Lancaster University

2022 EXAMINATIONS

PART II – Fourth Year / MSc – SPECIMEN

MATHEMATICS & STATISTICS

MATH401/MATH501: Statistical Fundamentals I **3 hours**

You should answer ALL questions. There is a total of 100 marks (Q1 38; Q2 38 ; Q3 56). You may only use the calculator that has been provided.

NOTE: this is a **specimen** paper written in January 2022 since the module is new for 2021-22. The total marks are deliberately > 100 as additional material has been included to give an idea of the range of possible questions that might be asked. Nonetheless, it is not exhaustive.

Questions start over the page

please turn over

1. Conditional on a parameter $\theta > 0$, a random variable, Y, has a density of

$$f(y|\theta) = \frac{y^5}{\theta^6} \exp\left\{-\frac{y^2}{\theta^2}\right\}$$

for $y > 0$; the density is 0 otherwise. A data set consists of n independent realisations of Y, denoted $y_1, \ldots, y_n$.

(a) Write down the log-likelihood for the data and hence find the maximum likelihood estimate for θ. [8]

(b) (i) Write down a formula for the observed information, $I_O(\theta; \mathbf{y})$. [4]

(ii) Use the fact that $\mathbb{E}[Y^a] = \frac{1}{2}\theta^a \Gamma(3 + a/2)$ to show that the expected information is $I_E(\theta) = 12n/\theta^2$. [4]

(c) You are told that $n = 100$ and:

$$\sum_{i=1}^{n} y_i = 341.2, \quad \sum_{i=1}^{n} y_i^2 = 1248.0, \quad \sum_{i=1}^{n} \exp(y_i) = 4533.7.$$

(i) Find a 95% Wald-based confidence interval for θ. [4]

(ii) Using your answer to c(i) or otherwise, find a 95% confidence interval for $\mathbb{E}[Y^2]$. [4]

(d) For $\theta_0 = 2.2$, conduct the following tests at the 1% level of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$; for each test, state your conclusion. For each test it is sufficient to compare against an appropriate critical value; it is not necessary to calculate a p-value.

(i) A Wald test using $I_E(\widehat{\theta})$. [4]

(ii) A score test using $I_E(\theta_0)$. [4]

(iii) A score test using $I_E(\widehat{\theta})$. [4]

(e) For a general single-parameter likelihood, state a logistical advantage of the test in (d)(ii) over the tests in (d)(i) and (d)(iii). [2]

please turn over

2. Conditional on two positive parameters, α and θ, a random variable Y has a density of

$$f(y|\alpha,\theta) = \frac{\theta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\theta y\}$$

for $y > 0$; the density is 0 otherwise. Conditional on the same two parameters, a random variable, Z has a density of

$$f(z|\alpha,\theta) = 2\theta^2 z \exp\left\{-\theta^2 z^2\right\}$$

for $z > 0$; the density is 0 otherwise. A data set consists of n independent realisations of Y, denoted $\mathbf{y} = (y_1, \ldots, y_n)$, and m independent realisations of z, denoted $\mathbf{z} = (z_1, \ldots, z_m)$. For part (c) you may use that $\log \Gamma(3.3) \approx 0.9871$; further, it is sufficient to compare against a suitable critical value - it is not necessary to calculate a p-value.

(a) (i) Show that, ignoring the additive constant, the log-likelihood can be written as

$$\ell(\alpha,\theta;\mathbf{y},\mathbf{z}) = -n\log\Gamma(\alpha) + (n\alpha + 2m)\log\theta + \alpha\sum_{i=1}^n \log y_i - \theta\sum_{i=1}^n y_i - \theta^2\sum_{i=1}^n z_i^2. \quad [2]$$

(ii) Hence find the maximum-likelihood estimate for θ conditional on a fixed value of α, $\widehat{\theta}_\alpha$.

[6]

(b) The maximum likelihood estimate is intractable so an `R` function, `Q2log.like()` is created to evaluate $\ell(\alpha,\theta)$ (with the parameters in that order) as given in Part(a)(i). The following call to `optim()` leads to the output below (with a few irrelevant lines removed for simplicity):

```
ctl=list(fnscale=-1)
a=optim(c(1,1),Q2log.like,ys=ys,zs=zs,hessian=TRUE,control=ctl); a
$par
[1] 2.8447060 0.4760389
$value
[1] -274.3389
$counts
function gradient
73       NA
$convergence
[1] 0
$hessian
        [,1]      [,2]
[1,] -63.05874  315.1008
[2,] 315.10076   BLANK
```

Further, $n = 150$, $m = 50$, $\sum_{i=1}^n \log y_i = 240.28$, $\sum_{i=1}^n y_i = 925.64$ and $\sum_{i=1}^n z_i^2 = 190.05$.

please turn over

(i) The [2,2] element of the Hessian has been deliberately left blank; derive it from the other information available to you. [4]

(ii) Find a 99% Wald-based confidence interval for α. [6]

(iii) Derive a 99% Wald-based confidence interval for $\alpha - 4\theta$. [10]

(c) For $\alpha_0 = 3.3$, conduct a likelihood-ratio test at the 2% level of $H_0 : \alpha = \alpha_0$ against $H_1 : \alpha \neq \alpha_0$. [10]

please turn over

3. (a) Consider a linear model for n response values, with a design matrix of X, a parameter vector of $\boldsymbol{\beta}$, and independent and identically distributed Gaussian residuals, $\epsilon_i \sim \mathsf{N}(0, \sigma^2)$ for $i = 1, \ldots, n$.

   (i) State the distribution of the vector of responses, $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$. [4]

   (ii) Hence derive, explaining each step, an expression for the variance of the maximum likelihood estimator of the coefficients: $\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \mathbf{Y}$. [8]

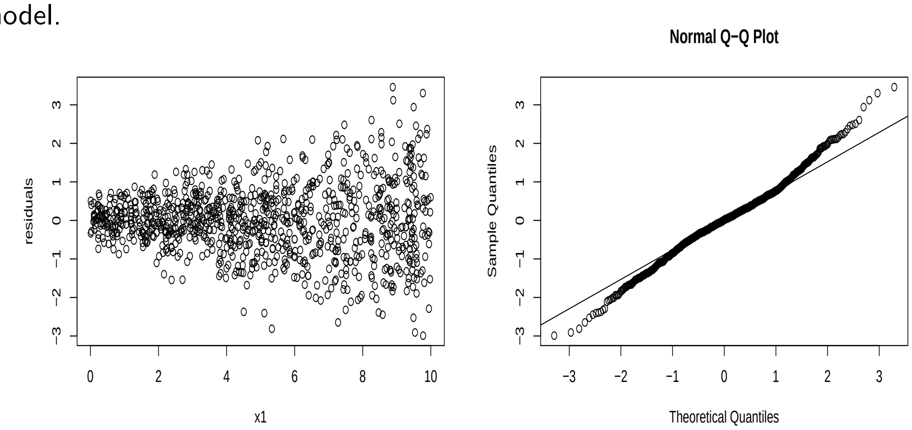
(b) A response y (measured in metres) is thought to depend on one continuous covariate, x_1, a binary covariate, x_2 (Y or N) and a categorical covariate with three categories (R, G and B), x_3 (**Aside**: in the real exam these would have names). A linear model is fitted to one thousand data records, leading to the following estimated coefficients for the best-fitting model.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.90217544	0.13444765	66.2129512	0.000000e+00
x1	0.71034832	0.00899967	78.9304875	0.000000e+00
x2Yes	0.96816329	0.13619988	7.1084004	2.240671e-12
x3G	-0.93183899	0.13528429	-6.8880060	1.002894e-11
x3R	-2.86780515	0.14054386	-20.4050551	1.452258e-77
x2Yes:x3G	-1.53329657	0.15482783	-9.9032362	4.100763e-22
x2Yes:x3R	-0.06939072	0.25015962	-0.2773858	7.815416e-01

   (i) Explain in words and numbers the effect of x_1 on y. [2]

   (ii) Explain in words and numbers the effects of x_2 and x_3 on y. [10]

   (iii) Boxplots of residuals against x_2 and x_3 show no pattern. The scatter plot of residuals against x_1 and the standard QQplot of the residuals are given below. Explain these two plots. How would you alter the model, introducing an additional parameter, γ, in light of this pattern? Write down and simplify the log-likelihood for the new model. [12]



please turn over

(c) A sample was taken from 2m below the surface in each of 75 places in the land owned by a particular farmer, and the fraction (by mass) of clay within the sample, y, was measured. The farmer also measured the GPS co-ordinates of each sample point, which have been converted to distances (in km) East (x_1) and North (x_2) from his farmhouse's chimney. He fits the following model:

$$\log\left(\frac{y_i}{1-y_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \epsilon_i,\ i = 1, \ldots, n$$

where the $\epsilon_i \sim \mathsf{N}(0, \sigma^2)$ are independent and identically distributed. The call to `lm()` and the resulting coefficient estimates are given below.

```
Mc=lm(log(y/(1-y))~x1+x2)
             Estimate   Std. Error  t value    Pr(>|t|)
(Intercept) -0.7776230 0.1546968  -5.0267546 3.516277e-06
x1           0.1140109 0.1287506   0.8855178 3.788254e-01
x2           0.1788088 0.1993881   0.8967876 3.728211e-01
```

(i) Explain why the above model is more appropriate than a linear model for $(y_1, \ldots, y_n)$. [4]

(ii) Predict the fraction of clay in the soil 2m under his (downstairs) fireplace and find a 95% CI for this fraction. [8]

(iii) Suggest an alternative model for the response that uses a *beta* distribution. (*Hint*: see the pages after the end of the exam for details of the beta distribution). [8]

end of exam

Formula sheet

Distribution quantiles

The 90%, 95%, 97.5%, 99% and 99.5% quantiles of the χ^2_d distribution for $d = 1$, 2 and 3 are:

d	90%	95%	97.5%	99%	99.5
1	2.706	3.841	5.024	6.635	7.879
2	4.605	5.991	7.378	9.210	10.597
3	6.251	7.815	9.348	11.345	12.838

The 90%, 95%, 97.5%, 99% and 99.5% quantiles of the standard normal distribution are:

90%	95%	97.5%	99%	99.5%
1.282	1.645	1.960	2.326	2.576

Matrix inverse

The inverse of a general 2×2 matrix is as follows:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

provided that $ad - bc \neq 0$.

Gamma function

The gamma function is defined as $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \exp(-x)\ \mathrm{d}x$. It satisfies $\Gamma(1) = 1$ and for $\alpha > 1$, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ so that $\Gamma(\alpha) = (\alpha-1)!$ for $\alpha = 1, 2, 3, \ldots$. Finally, $\Gamma(1/2) = \sqrt{\pi}$.

Standard distributions

The **normal distribution**, also known as the Gaussian distribution, with a parameters of μ and $\sigma > 0$ has a density of

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \quad (-\infty < y < \infty).$$

A random variable $Y \sim \mathsf{N}(\mu, \sigma^2)$ has: $\mathbb{E}[Y] = \mu$ and $\mathsf{Var}[Y] = \sigma^2$.

The **exponential distribution** with a rate parameter of $\beta > 0$ has a density of

$$f(y) = \beta \exp(-\beta y) \quad (y > 0).$$

A random variable $Y \sim \mathsf{Exp}(\beta)$ has: $\mathbb{E}[Y] = 1/\beta$ and $\mathsf{Var}[Y] = 1/\beta^2$.

The **gamma distribution** with a shape parameter of $\alpha > 0$ and a rate parameter of $\beta > 0$ has a density of

$$f(y) = \frac{1}{\Gamma(\alpha)} \beta^\alpha y^{\alpha-1} \exp(-\beta y) \quad (y > 0).$$

A random variable $Y \sim \mathsf{Gamma}(\alpha, \beta)$ has: $\mathbb{E}[Y] = \alpha/\beta$ and $\mathsf{Var}[Y] = \alpha/\beta^2$.

The **beta distribution** with shape parameters of $\alpha > 0$ and $\beta > 0$ has a density of

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \quad (0 < y < 1).$$

A random variable $Y \sim \mathsf{Beta}(\alpha, \beta)$ has: $\mathbb{E}[Y] = \alpha/(\alpha+\beta)$ and

$$\mathsf{Var}[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}.$$

The **non-standardised univariate Student-t distribution** with parameters of μ and $\lambda > 0$ and ν degrees of freedom has a density of

$$f(y) = \frac{\Gamma(\nu/2 + 1/2)}{\sqrt{\nu\pi\lambda^2}\,\Gamma(\nu/2)} \left\{1 + \frac{(y-\mu)^2}{\nu\lambda^2}\right\}^{-\frac{\nu+1}{2}} \quad (-\infty < y < \infty).$$

A random variable $Y \sim \mathsf{t}_\nu(\mu, \lambda)$ has: $\mathbb{E}[Y] = \mu$ provided $\nu > 1$.

The **Poisson distribution** with a mean parameters of $\lambda > 0$ has a mass function of

$$f(y) = \frac{\lambda^y}{y!} \exp(-\lambda) \quad (y = 0, 1, 2, \dots).$$

A random variable $Y \sim \mathsf{Pois}(\lambda)$ has: $\mathbb{E}[Y] = \lambda$ and $\mathsf{Var}[Y] = \lambda$.

The **binomial distribution** with a probability parameters of $p \in [0, 1]$ and a number of trials, n, has a mass function of

$$f(y) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y} \quad (y = 0, 1, 2, \dots, n).$$

A random variable $Y \sim \mathsf{Bin}(n, p)$ has: $\mathbb{E}[Y] = np$ and $\mathsf{Var}[Y] = np(1-p)$.

The **multivariate normal distribution**, also known as the multivariate Gaussian distribution, with expectation $\boldsymbol{\mu}$, a d-vector, and variance-covariance $\boldsymbol{\Sigma}$, a $d \times d$ positive-definite, symmetric matrix, has a density of

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{d/2} \mathsf{det}(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} \quad (\mathbf{y} \in \mathbb{R}^d).$$

Also, for some $d \times d$ positive-definite, symmetric matrix A and d-vector b, if $\mathbf{y}$ is a vector of dimension d with density

$$g(\mathbf{y}) \propto \exp(-\frac{1}{2}\mathbf{y}^\top A \mathbf{y} + \mathbf{y}^\top b),$$

then $\mathbf{Y} \sim \mathrm{N}_d(Ab, A^{-1})$, also sometimes written $\mathbf{Y} \sim \mathrm{MVN}(Ab, A^{-1})$.