M378K Introduction to Mathematical Statistics
Fall 2024
University of Texas at Austin
Practice for In-Term Exam III
Instructor: Milica Čudina

Notes: This is a closed book and closed notes exam. The maximal score on the real exam will be 100 points.
There are many ways in which any single problem can be solved. The solutions herein are just one possible way to tackle the given problems.
Time: 50 minutes

All written work handed in by the student is considered to be
their own work, prepared without unauthorized assistance.

The University Code of Conduct

"The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the university is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community. As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity."

"I agree that I have complied with the UT Honor Code during my completion of this exam."

Signature:

3.1. **Formulas.** If Y has the binomial distribution with parameters n and p, then $p_Y(k) = \mathbb{P}[Y = k] = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, \dots, n$, $\mathbb{E}[Y] = np$, $\mathrm{Var}[Y] = np(1-p)$. The binomial coefficients are defined as follows for integers $0 \le k \le n$, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. The moment generating function of Y is given by $m_Y(t) = (pe^t + q)^n$.

If Y has a geometric distribution with parameter p, then $p_Y(k) = p(1-p)^k$ for $k = 0, 1, \dots$, $\mathbb{E}[Y] = \frac{1-p}{p}$, $\mathrm{Var}[Y] = \frac{1-p}{p^2}$. Its mgf is $m_Y(t) = \frac{p}{1-qe^t}$ for t such that $qe^t < 1$.

If Y has a Poisson distribution with parameter λ, then $p_Y(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ for $k = 0, 1, \dots$, $\mathbb{E}[Y] = \mathrm{Var}[Y] = \lambda$. Its mgf is $m_Y(t) = e^{\lambda(e^t - 1)}$.

If Y has a uniform distribution on $[l, r]$, its density is

$$f_Y(y) = \frac{1}{r-l}\mathbf{1}_{(l,r)}(y),$$

its mean is $\frac{l+r}{2}$, and its variance is $\frac{(r-l)^2}{12}$. Let $U \sim U(0,1)$. The mgf of U is $m_U(t) = \frac{1}{t}(e^t - 1)$.

If Y has the standard normal distribution, then its mean is zero, its variance is one, and its density equals

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad x \in \mathbb{R}.$$

Its mgf is $m_Y(t) = e^{\frac{t^2}{2}}$.

If Y has the exponential distribution with parameter τ, then its cumulative distribution function is $F_Y(y) = 1 - e^{-\frac{y}{\tau}}$ for $y \ge 0$, its probability density function is $f_Y(y) = \frac{1}{\tau}e^{-y/\tau}$ for $y \ge 0$. Also, $\mathbb{E}[Y] = SD[Y] = \tau$. Its mgf is $m_Y(t) = \frac{1}{1-\tau t}$.

The mgf of $Y \sim \Gamma(k, \tau)$ is

$$m_Y(t) = \frac{1}{(1-\tau t)^k} \text{ for } t < 1/\tau.$$

Its expectation is $k\tau$ and its variance is $k\tau^2$. The χ^2−distribution with n degrees of freedom is the special case $\Gamma\left(\frac{n}{2}, 2\right)$

3.2. **DEFINITIONS.**

Problem 3.1. (10 points) Write down the definition of the **bias** of an estimator $\hat{\theta}$ of a parameter θ.

Solution:

$$\text{bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$$

Problem 3.2. (10 points) Write down the definition of the **mean squared error** of an estimator $\hat{\theta}$ of a parameter θ.

Solution:

$$\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2]$$

3.3. **TRUE/FALSE QUESTIONS.**

Problem 3.3. (5 points) Let $n \geq 2$ and let $Y_1, Y_2, \dots, Y_n$ be a random sample from $E(\tau)$. Then, in our usual notation,

$$Y_{(n)} \sim E(\tau/n).$$

True or false? Why?

Solution: FALSE
We know that

$$Y_{(1)} \sim E(\tau/n)$$

On the other hand, $Y_{(1)} < Y_{(n)}$ with a positive probability (see, e.g., Problem **5.6.6** from the lecture notes). So, the proposed claim must be false.

Problem 3.4. (5 points) Let Z be a standard normal random variable and let Q^2 have the χ^2-distribution with $\nu \geq 2$ degrees of freedom. Assume that Z and Q^2 are independent. Set

$$T = \frac{Z}{Q^2}.$$

Then, T has a $t-$distribution with ν degrees of freedom. *True or false? Why?*

Solution: From the definition of the $t-$distribution, we know that it's actually

$$T = \frac{Z}{\sqrt{\frac{Q^2}{\nu}}}$$

that has the $t-$ distribution.

3.4. **Free-response problems.**

Please, explain carefully all your statements and assumptions. Numerical results or single-word answers without an explanation (even if they're correct) are worth 0 points.

Problem 3.5. (20 points) Consider a random sample $Y_1, Y_2, \dots, Y_n$ from the Weibull distribution with parameters m and α, i.e., the distribution whose density is given by

$$f_Y(y) = \frac{m}{\alpha} y^{m-1} e^{-\frac{y^m}{\alpha}} \mathbf{1}_{(0,\infty)}(y)$$

where m and α are positive constants. What is the distribution of $Y_{(1)}$? If you recognize it as a named distribution, provide its name and its parameters in terms of $n, m,$ and α. If not, provide its density.

Solution: By definition, we know

$$Y_{(1)} = \min(Y_1, Y_2, \dots, Y_n)$$

The cumulative distribution function of the Weibull distribution is, for $y > 0$,

$$F_Y(y) = \int_0^y \frac{m}{\alpha} u^{m-1} e^{-\frac{u^m}{\alpha}}\, du.$$

We can use the substitution $v = \frac{u^m}{\alpha}$ to get

$$F_Y(y) = \int_0^{\frac{y^m}{\alpha}} e^{-v}\, dv = 1 - e^{-\frac{y^m}{\alpha}}.$$

Now, the cdf of the first order statistic is

$$F_{(1)}(y) = 1 - (1 - F_Y(y))^n = 1 - \left(e^{-\frac{y^m}{\alpha}}\right)^n = 1 - \left(e^{-\frac{y^m}{\alpha}}\right)^n = 1 - e^{-\frac{y^m}{\frac{\alpha}{n}}}.$$

So, $Y_{(1)}$ is again Weibull with parameters m and $\frac{\alpha}{n}$.

Problem 3.6. (10 points) Let $Y_1, Y_2, \dots, Y_n$ be a random sample from a population with mean μ and standard deviation σ. Let $\tilde{Y}_1, \tilde{Y}_2, \dots, \tilde{Y}_n$ be a random sample from a population with mean $\tilde{\mu}$ and $\tilde{\sigma}$. Assume that the two random samples are mutually independent random. Show that

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n (Y_i - \tilde{Y}_i)$$

is a consistent estimator for $\mu - \tilde{\mu}$.

Solution: We can use the same theorem we used in class, i.e., we can demonstrate that

- $\hat{\theta}_n$ is *unbiased*; and
- $$\mathrm{Var}[\hat{\theta}_n] \to 0 \quad \text{as } n \to \infty$$

To show that $\hat{\theta}_n$ is unbiased, we must prove that

$$\mathbb{E}[\hat{\theta}_n] = \mu - \tilde{\mu}.$$

From the given definition of $\hat{\theta}_n$, we have

$$\mathbb{E}[\hat{\theta}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)\right] = \frac{1}{n}\sum_{i=1}^{n}\Big(\mathbb{E}[Y_i] - \mathbb{E}[\tilde{Y}_i]\Big) = \frac{1}{n}(n)(\mu - \tilde{\mu}) = \mu - \tilde{\mu}.$$

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

As for the other claim, we have that

$$\mathrm{Var}[\hat{\theta}_n] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \tilde{Y}_i)\right].$$

Since $Y_1, \dots, Y_n$ and $\tilde{Y}_1, \dots, \tilde{Y}_n$ are random samples and also mutually independent, the additive formula for the variance applies. So, we get

$$\begin{aligned}\mathrm{Var}[\hat{\theta}_n] &= \frac{1}{n^2}\sum_{i=1}^{n}\Big(\mathrm{Var}[Y_i] + \mathrm{Var}[\tilde{Y}_i]\Big)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\left(\sigma^2 + \tilde{\sigma}^2\right) = \frac{\sigma^2 + \tilde{\sigma}^2}{n}.\end{aligned}$$

Since the second moments are finite, the above converges to 0 as $n \to \infty$. Thus, our proof is concluded!

Problem 3.7. (10 points) Let $Y_1, Y_2, \dots, Y_n$ be a random sample from $N(\mu, \sigma)$ for $n \geq 3$. Consider the following two estimators for μ:

$$\hat{\theta}_1 = \bar{Y} \quad \text{and} \quad \hat{\theta}_2 = \frac{Y_1 + Y_n}{2}$$

Are these estimators unbiased? If so, find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

Solution: We have shown in class (multiple times) that $\bar{Y}$ is unbiased for μ. So, $\hat{\theta}_1$ is unbiased. However, the other estimator is also a sample mean (but just for a portion of the original sample). So, it is unbiased as well.

By definition, the relative efficiency we are looking for is

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

We have shown in class that

$$\mathrm{Var}[\hat{\theta}_1] = \mathrm{Var}[\bar{Y}] = \frac{\sigma^2}{n}.$$

Using similar reasoning, we have that

$$\mathrm{Var}[\hat{\theta}_2] = \mathrm{Var}\left[\frac{Y_1 + Y_n}{2}\right] = \frac{\sigma^2}{2}.$$

So, the relative efficiency is

$$\mathrm{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\frac{\sigma^2}{2}}{\frac{\sigma^2}{n}} = \frac{n}{2}.$$

Problem 3.8. (10 points) Let $Y_1, \dots, Y_n$ be a random sample from a Poisson distribution with an unknown parameter λ. What is the maximum likelihood estimator for λ? Make sure that you prove your claim!

Solution: Let $y_1, \dots, y_n$ represent a set of observations of the above random sample. Then, the likelihood function is

$$L(\lambda; y_1, \dots, y_n) = \prod_{i=1}^{n} \left(e^{-\lambda} \frac{\lambda^{y_i}}{y_i!} \right) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} y_i}}{y_1! \dots y_n!}.$$

The log-likelihood function is

$$\ell(\lambda; y_1, \dots, y_n) = -n\lambda + \left(\sum_{i=1}^{n} y_i \right) \ln(\lambda) - \sum_{i=1}^{n} \ln(y_i!).$$

Differentiating with respect to λ, we get

$$\ell'(\lambda; y_1, \dots, y_n) = -n + \frac{\sum_{i=1}^{n} y_i}{\lambda}.$$

Equating the above to 0 and solving for λ, we get

$$\hat{\lambda}_{MLE} = \bar{Y}.$$

Problem 3.9. (10 points) Let $Y_1, \dots, Y_n$ be a random sample from $E(\tau)$. Find a sufficient statistic for τ and justify your answer.

Solution: Let $y_1, \dots, y_n$ represent a set of observations of the above random sample. Then, the likelihood function is

$$L(\tau; y_1, \dots, y_n) = \prod_{i=1}^{n} \left(\frac{1}{\tau} e^{-\frac{y_i}{\tau}} \right) = \frac{1}{\tau^n} e^{-\sum_{i=1}^{n} y_i}.$$

Using the Fisher-Neymann factorization criterion, and setting - in our notation from the lecture notes) -

$$g(\tau; t) = \frac{1}{\tau^n} e^{-t} \quad \text{and} \quad h(y_1, \dots, y_n) = 1,$$

we conclude that $T(Y_1, \dots, Y_n) = \sum_{i=1}^{n} Y_i$ is a sufficient statistic for τ.

3.5. MULTIPLE CHOICE QUESTIONS.

Problem 3.10. (5 points) In a sample $Y_1, \dots, Y_n$ from the exponential distribution $E(\tau)$ with parameter $\tau > 0$, $U = c\bar{Y}$ is a pivotal quantity if the value of the constant c is

(a) 1 (b) 2 (c) $2/\tau$ (d) τ (e) none of the above

Solution: The correct answer is **(c)**.
In cases (a), (b) and (d), the expected values of U are τ, 2τ and τ^2, respectively. All of those depend on τ. Consequently, the distribution of U must also depend on τ in those cases. When $c = 2/\tau$, $\frac{2}{\tau}\bar{Y}$ has the $\chi^2(2n)$-distribution.

Problem 3.11. (5 points) In a random sample of 100 voters 80 prefer candidate A and the rest prefer candidate B. The (approximate) $(1-\alpha)$-confidence interval for the parameter p (the population proportion of A voters) is of the form

$$[0.8 - z_{\alpha/2} \times c, 0.8 + z_{\alpha/2} \times c],$$

where $z_{\alpha/2} = \texttt{qnorm}(1-\alpha/2, 0, 1)$. The value of c is:

(a) 0.01 (b) 0.02 (c) 0.03 (d) 0.04 (e) ≥ 0.05

Solution: The correct answer is **(d)** since $c = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{0.8 \times 0.2/100} = 0.04$.

Problem 3.12. (5 points) A sample of size $n = 2$ from normal distribution with unknown μ and σ is collected and the data are

$$y_1 = 1 \text{ and } y_2 = 5.$$

The left end-point of a symmetric 95% confidence interval for σ^2 is

(a) 8/`qchisq(0.975,2)` (b) 16/`qchisq(0.975,1)` (c) 8/`qchisq(0.975,1)` (d) 16/`qchisq(0.975,2)` (e) **None of the above**

Solution: The correct answer is **(c)**.
The confidence interval is based on the pivotal quantity $(n-1)S^2/\sigma^2$ whose distribution is $\chi^2(n-1)$. In this case, $n = 2$, $\bar{Y} = 3$ so that $(n-1)s^2 = (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 = 2^2 + 2^2 = 8$, which produces the interval $[8/\texttt{qchisq(0.975,1)}, 8/\texttt{qchisq(0.025,1)}]$.

Problem 3.13. (5 points) Let $Y_1, \ldots, Y_n$ be a random sample from the uniform distribution $U(0, \theta)$, with parameter $\theta > 0$. The MSE (mean-squared error) of the estimator $\hat{\theta} = c\bar{Y}$ for θ is the smallest when the constant c equals

(a) $\frac{1}{2}$ (b) 2 (c) $\frac{6n}{3n+1}$ (d) $\frac{3n}{6n+1}$ (e) none of the above

Solution: The correct answer is **(c)**.
The bias of $c\bar{Y}$ is

$$\mathbb{E}[c\bar{Y}] - \theta = c \times \tfrac{1}{n} \times \textstyle\sum_i \mathbb{E}[Y_i] - \theta = (\tfrac{c}{2} - 1)\theta.$$

The variance of each Y_i is

$$\tfrac{1}{\theta} \int_0^\theta (x - \theta/2)^2 \, dx = \tfrac{1}{12}\theta^2,$$

so that

$$\mathrm{Var}[c\bar{Y}] = \tfrac{c^2}{n^2} \times n \times \tfrac{1}{12}\theta^2.$$

Using the formula that MSE = bias^2 + s.e.^{2}, we get

$$\mathrm{MSE}[c\bar{Y}] = \theta^2 \Big((\tfrac{c}{2} - 1)^2 + \tfrac{c^2}{12n} \Big).$$

To find the c that minimizes this expression we differentiate it with respect to c and set the derivative to 0, i.e., solve

$$\theta^2 \Big((c/2 - 1) + c/(6n) \Big) = 0, \text{ i.e., } c = 1/(1/2 + 1/6n) = \tfrac{6n}{3n+1}.$$

Problem 3.14. (5 points) Let $Y_1, \dots, Y_5$ be a random sample from the normal distribution $N(\mu, \sigma)$, with an <u>unknown</u> mean μ and an <u>unknown</u> standard deviation σ. The collected data turn out to be

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

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

(a) $3 - \frac{1}{2}\texttt{qnorm}(0.1, 0, 1)$.

(b) $3 - \frac{\sqrt{5}}{\sqrt{8}}\texttt{qt}(0.1, 4)$.

(c) $3 - \frac{1}{\sqrt{2}}\texttt{qt}(0.1, 5)$.

(d) $3 - \frac{1}{\sqrt{2}}\texttt{qt}(0.1, 4)$.

(e) none of the above

Solution: The correct answer is **(d)**.
The confidence interval is based on the pivotal quantity $\frac{\bar{Y}-\mu}{\sqrt{S^2/n}}$ which has the t distribution with 4 degrees of freedom. Therefore, for $b = \texttt{qt}(0.1, 4)$ we have

$$\mathbb{P}[\frac{(\bar{Y}-\mu)}{\sqrt{S^2/n}} \geq b] = 0.9.$$

We solve for μ to obtain

$$\mathbb{P}[\mu \leq \bar{Y} - b\sqrt{S^2/n}] = 0.9.$$

For our data set $\bar{y} = 3$ and $S^2 = 5/2$, so $\hat{\mu}_R = 3 - \frac{1}{\sqrt{2}}\texttt{qt}(0.1, 4)$.

Problem 3.15. (5 points) Let $Y_1, \dots, Y_5$ be a random sample from the normal distribution $N(\mu, 2)$, with an <u>unknown</u> mean μ and the <u>known</u> 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 <u>left</u> end-point $\hat{\mu}_L$ of the symmetric 90%-confidence interval $[\hat{\mu}_L, \hat{\mu}_R]$ for μ is

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

(b) $3 - \frac{2}{\sqrt{5}}\texttt{qnorm}(0.95, 0, 1)$.

(c) $3 - \frac{1}{\sqrt{5}}\texttt{qt}(0.95, 4)$.

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

(e) none of the above

Solution: The correct answer is **(b)**. The confidence interval in this case is based on the pivotal quantity $\sqrt{5}\left(\frac{\mu-\bar{Y}}{2}\right)$ which has the $N(0, 1)$ distribution. Therefore, for $a = -\texttt{qnorm}(0.95, 0, 1)$ we have

$$\mathbb{P}\left[a \leq \sqrt{5}\left(\frac{\mu-\bar{Y}}{2}\right)\right] = 0.95.$$

We solve for μ to obtain

$$\mathbb{P}[\bar{Y} + \frac{2}{\sqrt{5}}a \leq \mu] = 0.95.$$

For our data set $\bar{y} = 3$, so $\hat{\mu}_L = 3 - \frac{2}{\sqrt{5}}\texttt{qnorm}(0.95, 0, 1)$.