } Y = \prod_1^n X_i \leq c$$

where c is derived from $\alpha = P_{\theta'}(\prod_1^n X_i \leq c)$.

9.6.9. Show that

$$\sum_{i=1}^n [Y_i - \alpha - \beta(x_i - \overline{x})]^2 = n(\hat{\alpha} - \alpha)^2 + (\hat{\beta} - \beta)^2 \sum_{i=1}^n (x_i - \overline{x})^2 + \sum_{i=1}^n [Y_i - \hat{\alpha} - \hat{\beta}(x_i - \overline{x})]^2.$$

Solution.

$$\sum_{i=1}^n [Y_i - \alpha - \beta(x_i - \bar{x})]^2$$

$$= \sum_{i=1}^n Y_i^2 - 2\alpha Y_i - 2Y_i\beta(x_i - \bar{x}) + \alpha^2 + 2\alpha\beta(x_i - \bar{x}) + \beta^2(x_i - \bar{x})^2$$

Breaking up the summation into 2 separate summations: one containing alphas and one containing betas

$$= \sum_{i=1}^n Y_i^2 - 2\alpha Y_i + \alpha^2 - \sum_{i=1}^n 2Y_i\beta(x_i - \bar{x}) + 2\alpha\beta(x_i - \bar{x}) + \beta^2(x_i - \bar{x})^2$$

Recall that $\hat{\alpha} = \bar{Y} = \frac{1}{n}\sum_1^n Y_i$.

□

9.6.14. Fit $y = a + x$ to the data

x	0	1	2
y	1	3	4

by the method of least squares.

Solution.

a.

□

□