THE UNIVERSITY OF NEW SOUTH WALES

SCHOOL OF MATHEMATICS AND STATISTICS

Term One 2022

MATH5905
Statistical Inference

(1) TIME ALLOWED – THREE (3) HOURS

(2) TOTAL NUMBER OF QUESTIONS – 5

(3) ANSWER ALL QUESTIONS

(4) THE QUESTIONS ARE **NOT** OF EQUAL VALUE

(5) START EACH QUESTION ON A NEW PAGE.

(6) THIS PAPER MAY BE RETAINED BY THE CANDIDATE

YOU ARE TO COMPLETE THE TEST UNDER STANDARD EXAM CONDITIONS, WITH HANDWRITTEN SOLUTIONS.

YOU WILL THEN SUBMIT ONE OR MORE FILES CONTAINING YOUR SOLUTIONS. **MAKE SURE YOU SUBMIT ALL YOUR ANSWERS.**

ONE OF THE SUBMITTED FILES MUST INCLUDE A PHOTOGRAPH OF YOUR **STUDENT ID CARD** WITH THE **SIGNED**, HANDWRITTEN STATEMENT:
"I declare that this submission is entirely my own original work."

YOU CAN DELETE AND/OR RELOAD FILES UNTIL THE DEADLINE.

Start a new page clearly marked Question 1

1. **[12 marks]** Consider a decision problem with parameter space $\Theta = \{\theta_1, \theta_2\}$ and a set of non randomized decisions $D = \{d_i, 1 \le i \le 6\}$ with risk points $\{R(\theta_1, d_i), R(\theta_2, d_i)\}$ as follows:

i	1	2	3	4	5	6
$R(\theta_1, d_i)$	1	2	3	5	6	4
$R(\theta_2, d_i)$	7	4	6	2	6	7

a) **[1 mark]** Find the minimax rule(s) amongst the non-randomized rules in D.

b) **[1 mark]** Plot the risk set of all randomized rules $\mathcal{D}$ generated by the set of rules in D.

c) **[4 marks]** Find the risk point of the minimax rule in $\mathcal{D}$ and determine its minimax risk. Compare this risk with the risk found in the set of non-randomized decision rules D. Comment.

d) **[2 marks]** Define the minimax rule in the set $\mathcal{D}$ in terms of rules in D.

e) **[2 marks]** For which prior on $\{\theta_1, \theta_2\}$ is the minimax rule a Bayes rule?

f) **[2 marks]** Find the Bayes rule with respect to the prior $(1/5, 4/5)$ on (θ_1, θ_2) and determine its Bayes risk.

Start a new page clearly marked Question 2

2. **[11 marks]** We perform Bayesian estimation with quadratic loss $L(\theta, a) = (\theta - a)^2, \Theta = \mathcal{A} = (0, \infty)$. We are given a sample $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ of n i.i.d. random variables each with conditional distribution being Poisson: $f(x|\theta) = \frac{e^{-\theta}\theta^x}{x!}, x = 0, 1, 2, \ldots$

a) **[6 marks]** The prior on Θ is Gamma(2,1), with a density $\tau(\theta) = \begin{cases} \theta e^{-\theta} & \text{if } \theta > 0 \\ 0 & \text{elsewhere} \end{cases}$.
Show that the Bayes estimator of θ with respect to this prior is

$$\hat{\theta}_B = (\sum_{i=1}^{n} X_i + 2)/(n+1).$$

b) **[5 marks]** Compute the usual risk $R(\theta, \hat{\theta}_B)$ and then show that its Bayes risk is $r(\tau, \hat{\theta}_B) = \frac{2}{n+1}$.
Hint: Recalling the general definition of a Gamma(α, β) density:

$$f(x; \alpha, \beta) = \frac{e^{-\frac{x}{\beta}} x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}, x > 0,$$

where $\Gamma(\alpha) = \int_0^\infty \exp(-x) x^{\alpha-1} dx$.
For such a Gamma distributed random variable X, $E(X) = \alpha\beta, Var(X) = \alpha\beta^2$ holds.

Start a new page clearly marked Question 3

3. [**15 marks**]

Let $X_1, X_2, \ldots, X_n$ be a sample of i.i.d. random variables, each Raleigh-distributed, with a density

$$f(x;\theta) = \begin{cases} 2\frac{x}{\theta}e^{-\frac{x^2}{\theta}}, x > 0, \\ 0 \text{ elsewhere} \end{cases}$$

where $\theta > 0$ is an unknown parameter.

a) [**2 marks**] Show that $E(X_1^2) = \theta$ holds and, using this fact, calculate the Fisher information in this sample about θ.

b) [**2 marks**] Derive the Maximum Likelihood Estimator of θ.

c) [**2 marks**] Is the MLE unbiased for θ? Is it also an UMVUE of θ? Justify your answer.

d) [**2 marks**] Argue that the family has a monotone likelihood ratio in the statistic $T = \sum_{i=1}^n X_i^2$.

e) [**2 marks**] Show that $Y = \frac{2}{\theta}X_1^2$ has a χ_2^2 distribution with a density $f_Y(y) = \frac{1}{2}exp(-\frac{y}{2}), y > 0$.
Hence argue that $\frac{2T}{\theta}$ has a chi-square distribution with $2n$ degrees of freedom.
Hint. You may use the density transformation formula $f_Y(y) = f_X(x)|\frac{dx}{dy}|$.

f) [**2 marks**] Using d) and e), derive completely the uniformly most powerful size-α test φ^* of $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ (θ_0 is a specified constant). Justify all steps.

g) [**2 marks**] Justify the claim that for testing for testing $H_0 : \theta = \theta_0 = 2$ versus a two-sided alternative $H_1 : \theta \neq 2$, a Uniformly Most Powerful Unbiased (UMPU) α-size test exists and exhibit its structure.

h) [**1 mark**] It can be shown that $E(T) = \theta n, Var(T) = \theta^2 n$ holds for $T = \sum_{i=1}^n X_i^2$. Using Central Limit Theorem arguments, and the above $E(T)$ and $Var(T)$ values, specify completely (including the threshold constant) such an approximate UMPU α test.

Start a new page clearly marked Question 4

4. [**10 marks**] A random sample $\mathbf{X} = (X_1, X_2, X_3)$ of size $n = 3$ is taken from a population with density

$$f(x) = 2x, x \in [0, 1].$$

a) [**3 marks**] Find the marginal densities of $X_{(1)}$ and of $X_{(3)}$, as well as their joint density. Make sure that you define the support of the density in each case.

b) [**4 marks**] Find the density of the range $R = X_{(3)} - X_{(1)}$.

c) [**3 marks**] Find $E(X_{(1)}), E(X_{(3)})$, and $E(R)$.

Hint: The following formulae may be helpful:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} F(x)^{r-1}[1 - F(x)]^{n-r} f(x)$$

$$f_{X_{(i)},X_{(j)}}(u, v) =$$

$$\frac{n!}{(i-1)!(j-1-i)!(n-j)!} f(u)f(v)[F(u)]^{i-1}[F(v) - F(u)]^{j-1-i}[1 - F(v)]^{n-j}$$

for $-\infty < u < v < \infty$.

Start a new page clearly marked Question 5

5. [12 marks]

Consider the case where $X_i, i = 1, 2, \ldots, n$ are i.i.d. exponentially distributed with density $f(x) = \frac{1}{\theta}e^{-x/\theta}, x > 0$. Here $\theta > 0$ is the parameter of the distribution.

a) **[1 mark]** Show that the cumulant generating function $K_X(t)$ of a single observation from this distribution is defined whenever $t < \frac{1}{\theta}$ and that

$$K_X(t) = -\ln(1 - \theta t)$$

holds.

b) **[2 marks]** Compute the first and second derivative of the function $K_X(t)$.

c) **[4 marks]** Derive the first order saddlepoint approximation of the density $\hat{f}(\bar{x})$ of the arithmetic mean of n i.i.d. observations from the above exponential distribution.

d) **[2 marks]** Derive, using c) and the density transformation formula, the density of the sum of n i.i.d. exponentially distributed random variables.

e) **[3 marks]** Using Stirling's approximation

$$\Gamma(n+1) = n! \approx \sqrt{2\pi n}e^{-n}n^n$$

for large n, show that the approximation in d) coincides up to a normalising constant with the exact density (which is Gamma(n, θ)). Show that the normalising constant tends to 1 when $n \to \infty$.

Hint: The first order saddlepoint approximation is

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{n}{2\pi K_X''(\hat{t})}}e^{\{nK_X(\hat{t}) - n\hat{t}\bar{x}\}}.$$

Table of Common Distributions

Discrete Distributions

Bernoulli(p)
pmf $P(X = x|p) = p^x(1-p)^{1-x}; \quad x = 0, 1; \quad 0 \le p \le 1$
mean and variance $\mathbb{E}(X) = p, \quad \mathrm{Var}(X) = p(1-p)$
mgf $M_X(t) = (1-p) + pe^t$

Binomial(n, p)
pmf $P(X = x|n, p) = \binom{n}{x} p^x(1-p)^{n-x}; \quad x = 0, 1, \ldots n; \quad 0 \le p \le 1$
mean and variance $\mathbb{E}(X) = np, \quad \mathrm{Var}(X) = np(1-p)$
mgf $M_X(t) = [(1-p) + pe^t]^n$

Poisson(λ)
pmf $P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, \ldots; \quad 0 \le \lambda < \infty$
mean and variance $\mathbb{E}(X) = \lambda, \quad \mathrm{Var}(X) = \lambda$
mgf $M_X(t) = e^{\lambda(e^t-1)}$

Continuous Distributions

Beta(α, β)
pdf $f(x|\alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}; \quad 0 \le x \le 1; \quad \alpha, \beta > 0,$
$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and $\Gamma(n) = (n-1)!$
mean and variance $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
mgf $M_X(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}$

Cauchy (θ, σ)
pdf $f(x|\theta, \sigma) = \frac{1}{\pi\sigma} \frac{1}{1+\left(\frac{x-\theta}{\sigma}\right)^2}, \quad -\infty < x < \infty, -\infty < \theta < \infty, \sigma > 0$
mean and variance do not exist
mgf does not exist

Chi squared(p)
pdf $f(x|p) = \frac{1}{\Gamma(p/2)\, 2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad 0 \le x < \infty, \quad p = 1, 2, 3, \ldots$
mean and variance $\mathbb{E}(X) = p, \quad \mathrm{Var}(X) = 2p$
mgf $M_X(t) = \left(\frac{1}{1-2t}\right)^{p/2}, \, t < \frac{1}{2}$

Exponential(β)
pdf $f(x|\beta) = \frac{1}{\beta} e^{\frac{-x}{\beta}}; \quad 0 \le x \le \infty; \quad \beta > 0$
mean and variance $\mathbb{E}(X) = \beta, \quad \mathrm{Var}(X) = \beta^2$
mgf $M_X(t) = \frac{1}{1-\beta\, t}, \quad t < \frac{1}{\beta}$

Gamma(α, β)
pdf $f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}}\, x^{\alpha-1}\, e^{\frac{-x}{\beta}}; \quad 0 \le x \le \infty; \quad \alpha, \beta > 0$
mean and variance $\mathbb{E}(X) = \alpha\, \beta, \quad \mathrm{Var}(X) = \alpha\, \beta^2$
mgf $M_X(t) = \left(\frac{1}{1-\beta\, t}\right)^{\alpha}, \quad t < \frac{1}{\beta}$

Normal(μ, σ^2)
pdf $f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{\frac{-(x-\mu)^2}{2\sigma^2}}; \quad -\infty \le x \le \infty; \quad -\infty < \mu < \infty, \, \sigma > 0$
mean and variance $\mathbb{E}(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2$
mgf $M_X(t) = \exp\{\mu\, t + \frac{\sigma^2 t^2}{2}\}$

Uniform (a, b)
pdf $f(x|a, b) = \frac{1}{b-a}; \quad a \le x \le b$
mean and variance $\mathbb{E}(X) = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$
mgf $M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$