MTH 326 - Spring 2022
Exam 1 Corrections
Due: Monday March 21, 2022 (11:59pm)

1. (5 points) In a study to compare the perceived effects of two pain relievers, 200 patients were given medicine A, of whom 90% found relief, and 300 patients were given medicine B with 80% experiencing relief. Find a 95% confidence interval for the difference in population proportions experiencing relief between A and B.

   We want $\mu_A - \mu_B$ on a large sample $n > 30$ so we will need to use a $\mathcal{Z}$-test. For a 95% confidence interval ($\alpha = 0.05$) we use a $\mathcal{Z}$-score of $\mathcal{Z}_{0.025} = 1.960$.

   We have CI $\equiv \mu_1 - \mu_2 \pm \mathcal{Z}_{\alpha/2}\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$. Substituting in our given information with $A \equiv 1$ and $B \equiv 2$,

   $$\begin{aligned}\text{C.I.} &= 0.9 - 0.8 \pm 1.960\sqrt{\frac{0.9(1-0.9)}{200} + \frac{0.8(1-0.8)}{300}} \\ &\approx (0.03853,\ 0.16146).\end{aligned}$$

2. (10 points) Air trapped in amber from the Cretaceous era (75 million years ago) may suggest that the composition of our atmosphere has changed. Nine different samples have been obtained and the gas tested for the percentage of nitrogen in the atmosphere. We will treat these as a random sample.

   (a) Given that $\bar{X} = 59.6\%$ and $S^2 = 39.13$, compute a 99% confidence interval on the nitrogen level in the ancient atmosphere. (FYI, the nitrogen level of our air is 78.1% today.)

   We are given that $\alpha = 0.01$, $n = 9$, and therefore d.f. $= 8$. Since $n < 30$ we need a t-test and will use the following formula: CI $= \bar{X} \pm t_{\alpha/2}(\text{d.f.}) \cdot \text{SE}$, where SE $= \dfrac{\sigma}{\sqrt{n}} = \sqrt{\dfrac{\sigma^2}{n}}$. We have $t_{0.005}(8) = 3.355$ and hence

   $$\text{C.I.} = 59.6\% \pm 3.355 \cdot \sqrt{\frac{39.13\%}{9}} = (52.60438\%, 66.59561\%)\,.$$

   (b) Construct a 90% confidence interval for the population variance σ^2.

   For the 90% *lower* tail, using a modified formula $\left(0,\ \dfrac{(n-1)S^2}{\chi^2_\alpha}\right)$ and with $\alpha = 0.10$, $n = 9$, df $= 8$, $S^2 = 39.13\%$. Then $\chi^2_{0.1}(8) \approx 13.36$ and thus,

   $$\text{C.I.} \equiv \left(0,\ \frac{8 \cdot 39.13\%}{13.36}\right) \approx (0\%,\ 23.43113\%)\,.$$

3. (20 points) Suppose $Y_1,\ Y_2,\ Y_3$ and Y_4 is an iid random sample from an exponential distribution with unknown rate parameter $\beta > 0$:

$$f(y) = \frac{1}{\beta}e^{-y/\beta}, \quad 0 < y < \infty.$$

Consider the two estimators of β: $\widehat{\theta}_1 = \bar{Y}$ and $\widehat{\theta}_2 = \dfrac{2Y_1 + 3Y_2}{5}$.

(a) Show that $\widehat{\theta}_2$ is an unbiased estimator of β.

In order to show unbiased, we need $\mathrm{B}(\widehat{\theta}_2) = \mathrm{E}[\,\widehat{\theta}_2\,] - \mathrm{E}[\,\widehat{\theta}\,] = 0$. Exponential is a gamma with $\alpha = 1$. Hence $\mu = \alpha\beta = \beta$ and $\sigma^2 = \alpha\beta^2 = \beta^2$. Computing the expected value,

$$\mathrm{E}[\,\widehat{\theta}_2\,] \;=\; \mathrm{E}\left[\frac{2Y_1 + 3Y_2}{5}\right] \;=\; \frac{2}{5}\,\mathrm{E}\,[\,Y_1\,] + \frac{3}{5}\,\mathrm{E}\,[\,Y_2\,] \;=\; \frac{2}{5}\beta + \frac{3}{5}\beta \;=\; \beta$$

Thus $\mathrm{B}(\widehat{\theta}_2) = \mathrm{E}[\,\widehat{\theta}_2\,] - \mathrm{E}[\,\widehat{\theta}\,] = \beta - \beta = 0$. Hence, $\widehat{\theta}_2$ is unbiased.

(b) Determine the efficiency of $\widehat{\theta}_2$ relative to $\widehat{\theta}_1$.

The definition of efficiency is $\mathrm{eff}(\widehat{\theta}_2,\, \widehat{\theta}_1) = \dfrac{\mathrm{V}[\,\widehat{\theta}_1\,]}{\mathrm{V}[\,\widehat{\theta}_2\,]}$, so we will compute the respective variances.

$$\begin{aligned}\mathrm{V}[\,\widehat{\theta}_1\,] = \mathrm{V}[\,\bar{Y}\,] = \mathrm{V}\left[\frac{1}{4}\sum_{i=1}^{4} Y_i\right] &= \frac{1}{4^2}\,\mathrm{V}\,[\,Y_1 + Y_2 + Y_3 + Y_4\,] \\ &= \frac{\mathrm{V}\,[\,Y_1\,] + \mathrm{V}\,[\,Y_2\,] + \mathrm{V}\,[\,Y_3\,] + \mathrm{V}\,[\,Y_4\,]}{16} = \frac{4\beta^2}{16} = \frac{\beta^2}{4}\end{aligned}$$

$$\begin{aligned}\mathrm{V}[\,\widehat{\theta}_2\,] &= \mathrm{V}\left[\frac{2Y_1 + 3Y_2}{5}\right] = \mathrm{V}\left[\frac{2}{5}Y_1\right] + \mathrm{V}\left[\frac{3}{5}Y_2\right] \;=\; \frac{4}{25}\,\mathrm{V}\,[\,Y_1\,] + \frac{9}{25}\,\mathrm{V}\,[\,Y_2\,] \\ &= \frac{4}{25}\beta^2 + \frac{9}{25}\beta^2 \;=\; \frac{13}{25}\beta^2\end{aligned}$$

$$\mathrm{eff}(\widehat{\theta}_2,\, \widehat{\theta}_1) = \frac{\mathrm{V}[\,\widehat{\theta}_1\,]}{\mathrm{V}[\,\widehat{\theta}_2\,]} = \frac{\beta^2/4}{13\beta^2/25} = \frac{1}{4} \times \frac{25}{13} = \frac{25}{52}$$

(c) Now consider a third estimator of β, $\widehat{\theta}_3 = \min(Y_1,\ Y_2,\ Y_3,\ Y_4)$. Show that the distribution of $\widehat{\theta}_3$ is also exponentially distributed. (Recall the general density function for min order statistic is $f_{(1)}(y) = n\,[1 - F(y)]^{n-1}\, f(y)$.)

From gamma with $\alpha = 1$ we have $f(y) := \frac{1}{\beta} e^{-y/\beta}$, $y \in (0, \infty)$. Computing the antiderivitive,

$$\begin{aligned} F(y) &= \int_0^y \frac{1}{\beta} e^{-y/\beta}\, dy & g(y) = u = -\frac{y}{\beta} \\ &= \frac{1}{\beta}\int_{g(0)}^{g(y)} -\beta e^u du & du = -\frac{1}{\beta}dy \iff dy = -\beta du \\ &= -\int_0^{-y/\beta} e^u du = -\Big[e^u\Big]_{u=0}^{u=-\frac{y}{\beta}} = -\Big[e^{-\frac{y}{\beta}} - 1\Big] = -e^{-\frac{y}{\beta}} + 1. \end{aligned}$$

Substituting into the formula,

$$f_1(y) = 4\Big[1 - \left(-e^{-y/\beta} + 1\right)\Big]^3 \frac{e^{-y/\beta}}{\beta} = 4\left(e^{-y/\beta}\right)^3 \frac{e^{-y/\beta}}{\beta} = \frac{4}{\beta} e^{-4y/\beta}.$$

Hence it is still an exponential distribution with $\lambda = \dfrac{4}{\beta}$.

(d) Show that $\widehat{\theta}_3$ is a biased estimator and compute the mean square error of $\widehat{\theta}_3$.

$$\mathrm{E}\left[\, f_{(1)}(y)\,\right] = \mathrm{E}\left[\frac{4}{\beta} e^{-4y/\beta}\right] = \mathrm{E}\left[\frac{1}{\beta/4} e^{-y/(\beta/4)}\right] = \frac{\beta}{4}$$

$$\mathrm{B}(\widehat{\theta}_3) = \mathrm{E}[\,\widehat{\theta}_3\,] - \mathrm{E}[\,\widehat{\theta}\,] = \frac{\beta}{4} - \beta = -\frac{3}{4}\beta \neq 0 \quad \therefore \quad \widehat{\theta}_3 \text{ is biased}$$

$$\mathrm{V}[\,\widehat{\theta}_3\,] = \left(\frac{\beta}{4}\right)^2 = \frac{\beta^2}{16}$$

$$\begin{aligned} \mathrm{MSE}[\,\widehat{\theta}_3\,] &= \mathrm{V}[\,\widehat{\theta}_3\,] + \mathrm{B}[\,\widehat{\theta}_3\,]^2 \\ &= \frac{\beta^2}{16} + \left(-\frac{3}{4}\beta\right)^2 \\ &= \frac{1}{16}\beta^2 + \frac{9}{16}\beta^2 \\ &= \frac{5}{8}\beta^2 \end{aligned}$$

4. (25 points) Suppose that $X_1, \ldots, X_n$ is an iid sample from a Rayleigh distribution with parameter $\theta > 0$ unknown:

$$f(x) = \frac{2x}{\theta} e^{-x^2/\theta}, \quad 0 < x < \infty.$$

Note that $\mathrm{E}(X) = \frac{\sqrt{\pi\theta}}{2}$, $\mathrm{E}(X^2) = \theta$, $\mathrm{E}(X^3) = \frac{3\sqrt{\pi\theta^3}}{4}$, and $\mathrm{E}(X^4) = \frac{\theta^4}{2}$. (You do not need to prove these facts.)

(a) Find the method of moments estimator θ_{MOM} for θ.

$\mu_1' = \mathrm{E}[X] = \frac{\sqrt{\pi\theta}}{2}$ and $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$. Equating them,

$$\frac{\sqrt{\pi\theta}}{2} = \bar{X} \iff \sqrt{\pi\theta} = 2\bar{X} \iff \pi\theta = 4\bar{X}^2 \iff \theta_{\text{MOM}} = \frac{4\bar{X}^2}{\pi}$$

(b) Find and simplify the likelihood function $L(x_1, \ldots, x_n \mid \theta)$, complete the factorization, and determine a sufficient statistic for θ.

$$\begin{aligned} L(\vec{x} \mid \theta) &= \prod_{i=1}^{n} f(x_i \mid \theta) = \prod_{i=1}^{n} \frac{2x_i}{\theta} e^{-x_i^2/\theta} = \left(\frac{2}{\theta}\right)^n \prod_{i=1}^{n} x_i e^{-x_i^2/\theta} \\ &= \left(\frac{2}{\theta}\right)^n e^{-(\sum_{i=1}^{n} x_i^2)/\theta} \prod_{i=1}^{n} x_i = \left(\frac{2}{\theta}\right)^n e^{-S/\theta} \prod_{i=1}^{n} x_i \qquad S := \sum_{1}^{n} x_i^2 \end{aligned}$$

$$\therefore \quad g(S \mid \theta) = \left(\frac{2}{\theta}\right)^n e^{-S/\theta} \quad \text{and} \quad h(\vec{x}) = \prod_{i=1}^{n} x_i$$

And S is sufficient for θ by the Factorization Theorem.

(c) Find the maximum likelihood estimator θ_{MLE} for θ.

$$\begin{aligned} \ln(L(\theta)) &= n \ln\left(\frac{2}{\theta}\right) - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2 + \ln\left(\prod_{i=1}^{n} x_i\right) \\ &= \underbrace{n\ln 2 - n\ln\theta}_{\text{log division law}} - \frac{1}{\theta}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \ln x_i \end{aligned}$$

$$\frac{\partial}{\partial\theta}\left(\ln\left(L(\theta)\right)\right) = 0 - \frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2 + 0 = 0 \iff \frac{n}{\theta} = \frac{1}{\theta^2}\sum_{i=1}^{n} x_i^2$$

$$\iff n\theta = \sum_{i=1}^{n} x_i^2 \iff \theta_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

(d) Show that the maximum likelihood estimator of θ is consistent.

For a consistent estimator, we need $\lim_{n\to\infty} \mathrm{V}[\,\theta_{MLE}\,] = 0$. Computing the variance,

$$\mathrm{V}\left[\,\theta_{\mathrm{MLE}}\,\right] = \mathrm{V}\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2\right] = \frac{1}{n^2}\,\mathrm{V}\left[\sum_{i=1}^{n} x_i^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}\left[\,x_i^2\,\right] = \frac{1}{n}\,\mathrm{V}\left[\,x_i^2\,\right]$$

$$= \frac{1}{n}\left(\mathrm{E}\left[\,x_i^{2\cdot 2}\,\right] - \mathrm{E}\left[\,x_i^2\,\right]^2\right) = \frac{1}{n}\left(\frac{\theta^4}{2} - \theta^2\right)$$

$$\lim_{n\to\infty} \mathrm{V}[\,\theta_{\mathrm{MLE}}\,] = \lim_{n\to\infty}\frac{1}{n}\left(\frac{\theta^4}{2} - \theta^2\right) = \left(\frac{\theta^4}{2} - \theta^2\right)\lim_{n\to\infty}\frac{1}{n} = 0.$$

Therefore θ_{MLE} is consistent.

(e) Is the maximum likelihood estimator a minimum variance unbiased estimator? Briefly explain your answer.

$$\mathrm{E}\left[\,\theta_{\mathrm{MLE}}\,\right] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2\right] = \mathrm{E}\left[\,x_i^2\,\right] = \theta.$$

$$\mathrm{B}[\,\theta_{\mathrm{MLE}}\,] = \mathrm{E}\left[\,\theta_{\mathrm{MLE}}\,\right] - \theta = \theta - \theta = 0 \quad \therefore \quad \text{unbiased.}$$

Because θ_{MLE} is unbiased and S is a sufficient statistic (part b), then our θ_{MLE} is an MVUE via the Rao-Blackwell Theorem.