Autumn Semester 2021 **December 31, 2021**

IE 605: Engineering Statistics

Solutions of tutorial 11

Solution 1

1. **Given:** $L(x)$ and $U(x)$ satisfy $\mathbb{P}\{L(X) \leq \theta\} = 1-\alpha_1$ and $\mathbb{P}\{U(X) \geq \theta\} = 1-\alpha_2$, and $L(x) \leq U(x)$ for all x.

   To show: $\mathbb{P}\{L(X) \leq \theta \leq U(X)\} = 1-\alpha_1-\alpha_2$.

   Solution:

$$\begin{aligned} \text{Denote } A &= \{x : L(x) \leq \theta\} \\ \text{and } B &= \{x : U(x) \geq \theta\}. \\ \text{Then } A \cap B &= \{x : L(x) \leq \theta \leq U(x)\} \\ \text{and } 1 \geq \mathbb{P}\{A \cup B\} &= \mathbb{P}\{L(X) \leq \theta \text{ or } \theta \leq U(X)\} \\ &\geq \mathbb{P}\{L(X) \leq \theta \text{ or } \theta \leq L(X)\} = 1, \quad \text{since } L(x) \leq U(x), \forall x. \\ \text{Therefore, } P(A \cap B) &= P(A) + P(B) P(A \cup B) \\ &= 1-\alpha_1 + 1 - \alpha_2 - 1 = 1 - \alpha_1 - \alpha_2. \end{aligned}$$

2. **Given:** T is a continuous random variable with cdf $F_T(t|\theta)$ and $\alpha_1 + \alpha_2 = \alpha$.

   To show: An α level of acceptance region of the hypothesis $H_0 : \theta = \theta_0$ is $\{t : \alpha_1 \leq F_T(t|\theta_0) \leq 1-\alpha_2\}$, with associated confidence $(1-\alpha)$ set $\{\theta : \alpha_1 \leq F_T(t|\theta) \leq 1-\alpha_2\}$.

   Solution: Recall that if θ is the true parameter, then $F_T(T|\theta) \sim uniform(0,1)$.

   Thus, $\mathbb{P}_{\theta_0}(\{T : \alpha_1 \leq F_T(T|\theta_0) \leq 1-\alpha_2\}) = P(\alpha_1 \leq U \leq 1-\alpha_2) = 1-\alpha_2-\alpha_1,$

   where $U \sim uniform(0,1)$.

$$\text{Since } t \in \{t : \alpha_1 \leq F_T(t|\theta_0) \leq 1-\alpha_2\} \iff \theta \in \{\theta : \alpha_1 \leq F_T(t|\theta) \leq 1-\alpha_2\}$$

   the same calculation shows that the interval has confidence $1-\alpha_2-\alpha_1$.

Solution 2

Given: Let $X_1, \ldots, X_n$ be a random sample from a $N(0, \sigma_X^2)$ and $Y_1, \ldots, Y_n$ be a random sample from a $N(0, \sigma_Y^2)$, independent of the $X's$. Define $\lambda = \frac{\sigma_Y^2}{\sigma_X^2}$.

1. **To find:** The level α LRT of $H_0 : \lambda = \lambda_0$ vs $H_1 : \lambda \neq \lambda_0$.

   Solution: $\lambda(x, y) = \frac{\sup_{\lambda=\lambda_0} L(\sigma_X^2, \sigma_Y^2|x,y)}{\sup_{\lambda\in(0,+\infty)} L(\sigma_X^2, \sigma_Y^2|x,y)}$

   The unrestricted MLEs of σ_X^2 and σ_Y^2 are $\hat{\sigma}_X^2 = \frac{\sum X_i^2}{n}$ and $\hat{\sigma}_Y^2 = \frac{\sum Y_i^2}{m}$, as usual. Under the restriction, $\lambda = \lambda_0$, $\sigma_Y^2 = \lambda_0 \sigma_X^2$, and

   $$\begin{aligned} L(\sigma_X^2, \lambda_0\sigma_X^2|x, y) &= (2\pi\sigma_X^2)^{-n/2}(2\pi\lambda_0\sigma_X^2)^{-m/2}e^{-\sum x_i^2/(2\sigma_X^2)}.e^{-\sum y_i^2/(2\lambda_0\sigma_X^2)} \\ &= (2\pi\sigma_X^2)^{-(m+n)/2}\lambda_0^{-m/2}e^{-(\lambda_0\sum x_i^2+\sum y_i^2)/(2\lambda_0\sigma_X^2)} \end{aligned}$$

   Differentiating the log likelihood gives

   $$\begin{aligned} \frac{d\log L}{d(\sigma_X^2)^2} &= \frac{d}{d\sigma_X^2}\left[-\frac{m+n}{2}\log\sigma_X^2 - \frac{m+n}{2}\log(2\pi) - \frac{m}{2}\log\lambda_0 - \frac{\lambda_0\sum x_i^2 + \sum y_i^2}{2\lambda_0\sigma_X^2}\right] \\ &= -\frac{m+n}{2}(\sigma_X^2)^{-1} + \frac{\lambda_0\sum x_i^2 + \sum y_i^2}{2\lambda_0}(\sigma_X^2)^{-2} \quad \overset{\text{set}}{=} \quad 0 \end{aligned}$$

   which implies
   $$\hat{\sigma}_0^2 = \frac{\lambda_0\sum x_i^2 + \sum y_i^2}{\lambda_0(m+n)}$$

   To see this is a maximum, check the second derivative:

   $$\begin{aligned} \frac{d^2\log L}{d(\sigma_X^2)^2} &= \frac{m+n}{2}(\sigma_X^2)^{-2} - \frac{1}{\lambda_0}(\lambda_0\sum x_i^2 + \sum y_i^2)(\sigma_X^2)^{-3}\Bigg|_{\sigma_X^2=\hat{\sigma}_0^2} \\ &= -\frac{m+n}{2}(\hat{\sigma}_0^2)^{-2} < 0, \end{aligned}$$

   therefore $\hat{\sigma}_0^2$ is the MLE. The LRT statistic is

   $$\frac{(\hat{\sigma}_X^2)^{n/2}(\hat{\sigma}_Y^2)^{m/2}}{\lambda_0^{m/2}(\hat{\sigma}_0^2)^{(m+n)/2}},$$

   and the test is: Reject H_0 if $\lambda(x, y) < k$, where k is chosen to give the test size α.

2. **Express:** The rejection region of the LRT of part (1) in terms of F-distributed random variable.

   Under H_0, $\sum Y_i^2/(\lambda_0\sigma_X^2) \sim \chi_m^2$ and $\sum X_i^2/\sigma_X^2 \sim X_n^2$, independent. Also, we can write

   $$\begin{aligned} \lambda(X, Y) &= \left(\frac{1}{\frac{n}{m+n} + \frac{(\sum Y_i^2/\lambda_0\sigma_X^2)/m}{(\sum X_i^2/\sigma_X^2)/n}\cdot\frac{m}{m+n}}\right)^{n/2}\left(\frac{1}{\frac{m}{m+n} + \frac{(\sum X_i^2/\sigma_X^2)/n}{(\sum Y_i^2/\lambda_0\sigma_X^2)/m}\cdot\frac{n}{m+n}}\right)^{m/2} \\ &= \left[\frac{1}{\frac{n}{n+m} + \frac{m}{m+n}F}\right]^{n/2}\left[\frac{1}{\frac{m}{m+n} + \frac{n}{m+n}F^{-1}}\right]^{m/2} \end{aligned}$$

where $F = \frac{\sum Y_i^2/\lambda_0 m}{\sum X_i^2/n} \sim F_{m,n}$ under H_0. The rejection region is

$$\left\{(x,y) = \frac{1}{[\frac{n}{n+m} + \frac{m}{m+n}F]^{n/2}} \cdot \frac{1}{[\frac{m}{m+n} + \frac{n}{m+n}F^{-1}]^{m/2}} < c_\alpha\right\}$$

where c_α is chosen to satisfy

$$P\left\{\left[\frac{n}{n+m} + \frac{m}{m+n}F\right]^{-n/2}\left[\frac{m}{m+n} + \frac{n}{m+n}F^{-1}\right]^{-m/2} < c_\alpha\right\} = \alpha$$

3. To find: $(1-\alpha)$ confidence interval for λ.

   Solution: To ease notation, let $a = m/(n+m)$ and $b = a\sum y_i^2/\sum x_i^2$. From the duality of hypothesis tests and confidence sets, the set

   $$c(\lambda) = \left\{\lambda : \left(\frac{1}{a+b/\lambda}\right)^{n/2}\left(\frac{1}{(1-a)+\frac{a(1-a)}{b}\lambda}\right)^{m/2} \geq c_\alpha\right\}$$

   is a $1-\alpha$ confidence set for λ. We now must establish that this set is indeed an interval. To do this, we establish that the function on the left hand side of the inequality has only an interior maximum. That is, it looks like an upside-down bowl. Furthermore, it is straightforward to establish that the function is zero at both $\lambda = 0$ and $\lambda = \infty$. These facts imply that the set of λ values for which the function is greater than or equal to c_α must be an interval. We make some further simplifications. If we multiply both sides of the inequality by $[(1-a)/b]^{m/2}$, we need be concerned with only the behaviour of the function

   $$h(\lambda) = \left(\frac{1}{a+b/\lambda}\right)^{n/2}\left(\frac{1}{b+a\lambda}\right)^{m/2}$$

   Moreover, since we are most interested in the sign of the derivative of h, this is the same as the sign of the derivative of $\log h$, which is much easier to work with. We have

   $$\begin{aligned}\frac{d}{d\lambda}\log h(\lambda) &= \frac{d}{d\lambda}\left[-\frac{n}{2}\log(a+b/\lambda) - \frac{m}{2}\log(b+a\lambda)\right] \\ &= \frac{n}{2}\frac{b/\lambda^2}{a+b/\lambda} - \frac{m}{2}\frac{a}{b+a\lambda} \\ &= \frac{1}{2\lambda^2(a+b/\lambda)(b+a\lambda)}\left[-a^2m\lambda^2 + ab(n-m)\lambda + nb^2\right]\end{aligned}$$

   The sign of the derivative is given by the expression in square brackets, a parabola. It is easy to see that for $\lambda \geq 0$, the parabola changes sign from positive to negative. Since this is the sign change of the derivative, the function must increase then decrease. Hence, the function is an upside-down bowl, and the set is an interval.

Solution 3

Given: Let $\bar{X}$ be the mean of a random sample of size n from $N(\mu, 16)$.

To find: The smallest sample size n such that $(\bar{X} - 1, \bar{X} + 1)$ is a 0.90 level confidence interval for μ.

Solution: According to the question, we have

$$\begin{aligned}
\mathbb{P}\left\{\bar{X} - 1 \leq \mu \leq \bar{X} + 1\right\} &= 0.90 \\
\implies \mathbb{P}\left\{-1 \leq \bar{X} - \mu \leq 1\right\} &= 0.90 \\
\implies \mathbb{P}\left\{\frac{-1}{4/\sqrt{n}} \leq \frac{\bar{X} - \mu}{4/\sqrt{n}} \leq \frac{1}{4/\sqrt{n}}\right\} &= 0.90 \\
\implies \mathbb{P}\left\{\left|\frac{\bar{X} - \mu}{4/\sqrt{n}}\right| \leq \frac{1}{4/\sqrt{n}}\right\} &= 0.90 \\
\implies \mathbb{P}\left\{\left|Z\right| \leq \frac{1}{4/\sqrt{n}}\right\} &= 0.90, \qquad \text{where } Z = \frac{\bar{X} - \mu}{4/\sqrt{n}} \\
\implies 2\mathbb{P}\left\{Z \leq \frac{1}{4/\sqrt{n}}\right\} - 1 &= 0.90 \\
\implies \mathbb{P}\left\{Z \leq \frac{1}{4/\sqrt{n}}\right\} &= 0.95 \\
\implies \frac{1}{4/\sqrt{n}} &\approx 1.6 \\
\implies n &\approx 41
\end{aligned}$$

Solution 4

Given: Under the one-way ANOVA assumptions:

1. **To show:** The set of statistics $(\bar{Y}_1, \bar{Y}_2, \ldots, \bar{Y}_k, S_p^2)$ is sufficient for $(\theta_1, \theta_2, \ldots, \theta_k, \sigma^2)$.

   Solution: Under the ANOVA assumptions $Y_{ij} = \theta_i + \epsilon_{ij}$, where $\epsilon_{ij} \sim$ independent $N(0, \sigma^2)$, so $Y_{ij} \sim$ independent $N(\theta_i, \sigma^2)$. Therefore the sample pdf is

$$\begin{aligned}
&\prod_{i=1}^{k}\prod_{j=1}^{n_i}(2\pi\sigma^2)^{-1/2}e^{-\frac{y_{ij}-\theta_i)^2}{2\sigma^2}} \\
&= (2\pi\sigma^2)^{-\sum_i n_i/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \theta_i)^2\right\} \\
&= (2\pi\sigma^2)^{-\sum_i n_i/2} exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k} n_i\theta_i^2\right\} \times exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 + \frac{2}{2\sigma^2}\sum_{i=1}^{k}\theta_i n_i \bar{Y}_{i0}\right\}
\end{aligned}$$

Therefore, by the Factorization Theorem,

$$(\bar{Y}_{10}, \bar{Y}_{20}, \ldots, \bar{Y}_{k0}, \sum_i \sum_j Y_{ij}^2)$$

is jointly sufficient for $(\theta_1, \ldots, \theta_k, \sigma^2)$. Since $(\bar{Y}_{10}, \bar{Y}_{20}, \ldots, \bar{Y}_{k0}, S_p^2)$ is a 1-to-1 function of this vector. Hence $(\bar{Y}_{10}, \bar{Y}_{20}, \ldots, \bar{Y}_{k0}, S_p^2)$ is also jointly sufficient.

2. **To show:** $S_p^2 = \frac{1}{N-k} \sum_{i=1}^{k} (n_i - 1) S_i^2$ is independent of each $\bar{Y}_i, i = 1, 2 \ldots, k$.

   Solution: We can write

$$(2\pi\sigma^2)^{-\sum_i n_i/2} exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2 \right\}$$
$$= (2\pi\sigma^2)^{-\sum_i n_i/2} exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\left| y_{ij} - \bar{y}_{i0} \right| + \left| \bar{y}_{i0} - \theta_i \right|)^2 \right\}$$
$$= (2\pi\sigma^2)^{-\sum_i n_i/2} exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\left| y_{ij} - \bar{y}_{i0} \right|)^2 \right\} .exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{k} n_i (\left| \bar{y}_{i0} - \theta_i \right|)^2 \right\}$$

   so, by the Factorization Theorem, $\bar{Y}_{i0}, i = 1, \ldots, n$, is independent of $Y_{ij} - \bar{Y}_{i0}, j = 1, \ldots, n_i$, so S_p^2 is independent of each $\bar{Y}_{i0}$.

Solution 5

To show: To construct a t test for

1. $H_0 : \sum_i a_i\theta_i = \delta$ vs $H_1 : \sum_i a_i\theta_i \neq \delta$.

   Solution: Under the oneway ANOVA assumptions, we have $T = \sum_i a_i\bar{Y}_i \sim N(\sum_i a_i\theta_i, \sigma^2 \sum_i a_i^2/n_i)$,
   and under $H_0, \mathbb{E}[T] = \delta$. Thus, under H_0, $\frac{\sum_i a_i\bar{Y}_i - \delta}{\sqrt{S_p^2 \sum_i a_i^2/n_i}} \sim t_{N-k}$, where $N = \sum_i n_i$.

   Therefore, the test is to reject H_0 if $\frac{\left|\sum_i a_i\bar{Y}_i - \delta\right|}{\sqrt{S_p^2 \sum_i a_i^2/n_i}} > t_{\frac{\alpha}{2}, N-k}$.

2. $H_0 : \sum_i a_i\theta_i \leq \delta$ vs $H_1 : \sum_i a_i\theta_i > \delta$, where δ is a specified constant.

   Solution: Similarly for $H_0 : \sum_i a_i\theta_i \leq \delta$ vs $H_1 : \sum_i a_i\theta_i > \delta$, where δ is a specified constant, we reject H_0 if $\frac{\sum_i a_i\bar{Y}_i - \delta}{\sqrt{S_p^2 \sum_i a_i^2/n_i}} > t_{\alpha, N-k}$.

Solution 6

To show: For any set of constants $a = (a_1, \dots, a_k)$ and $b = (b_1, \dots, d_k)$, we have to show that under one-way ANOVA assumptions,

$$cov(\sum_i a_i \bar{Y}_i, \sum_i b_i \bar{Y}_i) = \sigma^2 \sum_i \frac{a_i b_i}{n_i}.$$

Solution: Under the ANOVA assumptions $Y_{ij} = \theta_i + \epsilon_{ij}$, where $\epsilon_{ij} \sim$ independent $N(0, \sigma^2)$, so $Y_{ij} \sim$ independent $N(\theta_i, \sigma^2)$. We have $\bar{Y}_{i0} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$.

$$\begin{aligned}
\bar{Y}_{i0} &\sim N\left(\theta_i, \frac{\sigma^2}{n_i}\right) \\
\implies \mathbb{E}\left[\bar{Y}_{i0}\right] &= \theta_i \text{ and } var(\bar{Y}_{i0}) = \frac{\sigma^2}{n_i}.
\end{aligned}$$

$$\begin{aligned}
cov(\sum_i a_i \bar{Y}_i, \sum_i b_i \bar{Y}_i) &= \sum_i a_i b_i var(\bar{Y}_{i0}) + \sum_{i \neq j} a_i b_j cov(\bar{Y}_{i0}, \bar{Y}_{j0}) \\
&= \sum_i a_i b_i var(\bar{Y}_{i0}) \qquad \text{since } Y_{ij} \text{ and } Y_{i'j'} \text{ are independent } \forall i \neq i', j \neq j' \\
&= \sigma^2 \sum_i \frac{a_i b_i}{n_i}.
\end{aligned}$$