Imperial College
London

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May – June 2013

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science.

Statistical Modelling I

Date: Friday, 24 May 2013. Time: 2.00pm. Time allowed: 2 hours.

This paper has FOUR questions.

Candidates should use TWO main answer books (A & B) for their solutions as follows:
book A - solutions to questions 1 & 2; book B - solutions to questions 3 & 4.

Supplementary books may only be used after the relevant main book(s) are full.

Statistical tables will not be provided.

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.
- Answer all the questions. Each question carries equal weight.
- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers.
- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.
- Calculators may not be used.

1. (i) Consider a statistical model in which the observations $X_1, \ldots, X_n$ are independent and identically distributed on $(0, \infty)$ with probability density function (pdf)

$$f(x) = \frac{\sqrt{\beta}}{x\sqrt{2\pi}} \exp\left(-\frac{\beta}{2}(\log(x) - \alpha)^2\right),$$

where $\log$ denotes the natural logarithm and $\alpha \in \mathbb{R}$ and $\beta > 0$ are parameters.

Write down the likelihood function.

Compute the MLE for α, assuming that β is a known constant

(ii) Suppose that in the above model, α was known. We want to use a Bayesian approach to estimate β. Suppose that a-priori, $\beta \sim Gamma(2, 2)$. What is the posterior distribution of β given $X_1, \ldots, X_n$?

(Recall: If $Z \sim Gamma(c, d)$ then Z has the pdf $f(z) = \frac{d^c}{\Gamma(c)} z^{c-1} e^{-dz}, \quad z > 0$).

(iii) Consider the statistical model $X \sim \text{Exp}(\lambda)$ for unknown $\lambda > 0$. In other words, we observe one realisation of X with $\mathrm{P}_\lambda(X \le t) = 1 - \exp(-\lambda t)$ for $t > 0$.

Construct an appropriate test for

$$H_0 : \lambda \le 5 \quad \text{against} \quad H_1 : \lambda > 5$$

at the level 5%. Clearly state the decision rule.

Define, compute and sketch the power function of this test.

Suppose we observe the realisation $x = 0.2$. Compute the p-value of the test. Using this, decide if the test rejects H_0 at the level 10%.

2. Consider a statistical model in which $X \sim \text{Binomial}(n, \theta)$ with unknown parameter $\theta \in (0, 1)$. The number $n \in \mathbb{N}$ of trials is known.

(i) Work out the Rao-Cramer lower bound in this situation.
State for which estimators it applies.

(ii) Prove that the Rao-Cramer lower bound is attainable in this situation.

(iii) Give an example of an estimator which has a variance smaller than the Rao-Cramer bound.

3. Consider the following linear model.

$$\boldsymbol{Y} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\beta} \in \mathbb{R}^2, \quad \mathrm{E}\, \boldsymbol{\epsilon} = 0, \quad \mathrm{cov}(\boldsymbol{\epsilon}) = \sigma^2 I \text{ for some } \sigma^2 > 0.$$

Suppose we observe $\boldsymbol{Y} = (3, 1, -1, -1)^T$.

(i) Compute the least squares estimator of $\boldsymbol{\beta}$ in this linear model as a function of $\boldsymbol{Y}$. Use the observed $\boldsymbol{Y}$ to calculate the estimate of $\boldsymbol{\beta}$ in numerical form.

(ii) Define the vector of residuals and compute it for the above observation.

(iii) State an unbiased estimator $\widehat{\sigma}^2$ of σ^2 for a linear model with the standard second-order assumptions.

(iv) Evaluate this estimator in the above situation.

(v) Prove that the estimator $\widehat{\sigma}^2$ you have given in part (iii) is unbiased.
You can use alternative forms of $\widehat{\sigma}^2$ without proof. Clearly indicate what results concerning projection matrices and linear algebra you are using.
Hint: You can either prove the result for an arbitrary linear model or show the result just for this specific model.

4. Consider a linear model satisfying the normal theory assumptions (NTA) given in the lectures where the design matrix $X \in \mathbb{R}^{n \times p}$ is of rank p. We will use the usual notation from the lecture for linear models. In particular, $\widehat{\boldsymbol{\beta}}$ denotes the least squares estimator of $\boldsymbol{\beta}$. Let $\boldsymbol{c} \in \mathbb{R}^p$ be a known constant.

(i) Derive the distribution of $\boldsymbol{c}^T \widehat{\boldsymbol{\beta}}$.

(ii) State the distribution of RSS/σ^2 without proof.

(iii) Derive the distribution of

$$T = \frac{\boldsymbol{c}^T \widehat{\boldsymbol{\beta}} - \boldsymbol{c}^T \boldsymbol{\beta}}{\sqrt{\boldsymbol{c}^T (X^T X)^{-1} \boldsymbol{c} \frac{\mathrm{RSS}}{n-p}}}.$$

If you need a lemma from the lectures to establish independence of certain components of T then state this lemma fully.

(iv) Suppose $p = 3$. Describe in detail how you would construct a test with level 5% for

$$H_0 : \beta_3 = 2 \quad \text{against} \quad H_1 : \beta_3 \neq 2.$$

What is the distribution of the test statistic you are using under H_1?

You may use the results from the lectures about projection matrices and multivariate normal distributions without proof.

1. (i) The likelihood is $\boxed{\text{seen/sim.seen} \Downarrow}$

$$L(\alpha, \beta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{\sqrt{\beta}}{x_i\sqrt{2\pi}} \exp\left(-\frac{\beta}{2}(\log(x_i) - \alpha)^2\right).$$

Thus the log-likelihood is

$$\begin{aligned} l(\alpha, \beta) = \log L(\alpha, \beta) &= \sum_{i=1}^{n} \log\left(\frac{\sqrt{\beta}}{x_i\sqrt{2\pi}} \exp\left(-\frac{\beta}{2}(\log(x_i) - \alpha)^2\right)\right) \\ &= -\sum_{i=1}^{n} \log(x_i) - \frac{n}{2}\log(2\pi/\beta) - \sum_{i=1}^{n} \frac{\beta}{2}(\log(x_i) - \alpha)^2 \end{aligned}$$

Differentiating with respect to α gives

$$\frac{\partial}{\partial\alpha} l(\alpha, \beta) = \beta \sum_{i=1}^{n} (\log(x_i) - \alpha) = \beta(-n\alpha + \sum_{i=1}^{n} \log(x_i))$$

Equating this to 0 and solving for α gives $\widehat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \log(x_i)$ as candidate for the MLE. $\boxed{5}$

This is indeed the MLE because

$$\left(\frac{\partial}{\partial\alpha}\right)^2 l(\alpha, \beta) = -n\beta < 0.$$

$\boxed{2}$

(ii)

$$\begin{aligned} f(\beta|x_1, \ldots, x_n) &\propto \left(\prod_{i=1}^{n} f(x_i|\beta)\right)\pi(\beta) \\ &\propto \beta^1 e^{-2\beta} \prod_{i=1}^{n} \sqrt{\beta} \exp\left(-\frac{\beta}{2}(\log(x_i) - \alpha)^2\right) \\ &= \beta^{1+n/2} \exp(-\beta(2 + \frac{1}{2}\sum_{i=1}^{n}(\log(x_i) - \alpha)^2)). \end{aligned}$$

Thus $\beta|x_1, \ldots, x_n \sim Gamma(2 + n/2, 2 + \frac{1}{2}\sum_{i=1}^{n}(\log(x_i) - \alpha)^2)$. $\boxed{3}$

(iii) *Students have seen constructions of tests and power functions for normal distributed observations, but not for this specific situation.*

Note: large values of λ lead to small realisations of X.

Thus it makes sense to reject H_0 for small values of X, more precisely, we reject H_0 if $X \leq c$ for $c \in \mathbb{R}$ such that $\mathrm{P}_\lambda(X \leq c) \leq 0.05$ for all $\lambda \leq 5$.

Note that $\forall \lambda \leq 5$,

$$\mathrm{P}_\lambda(X \leq c) = 1 - \exp(-\lambda c) \leq 1 - \exp(-5c).$$

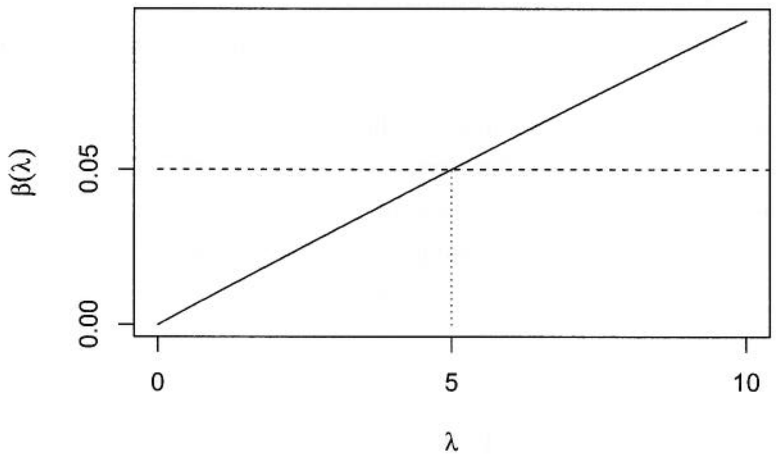
Thus, solving $1 - \exp(-5c) = 0.05$ will give such a c. This leads to

$$c = -\frac{\log(0.95)}{5}$$

To summarise, we reject H_0 if $X \leq -\frac{\log(0.95)}{c}$. $\boxed{4}$

The power function is $\beta : (0, \infty) \to \mathbb{R}, \beta(\lambda) = \mathrm{P}_\lambda(\text{reject } H_0)$. For this specific test,

$$\beta(\lambda) = \mathrm{P}_\lambda(X \leq c) = 1 - \exp\left(\frac{\lambda}{5}\log(0.95)\right).$$



The main points the plot should show is that $\beta(\lambda)$ is increasing, less or equal to 0.05 for $\lambda \leq 5$ and greater than 0.05 for $\lambda > 5$. $\boxed{3}$

Computation of the p-value:

$$p = \sup_{\lambda \leq 5} \mathrm{P}_\lambda(X \text{ "more extreme" than } 0.2) = \sup_{\lambda \leq 5} \mathrm{P}_\lambda(X \leq 0.2)$$
$$= \sup_{\lambda \leq 5}(1 - \exp(-0.2\lambda)) = 1 - \exp(-0.2 \cdot 5) = 1 - \exp(-1)$$

Using $e \geq 2$, this implies $p \geq 1/2$. As $p \geq 0.1$ this shows that the test does not reject H_0 at the level 10%. $\boxed{3}$

2. (i) Let $p_\theta(x) = \mathrm{P}_\theta(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$ for $x = 0, \dots, n$. seen ⇓

Will compute the Fisher Information as $I(\theta) = \mathrm{E}_\theta(-\left(\frac{\partial}{\partial\theta}\right)^2 \log p_\theta(X))$. To compute this, $\frac{\partial}{\partial\theta}\log p_\theta(x) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$ and thus

$$-\left(\frac{\partial}{\partial\theta}\right)^2 \log p_\theta(x) = \frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2}$$

As $\mathrm{E}(X) = n\theta$, this implies

$$I(\theta) = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}.$$

8

The Rao-Cramer theorem states that for all unbiased estimators T of θ

$$\mathrm{Var}_\theta(T) \geq \frac{1}{I(\theta)} = \frac{\theta(1-\theta)}{n}$$

2

(ii) Students have to give an estimator that is unbiased and whose variance attains the lower bound for all p. The maximum likelihood estimator in this situation is $T = X/n$. This is an unbiased estimator as $\mathrm{E}(T) = \frac{\mathrm{E}(X)}{n} = \frac{np}{n} = p$ for all p. Its variance is $\mathrm{var}(T) = \frac{\mathrm{var}(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$ which is equal to the Rao-Cramer bound. sim. seen ⇓

6

(iii) To find an example of an estimator which has smaller variance than the above Rao-Cramer bound, biased estimators have to be considered. unseen ⇓

The easiest example would be constant estimators (i.e. estimators that always return the same value without taking the observation into account). An example would be the estimator $S = 0.5$ with variance

$$\mathrm{Var}(S) = 0 < \frac{p(1-p)}{n} \quad \forall p.$$

A more sensible estimator would be the estimator $T = \frac{X+1}{n+2}$ which has variance $\mathrm{var}(T) = \frac{p(1-p)n}{(n+2)^2} < \frac{p(1-p)}{n}$.

4

3. (i) Let X denote the design matrix. Then

sim. seen ⇓

$$X^TX = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix},\ (X^TX)^{-1} = \tfrac{1}{8-4}\begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1 \end{pmatrix},$$

$$X^T\boldsymbol{Y} = \begin{pmatrix} Y_1 + Y_2 + Y_3 + Y_4 \\ Y_1 + Y_2 \end{pmatrix}$$ and thus

$$\widehat{\boldsymbol{\beta}} = (X^TX)^{-1}X^T\boldsymbol{Y} = \begin{pmatrix} \frac{1}{2}(Y_3 + Y_4) \\ \frac{1}{2}(Y_1 + Y_2) - \frac{1}{2}(Y_3 + Y_4) \end{pmatrix}$$

4

Plugging in the observation $\boldsymbol{Y} = (3, 1, -1, -1)^T$ gives $\widehat{\boldsymbol{\beta}} = (-1, 3)^T$.

2

(ii) The vector of residuals is $\boldsymbol{e} = \boldsymbol{Y} - X\widehat{\boldsymbol{\beta}}$, which in this case gives

$$\boldsymbol{e} = \begin{pmatrix} 3 \\ 1 \\ -1 \\ -1 \end{pmatrix} - \begin{pmatrix} 2 \\ 2 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}.$$

3

(iii) An unbiased estimator of σ^2 is given by

$$\widehat{\sigma}^2 = \mathrm{RSS}/(n - r)$$

where $\mathrm{RSS} = \boldsymbol{e}^T\boldsymbol{e}$, n is the number of rows of $\boldsymbol{X}$ and $r = \mathrm{rank}(X)$.

3

(iv) Here: $n = 4$ and $r = 2$. Thus

$$\widehat{\sigma}^2 = 2/2 = 1.$$

2

(v) *(Students have seen the general proof but not the proof for a specific situation such as this one.)*

seen ⇓

Let P be the projection matrix onto the space spanned by the columns of $\boldsymbol{X}$. Let $Q = I - P$. We know from the lectures that $\mathrm{RSS} = Y^TQY$. Hence,

$$\begin{aligned}
\mathrm{E}(\mathrm{RSS}) &= \mathrm{E}\,\mathrm{trace}\,\mathrm{RSS} \quad [\mathrm{RSS} \in \mathbb{R}] \\
&= \mathrm{E}\,\mathrm{trace}(\boldsymbol{Y}^TQ\boldsymbol{Y}) = \mathrm{E}\,\mathrm{trace}(Q\boldsymbol{Y}\boldsymbol{Y}^T) \quad [\mathrm{trace}(AB) = \mathrm{trace}(BA)] \\
&= \mathrm{trace}(Q\,\mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}^T)) \\
&= \mathrm{trace}(Q[\mathrm{cov}\,\boldsymbol{Y} + \mathrm{E}(\boldsymbol{Y})\,\mathrm{E}(\boldsymbol{Y})^T]) \quad [\text{Def cov}] \\
&= \mathrm{trace}(Q\sigma^2) + \mathrm{trace}(QX\boldsymbol{\beta}(X\boldsymbol{\beta})^T) \quad [\text{model assumptions}] \\
&= \sigma^2\,\mathrm{trace}(I - P) + 0 = \sigma^2(n - \mathrm{trace}(P)) \\
&= \sigma^2(n - \mathrm{rank}(P)) \quad [\text{trace of a projection matrix is equal to its rank}] \\
&= \sigma^2(n - r).
\end{aligned}$$

Thus $\mathrm{E}(\widehat{\sigma}^2) = \sigma^2$.

Alternative solution via direct computations for this example: $\boxed{\text{unseen} \Downarrow}$

In this case $n = 4$ and $r = 2$ and thus the estimator is $\widehat{\sigma}^2 = \frac{1}{2}\,\mathrm{RSS}$.

The general form for the residual is

$$\boldsymbol{e} = \boldsymbol{Y} - X\widehat{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{Y}_1 - \widehat{\beta}_1 - \widehat{\beta}_2 \\ \boldsymbol{Y}_2 - \widehat{\beta}_1 - \widehat{\beta}_2 \\ \boldsymbol{Y}_3 - \widehat{\beta}_1 \\ \boldsymbol{Y}_4 - \widehat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_1 - \frac{1}{2}(Y_1 + Y_2) \\ \boldsymbol{Y}_2 - \frac{1}{2}(Y_1 + Y_2) \\ \boldsymbol{Y}_3 - \frac{1}{2}(Y_3 + Y_4) \\ \boldsymbol{Y}_4 - \frac{1}{2}(Y_3 + Y_4) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \boldsymbol{Y}_1 - Y_2 \\ \boldsymbol{Y}_2 - Y_1 \\ \boldsymbol{Y}_3 - Y_4 \\ \boldsymbol{Y}_4 - Y_3 \end{pmatrix}$$

Thus

$$\mathrm{RSS} = \boldsymbol{e}^T\boldsymbol{e} = \frac{1}{4}(2(Y_1 - Y_2)^2 + 2(Y_3 - Y_4)^2) = \frac{1}{2}((Y_1 - Y_2)^2 + (Y_3 - Y_4)^2)$$

For this particular model, $\mathrm{E}\,Y_1 = \mathrm{E}\,Y_2$ and $\mathrm{E}\,Y_3 = \mathrm{E}\,Y_4$, thus

$$\mathrm{RSS} = \frac{1}{2}((\epsilon_1 - \epsilon_2)^2 + (\epsilon_3 - \epsilon_4)^2)$$

Hence, using that the ϵ_i are uncorrelated, have mean 0 and variance σ^2,

$$\mathrm{E}\,\widehat{\sigma}^2 = \mathrm{E}\,\frac{1}{2}\,\mathrm{RSS} = \frac{1}{4}\left(\mathrm{E}(\epsilon_1 - \epsilon_2)^2 + \mathrm{E}(\epsilon_3 - \epsilon_4)^2\right) = \frac{4}{4}\,\mathrm{E}\,\epsilon_1^2 = \sigma^2.$$

Thus $\widehat{\sigma}^2$ is unbiased. $\boxed{6}$

4. (i) Since $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}} = \boldsymbol{c}^T(X^TX)^{-1}X^T\boldsymbol{Y}$ and $\boldsymbol{Y} \sim N(X\boldsymbol{\beta}, \sigma^2 I)$, $\boxed{\text{seen} \Downarrow}$

$$\mathrm{E}\,\boldsymbol{c}^T\widehat{\boldsymbol{\beta}} = \boldsymbol{c}^T\boldsymbol{\beta}$$

$$\begin{aligned}\mathrm{Var}(\boldsymbol{c}^T\widehat{\boldsymbol{\beta}}) &= \mathrm{Var}(\boldsymbol{c}^T(X^TX)^{-1}X^T\boldsymbol{Y}) = \boldsymbol{c}^T(X^TX)^{-1}X^T\,\mathrm{cov}(\boldsymbol{Y})X(X^TX)^{-1}\boldsymbol{c}\\ &= \boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}\sigma^2\end{aligned}$$

and thus $\boldsymbol{c}^T\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{c}^T\boldsymbol{\beta}, \boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}\sigma^2)$. $\boxed{5}$

(ii) $\mathrm{RSS}/\sigma^2 \sim \chi^2_{n-p}$. $\boxed{2}$

(iii) Let $U = \frac{\boldsymbol{c}^T\widehat{\boldsymbol{\beta}} - \boldsymbol{c}^T\boldsymbol{\beta}}{\sqrt{\boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}\sigma^2}}$.
By the previous part, $U \sim N(0,1)$.
Furthermore, $T = U/\sqrt{\mathrm{RSS}/\sigma^2/(n-p)}$
$\widehat{\boldsymbol{\beta}} = (X^TX)^{-1}X^T\boldsymbol{Y}$ and $\mathrm{RSS} = \boldsymbol{Y}^TQ\boldsymbol{Y}$ are independent by using the following Lemma from the lectures:
Let $\boldsymbol{Z} \sim N(\boldsymbol{\mu}, I)$, $A \in \mathbb{R}^{n,n}$ pos. semidefinite symmetric and let B be a matrix such that $BA = 0$. Then $\boldsymbol{Z}^TA\boldsymbol{Z}$ and $B\boldsymbol{Z}$ are independent.
Using $\boldsymbol{Z} = \boldsymbol{Y}/\sigma$, $Q = A$ and $\boldsymbol{c}^T(X^TX)^{-1}X^T = B$, we see that $BA = \boldsymbol{c}^TX^TX)^{-1}X^TQ = \boldsymbol{c}^T(X^TX)^{-1}\underbrace{(QX)}_{=0}{}^T = 0$. Hence, $\frac{\mathrm{RSS}}{\sigma^2}$ and $c^T\widehat{\beta}X^T = B\boldsymbol{Y}$ are independent. This implies that U and RSS/σ^2 are independent. Thus, by the definition of the t-distribution,

$$T \sim t_{n-p}.$$

$\boxed{6}$

(iv) *(Students have only seen tests for components of β being equal to 0)* $\boxed{\text{sim. seen} \Downarrow}$
Let $\boldsymbol{c} = (0,0,1)^T$. From part (iii) we know that under H_0,

$$S = \frac{\widehat{\beta}_3 - 2}{\sqrt{\boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}\frac{\mathrm{RSS}}{n-p}}} \sim t_{n-3}.$$

Let q_1 and q_2 be such that $\mathrm{P}(q_1 \le Z \le q_2) = 0.95$ for $Z \sim t_{n-p}$.
We reject H_0 if the observed value s of S is outside the interval (q_1, q_2). $\boxed{4}$
(*The fact that this is noncentral t under alternatives has been mentioned in class; students have not seen the value of the noncentrality parameter.*)
Under H_1, $\boldsymbol{c}^T\beta = \beta_3 \neq 2$. The arguments in part (iii) with S replacing T still work with the exception of the distribution of U. Now $U \sim N(\delta, 1)$ with

$$\delta = \frac{\beta_3 - 2}{\sqrt{\boldsymbol{c}^T(X^TX)^{-1}\boldsymbol{c}\sigma^2}}.$$

Hence, $S \sim t_{n-3}(\delta)$. $\boxed{3}$