BSDS (2024, Semester II): Statistics II

End-semester Examination

NAME :	ROLL :

Time: 180 minutes **Total attainable marks:** 50

1. Let $X_1, \ldots, X_n$ be a random sample from a one-parameter beta distribution with parameter θ and the probability density function (PDF) as follows

$$f_\theta(x) = \frac{\Gamma(\theta+1)}{\Gamma(\theta)}(1-x)^{\theta-1}, \quad 0 < x \le 1, \quad \theta > 0.$$

   (a) Based on the n samples, find the MLE of $1/\theta$.

   (b) Is the MLE same as the UMVUE of $1/\theta$? [4 + 6]

(a) $\frac{\Gamma(\theta+1)}{\Gamma(\theta)} = \frac{\theta\,\Gamma(\theta)}{\Gamma(\theta)} = \theta$

Thus, the log-likelihood fn. : $\ell(\theta) = \log\left(\theta^n \prod_{i=1}^n (1-x_i)^{\theta-1}\right)$

$= n\log\theta + (\theta-1)\sum_{i=1}^n \log(1-x_i)$

$\frac{\partial \ell(\theta)}{\partial\theta} = 0$ provides $\frac{n}{\theta} + \sum_{i=1}^n \log(1-x_i) = 0 \Rightarrow \theta = -\left[\frac{1}{n}\sum_{i=1}^n \log(1-x_i)\right]^{-1}$ — (FOC)

Further, $\frac{\partial^2 \ell(\theta)}{\partial\theta^2} = -\frac{n}{\theta^2} < 0 \quad \forall\, \theta$

$\therefore \quad \hat{\theta}_{MLE} = -\left\{\frac{1}{n}\sum_{i=1}^n \log(1-x_i)\right\}^{-1}$

By invariance property of MLE, if $\psi = \psi(\theta) = 1/\theta$, then

$\hat{\psi}_{MLE} = 1/\hat{\theta}_{MLE} = -\frac{1}{n}\sum_{i=1}^n \log(1-x_i)$.

(b) Observe that $f_\theta(\underset{\sim}{x}) = \exp\left\{\underbrace{n\log\theta}_{A(\theta)} + \underbrace{(\theta-1)}_{B(\theta)}\underbrace{\sum_{i=1}^n \log(1-x_i)}_{T(\underset{\sim}{x})}\right\}$

Thus, $f_\theta(\underset{\sim}{x})$ belongs to the exponential family and $T(\underset{\sim}{x}) = \sum_{i=1}^n \log(1-x_i)$ is a complete sufficient statistic.

Now, $E[T(\underset{\sim}{x})] = \sum_{i=1}^{n} E[\log(1-x_i)] = n\, E[\log(1-x_1)]$,

and $E[\log(1-x_1)] = \theta \int_0^1 \log(1-u)\,(1-u)^{\theta-1}\,du$

let $t = (1-u)^\theta$ then

u	0	1
t	1	0

$\partial t = -\theta(1-u)^{\theta-1}$

$$\therefore \quad E[\log(1-X_1)] = -\frac{1}{\theta}\int_1^0 \log t\, dt = \frac{1}{\theta}\int_0^1 \log t\, dt$$

$$= \frac{1}{\theta}\left[t\log t - \int dt\right]_0^1 = \frac{1}{\theta}\left[t\log t - t\right]_0^1$$

$$= -\frac{1}{\theta}$$

$$\therefore \quad E[\hat{\psi}_{MLE}] = \frac{1}{\theta} = \psi.$$

Thus, $\hat{\psi}_{MLE}$, being unbiased and function of CSS, is the UMVUE of $\psi = 1/\theta$.

2. Suppose it is known that the number of items produced in a factory during a week is a random variable with mean 50.

   (a) Provide an upperbound on the probability that this week's production will exceed 75?

   (b) If the variance of a week's production is known to equal 25, then can you have a better upperbound compared to that in part (a)?

   (c) Further, provide a lowerbound of the probability that this week's production will be between 40 and 60? [3 + 4 + 3]

(a) By Markov inequality $P(X>t) \leq \frac{E(X)}{t}$ (as X is non-neg.)

Thus, $P(X>75) \leq \frac{E(X)}{75} = \frac{50}{75} = \frac{2}{3} \approx 0.67$

(b) Again by Markov inequality,

$$P(X>t) = P(X^2 > t^2) \leq \frac{E(X^2)}{t^2} = \frac{var(X) + E^2(X)}{t^2}.$$

Using this, $P(X>75) \leq \frac{25+50^2}{75^2} = \frac{1+100}{3\times 75} \approx 0.45 < 0.67$

So, we get a sharper bound.

(c) By Chebyshev's inequality,

$$P(|X-E(X)| > t) \leq \frac{var(X)}{t^2}$$

Using this, $P(|X-50| > 10) = P(40 < X < 60) \leq \frac{25}{100}$

$= 1/4$.

3. Historical data indicate that 4% of the components produced at a certain manufacturing facility are defective. A labor dispute has recently been started, and management is curious about whether it will result in any change in this figure of 4%. If a random sample of 500 items indicated 25 defectives, is this significant evidence to conclude that products quality is now depreciated?

[Write explicitly the modeling assumptions (if any), hypotheses to be tested, test statistic, test function and the conclusion obtained after testing. You may use the facts: $\tau_{0.05} = 1.64$, $\tau_{0.01} = 2.32$, where τ_α is the upper α point of $N(0,1)$ distribution.] [10]

Let X be the random variable indicating the number of defectives.

$$X \sim Bin(p).$$

To test $H_0 : p = 0.04$ vs $H_1 : p > 0.04$

Let $X_1, \cdots, X_n$ with $n = 500$ be the status of 500 items examined.

$$X_i = \begin{cases} 1 & \text{if the } i\text{th item is defective} \\ 0 & \text{otherwise.} \end{cases}$$

$\sum_{i=1}^{n} X_i = 25$ is obtained.

As n is large, we consider a large sample test for testing H_0 vs H_1.

The large sample version of the UMP test is:

$$\phi_{LS}(\underset{\sim}{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i > c_\alpha \\ 0 & \text{otherwise,} \end{cases}$$

and it satisfies $E_{H_0}\left[\phi_{LS}(\underset{\sim}{x})\right] = P_{H_0}\left(\sum_{i=1}^{n} X_i > c_\alpha\right) = \alpha$

$$\Rightarrow P_{H_0}\left(\frac{\sum_{i=1}^{n} X_i - n \times 0.04}{\sqrt{n \times 0.04 \times 0.96}} > \frac{c_\alpha - n \times 0.04}{\sqrt{n \times 0.04 \times 0.96}}\right) = \alpha$$

$$\Rightarrow 1 - \Phi\left(\frac{c_\alpha - 20}{\sqrt{19.2}}\right) = \alpha$$

$$\therefore \tau_\alpha = \frac{c_\alpha - 20}{\sqrt{19.2}} \Rightarrow c_\alpha = \tau_\alpha \times \sqrt{19.2} + 20$$

$$= \begin{cases} 20 + \sqrt{19.2} \times 1.64 & \text{for } \alpha = 0.05 \\ 20 + \sqrt{19.2} \times 2.32 & \text{for } \alpha = 0.01 \end{cases}$$

$$= \begin{cases} 22.27 & \text{for } \alpha = 0.05 \\ 23.21 & \text{for } \alpha = 0.01 \end{cases}$$

As $\sum_{i=1}^{n} X_i$ is larger than C_α for both the choices of α, we reject H_0 at both the significance level.

Conclusion: Based on the data we may conclude that the product quality is now significantly depreciated.

4. Let $X_1, \ldots, X_n$ be a random sample from `uniform`$(0, \theta)$ distribution. Consider the problem of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ $(\theta_1 > \theta_0)$.

   (a) Draw the graph of $\lambda(\mathbf{x}) = f_1(\mathbf{x})/f_0(\mathbf{x})$ with respect to the statistic $X_{(n)}$, where f_0 and f_1 are joint pdfs of $\{X_1, \ldots, X_n\}$ under H_0 and H_1, respectively. Hence or otherwise, determine if $\lambda(\mathbf{x})$ is monotone with respect to $X_{(n)}$.
   [It is enough to consider the behavior of $\lambda(\mathbf{x})$ in the space of $\mathbf{x}$ where at least one of $f_0(\mathbf{x})$ or $f_1(\mathbf{x})$ is positive.]

(a) $X_1, \cdots, X_n \overset{\text{IID}}{\sim}$ Uniform $(0, \theta)$.

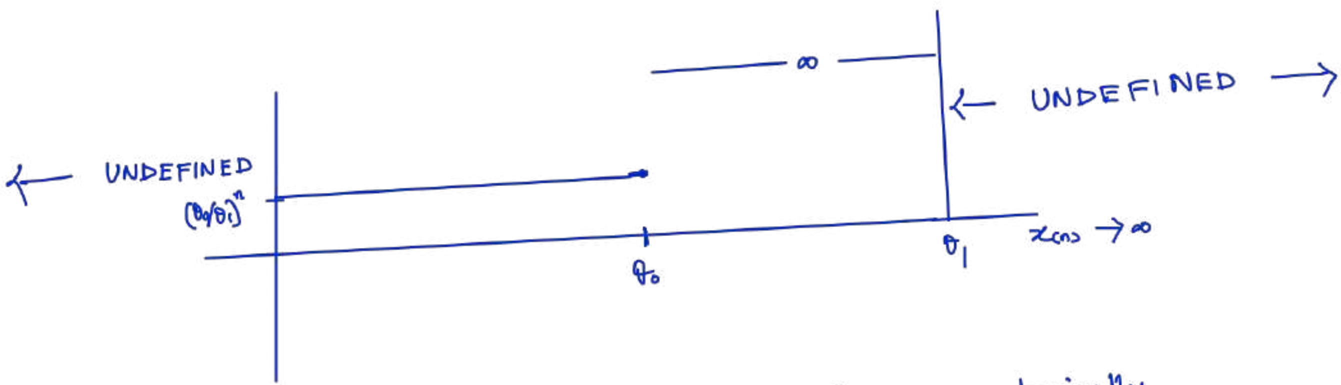
$H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$ $(\theta_1 > \theta_0)$

$$\lambda(\underset{\sim}{x}) = \frac{f_1(\underset{\sim}{x})}{f_0(\underset{\sim}{x})} = \frac{1}{\theta_1^n} \cdot \mathbb{I}\left(x_{(n)} \leq \theta_1\right) \left\{\frac{1}{\theta_0^n} \mathbb{I}\left(x_{(n)} \leq \theta_0\right)\right\}^{-1}$$

$$= \left(\frac{\theta_0}{\theta_1}\right)^n \frac{\mathbb{I}\left(x_{(n)} \leq \theta_1\right)}{\mathbb{I}\left(x_{(n)} \leq \theta_0\right)}$$

$\therefore$

$x_{(n)}$	$[-\infty, 0)$	$[0, \theta_0]$	$(\theta_0, \theta_1]$	(θ_1, ∞)
$\lambda(\underset{\sim}{x})$	undefined	$(\theta_0/\theta_1)^n$	∞	undefined



From the graph it is clear that $\lambda(\underset{\sim}{x})$ is a monotonically non-decreasing function of $X_{(n)}$.

(b) Consider the following two tests for testing H_0 against H_1:

$$\phi_0(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad x_{(n)} \geq k, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad \phi_1(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad x_{(n)} \geq \theta_0 \quad \text{or} \quad x_{(n)} \leq c, \\ 0 & \text{otherwise,} \end{cases} .$$

Find k and c such that both ϕ_0 and ϕ_1 are size α tests.

(b) <u>To find k:</u>

$$E_{H_0}\left[\phi_0(\underset{\sim}{x})\right] = P_{H_0}\left(x_{(n)} \geq k\right) = 1 - \left(\frac{k}{\theta_0}\right)^n = \alpha$$

$$\Rightarrow \quad k^n = \theta_0^n (1-\alpha) \quad \Rightarrow k = \theta_0 (1-\alpha)^{1/n}.$$

<u>To find c:</u>

$$E_{H_0}\left[\phi_1(\underset{\sim}{x})\right] = \underbrace{P_{H_0}\left(x_{(n)} \geq \theta_0\right)}_{0 \text{ under } H_0} + P_{H_0}\left(x_{(n)} \leq c\right) = \alpha$$

$$\Rightarrow \quad \left(\frac{c}{\theta_0}\right)^n = \alpha \quad \Rightarrow \quad c = \theta_0 \alpha^{1/n}.$$

(c) Find the power functions of the tests ϕ_0 and ϕ_1. Which one has higher power under H_1.

Power fn. of ϕ_0

$$\beta_{\phi_0}(\theta) = P_\theta\left(X_{(n)} \geq (1-\alpha)^{1/n}\theta_0\right)$$
$$= 1 - \left[\frac{(1-\alpha)^{1/n}\theta_0}{\theta}\right]^n = 1 - (1-\alpha)\left(\frac{\theta_0}{\theta}\right)^n$$

Power fn. of ϕ_1 :

$$\beta_{\phi_1}(\theta) = P_\theta\left(X_{(n)} \leq \alpha^{1/n}\theta_0\right) + P_\theta\left(X_{(n)} \geq \theta_0\right)$$
$$= \left(\frac{\alpha^{1/n}\theta_0}{\theta}\right)^n + 1 - \left(\frac{\theta_0}{\theta}\right)^n = 1 - \left(\frac{\theta_0}{\theta}\right)^n (1-\alpha).$$

Thus, both the power functions are same.

(d) Verify if ϕ_0 remains a size-α test for the choice of k obtained in part (b) if the null hypothesis is modified to $H_0' : \theta \leq \theta_0$. [3 + (1.5 + 1.5) + (1.5 + 1.5) + 3]

If $H_0' : \theta \leq \theta_0$ is considered, then size of ϕ_0 will be :

$$\begin{aligned}
\text{Size of } \phi_0 &= \sup_{\theta \leq \theta_0} \beta_{\phi_0}(\theta) \\
&= \sup_{\theta \leq \theta_0} \left[1 - (1-\alpha) \left(\frac{\theta_0}{\theta}\right)^n \right] \\
&= 1 - (1-\alpha) \inf_{\theta \leq \theta_0} \left(\frac{\theta_0}{\theta}\right)^n \\
&= 1 - 1 + \alpha = \alpha,
\end{aligned}$$

as $\left(\theta_0/\theta\right)^n$ is a strictly decreasing fn. of θ.

Thus, ϕ_0 will remain a size-α test under H_0'.

5. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{normal}(0, \sigma^2)$ distribution.

(a) Find the UMVUE of σ^2, say $\widehat{\sigma}_n^2$.

The joint density of $X_1, \cdots, X_n$ is

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \exp\left\{ \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right)\right)\right\}$$

$$= \exp\left\{ \underbrace{-\frac{n}{2}\log(2\pi)}_{A(\underset{\sim}{x})} - \underbrace{\frac{n}{2}\log\sigma^2}_{B(\theta)} - \underbrace{\frac{1}{2\sigma^2}}_{C(\theta)} \underbrace{\sum_{i=1}^{n} x_i^2}_{T(\underset{\sim}{x})} \right\}$$

$\therefore$ $f_{\underset{\sim}{X}}(\underset{\sim}{x})$ belongs to the exponential family and $T(\underset{\sim}{X}) = \sum_{i=1}^{n} X_i^2$ is a complete sufficient statistic (CSS).

$$E\left(T(\underset{\sim}{X})\right) = \sum_{i=1}^{n} E(X_i^2) = n\,E(X_1^2) = n\,var(X_1) + n\,E^2(X_1) = n\sigma^2.$$

$\therefore$ $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, being an unbiased function of CSS, is the UMVUE.

(b) Using a function of $\widehat{\sigma}_n^2$ as pivot, find a symmetric $(1-\alpha)$-confidence interval for σ^2.
[Note: By symmetry we indicate the following: If $T(\mathbf{X}, \sigma^2)$ is the pivot, then start with the points (a, b) such that $P\left(T(\mathbf{X}, \sigma^2) < a\right) = \alpha/2$ and $P\left(T(\mathbf{X}, \sigma^2) > b\right) = \alpha/2$.]

As $\frac{X_i}{\sigma} \sim N(0,1)$, we have $\sum_{i=1}^{n} \frac{X_i^2}{\sigma^2} \sim \chi^2_{(n)}$ distn.

Thus, $T(\underset{\sim}{X}, \sigma^2) = \sum_{i=1}^{n} X_i^2/\sigma^2$ has (i) a completely known distribution free of σ^2; (ii) is a strictly decreasing function of σ^2, and (iii) $T(\underset{\sim}{X}, \sigma^2) = a$ is solvable for σ^2.

So, $T(\underset{\sim}{X}, \sigma^2)$ is a valid pivot.

Let $\chi^2_{\alpha;n}$ be the upper-α point of $\chi^2_{(n)}$ distribution.

Then $P_{\sigma^2}\left(\chi^2_{(1-\alpha/2);n} \le T(\underset{\sim}{X}, \sigma^2) \le \chi^2_{\alpha/2;n}\right) = 1-\alpha$

$$\Rightarrow P_{\sigma^2}\left(\chi^2_{(1-\alpha/2);n} \le \frac{\sum_{i=1}^{n} X_i^2}{\sigma^2} \le \chi^2_{\alpha/2;n}\right) = (1-\alpha)$$

$$\Rightarrow P_{\sigma^2}\left(\frac{\sum_{i=1}^{n} X_i^2}{\chi^2_{\alpha/2;n}} \le \sigma^2 \le \frac{\sum_{i=1}^{n} X_i^2}{\chi^2_{1-\alpha/2;n}}\right) = (1-\alpha)$$

Thus, $\underbrace{\left[\frac{\sum_{i=1}^{n} X_i^2}{\chi^2_{\alpha/2;n}}, \frac{\sum_{i=1}^{n} X_i^2}{\chi^2_{1-\alpha/2;n}}\right]}_{\left[\frac{n\hat{\sigma}_n^2}{\chi^2_{\alpha/2;n}}, \frac{n\hat{\sigma}_n^2}{\chi^2_{1-\alpha/2;n}}\right]}$ is a $(1-\alpha)$-confidence interval based on $\hat{\sigma}_n^2$.

(c) Find the large sample distribution of the pivot $T_n = T(\mathbf{X}, \sigma^2)$, i.e., find sequence of real numbers $\{a_n\}$ and $\{b_n\}$ such that

$$\frac{T_n - a_n}{b_n} \xrightarrow{d} Z, \quad \text{as } n \to \infty,$$

where Z is a non-degenerate distribution.

As $X_1, \cdots, X_n$ are IID, by CLT, $\dfrac{\sum_{i=1}^{n} X_i^2 - E\left(\sum_{i=1}^{n} X_i^2\right)}{\sqrt{\operatorname{var}\left(\sum_{i=1}^{n} X_i^2\right)}} \xrightarrow{d} Z$, where $Z \sim N(0,1)$

Here $E\left(\sum_{i=1}^{n} X_i^2\right) = n\sigma^2$ and

$\operatorname{var}\left(\sum_{i=1}^{n} X_i^2\right) = n \operatorname{var}\left(X_1^2\right) = 2n\sigma^4$

as $\dfrac{X_1^2}{\sigma^2} \sim \chi^2_{(1)}$

and $\operatorname{var}\left(\dfrac{X_1^2}{\sigma^2}\right) = 2$

$\therefore \dfrac{\sum_{i=1}^{n} X_i^2 - n\sigma^2}{\sqrt{2n\sigma^4}}$

$= \dfrac{T(\underset{\sim}{X}, \sigma^2) - n}{\sqrt{2n}} \xrightarrow{d} Z$ as $n \to \infty$ where $Z \sim N(0,1)$.

$\therefore a_n = n$ and $b_n = \sqrt{2n}$ $\forall n$.

(d) Based on the large sample distribution obtained in (c) derive an approximate symmetric $(1-\alpha)$-confidence interval for σ^2.

Let τ_α be the upper α point of $N(0,1)$ distribution.

Then by above large sample distribution

$$P_{\sigma^2}\left(-\tau_{\alpha/2} \leq \frac{T(\underset{\sim}{X},\sigma^2)-n}{\sqrt{2n}} \leq \tau_{\alpha/2}\right) = (1-\alpha)$$

$$\Rightarrow P_{\sigma^2}\left(n-\sqrt{2n}\,\tau_{\alpha/2} \leq \frac{\sum_{i=1}^n X_i^2}{\sigma^2} \leq n+\tau_{\alpha/2}\sqrt{2n}\right) = (1-\alpha)$$

$$\Rightarrow P_{\sigma^2}\left(\frac{\sum_{i=1}^n X_i^2}{n+\tau_{\alpha/2}\sqrt{2n}} \leq \sigma^2 \leq \frac{\sum_{i=1}^n X_i^2}{n-\tau_{\alpha/2}\sqrt{2n}}\right) = (1-\alpha).$$

Thus, $\left[\frac{\sum_{i=1}^n X_i^2}{n+\tau_{\alpha/2}\sqrt{2n}}, \frac{\sum_{i=1}^n X_i^2}{n-\tau_{\alpha/2}\sqrt{2n}}\right] \equiv \left[\frac{n\hat{\sigma}_n^2}{n+\tau_{\alpha/2}\sqrt{2n}}, \frac{n\hat{\sigma}_n^2}{n-\tau_{\alpha/2}\sqrt{2n}}\right]$

is a symmetric $(1-\alpha)$ - confidence interval based on the large sample distn.

(e) For $\alpha = 0.05$ and $n = 25$ compare the two confidence intervals obtained in parts (b) and (d) in terms of the ratio of the upper confidence limit (UCL) and the lower confidence limit (LCL). Which one provides a sharper confidence interval?

You may use the following facts: Let τ_α and $\chi^2_{\alpha,r}$ be the upper-α points of $N(0,1)$ and χ^2_r distributions. Then

α	τ_α	$\chi^2_{\alpha,24}$	$\chi^2_{\alpha,25}$	$\chi^2_{1-\alpha,24}$	$\chi^2_{1-\alpha,25}$
0.025	1.64	39.36	40.64	12.40	13.12
0.05	1.96	36.42	37.65	13.85	14.61

For $\alpha = 0.05$, $n = 25$, $\frac{UCL}{LCL}$ of the exact CI is:

$$\frac{\chi^2_{\alpha/2;n}}{\chi^2_{1-\alpha/2;n}} = \frac{40.64}{13.12} \approx 3.09$$

$\frac{UCL}{LCL}$ of the large sample CI is:

$$\frac{n + \tau_{\alpha/2}\sqrt{2n}}{n - \tau_{\alpha/2}\sqrt{2n}} = \frac{25 + 1.64 \times 5 \times \sqrt{2}}{25 - 1.64 \times 5 \times \sqrt{2}}$$

$$= \frac{36.5966}{13.4034} \approx 2.73$$

$\therefore$ The large sample CI is sharper.

(f) Find the minimum sample size n_0 ensuring that the ratio of UCL and LCL obtained in part (d) is at most 1.1. [$3 \times 6 = 18$]

The ratio of UCL and LCL obtained in part (d):

$$\frac{n + 1 \cdot 64 \sqrt{2n}}{n - 1 \cdot 64 \sqrt{2n}} \leq 1 \cdot 1$$

$$\Rightarrow \quad n + 1 \cdot 64 \sqrt{2n} \leq 1 \cdot 1 n - 1 \cdot 1 \times 1 \cdot 64 \times \sqrt{2n}$$

$$\Rightarrow \quad 0 \cdot 1 n \geq 1 \cdot 64 \times \sqrt{2n} \times 2 \cdot 1$$

$$\Rightarrow \quad \sqrt{n} \geq \frac{2 \cdot 1 \times 1 \cdot 64 \times \sqrt{2}}{0 \cdot 1} \quad \Rightarrow \quad n \geq 2372 \cdot 23$$

$\therefore$ n is at least 2373.

x——x