THE UNIVERSITY OF NEW SOUTH WALES

SCHOOL OF MATHEMATICS AND STATISTICS

Term One 2021

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 $D = \{d_i, 1 \le i \le 7\}$ with risk points $\{R(\theta_1, d_i), R(\theta_2, d_i)\}$ as follows:

i	1	2	3	4	5	6	7
$R(\theta_1, d_i)$	0	1	3	4	6	5	3
$R(\theta_2, d_i)$	5	2	4	1	4	6	3

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

b) **[1 mark]** Sketch the risk set of all randomized decision rules $\mathcal{D}$ generated by the set of non-randomized decision rules D.

c) **[3 marks]** Find the risk point of the minimax rule in $\mathcal{D}$ and determine its minimax risk. Compare this risk with the minimax risk found in the set of non-randomized decision rules D.

d) **[3 marks]** Define the minimax rule in the set $\mathcal{D}$ in terms of rules in D.

e) **[3 marks]** Find the prior on $\{\theta_1, \theta_2\}$ such that the minimax rule in the set $\mathcal{D}$ is also a Bayes rule. Computes it's Bayes risk.

f) **[1 mark]** For a small positive ϵ, illustrate on the risk set the risk points of all rules which are ϵ-Bayes for the prior found in part (e).

Start a new page clearly marked Question 2

2. **[12 marks]** Let $X_1, X_2, \ldots, X_n$ be a random sample with the common density function

$$f(x, \theta) = \theta^{-1} e^{-x/\theta}, \qquad x > 0$$

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

a) **[2 marks]** Provide justification that the statistic $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ is a complete and sufficient statistic for θ.

b) **[1 mark]** Determine the UMVUE of θ.

c) **[1 mark]** Determine the MLE of θ.

d) **[1 mark]** The density can also be written in the form

$$f(x, \tau) = \tau e^{-\tau x}, \qquad x > 0$$

where $\tau = \frac{1}{\theta} > 0$ is an unknown parameter. Determine the MLE of τ.

e) **[3 marks]** Show that the UMVUE of τ is given by

$$\hat{\tau}_{\text{umvue}} = \frac{n-1}{n\bar{X}}.$$

Hence, as usual the reciprocal of the MLE is the MLE of $1/\theta$, but, in this situation, the reciprocal of the UMVUE is not the UMVUE of $1/\theta$.
Hint: consider $\mathbb{E}(\bar{X}^{-1})$ and note that $\Gamma(n) = (n-1)\Gamma(n-1)$ and

$$T = \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \theta).$$

f) **[2 marks]** Determine the MLE of $h(\theta) = P(X \leq 2)$.

g) **[2 marks]** Determine the asymptotic distribution of $h(\theta) = P(X \leq 2)$.

Start a new page clearly marked Question 3

3. **[13 marks]** Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be independent and identically distributed random variables from a population with density

$$f(x, \theta, \phi) = \frac{\phi}{\theta}\left(\frac{\theta}{x}\right)^{\phi+1}, \quad x \geq \theta,\ \theta > 0,\ \phi > 0,$$

where θ and ϕ are unknown parameters.

a) **[2 marks]** Apply the Factorization Criterion to show that a two-dimensional sufficient statistic for (θ, ϕ) is

$$T = (T_1, T_2) = \left(\prod_{i=1}^{n} X_i, X_{(1)}\right).$$

b) **[2 marks]** Show that the sufficient statistic T in part (a) is also minimal sufficient statistic for (θ, ϕ).

c) **[2 marks]** Show that the density of the statistic $T = X_{(1)}$ is given by

$$f_{X_{(1)}}(x) = \frac{n\phi}{\theta}\left(\frac{\theta}{x}\right)^{n\phi+1} \qquad \text{for} \qquad x \geq \theta$$

otherwise zero.
Hint: The density for the minimum is given by

$$f_{X_{(1)}}(y_1) = n[1 - F_X(y_1)]^{n-1} f_X(y_1).$$

d) **[2 marks]** For the remainder of this question assume that the parameter ϕ is known. Find the MLE of θ and provide appropriate justification.

e) **[2 marks]** Show that the MLE is a biased estimator for θ.

f) **[3 marks]** Show that $T = X_{(1)}$ is complete and hence find the UMVUE of θ.

Start a new page clearly marked Question 4

4. **[12 marks]** Let $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ be a random sample from a normal distribution $\mathrm{N}(\theta,\, 8)$.

a) **[1 mark]** Compute the likelihood function for this sample.

b) **[2 marks]** Find a complete and minimal sufficient statistic for this family.

c) **[3 marks]** From the definition of monotone likelihood ratio, show that this family has a monotone likelihood ratio in its complete and sufficient statistic.

d) **[2 marks]** Provide the structure of the UMP α-test for testing H_0: $\theta \geq 10$ versus $H_1 : \theta < 10$ giving appropriate justification.

e) **[4 marks]** Find the sample size n with power function $\gamma(\theta)$ so that approximately $\gamma(8) = 0.95$ and $\gamma(10) = 0.05$.
Hint: The distribution of $\sum_{i=1}^{n} X_i$ is normal with mean θn and variance $8n$ and the upper 0.05 percentile of a standard normal is 1.65.

Start a new page clearly marked Question 5

5. **[11 marks]** Consider n independent identically distributed (iid) observations from a χ^2_p distribution.

a) **[1 mark]** Determine the cumulant generating function for a single observation from a χ^2_p distribution.

b) **[1 mark]** Compute the first and second derivative of the cumulant generating function.

c) **[1 mark]** Find the saddlepoint $\hat{t}$.

d) **[2 marks]** Evaluate $K_X(\hat{t})$ and $K''_X(\hat{t})$ and simplify as much as possible.

e) **[3 marks]** Show that the first order saddlepoint approximation for the density of $\bar{X}$ is

$$\hat{f}(\bar{x}) \approx \sqrt{\frac{np}{4\pi}} p^{-np/2} e^{np/2} \bar{x}^{np/2-1} e^{-n\bar{x}/2}$$

Hint: The first order saddlepoint approximation is

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

f) **[2 marks]** Hence, determine the first order saddlepoint approximation for the density of the sum of n iid observations $S = \sum_{i=1}^{n} X_i$ from this distribution.
Hint: Consider using the density transformation formula

$$f_S(s) = f_X(x(s)) \left| \frac{\mathrm{d}x}{\mathrm{d}s} \right|.$$

g) **[1 mark]** What is the approximate distribution of this saddlepoint approximation? Keep in mind the famous *Stirling approximation* of the Gamma function that states

$$\Gamma\left(\frac{np}{2}\right) \approx \sqrt{\frac{4\pi}{np}} \left(\frac{np}{2}\right)^{np/2} e^{-np/2}.$$

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 $P(X=x|n,p)=\binom{n}{x}p^x(1-p)^{n-x};\quad x=0,1,\dots n;\quad 0\le p\le 1$
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 $P(X=x|\lambda)=\frac{e^{-\lambda}\lambda^x}{x!};\quad x=0,1,\dots;\quad 0\le\lambda<\infty$
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 $f(x|p) = \frac{1}{\Gamma(p/2)\, 2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad 0 \le x < \infty, \quad p = 1, 2, 3, \dots$
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 $M_X(t) = \frac{e^{bt} - e^{at}}{(b-a)t}$

END OF EXAMINATION