M378K Introduction to Mathematical Statistics

Homework assignment #10

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

Problem 10.1. *(10 points) Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be independent unbiased estimators of θ. You are given that $Var[\hat{\theta}_1] = 4$ and $Var[\hat{\theta}_2] = 9$.*

(i) (8 pts) Consider the class of estimators of the form $\hat{\theta} = \alpha\hat{\theta}_1 + \beta\hat{\theta}_2$. Find the UMVUE in this class of estimators.

(ii) (2 pts) Calculate the MSE of the estimator $\hat{\theta}$ you found in part (i) as an estimator of θ.

Solution:

(i) In order for the new estimators $\hat{\theta}$ to be unbiased, it is necessary that $\alpha + \beta = 1$. So,

$$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}[\alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2].$$

Due to independence of $\hat{\theta}_1$ and $\hat{\theta}_2$, the above equals

$$\alpha^2 \mathrm{Var}[\hat{\theta}_1] + (1-\alpha)^2 \mathrm{Var}[\hat{\theta}_2] = 4\alpha^2 + 9(1-\alpha)^2.$$

Our aim is to minimize the above. As a function of α, the above is evidently a parabola facing up. So, if we differentiate with respect to α, set the derivative to 0, and solve for α, we will have identified the minimum of the function. We get

$$8\alpha - 18(1-\alpha) = 0 \quad \Rightarrow \quad 4\alpha - 9 + 9\alpha = 0 \quad \Rightarrow \quad \alpha = \frac{9}{13}.$$

(ii) Since the estimator $\hat{\theta}$ is unbiased, its MSE is equal to its variance. We have

$$\mathrm{Var}[\hat{\theta}] = \left(\frac{9}{13}\right)^2 (4) + \left(\frac{4}{13}\right)^2 (9) = \frac{36}{13}.$$

Problem 10.2. *(5 points) For a fixed confidence level, a broader confidence interval is preferred since it is more likely to contain the true value of the parameter of interest.* True or false? Why?

Solution: FALSE
At a fixed confidence level $C = 1 - \alpha$, the probability that a confidence interval contains the true parameter is exactly $C = 1 - \alpha$.

Problem 10.3. *(5 points) You model the weights of individual boxes of Turkish delight as normally distributed with an unknown mean μ and a known standard deviation of 10 grams. You gather a random sample of 16 boxes of Turkish delight and carefully weigh them. You obtain the following values*

$$97.14, 107.05, 103.17, 106.27, 98.63, 90.66, 105.29, 87.12,$$
$$112.40, 95.61, 95.19, 114.14, 94.47, 112.89, 105.80, 92.96.$$

Provide the 80% confidence interval for μ based on the above data.

Solution: The sample average for the above values is $\bar{y} = 101.1744$.

The critical value of the standard normal distribution at the 80% confidence level is $z^* = \Phi^{-1}(0.90) = 1.28$. So, our confidence interval is of the form

$$\mu = 101.1744 \pm 1.28 \left(\frac{10}{\sqrt{16}} \right) = 101.1744 \pm 3.2.$$

Problem 10.4. *(5 points) A pollster is trying to estimate the proportion of the population in favor of candidate A (in a two-way race between A and B). The quality of her sample is such that it can be safely assumed to be a random sample from the Bernoulli distribution with the unknown parameter p. She is interested in the smallest sample size she will need in order to be able to pinpoint the value of p with $\pm 1\%$ accuracy, with 95% confidence.*

Basing your analysis on the estimator $\hat{p} = Y/n$, where Y is the number of supporters of candidate A in the sample, find the smallest such n under the assumption that the sampling distribution of $\hat{p}$ is well approximated by a normal distribution (of appropriate mean and variance left for you to figure out).

Solution: The sampling distribution of Y is binomial, with parameters n and p, and, so

$$\mathbb{E}[Y] = np, \text{ and } \operatorname{Var}[Y] = np(1-p). \text{ Therefore, } \mathbb{E}[\hat{p}] = p \text{ and } \sqrt{\operatorname{Var}[\hat{p}]} = \sqrt{\tfrac{p(1-p)}{n}}.$$

We take $\mu = \mathbb{E}[\hat{p}]$ and $\sigma = \sqrt{\frac{p(1-p)}{n}}$, so that the random variable

$$\tfrac{\hat{p}-\mu}{\sigma} = \tfrac{\hat{p}-p}{\sqrt{p(1-p)/n}}$$

is an approximate pivotal quantity with the approximate distribution $N(0,1)$, by the Central Limit Theorem. This random variable takes values in $[-1.96, 1.96]$ with the (approximate) probability of 95%, so that

$$0.95 \approx \mathbb{P}\left[\tfrac{|\hat{p}-p|}{\sqrt{p(1-p)/n}} \leq 1.96 \right] = \mathbb{P}[|\hat{p}-p| \leq 1.96\sqrt{p(1-p)/n}].$$

We want $1.96\sqrt{p(1-p)/n}$ to be equal to or smaller than 1%, but we do not know the value of p. To guarantee this for all values of p, we note that $p(1-p)$ is largest when $p = \frac{1}{2}$. Therefore, if we choose n such that $1.96\sqrt{1/2(1-1/2)/n} = 1\%$, we are done. The algebra gives

$$n = \left(\tfrac{1.96(0.5)}{0.01}\right)^2 = 9604.$$

Problem 10.5. *(25 points)*
Source: "Mathematical Statistics with Applications" by Wackerley, Mendenhall, and Sheaffer.
Let the random variable Y be gamma distributed with parameters $k = 4$ and τ unknown.

(i) *(10 points) Show, using moment generating functions, that*

$$U := \frac{2Y}{\tau} \sim \chi^2(df = 8).$$

(ii) *(15 points) Using $U = 2Y/\tau$ as a pivotal quantity, construct a 90%-confidence interval for τ.*

Solution: It's given that $Y \sim \Gamma(k = 4, \tau = ?)$. Using the formula from the lecture notes, we know that the moment generating function of Y has the form

$$m_Y(t) = (1 - \tau t)^{-4}.$$

In general, for any linear transform $\tilde{Y} = \alpha Y + \beta$, its moment generating function is

$$m_{\tilde{Y}}(t) = \mathbb{E}[e^{\tilde{Y}t}] = \mathbb{E}[e^{(\alpha Y + \beta)t}] = e^{\beta t}\mathbb{E}[e^{(\alpha t)Y}] = e^{\beta t} m_Y(\alpha t).$$

So, since in our problem, U is a linear transform of Y with $\alpha = \frac{2}{\tau}$ and $\beta = 0$, we have that

$$m_U(t) = m_Y\left(\frac{2t}{\tau}\right) = \left(1 - \tau\frac{2t}{\tau}\right)^{-4} = (1 - 2t)^{-4}.$$

Again, consulting the moment generating function for the gamma distribution, we realize that U is $\Gamma(4, 2)$ - which is the same thing as $\chi^2(df = 8)$.

The random variable $U = 2Y/\tau$ is, thus, a suitable pivotal quantity in the task of figuring out a 90%-confidence interval for the target parameter τ. As is usually the case, we want a "symmetric" confidence interval, i.e., we need constants a and b such that

$$\mathbb{P}[U < a] = \mathbb{P}[U > b] = 0.05.$$

Using the χ^2-tables with 8 degrees of freedom, we obtain $a = 2.73264$ and $b = 15.5073$. So,

$$\mathbb{P}[a \le U \le b] = 0.90 \quad \Leftrightarrow \mathbb{P}[2Y/15.5073 \le \tau \le 2Y/2.7326]$$

In the form of an interval, we can write

$$\left(\frac{2Y}{15.5073}, \frac{2Y}{2.7326}\right).$$

Problem 10.6. *(10 points)*
Let the random variable Y be normally distributed with mean zero and the standard deviation σ unknown. Using Y^2/σ^2 as a pivotal quantity, construct a 95%-confidence interval for σ^2.

Solution: We use the χ^2-tables for $Y^2/\sigma^2 \sim \chi^2(1)$. We have

$$0.95 = \mathbb{P}[0.0009821 \leq Y^2/\sigma^2 \leq 5.02389].$$

So, the required confidence interval is

$$\left(\frac{Y^2}{5.02389}, \frac{Y^2}{0.0009821} \right).$$