Jane Downer
MATH 476
March 6, 2021

HOMEWORK 3

1. Refer to Exercise 8.28. If the researcher wants to estimate the difference in proportions to within .05 with 90% confidence, how many graduates and nongraduates must be interviewed? (Assume that an equal number will be interviewed from each group.)

> In a study of the relationship between birth order and college success, an investigator found that 126 in a sample of 180 college graduates were firstborn or only children; in a sample of 100 nongraduates of comparable age and socioeconomic background, the number of firstborn or only children was 54. Estimate the difference in the proportions of firstborn or only children for the two populations from which these samples were drawn. Give a bound for the error of estimation.

$n_1 = 180$

$\hat{p}_1 = 126/180 = 0.7$

$n_2 = 100$

$\hat{p}_2 = 54/100 = 0.54$

$z_{\alpha/2} = z_{0.05} = 1.645$

margin of error $z_{\alpha/2}\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 0.05$

$$1.645\sqrt{\frac{(0.7)(0.3) + (0.54)(0.46)}{n}} = 0.05$$

$$\frac{(0.7)(0.3) + (0.54)(0.46)}{n} = (0.05/1.645)^2$$

$$n = \frac{(0.7)(0.3) + (0.54)(0.46)}{(0.05/1.645)^2}$$

$= 496.177$

round up: $\boxed{n = 497}$

2. Suppose that we obtain independent samples of sizes n_1 and n_2 from two normal populations with equal variances. Use the appropriate pivotal quantity from Section 8.8 to derive a $100(1-\alpha)\%$ *upper confidence bound* for $\mu_1 - \mu_2$.

$$C.I._{\mu_1 - \mu_2} = (\bar{X}_1 - \bar{X}_2) \pm z_\alpha \cdot \sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

variance unknown ⟶ replace w/ t_α replace w/ S_p

$$C.I. = (\bar{X}_1 - \bar{X}_2) + t_\alpha \cdot S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

$$\boxed{C.I. = (\bar{X}_1 - \bar{X}_2) + t_\alpha \left[\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}\right]\left[\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]}$$

3. $Y_1, \dots, Y_n \sim N(\mu, \sigma^2)$ sample variance: S^2

both unknown

Part 1: $100(1-\alpha)\%$ upper confidence bound for σ^2

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$P\left(\chi^2_{n-1,1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2}\right) = 1-\alpha$$

$$\boxed{\sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}}$$

Part 2: $100(1-\alpha)\%$ upper confidence bound for σ

$$P\left(\chi^2_{n-1,1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2}\right) = P\left(\sqrt{\chi^2_{n-1,1-\alpha/2}} \le \sqrt{\frac{(n-1)S^2}{\sigma^2}}\right)$$

$$\boxed{\sigma \le S \cdot \sqrt{\frac{(n-1)}{\chi^2_{n-1,1-\alpha/2}}}}$$

4. $Y_1, \ldots, Y_n \sim \text{uniform}(\theta, \theta+1)$

find $\begin{cases} \hat{\theta}_1 : \text{function of } \bar{Y} \\ \hat{\theta}_2 : \text{function of } Y_{(n)} \end{cases}$

$$\left[\begin{array}{l} \hat{\theta}_1 = \bar{Y} - \frac{1}{2} \\ \hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1} \end{array}\right]$$

part 1

$$E(\bar{Y}) = E\left(\frac{\sum Y_i}{n}\right)$$

$$= \frac{1}{n} \cdot E\left(\sum Y_i\right)$$

$$= \frac{1}{2} \cdot \left(E(y_1) + \ldots + E(y_n)\right)$$

$$= \frac{1}{n} \cdot n\left(\theta + \frac{1}{2}\right)$$

$$= \theta + \frac{1}{2}$$

$$E\left(\bar{Y} - \frac{1}{2}\right) = E(\bar{Y}) - E\left(\frac{1}{2}\right)$$

$$= \theta + \frac{1}{2} - \frac{1}{2}$$

$$= \theta \ \checkmark$$ unbiased

$$E(Y_i) = \text{halfway between } a \,\&\, b$$

$$= \frac{\theta + (\theta+1)}{2}$$

$$= \theta + \frac{1}{2}$$

$$F(y) = \frac{y-a}{b-a} \qquad \theta \le y \le \theta+1$$

$$= \frac{y-\theta}{\theta+1-\theta}$$

$$= y - \theta$$

$$F_{Y_{(n)}}(y) = \left(F(y)\right)^n$$

$$f_{Y_{(n)}}(y) = n \cdot F(y)^{n-1}$$

$$= n \cdot (y-\theta)^{n-1}$$

$$E(Y_{(n)}) = \int_{\theta}^{\theta+1} y \cdot n \cdot (y-\theta)^{n-1}\, dy$$

$\int u \cdot dv = uv - \int v \cdot du$

$u = ny \qquad v = \frac{1}{n}(y-\theta)^n$

$du = n \qquad dv = (y-\theta)^{n-1}$

$$= ny \cdot \frac{1}{n}(y-\theta)^n - \int n \cdot \frac{1}{n}(y-\theta)^n\, dy \Big|_{\theta}^{\theta+1}$$

$$= ny \cdot \frac{1}{n}(y-\theta)^n - \frac{1}{n+1}(y-\theta)^{n+1} \Big|_{\theta}^{\theta+1}$$

$$= (\theta+1)(\theta+1-\theta)^n - \frac{1}{n+1}(\theta+1-\theta)^{n+1}$$

$$= \theta + 1 - \frac{1}{n+1}$$

$$= \theta + \frac{n}{n+1}$$

$$E\left(Y_{(n)} - \frac{n}{n+1}\right) = E(Y_{(n)}) - E\left(\frac{n}{n+1}\right)$$

$$= \theta + \frac{n}{n+1} - \frac{n}{n+1}$$

$$= \theta \ \checkmark$$ unbiased

part 2

$$\text{eff}(\hat\theta_1, \hat\theta_2) = \frac{\text{var}(\hat\theta_2)}{\text{var}(\hat\theta_1)}$$

$$\text{var}(Y_i) = \frac{(b-a)^2}{12}$$ (from definition of uniform distribution)

$$= \frac{(\theta + 1 - \theta)^2}{12}$$

$$= \frac{1}{12}$$

$$\text{var}(\bar{Y} - \tfrac{1}{2}) = \text{var}(\bar{Y})$$

$$= \text{var}\left(\frac{1}{n}\Sigma Y_i\right)$$

$$= \frac{1}{n^2}\left(n \cdot \frac{1}{12}\right)$$

$$= \frac{1}{12n}$$

$$\text{var}\left(Y_{(n)} - \frac{n}{n+1}\right) = \text{var}(Y_{(n)})$$

$$= E(Y_{(n)}^2) - (E(Y_{(n)}))^2$$

$$E(Y_{(n)}^2) = \int_\theta^{\theta+1} ny^2(y-\theta)^{n-1}\,dy$$

$$= \int_\theta^{\theta+1} n((y-\theta) + \theta)^2 (y-\theta)^{n-1}\,dy$$

$$= \int_\theta^{\theta+1} n((y-\theta)^2 + 2\theta(y-\theta) + \theta^2)(y-\theta)^{n-1}\,dy$$

$$= \int_\theta^{\theta+1} n(y-\theta)^2(y-\theta)^{n-1}\,dy + \int_\theta^{\theta+1} n(2\theta)(y-\theta)(y-\theta)^{n-1}\,dy + \int_\theta^{\theta+1} n\cdot\theta^2(y-\theta)^{n-1}\,dy$$

continued

$$= \int_{\theta}^{\theta+1} n(y-\theta)^{n+1}\,dy + \int_{\theta}^{\theta+1} 2n\theta(y-\theta)^{n}\,dy + \int_{\theta}^{\theta+1} n\cdot\theta^2(y-\theta)^{n-1}\,dy$$

$$= \frac{n}{n+2}(y-\theta)^{n+2} + \frac{2n\theta}{n+1}(y-\theta)^{n+1} + \frac{n\theta^2}{n}(y-\theta)^{n}\Big|_{\theta}^{\theta+1}$$

$$= \frac{n}{n+2} + \frac{2n\theta}{n+1} + \theta^2$$

$$= \left(\theta + \frac{n}{n+1}\right)^2$$

$$var(\hat{\theta}_2) = \left(\frac{n}{n+2} + \frac{2n\theta}{n+1} + \theta^2\right) - E(Y_{(n)})^2$$

$$= \frac{n}{n+2} + \frac{2n\theta}{n+1} + \theta^2 - \left(\theta^2 + \frac{2\theta n}{n+1} + \frac{n^2}{(n+1)(n+1)}\right)$$

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

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

$$= \frac{n^3+2n^2+n-n^3-2n^2}{(n+2)(n+1)(n+1)}$$

$$= \frac{n}{(n+2)(n+1)(n+1)}$$

$$\therefore\ eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{n}{(n+2)(n+1)(n+1)} \Big/ \frac{1}{12n}$$

$$\boxed{eff(\hat{\theta}_1, \hat{\theta}_2) = \frac{12n^2}{(n+2)(n+1)(n+1)}}$$

part 3

absolute efficiency: CRLB

$var(\hat{\theta}) \geq I(\theta)$?

$$I(\theta) = n \cdot E\left[\frac{-\partial^2 \ln(f(y))}{\partial \theta^2}\right]^{-1}$$

$$= n \cdot E\left(\frac{-\partial^2 \ln(1)}{\partial \theta^2}\right)^{-1}$$

$$= n \cdot E\left(\frac{-\partial^2}{\partial \theta^2}(0)\right)^{-1}$$

= undefined

$$L(\theta) = \prod_{i=1}^{n} f(y)$$

$$= \prod_{i=1}^{n} 1$$

$$= 1 = 1 \times 1$$

$$= g(\mu, Y_{(n)}) \times h(y_1, \ldots, y_n)$$

$$= g(\mu, \bar{Y}) \times h(y_1, \ldots, y_n)$$

$\hat{\theta}_1$ and $\hat{\theta}_2$ are each functions of sufficient estimators, and the expected value of each estimator is equal to θ. Therefore, $\hat{\theta}_1$ and $\hat{\theta}_2$ are MVUE's, and thus, they are efficient.

5. Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution over the interval (θ_1, θ_2). Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ and $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ are jointly sufficient for θ_1 and θ_2.

$Y_1, \ldots, Y_n \sim \text{uniform}(\theta_1, \theta_2)$

$$f_X = \begin{cases} \dfrac{1}{\theta_2 - \theta_1} & \theta_1 \le y \le \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} f_X(y)$$

$$= \begin{cases} \dfrac{1}{(\theta_2 - \theta_1)^n} & \text{if } y_{(1)} \text{ and } y_{(n)} \text{ fall within } \theta_1 \text{ and } \theta_2 \\ 0 & \text{if either } y_{(1)} < \theta_1 \text{ or } y_{(n)} > \theta_n \end{cases}$$

$$= \frac{1}{(\theta_2 - \theta_1)^n} \times 1$$

$$= g(\theta_1, \theta_2, y_{(1)}, y_{(n)}) \times h(y_1, \ldots, y_n)$$

$L(\theta_1, \theta_2)$ can be written as the product of a function $g(\theta_1, \theta_2, y_{(1)}, y_{(n)})$ and a function $h(y_1, \ldots, y_n)$. The function g depends on $\theta_1, \theta_2, y_{(1)}$, and $y_{(n)}$, whereas h is independent of θ_1 and θ_2.

Therefore, $Y_{(1)}$ and $Y_{(n)}$ are jointly sufficient for θ_1 and θ_2.

6. $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$

Part 1 σ^2 known, μ unknown $\longrightarrow$ find sufficient statistic for μ

$$L(\mu) = \prod_{i=1}^{n} f_y(y)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \right)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

goes to zero when summed

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}\left((y_i-\bar{y})^2 + \overbrace{2(y_i-\bar{y})(\bar{y}-\mu)} + (\bar{y}-\mu)^2\right)}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}\cdot\left[\sum_{i=1}^{n}(y_i-\bar{y})^2 + n(\bar{y}-\mu)^2\right]}$$

$$= \left[e^{-\frac{1}{2\sigma^2}\cdot n(\bar{y}-\mu)^2} \right] \times \left[\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2} \right]$$

$$= g(\mu, \bar{y}) \times h(y_1, \ldots, y_n)$$

$\boxed{\bar{Y} \text{ is a sufficient statistic for } \mu.}$

Part 2 μ is known, σ^2 is unknown $\longrightarrow$ find sufficient statistic for σ^2

$$L(\sigma^2) = \prod_{i=1}^{n} f_y(y)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(y_i-\mu)^2} \right)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

$$= \left[\frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2} \right] \times 1$$

$$= g\left(\sigma^2, \sum_{i=1}^{n}(y_i-\mu)^2\right) \times h(y_1, \ldots, y_n)$$

$$\boxed{\sum_{i=1}^{n}(y_i-\mu)^2 \text{ is a sufficient statistic for } \sigma^2.}$$