Student ID:

The University of Melbourne

Semester 1 Assessment — 2019

School of Mathematics and Statistics

MAST90105 Methods of Mathematical Statistics

Exam duration: **3 hours**
Reading time: **15 minutes**
This paper has **6 pages** including this page

Authorised materials:
The calculator authorised at the University of Melbourne is the CASIO FX82 and this is permitted.
Two A4 double-sided handwritten sheets of notes.

Instructions to invigilators:
Sixteen-page script books shall be supplied to each student.
Students may not take this paper with them at the end of the exam.

Instructions to students:
There are 9 questions. All questions may be attempted.
The number of marks for each question is indicated after the question.
The total number of marks available is 100.
Your raw mark of this exam will be multiplied with 0.35 before being added to your final subject mark.

This exam paper is not to be held by Baillieu Library.

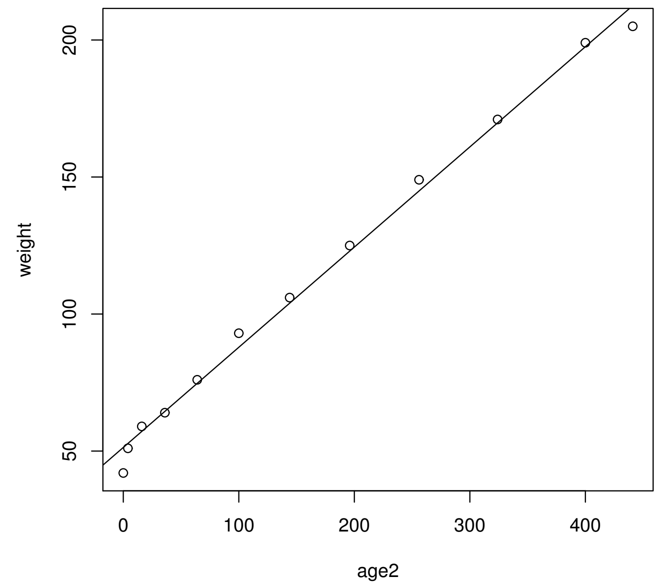

Figure 1: see Question 1(a)

1. The body weight of a chick was measured at birth and every second day thereafter until day 20 and also on day 21. The data are recorded in the following table:

weight (grams)	42	51	59	64	76	93	106	125	149	171	199	205
age (days)	0	2	4	6	8	10	12	14	16	18	20	21

We want to model the weight as a quadratic function of age:

$$\text{weight} = \alpha + \beta \cdot \text{age}^2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2).$$

(a) What do the following R commands and output, including Figure 1, reveal about the relationship between weight and age?

```
> age2=age^2
> out=lm(weight~age2)
> plot(age2,weight)
> abline(out)
> out$coefficients
(Intercept)        age2
51.3162176   0.3655757
```

(b) Find the predicted value of weight if age is (i) 2 weeks; (ii) 1 year. Do you think the predicted values are reliable? Discuss possible limitations of this quadratic model.

(c) Define residuals for this model and explain the plots in the following R commands and what their output (see Figure 2) demonstrates.

```
> qqnorm(out$residuals)
> qqline(out$residuals)
```

[4 + 4 + 4 = 12 marks]

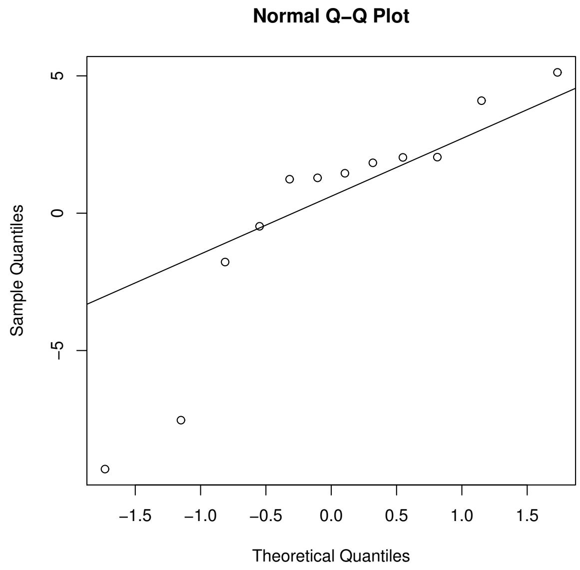

Figure 2: see Question 1(c)

2. Let $X_1, \ldots, X_n$ be a random sample from a distribution with the cumulative distribution function (cdf)

$$F(x;\theta) = \begin{cases} 0, & x < 0, \\ \left(\frac{x}{3}\right)^{\theta}, & 0 \le x < 3, \\ 1, & x \ge 3, \end{cases}$$

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

(a) Find the probability density function (pdf) for this distribution.

(b) Write down the log-likelihood function and simplify it.

(c) Find the maximum likelihood estimator of θ and check that your estimator gives the maximum of the likelihood function.

(d) Find the Fisher Information and give the Cramér-Rao lower bound for the variance of unbiased estimators of θ.

(e) Suppose $n = 20$ and the observations are:

```
2.698 1.758 2.706 2.203 1.162 1.973 2.442 1.993 2.201 2.899
2.855 0.917 2.871 2.130 1.570 0.962 2.475 2.588 2.293 2.161
```

with

$$\sum_{i=1}^{20} X_i = 42.857, \quad \sum_{i=1}^{20} \ln(X_i) = 14.251.$$

i. Find the maximum likelihood estimate of θ.

ii. Construct an approximate 95% confidence interval for θ.

Some R code output that may help:

```
qnorm(c(0.5,0.9,0.975,0.99))
[1] 0.000000 1.281552 1.959964 2.326348
log(c(2,3,4,5))
[1] 0.6931472 1.0986123 1.3862944 1.6094379
```

[2 + 4 + 4 + 3 + 3 = 16 marks]

3. Let $X_1, \ldots, X_n$ be a random sample from the probability density function

$$f(x;\theta) = \begin{cases} \frac{3}{8}(x^2 + \theta x + 1), & -1 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $-2 \le \theta \le 2$ is the unknown parameter.

(a) Find the method of moments estimator of θ.

(b) Is the estimator of θ you found in (a) unbiased or not? Explain why.

(c) Find the variance of this estimator as a function of θ.

[4 + 2 + 4 = 10 marks]

4. The number of telephone calls received by a call center is a Poisson process with the rate of λ calls per minute. We assume that the prior distribution of λ has the gamma density with shape and rate parameters α and τ, respectively:

$$f(\lambda|\alpha,\tau) = \frac{\tau^\alpha}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\tau\lambda}, \quad \lambda, \alpha, \tau > 0.$$

Assume that the call center received m calls in 10 minutes.

(a) Find the posterior distribution of λ.

(b) What is the Bayes estimator of λ using the quadratic loss?

(c) Assume now that the prior distribution of λ has the exponential density

$$f(\lambda) = 0.2e^{-0.2\lambda}, \quad \lambda > 0.$$

The call center received 65 calls in 10 minutes. Find the 95% posterior probability interval for the rate parameter λ. The following R code output may be helpful:

```
qgamma(c(0.01, 0.025, 0.975, 0.99), 65, 10.2)
[1] 4.678972 4.918199 8.012409 8.354075
qgamma(c(0.01, 0.025, 0.975, 0.99), 66, 10.2)
[1] 4.762914 5.004352 8.122337 8.466217
qexp(c(0.01, 0.025, 0.975, 0.99), 0.2)
[1] 0.05025168 0.12658904 23.02585093 18.44439727
```

[5 + 2 + 3 = 10 marks]

5. Let $x_1, \ldots, x_n$ be some positive constants and assume that $Y_1, \ldots, Y_n$ are independent and $Y_i \sim N(\beta x_i, \frac{\sigma^2}{x_i})$.

(a) Show that the maximum likelihood estimators for β and σ^2 are

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i^2 Y_i}{\sum_{i=1}^n x_i^3}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n x_i(Y_i - \hat{\beta}x_i)^2.$$

You are not required to demonstrate that the stationary points give the maximum of the log-likelihood function.

(b) Show that $\hat{\beta}$ is an unbiased estimator of β.

(c) Show that $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^n x_i^3}$.

(d) Show the following equality:

$$\sum_{i=1}^{n} x_i(Y_i - \beta x_i)^2 = \sum_{i=1}^{n} x_i(Y_i - \hat{\beta} x_i)^2 + (\beta - \hat{\beta})^2 \sum_{i=1}^{n} x_i^3.$$

(e) Hence show that $\mathrm{E}\left\{\sum_{i=1}^{n} x_i(Y_i - \hat{\beta} x_i)^2\right\} = (n-1)\sigma^2$. Is the maximum likelihood estimator $\hat{\sigma}^2$ unbiased? If not, deduce an unbiased estimator of σ^2.

[5 + 2 + 2 + 5 + 4 = 18 marks]

6. Let Z_1, Z_2, Z_3 be independent random variables with the cdf

$$F(z) = \begin{cases} 0, & z < 0, \\ \left(\frac{z}{\theta}\right)^2, & 0 \le z \le \theta, \\ 1, & z > \theta. \end{cases}$$

(a) Find the cdf of $Y_3 = \max\{Z_1, Z_2, Z_3\}$.

(b) Show that the cdf of $Y^* = Y_3/\theta$ does not depend on θ (i.e., Y^* is a pivot). Use the 2.5% and 97.5% quantiles of this distribution to construct the 95% confidence interval for θ as a function of Y_3.

[4 + 5 = 9 marks]

7. A random variable Z has the probability density function

$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

A single observation of Z is used to test $H_0 : \theta = 1$ against $H_1 : \theta = 2$. The null hypothesis is rejected if the observed value of Z is greater than 0.9.

(a) Find the probability of type 1 error for this test.

(b) Find the probability of type 2 error for this test.

(c) Find the power of this test.

(d) Assume now that $H_0 : \theta = 1$ is rejected in favor of $H_1 : \theta = 2$ if $Z > c$. Find c such that the probability of type 2 error is 0.05 for this test.

[4 + 2 + 2 + 3 = 11 marks]

8. Reaction times (in seconds) to two different stimuli, A and B, for five patients are recorded in the table below.

Stimulus	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
A	0.8	0.6	0.7	0.7	0.8
B	1.0	0.9	0.8	0.9	1.0

Let μ_A and μ_B be the mean reaction times to stimuli A and B, for all patients.

(a) Assuming normal distribution for the paired differences, construct the 95% confidence interval for $\delta = \mu_A - \mu_B$. The following R code output may be useful:

```
qnorm(0.975)
[1] 1.959964
qt(0.975,df=c(2,3,4,5))
[1] 4.302653 3.182446 2.776445 2.570582
```

(b) Do we reject $H_0 : \delta = 0$ in favor of $H_1 : \delta \neq 0$ at the 5% significance level?

[5 + 2 = 7 marks]

9. A dice has 4 sides. Let π_k be the probability that the dice shows k, $k = 1, 2, 3, 4$. We throw this dice $n = 100$ times and count how many times each side shows up, which gives $y_1 = 14, y_2 = 24, y_3 = 30, y_4 = 32$. We think that the dice is biased and let the null hypothesis $H_0 :\ \pi_1 : \pi_2 : \pi_3 : \pi_4 = 1 : 2 : 3 : 4$. Test this hypothesis at the 5% significance level. Find an approximate p-value of your test. The following R code output may be helpful:

```
qchisq(c(0.01, 0.025, 0.95, 0.975), 3)
[1] 0.1148318 0.2157953 7.8147279 9.3484036
qchisq(c(0.01, 0.025, 0.95, 0.975), 4)
[1] 0.2971095 0.4844186 9.4877290 11.1432868
1-pchisq(c(2,3,4,5), 3)
[1] 0.5724067 0.3916252 0.2614641 0.1717971
1-pchisq(c(2,3,4,5), 4)
[1] 0.7357589 0.5578254 0.4060058 0.2872975
```

[7 = 7 marks]

`Total marks = 100`

End of the Questions