Solutions to Assignment of Chapter 5

D. Ding

April 2022

5.1 A single observation of a random variable having a hypergeometric distribution with $N = 7$ and $n = 2$ is used to test the null hypothesis $k = 2$ against the alternative hypothesis $k = 4$. If the null hypothesis is rejected if and only if the value of the random variable is 2, find the probabilities of type I and type II errors. $\frac{C_k^x}{C}$

Solution. Note that the probability distribution of X is given by

$$\mathbb{P}(X = x) = \frac{C_k^x \cdot C_{N-k}^{n-x}}{C_N^n}, \quad x = 0, 1, \ldots .n, \ x \le k, \ n - x \le N - k,$$

where $N = 7$ and $n = 2$. Since the null hypothesis $H_0 : k = 2$ is to be rejected in favor of the alternative hypothesis $H_1 : k = 4$ when $x = 2$, the probabilities of types I and II errors, are given respectively by

$$\alpha = \mathbb{P}(X = 2; k = 2) = \frac{C_2^2 \cdot C_5^0}{C_7^2} = \frac{1}{21},$$

and

$$\beta = \mathbb{P}(X < 2; k = 4) = \frac{C_4^0 \cdot C_3^2}{C_7^2} + \frac{C_4^1 \cdot C_3^1}{C_7^2} = \frac{3}{31} + \frac{12}{21} = \frac{5}{7}. \qquad \square$$

5.2 A single observation of a random variable X having a geometric distribution is used to test the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1 > \theta_0$. If the null hypothesis H_0 is rejected if and only if the observed value of the random variable X is greater that or equal to the positive integer k, find expressions for the probabilities of type I and type II errors.

Solution. Note that the probability distribution of X is given by

$$\mathbb{P}(X = x) = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, 3, \ldots .$$

Since the null hypothesis $H_0 : \theta = \theta_0$ is rejected if and only if $x \ge k$, the expressions for α and β, the probabilities of types I and II errors, are given respectively by

$$\alpha = \mathbb{P}(X \ge k; \theta_0) = \sum_{x=k}^{\infty} \theta_0 (1 - \theta_0)^{x-1} = (1 - \theta_0)^{k-1},$$

and

$$\beta = \mathbb{P}(X < k; \theta_1) = \sum_{x=1}^{k-1} \theta_1 (1 - \theta_1)^{x-1} = 1 - (1 - \theta_1)^{k-1}. \qquad \square$$

5.3 Let X_1, X_2, X_3, X_4 constitute a random sample from a normal population with $\sigma^2 = 4$. Assume that the null hypothesis $H_0 : \mu = \mu_0$ is to be rejected in favor of the alternative hypothesis $H_1 : \mu = \mu_1 > \mu_0$ when $\bar{x} > \mu_0 + 1.96$.

(1) Find the probability of committing type error I.

(2) If $\mu_1 = \mu_0 + 3.5$, find the probability of committing type II error.

Solution. (1) Since the population has the normal distribution, according to proposition 2.1.2 (Theorem 8.4), we have that $\bar{X} \sim N(\mu, \sigma^2/n)$. Thus, the probability of committing type I error, we have

$$\alpha = \mathbb{P}\big(\bar{X} > \mu_0 + 1.96; \mu_0\big) = \mathbb{P}\Big(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{(\mu_0 + 1.96) - \mu_0}{\sigma/\sqrt{n}}\Big) = \mathbb{P}\Big(Z > \frac{1.96\sqrt{n}}{\sigma}\Big),$$

where Z has the standard normal distribution. Substituting $\sigma = 2$ and $n = 4$, we get

$$\alpha = \mathbb{P}\big(Z > 1.96\big) = 0.025.$$

(2) The probability β of committing type II error is given by

$$\beta = \mathbb{P}\big(\bar{X} \leq \mu_0 + 1.96; \mu_1\big) = \mathbb{P}\Big(\frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} > \frac{(\mu_0 + 1.96) - \mu_1}{\sigma/\sqrt{n}}\Big) = \mathbb{P}\Big(Z > \frac{1.96 - (\mu_1 - \mu_0)}{\sigma/\sqrt{n}}\Big),$$

Substituting $\sigma = 2$, $n = 4$, and $\mu_1 - \mu_0 = 3.5$, we get

$$\beta = \mathbb{P}\big(Z \leq -1.54\big) = 0.0616. \qquad \square$$

5.4 A single observation of a random variable X having a uniform density with $\alpha = 0$ is used to test the null hypothesis $H_0 : \beta = \beta_0$ against the alternative hypothesis $H_1 : \beta = \beta_0 + 2$. Assume that the null hypothesis H_0 is rejected if and only if the random variable X takes on a value greater than $\beta_0 + 1$. Find the probabilities of type I and type II errors.

Solution. Note that the density of X is given by

$$f(x;\beta) = 1/\beta, \quad \text{if} \quad 0 \leq x \leq \beta, \quad \text{and} \quad f(x;\beta) = 0, \quad \text{elsewhere.}$$

The probability of type I error is given by

$$\mathbb{P}\big(X \geq \beta_0 + 1; \beta_0\big) = \int_{\beta_0+1}^{\infty} f(x;\beta_0)\mathrm{d}x = 0,$$

and the probability of type II is given by

$$\mathbb{P}\big(X < \beta_0 + 1; \beta_0 + 2\big) = \int_{-\infty}^{\beta+1} f(x;\beta_0 + 2)\mathrm{d}x = \frac{\beta_0 + 1}{\beta_0 + 2}. \qquad \square$$

5.5 A random sample of size n from an exponential population is used to test the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta = \theta_1 > \theta_0$. Find the most powerful critical region of size α.

Solution. Two likelihood functions are given by

$$L_0 = \Big(\frac{1}{\theta_0}\Big)^n \exp\Big\{ -\frac{1}{\theta_0}\sum_{k=1}^{n} x_k\Big\} \quad \text{and} \quad L_1 = \Big(\frac{1}{\theta_1}\Big)^n \exp\Big\{ -\frac{1}{\theta_1}\sum_{k=1}^{n} x_k\Big\}.$$

According to the Neyman-Pearson's lemma, the most powerful critical region C must be given by

$$\frac{L_0}{L_1} \leq k, \quad \text{inside } C; \quad \text{and} \quad \frac{L_0}{L_1} > k, \quad \text{outside} \quad C,$$

where k is a constant. Thus, we get

$$\frac{L_0}{L_1} = \left(\frac{\theta_1}{\theta_0}\right)^n \exp\left\{-\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{k=1}^{n} x_k\right\} \le k, \quad \text{inside} \quad C;$$

or,

$$n \log\left(\frac{\theta_1}{\theta_0}\right) - \left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum_{k=1}^{n} x_k \le \log k, \quad \text{inside} \quad C.$$

Since $\theta_1 > \theta_0$, we have

$$n\bar{x} = \sum_{k=1}^{n} x_k \ge K, \quad \text{inside} \quad C,$$

where K is an expression in k, n, θ_0, and θ_1, i.e.,

$$K = \frac{\theta_0\theta_1}{\theta_1 - \theta_0}\left[n \log\left(\frac{\theta_0}{\theta_1}\right) - \log k\right].$$

Since the size of the critical region is α, the constant K satisfies the following equation:

$$\mathbb{P}\big(n\bar{X} \le K; \theta_0\big) = \alpha.$$

From Example 7.16 in Section 1.4, we have known that the random variable $n\bar{X}$ has a gamma distribution with the parameters $\alpha = n$ and $\beta = \theta_0$. Thus, we can find the constant K by solving the following equation:

$$\alpha = \int_K^\infty \frac{1}{\theta_0^n \Gamma(n)} x^{n-1} e^{-x/\theta_0} \mathrm{d}x.$$

Since the chi-square distribution is a gamma distribution with $\beta = 2$, if $\theta_0 = 2$ then the random variable $n\bar{X}$ has a chi-square distribution with the $\nu = 2n$ degrees of freedom, and so that we have

$$K = \chi^2_{\alpha,2n}.$$

□

5.6 A single observation of a random variable having a geometric distribution is to be used to test the null hypothesis that its parameter equals θ_0 against the alternative that it equals $\theta_1 > \theta_0$. Use the NeymanPearson lemma to find the best critical region of size α.

Solution. Since the probability distribution of geometric distribution is given by

$$g(x; \theta) = \theta(1-\theta)^{x-1}, \quad x = 1, 2, \ldots,$$

two likelihood functions are given by

$$L_0 = \theta_0(1-\theta_0)^{x-1}, \quad \text{and} \quad L_1 = \theta_1(1-\theta_1)^{x-1}.$$

According to the Neyman-Pearson's lemma, the most powerful critical region C must be given by

$$\frac{L_0}{L_1} \le k, \text{ inside } C; \quad \frac{L_0}{L_1} > k, \text{ outside } C,$$

where k is a constant. Thus, we get

$$\frac{L_0}{L_1} = \left(\frac{\theta_0(1-\theta_1)}{\theta_1(1-\theta_0)}\right)\left(\frac{1-\theta_0}{1-\theta_1}\right)^x \le k, \quad \text{inside} \quad C.$$

Thus, we have

$$\log\left(\frac{\theta_0(1-\theta_1)}{\theta_1(1-\theta_0)}\right) + x\cdot\log\left(\frac{1-\theta_0}{1-\theta_1}\right) \le \log k, \quad \text{inside} \quad C.$$

$x\log(\quad) \le \log k -$

Therefore, we have

$$x \le K, \text{ inside } C,$$

where K is an expression in k, θ_0, and θ_1, i.e.,

$$K = \frac{\log k - \log\left(\frac{\theta_0(1-\theta_1)}{\theta_1(1-\theta_0)}\right)}{\log\left(\frac{1-\theta_0}{1-\theta_1}\right)}. \quad \square$$

5.7 Given a random sample of size n from a normal population with $\mu = 0$, construct the most powerful critical region of size α to test the null hypothesis $H_0 : \sigma = \sigma_0$ against the alternative hypothesis $H_1 : \sigma = \sigma_1 > \sigma_0$.

Solution. Two likelihood functions are given by

$$\begin{aligned} L_0 &= \left(\frac{1}{\sigma_0\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{k=1}^n x_k^2\right\}, \\ L_1 &= \left(\frac{1}{\sigma_1\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma_1^2}\sum_{k=1}^n x_k^2\right\}. \end{aligned}$$

According to the Neyman-Pearson's lemma, the most powerful critical region C must be given by

$$\frac{L_0}{L_1} \le k, \text{ inside } C; \quad \frac{L_0}{L_1} > k, \text{ outside } C,$$

where k is a constant. Thus, we get

$$\frac{L_0}{L_1} = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left\{-\left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right)\sum_{k=1}^n x_k^2\right\} \le k, \quad \text{inside} \quad C.$$

Thus, we have

$$n\log\left(\frac{\sigma_1}{\sigma_0}\right) - \left(\frac{1}{2\sigma_0^2} - \frac{1}{2\sigma_1^2}\right)\sum_{k=1}^n x_k^2 \le \log k, \quad \text{inside} \quad C.$$

Since $\sigma_1 > \sigma_0$, we have

$$\sum_{k=1}^n x_k^2 \ge K, \text{ inside } C,$$

where K is an expression in k, n, σ_0, and σ_1, i.e.,

$$K = \frac{2(\sigma_0\sigma_1)^2}{\sigma_1^2 - \sigma_0^2}\left[n\log\frac{\sigma_1}{\sigma_0} - \log k\right].$$

On the other hand, if the null hypothesis $H_0 : \sigma = \sigma_0$ is true, according to Proposition 2.3.3 (Theorem 8.8) in Section 2.3, we see that

$$\frac{1}{\sigma_0^2}\sum_{k=1}^n X_k^2 \sim \chi^2(n).$$

$$\sum\left(\frac{X_k^2 - 0}{\sigma_0^2}\right) = \sum Y_k^2 \qquad \xi = \sum Y_k^2 \sim$$

Therefore, we have

$$\alpha = \mathbb{P}\Big(\sum_{k=1}^{n} X_k^2 \geq K; \sigma_0\Big) = \mathbb{P}\Big(\frac{1}{\sigma_0^2}\sum_{k=1}^{n} X_k^2 \geq \frac{K}{\sigma_0^2}; \sigma_0\Big),$$

since the size of the critical region C is α. This implies

$$K = \sigma_0^2 \chi^2_{\alpha,n}. \qquad \square$$

5.8 Suppose that a random sample $X_1, \ldots, X_n$ of size $n = 16$ from a normal population with the known variance $\sigma = 3$ is used to test the null hypothesis $H_0 : \mu = 5.125$ against the alternative hypothesis $H_1 : \mu = 2$.

(1) Find the most powerful critical region C of size $\alpha = 0.05$.

(2) Find the probabilities of committing the type I and type II errors, respectively.

Solution. (1) The two likelihood functions are given by

$$L_0 = \Big(\frac{1}{\sigma\sqrt{2\pi}}\Big)^n \exp\Big\{-\frac{1}{2}\sum_{k=1}^{n}\frac{(x_k-\mu_0)^2}{\sigma^2}\Big\},$$

$$L_1 = \Big(\frac{1}{\sigma\sqrt{2\pi}}\Big)^n \exp\Big\{-\frac{1}{2}\sum_{k=1}^{n}\frac{(x_k-\mu_1)^2}{\sigma^2}\Big\}.$$

1

According to the Neyman-Pearson's lemma, the most powerful critical region C must be given by

$$\frac{L_0}{L_1} \leq k, \text{ inside } C; \quad \frac{L_0}{L_1} > k, \text{ outside } C,$$

where k is a constant. Thus, we get

2

$$\exp\Big\{-\frac{n}{2}(\mu_0^2-\mu_1^2) + (\mu_0-\mu_1)\sum_{k=1}^{n} x_k\Big\} \leq k, \quad \text{inside } C.$$

Since $(\mu_0 - \mu_1) = 3.125 > 0$, we have

$e^{\{\ \}} \leq k$ $\quad \{\ \} \leq \log k$ $\quad \bar{X} = \frac{1}{n}\sum x$

3

$$\bar{x} \leq K, \text{ inside } C,$$

where K is an expression in k, n, μ_0, and μ_1, i.e.,

$$K = \frac{\log k + \frac{1}{2}n(\mu_0^2-\mu_1^2)}{n(\mu_0-\mu_1)}.$$

Particularly

Since the size of C is $\alpha = 0.05$, $\sigma = 3$, $\mu_0 = 5.125$, and $n = 16$, according to Theorem 8.4 in Section 2.1, if the null hypothesis H_0 is true, we have

$$\bar{X} \sim N\big(\mu_0, \frac{\sigma}{\sqrt{n}}\big) = N\big(5.125, \frac{3}{4}\big).$$

Thus, we get

$$\alpha = \mathbb{P}(\bar{X} \leq K; \mu_0) = \mathbb{P}\Big(\frac{\bar{X}-\mu_0}{3/4} > \frac{K-\mu_0}{3/4}\Big) = \mathbb{P}\Big(Z \leq \frac{K-\mu_0}{(3/4)}\Big),$$

½ tail

where Z is a standard normal random variable. From the statistical table of the standard normal distribution, we obtain

$$\cancel{\frac{3}{4}}(K - 5.125) = -z_{0.05} = -1.645$$

$\frac{4}{3}(k - 5.125) =$

and so that $K = 3.891$. Hence, the most powerful critical region C is given by

$$\cancel{\bar{x} \leq} 3.891.$$

hence $\bar{x}$

(2) According to the definition, the probability of committing the type I error is the size of the critical region C. Thus, the probability $\alpha = 0.05$; On the other hand, the probability of committing the type II error is given by

$$\beta = \mathbb{P}(\bar{X} > K; \mu_1) = \mathbb{P}\Big(\frac{\bar{X} - 2}{3/4} > \frac{K - \mu_1}{3/4}\Big) = \mathbb{P}\Big(Z > \frac{3.891 - 2}{3/4}\Big) = \mathbb{P}(Z > 2.521),$$

$\geq 0.$

and so that $\beta = 0.0059$. The problem is solved. □

5.9 A biologist wants to test the null hypothesis that the mean wingspan of a certain kind insect is 12.3 mm against the alternative that it is not 12.3 mm. If she takes a random sample and decides to accept the null hypothesis if and only if the mean of the sample falls between 12.0 mm and 12.6 mm, what decision will she make if she gets the sample mean of 12.9 mm and will it be in error if (1) $\mu = 12.5$ mm? and (2) $\mu = 12.3$ mm?

Solution. (1) She will reject the null hypothesis, and her decision is correct.

(2) She will reject the null hypothesis, but her decision is not correct (type I error). □

5.10 An employee of a bank wants to test the null hypothesis that on the average the bank cashes 10 bad checks per day against the alternative that this figure is too small. If he takes a random sample and decides to reject the null hypothesis if and only if the mean of the sample exceeds 12.5, what decision will he make if he gets the sample mean of 11.2, and will it be in error if (1) $\lambda = 11.5$? and (2) $\lambda = 10.0$?

Solution. (1) He will accept the null hypothesis, but his decision is not correct (type II error).

(2) He will accept the null hypothesis, and his decision is correct. □

5.11 A random sample of size n is to be used to test the null hypothesis that the parameter θ of an exponential population equals θ_0 against the alternative that it does not equal θ_0.

(1) Find an expression for the likelihood ratio statistic.

(2) Use the result of part (1) to show that the critical region of the likelihood ratio test can be written as

$$\bar{x} \cdot e^{-\bar{x}/\theta_0} \leq K.$$

Solution. (1) We shall test the null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta \neq \theta_0$. Since the density of an exponential distribution is given by

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & \text{for } x > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\theta > 0$ is the parameter. Thus, we have

$L(\theta)$

then

$$\begin{aligned}
\max L_0(x_1, \ldots, x_n) &= \prod_{k=1}^{n} f(x_k; \theta_0) = \frac{1}{\theta_0^n} \exp\Big\{-\frac{1}{\theta_0} \sum_{k=1}^{n} x_k\Big\}, \\
\max L(x_1, \ldots, x_n) &= \prod_{k=1}^{n} f(x_k; \hat{\theta}) = \frac{1}{\hat{\theta}^n} \exp\Big\{-\frac{1}{\hat{\theta}_0} \sum_{k=1}^{n} x_k\Big\},
\end{aligned}$$

where $\hat{\theta}$ is the maximum likelihood estimation of θ. According to Example 10.16 in Section 3.3, we have known that $\hat{\theta} = \bar{x}$. Thus, we get the expression for the likelihood ratio statistic

$$\begin{aligned}
\lambda(x_1, \ldots, x_n) &= \frac{\max L_0(x_1, \ldots, x_n)}{\max L(x_1, \ldots, x_n)} = \frac{\frac{1}{\theta_0^n} \exp\Big\{-\frac{1}{\theta_0} \sum_{k=1}^{n} x_k\Big\}}{\frac{1}{\bar{x}^n} \exp\Big\{-\frac{1}{\bar{x}} \sum_{k=1}^{n} x_k\Big\}} \\
&= \Big(\frac{\bar{x}}{\theta_0}\Big)^n \exp\Big\{-\frac{n\bar{x}}{\theta_0} + n\Big\}.
\end{aligned}$$

(2) The critical region of this likelihood ratio test is given by

$$\lambda(x_1, \ldots, x_n) = \Big(\frac{\bar{x}}{\theta_0}\Big)^n \exp\Big\{-\frac{n\bar{x}}{\theta_0} + n\Big\} \leq k,$$

or

$$\bar{x}e^{-\bar{x}/\theta_0} \leq \theta_0 \sqrt[n]{ke^{-n}} = K. \quad \square$$

5.12 The number of successes in n trials is to be used to test the null hypothesis that the parameter θ of a binomial population equals $\frac{1}{2}$ against the alternative that it does not equal $\frac{1}{2}$.

(1) Find an expression for the likelihood ratio statistic.

(2) Use the result of part (1) to show that the critical region of the likelihood ratio test can be written as

$$x \cdot \log x + (n-x) \cdot \log(n-x) \geq K,$$

where x is the observed number of successes, and K is a constant that depends on the size of the critical region.

(3) Study the graph of $f(x) = x \cdot \log x + (n-x) \cdot \log(n-x)$, in particular its minimum and its symmetry, to show that the critical region of this likelihood ratio test can also be written as

$$\Big|x - \frac{n}{2}\Big| \geq C,$$

where C is a constant such that $f(C) = K$.

Solution. (1) We shall test the null hypothesis $H_0 : \theta = \frac{1}{2}$ against the alternative hypothesis $H_1 : \theta \neq \frac{1}{2}$. Note that the likelihood function is given by

$$L(x; \theta) = C_n^x \theta^x (1-\theta)^{n-x}.$$

We have

$$\max L_0(x) = L\Big(x; \frac{1}{2}\Big) = C_n^x \Big(\frac{1}{2}\Big)^n.$$

and $\max L(x) = L(x; \hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimation of θ. Since

$$\log L(\theta) = \log C_n^x + x \log \theta + (n-x) \log(1-\theta).$$

Differentiating the function $\log L(\theta)$ yields

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\log L(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta}.$$

Equating the derivative to zero and solving the equation, we get that $\hat{\theta} = x/n$. Thus, we have

$$\max L(x) = L\big(x;\hat{\theta}\big) = C_n^x\left(\frac{x}{n}\right)^x\left(\frac{n-x}{n}\right)^{n-x}.$$

Therefore, we get the expression for the likelihood ratio statistic

$$\lambda(x) = \frac{\max L_0(x)}{\max L(x)} = \frac{C_n^x\left(\frac{1}{2}\right)^n}{C_n^x\left(\frac{x}{n}\right)^x\left(\frac{n-x}{n}\right)^{n-x}} = \frac{(n/2)^n}{x^x(n-x)^{n-x}}.$$

(2) The critical region of this likelihood ratio test is given by

$$\lambda(x) = \frac{(n/2)^n}{x^x(n-x)^{n-x}} \le k,$$

So, we get

$$-n\log 2 + n\log n - x\log x - (n-x)\log(n-x) \le \log k.$$

or

$$x\log x + (n-x)\log(n-x) \ge n\log n - n\log 2 - \log k = K,$$

K is a constant that depends on the size of the critical region.

(3) Consider the function:

$$f(x) = x\log x + (n-x)\log(n-x).$$

From the equation

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \log x + 1 - \log(n-x) - 1 = \log x - \log(n-x) = 0,$$

we get that $x = n - x$, and so that $x = n/2$ is the minimum point of $f(x)$. Also, we have that $f(n-x) = f(x)$, i.e. $f(x)$ is symmetrical about $x = n/2$. Hence, we conclude that there exists a constant $C = f(K)$ such that

$$f(x) = x\log x + (n-x)\log(n-x) \ge K \Longleftrightarrow \left|x - \frac{n}{2}\right| \ge C.$$

Therefore, the critical region of this likelihood ratio test can also be written as

$$\left|x - \frac{n}{2}\right| \ge C,$$

where C is a constant such that $f(C) = K$. □