M378K Introduction to Mathematical Statistics
Homework assignment #10

Please, provide your **final answer only** to the following problems.

Problem 10.1. *(5 points) Let $Y_1, \dots, Y_5$ be a random sample from the normal distribution $N(\mu, 2)$, with an unknown mean μ and the known standard deviation $\sigma = 2$. The collected data turn out to be*

$$y_1 = 2, y_2 = 5, y_3 = 1, y_4 = 4, y_5 = 3.$$

The right end-point of the one-sided 90%-confidence interval $(-\infty, \hat{\mu}_R]$ for μ is

(a) $3 + \frac{2}{\sqrt{5}} qnorm(0.9, 0, 1)$.

(b) $3 + \frac{2}{5} qnorm(0.9, 0, 1)$.

(c) $3 + \frac{1}{\sqrt{5}} qt(0.9, 4)$.

(d) $3 + \frac{1}{5} qnorm(0.9, 5)$.

(e) **None of the above.**

Solution: The correct answer is **(a)**.
The confidence interval in this case is based on the pivotal quantity $\sqrt{5}\frac{\mu - \bar{Y}}{2}$ which has the $N(0, 1)$ distribution. Therefore, for $b = \mathtt{qnorm}(0.9, 0, 1)$ we have

$$\mathbb{P}[\sqrt{5}\tfrac{\mu - \bar{Y}}{2} \leq b] = 0.9.$$

We solve for μ to obtain

$$\mathbb{P}[\mu \leq \bar{Y} + \tfrac{2}{\sqrt{5}} b] = 0.9.$$

For our data set $\bar{y} = 3$, so $\hat{\mu}_R = 3 + \frac{2}{\sqrt{5}} \mathtt{qnorm}(0.9, 0, 1)$.

Problem 10.2. *(5 points) A random sample of size $n = 5$ from the normal distribution with unknown mean μ and an unknown standard deviation σ yielded the values $y_1, \dots, y_5$ such that*

$$\sum y_i = 10 \text{ and } \sum_{i=1}^{5} (y_i - 2)^2 = 4.$$

The value of $\hat{\mu}_L$ such that $(\hat{\mu}_L, \infty)$ is (an asymmetric) 95%-confidence interval for μ is

(a) $2 - \frac{1}{\sqrt{5}} \mathtt{qt(0.95,4)}$

(b) $2 - \frac{1}{5} \mathtt{qnorm(0.95)}$

(c) $2 - \frac{1}{5}\texttt{qt(0.95,4)}$

(d) $2 - \frac{1}{\sqrt{5}}\texttt{qnorm(0.95)}$

(e) $2 - \frac{1}{\sqrt{5}}\texttt{qt(0.975,4)}$

Solution: The correct answer is **(a)**.
Since both μ and σ are unknown, the interval is based on the t-distribution with $4 = 5-1$ degrees of freedom and has the form

$$(\bar{Y} - \texttt{qt}(1-\alpha, n-1)\sqrt{S^2/n}, \infty)$$

In our case, $\bar{Y} = 2$ and $S^2 = 1$, so the interval is given by

$$(2 - \frac{1}{\sqrt{5}}\texttt{qt}(0.95, 4), \infty) \approx (1.05, \infty)$$

Please, provide your **complete solutions** to the following problems. Final answers only, even if correct will earn zero points for those problems.

Problem 10.3. *(10 points) 5 astronomy teams from across the world measured the distance to Proxima Centauri using a new method. It is reasonable to assume that the error of this method is normally distributed, but, since it is new, there is no information about its standard variation. Find a 95%-confidence interval for the distance if the obtained measurements are (in light years)*

$$4.20, 4.16, 4.04, 4.01, 4.20.$$

What would your confidence interval look like if they used an established method whose standard deviation of the measurement error is 0.1?

Solution: In the first situation, we are dealing with an unknown mean and an unknown standard deviation, so we use a confidence interval based on the t-distribution (here $t_{\alpha/2,n-1}$ denotes the $1-\alpha/2$-quantile of the t-distribution with $n-1$ degrees of freedom, i.e. $\mathbb{P}[X \leq t_{\alpha/2,n-1}] = 1-\alpha/2$ where X is t-distributed with $df = n-1$.)

$$[\overline{Y} - t_{\alpha/2,n-1}S/\sqrt{n}, \overline{Y} + t_{\alpha/2,n-1}S/\sqrt{n}]$$

For the data-set in the problem, we have

$$\bar{Y} = 4.122, S = 0.091, n = 5, \alpha = 0.05, t_{\alpha/2,4} = 2.776,$$

and, so, the required interval is

$$[4.009, 4.235]$$

If the standard deviation were known, we would construct an interval based on the normal distribution

$$[\overline{Y} - z_{\alpha/2}\sigma/\sqrt{n}, \overline{Y} + z_{\alpha/2}\sigma/\sqrt{n}] = [4.034, 4.210]$$

Problem 10.4. *(30 points) Let $Y_1, \ldots, Y_n$ be a random sample from $U(0, \theta)$ with θ unknown. Consider the following two estimators:*

$$\hat{\theta}_1 = 2\bar{Y} \quad \textit{and} \quad \hat{\theta}_2 = \left(\frac{n+1}{n}\right) Y_{(n)}$$

(i) (5 points) Prove that $\hat{\theta}_1$ is unbiased.

(ii) (10 points) Prove that $\hat{\theta}_2$ is unbiased.

(iii) (15 points) Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

Solution:

(i)

$$\mathbb{E}[\hat{\theta}_1] = \mathbb{E}[2\bar{Y}] = 2\mathbb{E}[\bar{Y}] = 2\mathbb{E}[Y_1] = 2\left(\frac{\theta}{2}\right) = \theta$$

(ii) *Note: This calculation is a* **special case** *of the one we did in Problem #13.3 from Problem Set #13.*

$$\mathbb{E}[\hat{\theta}_2] = \mathbb{E}\left[\left(\frac{n+1}{n}\right) Y_{(n)}\right] = \left(\frac{n+1}{n}\right) \mathbb{E}\left[Y_{(n)}\right]$$

As we have shown in class, the density $g_{(n)}$ of the n^{th} order statistic can be expressed as

$$g_{(n)}(y) = n f_Y(y)[F_Y(y)]^{n-1}$$

where f_Y and F_Y stand for the common pdf and cdf of the random variables in the random sample $Y_1, \ldots, Y_n$. Since the population distribution is given to be uniform on $(0, \theta)$, we can write

$$g_{(n)}(y) = n\left(\frac{1}{\theta}\right)\left(\frac{y}{\theta}\right)^{n-1} \mathbf{1}_{(0,\theta)}(y).$$

So,

$$\mathbb{E}[Y_{(n)}] = \int_0^\theta y g_{(n)}(y)\, dy = \int_0^\theta \frac{n y^n}{\theta^n}\, dy = \left(\frac{n}{n+1}\right)\theta.$$

Hence, $\hat{\theta}_2$ is, indeed, unbiased.

(iii) By definition, the relative efficiency we are looking for is

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{Var}[\hat{\theta}_2]}{\text{Var}[\hat{\theta}_1]}.$$

We can easily get $\text{Var}[\hat{\theta}_1]$ since we know that the variance of a single Y_1 which is uniform across $(0, \theta)$ is $\frac{\theta^2}{12}$. So,

$$\text{Var}[\hat{\theta}_1] = \text{Var}[2\bar{Y}] = 4\,\text{Var}[\bar{Y}] = 4\left(\frac{\text{Var}[Y_1]}{n}\right) = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}.$$

As for $\hat{\theta}_2$, we can again use *Problem #13.3 from Problem Set #13* as a guide. First, we calculate the second moment of $Y_{(n)}$.

$$\mathbb{E}\left[\left(Y_{(n)}\right)^2\right] = \int_0^\theta y^2 g_{(n)}(y)\,dy = \int_0^\theta \frac{ny^{n+1}}{\theta^n}\,dy = \left(\frac{n}{n+2}\right)\theta^2.$$

So,

$$\begin{aligned}\mathrm{Var}[Y_{(n)}] = \mathbb{E}\left[Y_{(n)}^2\right] - \left(\mathbb{E}\left[Y_{(n)}\right]\right)^2 &= \left(\frac{n}{n+2}\right)\theta^2 - \left(\left(\frac{n}{n+1}\right)\theta\right)^2\\ &= \left(\left(\frac{n}{n+2}\right) - \left(\frac{n}{n+1}\right)^2\right)\theta^2.\end{aligned}$$

While you might be tempted to simplify this further, **resist the temptation!** The variance of $\hat{\theta}_2$ equals

$$\begin{aligned}\mathrm{Var}[\hat{\theta}_2] = \mathrm{Var}\left[\left(\frac{n+1}{n}\right)Y_{(n)}\right] &= \left(\frac{n+1}{n}\right)^2 \mathrm{Var}\left[Y_{(n)}\right]\\ &= \left(\frac{n+1}{n}\right)^2\left(\left(\frac{n}{n+2}\right) - \left(\frac{n}{n+1}\right)^2\right)\theta^2\\ &= \left(\frac{(n+1)^2}{n(n+2)} - 1\right)\theta^2 = \frac{\theta^2}{n(n+2)}\,.\end{aligned}$$

Finally,

$$\mathrm{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\mathrm{Var}[\hat{\theta}_2]}{\mathrm{Var}[\hat{\theta}_1]} = \frac{\frac{\theta^2}{n(n+2)}}{\frac{\theta^2}{3n}} = \frac{3}{n+2}\,.$$