Name:

UTeid:

Notes: This is a closed book and closed notes exam. The maximal score on the exam is 100 points.
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 $k = 0, \ldots, n$, $\mathbb{E}[Y] = np$, $\mathrm{Var}[Y] = np(1-p)$. 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 $k = 0, 1, \ldots$, $\mathbb{E}[Y] = \frac{1-p}{p}$, $\mathrm{Var}[Y] = \frac{1-p}{p^2}$. 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 $k = 0, 1, \ldots$, $\mathbb{E}[Y] = \mathrm{Var}[Y] = \lambda$. 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 **random sample** of size n from a distribution D.

Solution: A *random sample* of size n from a distribution D is a random vector

$$(Y_1, Y_2, \ldots, Y_n)$$

such that:

1. $Y_1, Y_2, \ldots, Y_n$ are **independent**, and
2. Y_i has the distribution D for every $i = 1, 2, \ldots, n$.

Problem 3.2. (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 3.3. (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 3.4. (5 points) Let Z be a standard normal random variable and let Q^2 have the χ^2-distribution with ν degrees of freedom. Assume that Z and Q^2 are independent. Set

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

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

Solution: TRUE
From the definition of the $t-$distribution, we know that

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

has the $t-$ distribution with ν degrees of freedom.

Problem 3.5. (5 points) Let $Y_1, Y_2, \ldots, Y_n$ be a random sample. Then,

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(0,\infty)}(Y_i)$$

is a well-defined statistic. *True or false?*

Solution: TRUE

Problem 3.6. (5 points) Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from $N(\mu, \sigma)$ with μ **known** and σ **unknown**. Then,

$$\frac{1}{n} \sum_{i=1}^{n} (Y_i - \mu)^2$$

is a well-defined estimator for σ^2. *True or false?*

Solution: TRUE

3.4. **Free-response problems.**

Problem 3.7. (15 points) Let $Y_1, \ldots, Y_n$ be a random sample from an exponential distribution with an unknown parameter τ. What is the maximum likelihood estimator for τ? Make sure that you prove your claim!

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

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

The log-likelihood function is

$$\ell(\tau; y_1, \ldots, y_n) = -n \ln(\tau) - \frac{1}{\tau} \sum_{i=1}^{n} y_i.$$

Differentiating with respect to τ, we get

$$\ell'(\tau; y_1, \ldots, y_n) = -\frac{n}{\tau} + \frac{\sum_{i=1}^{n} y_i}{\tau^2}.$$

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

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

Problem 3.8. (20 points) Let the time it takes for a single elf to construct a toy engine be a continuous random variable with the following probability density function

$$f_Y(y) = e^{-(y-\theta)}\mathbf{1}_{(\theta,\infty)}(y).\tag{3.1}$$

The parameter θ is assumed positive and corresponds to the minimum required engine construction time. The group of n independent elves is competing about who will construct their toy engine first. What is the density of the time until the first toy engine is constructed?

Solution: Let $Y_1, \ldots, Y_n$ denote the individual times until each elf is done constructing their toy engine. The time at which the fastest elf is done is, then, the first order statistic

$$Y_{(1)} = \min(Y_1, \ldots, Y_n).\tag{3.2}$$

Let's start by figuring out its cumulative distribution function. For each $y \leq \theta$, the cumulative distribution function equals 0. For $y > \theta$, we have

$$
\begin{aligned}
F_{(1)}(y) &= \mathbb{P}[Y_{(1)} \leq y]\\
&= \mathbb{P}[\min(Y_1, \ldots, Y_n) \leq y]\\
&= 1 - \mathbb{P}[\min(Y_1, \ldots, Y_n) > y]\\
&= 1 - \mathbb{P}[Y_1 > y, \ldots, Y_n > y].
\end{aligned}\tag{3.3}
$$

Since the elves are independent, the above equals

$$
\begin{aligned}
F_{(1)}(y) &= 1 - \mathbb{P}[Y_1 > y] \ldots \mathbb{P}[Y_n > y]\\
&= 1 - (1 - \mathbb{P}[Y_1 \leq y]) \ldots (1 - \mathbb{P}[Y_n \leq y])\\
&= 1 - (1 - F_Y(y))^n.
\end{aligned}\tag{3.4}
$$

So, the density of $Y_{(1)}$ is

$$g_{(1)}(y) = nf_Y(y)(1 - F_Y(y))^{n-1}.\tag{3.5}$$

Of course, some students would remember the above expression for the density of the first order statistic from class without needing to re-derive it.

Now, for each $y > \theta$, the cdf of Y can be obtained by integration as

$$F_Y(y) = \int_\theta^y f_Y(u)\,du = \int_\theta^y e^{-(u-\theta)}\,du = e^\theta \int_\theta^y e^{-u}\,du = e^\theta(e^{-\theta} - e^{-y}) = 1 - e^{-(y-\theta)}.\tag{3.6}$$

So, for $y > \theta$, we have

$$g_{(1)}(y) = ne^{-(y-\theta)}(e^{-(y-\theta)})^{n-1} = ne^{-n(y-\theta)}.\tag{3.7}$$

Alternatively: You could have noticed that an individual elf's work time could be expressed as

$$Y = \theta + Y'\tag{3.8}$$

where $Y' \sim E(1)$. Then, the minimum of $Y_1', \ldots, Y_n'$ is $E(1/n)$. So, the minimum work time would be

$$Y_{(1)} = \theta + Y_{(1)}'\tag{3.9}$$

with the density

$$g_{(1)}(y) = ne^{-n(y-\theta)}.\tag{3.10}$$

3.5. MULTIPLE CHOICE QUESTIONS.

Problem 3.9. (5 points) In a sample $Y_1, \ldots, Y_n$ from the uniform distribution on $(0, \theta)$ with parameter $\theta > 0$, $U = cY_{(n)}$ is a pivotal quantity if the value of the constant c is

(a) $\frac{1}{2}$

(b) $1/\theta$

(c) θ

(d) $\frac{n+1}{n}$

(e) none of the above

Solution: The correct answer is **(b)**.
The pivotal quantity

$$U = \frac{Y_{(n)}}{\theta} = \max(Y_1, \ldots, Y_n)/\theta$$

has the distribution with cdf y^n on $[0, 1]$.

Problem 3.10. (5 points) In a random sample of 100 voters 64 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

$$[0.64 - z_{\alpha/2} \times c, 0.64 + z_{\alpha/2} \times c],$$

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

(a) 0.016

(b) 0.024

(c) 0.036

(d) 0.048

(e) ≥ 0.05

Solution: The correct answer is **(d)** since $c = \sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{0.64 \times 0.36/100} = 0.048$.

Problem 3.11. (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 right end-point of a symmetric 95% confidence interval for σ^2 is

- (a) $8/\texttt{qchisq(0.975,2)}$
- (b) $16/\texttt{qchisq(0.025,1)}$
- (c) $8/\texttt{qchisq(0.025,1)}$
- (d) $16/\texttt{qchisq(0.975,2)}$
- (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 3.12. (5 points)Let $Y_1, \ldots, Y_5$ 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>right</u> end-point $\hat{\mu}_R$ of the symmetric 90%-confidence interval $[\hat{\mu}_L, \hat{\mu}_R]$ for μ is

- (a) $3 + \frac{2}{\sqrt{5}}\texttt{qnorm}(0.95, 0, 1)$.
- (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 **(a)**.
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
$$\mathbb{P}\left[-a \leq \sqrt{5}\left(\frac{\mu - \bar{Y}}{2}\right) \leq a\right] = 0.90.$$
We solve for μ to obtain
$$\mathbb{P}[\bar{Y} - \frac{2}{\sqrt{5}}a \leq \mu \leq \bar{Y} + \frac{2}{\sqrt{5}}a] = 0.90.$$
For our data set $\bar{y} = 3$, so $\hat{\mu}_R = 3 + \frac{2}{\sqrt{5}}\texttt{qnorm}(0.95, 0, 1)$.

Problem 3.13. (5 points) You are monitoring a cash register at a store. A seemingly endless queue of customers is waiting. The times it takes for Cassie the Cashier to check out a single customer is exponential with mean 5 minutes for each customer. Moreover, the customer service times are independent. Cassie the cashier can go on a break after every batch of 10 customers leave. She just came in from a break. What is the distribution of her waiting time until the next break?

- (a) $\Gamma(10, 5)$, i.e., $k = 10$, $\tau = 5$

(b) $\Gamma(50, 1)$, i.e., $k = 50$, $\tau = 1$

(c) $\chi^2(25)$

(d) $E(50)$

(e) $E(1/50)$

Solution: The correct answer is **(a)**.

We need the distribution of the sum of 10 independent exponential random variables $E(\tau)$ with parameter $\tau = 5$. Each $E(5)$ is a special case of the gamma distribution with parameters $k = 1$ and $\tau = 5$. Gammas are additive in the shape parameter, so the result is $\Gamma(10, 5)$.