NATURAL SCIENCES TRIPOS **Part IA**

Friday 7th June 2024 09:00 to 12:00

MATHEMATICAL BIOLOGY

MB/1

You have ***three hours*** *to answer this paper (plus any pre-agreed individual adjustments).*

This paper consists of 10 questions grouped in 5 sections (A, B, C, D, E). You must answer eight questions, including at least one question from each of the sections A to E. Indicate your question selection on the form to the right. It is imperative that your selection matches the set of 8 answers you are submitting: any extra answers will not be marked.

All 10 questions carry equal weight. Indicative marks (out of 20) are given for each question part. For full credit, calculation steps must be shown where necessary to explain your answers, ensuring that the final answer is clearly visible.

This examination is ***closed book****, so access to the internet and use of mobile devices is prohibited. You may use an* ***approved calculator*** *and the* ***Mathematical Biology Formula Handbook*** *attached to this paper on Inspera.*

All answers should be ***hand-written on Inspera Scan paper, as outlined below****. Make sure that your handwriting and any diagrams or equations are clear and legible in your submitted files. Inspera Scan graph paper is available for drawing diagrams or figures.*

You must start each question on a new piece of Scan paper*. If your answer to a question spans more than one page, please write the question number at the top of each page. Once you finish a question,* ***you should number all the pages used for that question*** *(e.g. if you used 3 pages for question A1, they should be numbered "A1 – 1/3", "A1 – 2/3" and "A1 – 3/3").*

You must also label each page of Scan paper with your Blind Grade number and the Inspera Scan Code provided for each question (shown at the bottom of each page on Inspera).

After you have completed the assessment:

- Check that each Inspera Scan page is labelled appropriately, including your Blind Grade Number.
- Inspera Scan pages **MUST** be placed in the Inspera Scan folder at the completion of the exam, and left on your desk.
- Inspera Scan pages will be scanned by administrators after the completion of the exam.

Section A – Question A1

Parts (a), (b) and (c) are independent of each other but must all be answered.

(a) A mail order company sells cucumber seeds in packets of 10 seeds for £2.49. Each packet costs the company £1.75 to produce and deliver to a customer. Each seed has an 87% chance of germination, independently of all others. The company promises to replace, at no cost to the buyer, any packet that has 3 or more seeds that do not germinate.

(i) What is the probability that, in a randomly selected packet of cucumber seeds, 3 or more seeds will not germinate? For full credit you must define the random variable and state the distribution used.

[4 marks]

(ii) Suppose 8% of all customers request a replacement packet of seeds. What is the expected net income per packet to the company?

[3 marks]

(b) An online pest alert system has been set up to enable anyone to report a potential outbreak of oak processionary moth (OPM). Reports are labelled on the system as submitted by: P (members of the public); V (trained volunteers), and F (forestry workers). Analysis of historic data indicates that the accuracy of reports varies among the three groups as follows:

- P: 15% of reports are genuine OPM, 85% of reports are something else.
- V: 65% of reports are genuine OPM, 35% of reports are something else.
- F: 95% of reports are genuine OPM, 5% of reports are something else.

(i) On a typical day, there are 200 distinct reports of potential outbreaks of OPM, of which 68% come from group P, 17% come from group V, and the remainder from group F. What is the probability that a report chosen at random is **not** a genuine OPM?

[4 marks]

(ii) A field visit following up a randomly selected online submission finds that the report was correct and that OPM was indeed present. What is the probability that the report was made by a member of the public (P)?

[4 marks]

The question continues on the next page

(c) A continuous random variable *X* has the following probability density function:

$$f(x) = \begin{cases} \dfrac{8x - x^3}{12} & \text{if } 0 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the mode of the distribution of *X*.

[5 marks]

Section A – Question A2

A researcher is studying a population of stem cells, in which each cell can switch between two states A and B. By measuring the expression of a set of marker genes every hour, the researcher estimates that on average, from one hour to the next, 40% of cells in state A switch to state B, whereas 20% of cells in state B switch to state A.

(a) Let $\boldsymbol{d}_t$ be the vector of cell densities (cells per ml) in states A and B at time t, where t is in units of hours. Assuming that there are no other states involved, write down a matrix **M** representing the state dynamics of the population of cells, such that $\boldsymbol{d}_{t+1} = \mathbf{M}\,\boldsymbol{d}_t$.

[5 marks]

(b) Calculate the eigenvalues and eigenvectors of **M**, and use them to construct the diagonalisation of **M**.

[10 marks]

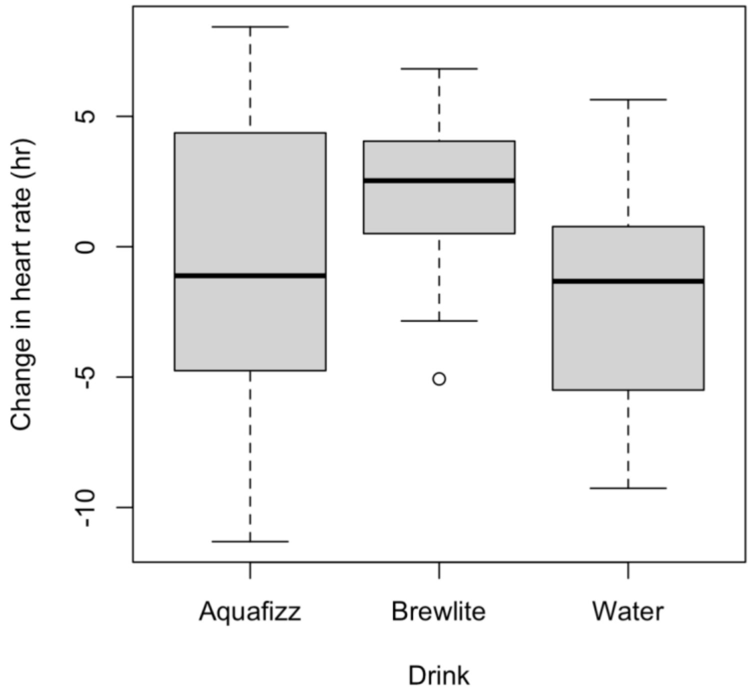
(c) At a particular time, the researcher estimates 3×10^6 cells per ml in state A and 0.7×10^6 cells per ml in state B. Assuming that the switching probabilities remain constant, what will the cell densities be 8 hours later?

[5 marks]

Section B – Question B3

A study into the effect of fizzy drinks on heart rate was carried out on a sample of volunteers. Two fizzy drinks, *Aquafizz* and *Brewlite*, were compared against water as a control. The participants were randomly split into three groups, each receiving the same volume of one of the three drinks. Their heart rate, measured in beats per minute, was recorded 2 minutes before and 2 minutes after consumption. In the figures and analyses below, the variable **hr** represents the net change in heart rate between the two successive measurements.

The figure below shows the distribution of the net change in heart rates in each group:



An analysis of variance was then conducted using R, and the output is shown below with two of the values replaced with letters **x** and **y**:

```
Call:
lm(formula = hr ~ drink, data = hr)

Residuals:
     Min       1Q   Median       3Q      Max
-10.7830  -3.7038   0.2436   2.6322   8.9579

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.5267     0.9741  -0.541    0.591
drink_brewlite   2.4556     1.3777   1.782    0.080 .
drink_water     -1.1450     1.3777  -0.831    0.409
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.357 on 57 degrees of freedom
Multiple R-squared:  0.1112,  Adjusted R-squared:  0.08003
F-statistic: 3.566 on 2 and 57 DF,  p-value: 0.03473
```

```
Analysis of Variance Table

Response: hr
          Df  Sum Sq Mean Sq F value  Pr(>F)
drink      2  135.37       x  3.5662       y    *
Residuals 57 1081.82  18.979
---
```

(a) How many participants were there in the sample?

[1 mark]

(b) Write down the null hypothesis being tested in the Analysis of Variance.

[1 mark]

(c) According to the model, what is the predicted value of **hr** for Aquafizz? Give your answer to 1 decimal place.

[1 mark]

(d) Calculate the missing values **x** and **y** in the ANOVA table.

[2 marks]

(e) Write a sentence suitable for a scientific report to summarise the findings, quoting appropriate summary statistics.

[3 marks]

(f) Explain why using the Pr(>|t|) column in the R output would give a different conclusion to the ANOVA output. Which output should you use and why?

[2 marks]

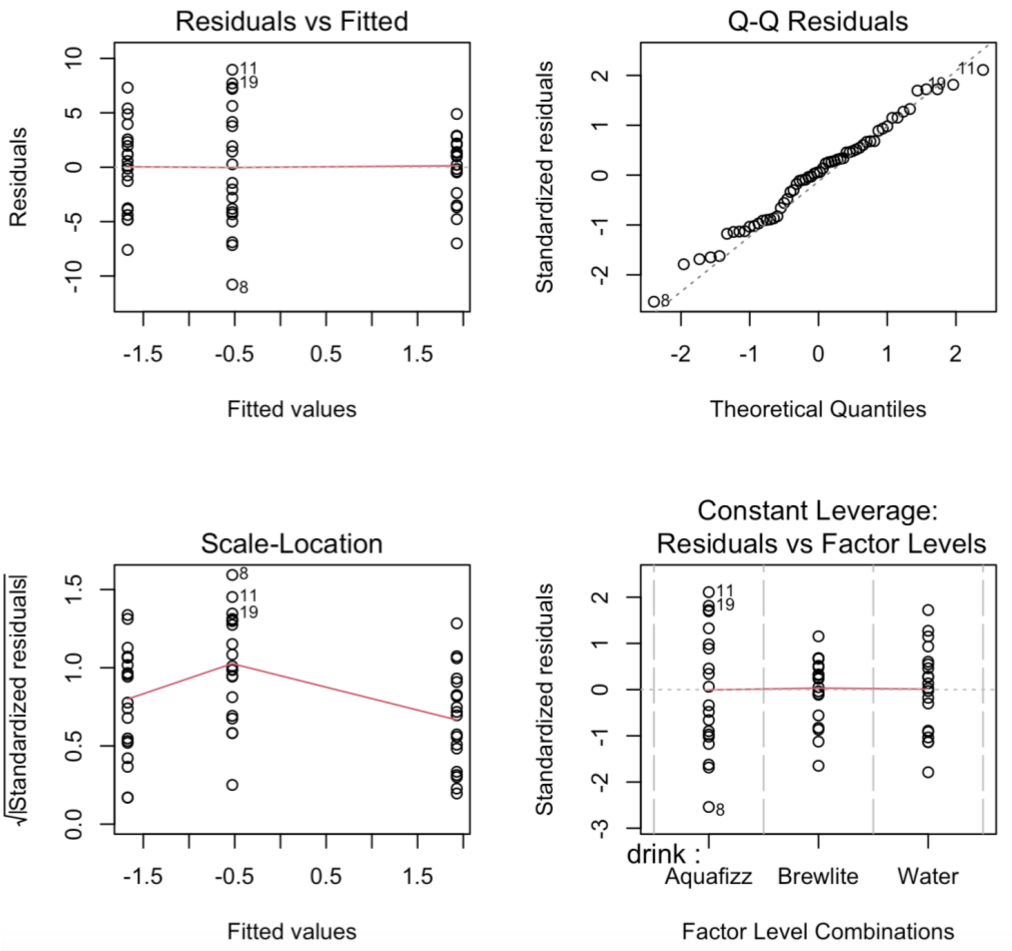
(g) State the 4 assumptions of this type of ANOVA.

[4 marks]

(h) Use the figure below to check the relevant assumptions of ANOVA. Is the ANOVA appropriate?

[3 marks]

The question continues on the next page



A reviewer suggests that age is an important factor in blood pressure, so this should be considered too. A model containing age, drink and their interaction is then considered in a backwards stepwise process.

(i) Describe what real-world feature the interaction term is designed to measure.

[1 mark]

The first step of the backwards stepwise elimination in R is shown below:

```
Model 1: hr ~ drink + age
Model 2: hr ~ drink + age + drink:age

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     56 969.13
2     54 890.07  2     79.06 2.3983 0.1005
```

(j) Should the process conclude at this step or should it continue? Justify your answer. If it should conclude, write the appropriate conclusion. If it should continue, state which models should be compared in the next step.

[2 marks]

END OF QUESTION B3

Section B – Question B4

This question describes a study of the reproductive habits of crabs. A marine biologist collected a sample of 100 male crabs, measured their length in cm, and recorded whether they had consumed food in the last week. The scientist then observed whether they were successful or rejected in a mating attempt.

The data was analysed using a generalised linear model in R, and the output is shown below:

```
Call:
glm(formula = mate ~ eaten + length, family = binomial(), data = mate)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.66569    1.14293  -3.207  0.00134 **
eaten_no-food -0.51659    0.54937  -0.940  0.34704
length         0.15680    0.06389   2.454  0.01412 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Null deviance: 97.245  on p  degrees of freedom
Residual deviance: 88.362  on q  degrees of freedom
AIC: 94.362
```

(a) What is the dependent variable in this model?

[1 mark]

(b) Explain why a logistic generalised linear model is more appropriate in this situation than the usual linear model.

[1 mark]

(c) What is the probability that a crab of length 20cm who has eaten no food will be successful in a mating attempt?

[2 marks]

Reminder: the logistic function is $f(x) = \frac{1}{1+e^{-x}}$

(d) Explain what is meant by the null model.

[1 mark]

(e) Calculate the value of ***p*** and ***q*** in the output above.

[2 marks]

(f) Is there significant evidence that the fitted model is different from the null model? Use a 5% significance level.

[5 marks]

The question continues on the next page.

For a random event with probability of success *P* and probability of failure 1–*P*, the odds of success are defined as *P*/(1–*P*).

(g) Show that the ratio

$$\frac{\text{Odds of mating having eaten food}}{\text{Odds of mating having eaten no food}}$$

is independent of the length of the crab.

[5 marks]

Next, each variable was dropped in turn from the model. The ouput is:

```
Single term deletions

Model:
mate ~ eaten + length
       Df Deviance
<none>      88.362
eaten   1   89.257
length  1   95.604
```

(h) What terms should be dropped from the model? Use a 5% significance test to justify your answer.

[3 marks]

Section C – Question C5

The density of a plant population Y at time t is modelled using the differential equation:

$$\frac{dY}{dt} = bY\left(1 - \frac{Y}{k}\right)$$

in which b and k are positive constants.

(a) What is the usual name of this model?

[1 mark]

(b) Suggest a biological interpretation of the parameters b and k.

[2 marks]

(c) The model includes density dependence. What, precisely, does this mean?

[1 mark]

(d) Find the two equilibria of the model. By sketching a graph of dY/dt as a function of Y, or otherwise, investigate the local stability of the equilibria.

[4 marks]

(e) Treating the model as a Bernoulli equation, use the substitution $Z = 1/Y$ to show that the differential equation can be transformed to:

$$\frac{dZ}{dt} = a - cZ$$

where you should find a and c in terms of b and k.

[4 marks]

(f) Use your answer to part (e) to find an expression for Y as a function of t, given the density of the plant population at $t = 0$ is Y_0 and assuming $0 < Y_0 < k$.

[6 marks]

(g) Sketch Y as a function of t, using two different initial conditions Y_0.

[2 marks]

Section C – Question C6

The density of organic matter in soil X at time t is modelled using the differential equation

$$\frac{dX}{dt} = -c\,X$$

where c is a positive constant. The density of organic matter at $t = 0$ is X_0.

(a) Find X as a function of t and sketch a labelled graph of the function, indicating how its behaviour depends on the parameters c and X_0.

[4 marks]

The per capita reproduction rate of a soil-borne fungus at time t is modelled using the equation

$$r(t) = a\,X(t) - b$$

where a and b are positive constants and $X(t)$ is the density of organic matter at time t.

For the remainder of this question, we assume that parameter values are such that $X_0 > b/a$.

(b) Sketch a labelled graph of r as a function of t, showing clearly how your sketch depends on the parameters a and b.

[2 marks]

(c) Find the time at which $r(t) = 0$, and show that this time is always positive.

[3 marks]

The density of soil-borne fungus Y at time t is modelled using the differential equation

$$\frac{dY}{dt} = r(t)\,Y$$

The initial density of fungus is $Y_0 > 0$.

(d) Using your solution to part (a) and the definition of $r(t)$ given above, solve the differential equation to find Y as a function of t.

[6 marks]

(e) Sketch a graph of Y as a function of t.
Hint: Your answer to part (c) provides useful information to help with this sketch.

[5 marks]

Section D – Question D7

Consider the nonlinear system of equations:

$$\frac{dx}{dt} = x\,y - 2\,y$$

$$\frac{dy}{dt} = (y+3)(y-x-2)$$

(a) Find and classify all equilibria.

[10 marks]

(b) Sketch the phase plane, including the nullclines, equilibrium points, direction field (with justification), and sufficient trajectories to illustrate the behaviour of the system.

[6 marks]

(c) For $(x_0, y_0) = (1, 2)$ sketch the trajectory associated with these initial conditions in the phase plane using a different colour. In a separate figure, sketch the graphs of $x(t)$ and $y(t)$ against time on the same set of axes.

[4 marks]

Section D – Question D8

In this question, we consider a compartmental model with two age groups to represent the population dynamics of a species of bats. In the system of differential equations below, $A(t)$ and $J(t)$ represent the densities of adults and juveniles, respectively, and $f(A)$ is a positive function that satisfies $f(0) = 1$ and $f'(A) < 0$ for all $A \geq 0$. We assume that $\alpha > 0, m > 0$ and $b > 0$.

$$\frac{dA}{dt} = \alpha\, J - m\, A$$

$$\frac{dJ}{dt} = b\, A\, f(A) - m\, J - \alpha\, J$$

(a) Propose a biological interpretation for each of the parameters α, m, b and for the function $f(A)$.

[4 marks]

We now assume that $f(A) = \frac{c}{c+A}$ with parameter $c > 0$.

(b) Sketch $f(A)$ and propose a biological interpretation for parameter c.

[3 marks]

(c) Calculate the nullclines and stationary points of the model and show that a positive equilibrium (A^*, J^*) only exists if

$$\frac{b}{m}\frac{\alpha}{\alpha + m} > 1$$

Hint: it is easier to consider the phase plane with *A* on the x-axis and *J* on the y-axis and derive the equations of the nullclines accordingly.

[5 marks]

(d) Numerical application: $b = 2.5, \alpha = 4, m = 1$. Let $K = A^* + J^*$ be the carrying capacity of this population. Calculate K as a function of c, and deduce the value of c to get a carrying capacity $K = 50$.

[2 marks]

(e) Sketch the phase plane using the above numerical values, with A on the x-axis and J on the y-axis. You should clearly indicate the nullclines, stationary points, and (with a brief justification) the direction field.

[6 marks]

Section E – Question E9

It is possible for two DNA-binding proteins A and C to bind to adjacent stretches of a DNA molecule B, while also forming protein-protein interactions with each other. The two proteins only interact weakly in the absence of DNA. In the presence of DNA, the formation of a three-component complex ABC can be modelled by the following two sets of reactions:

$$\mathrm{A} + \mathrm{B} + \mathrm{C} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \mathrm{AB} + \mathrm{C} \underset{k_{-2}}{\overset{k_{+2}}{\rightleftharpoons}} \mathrm{ABC}$$

$$\mathrm{A} + \mathrm{B} + \mathrm{C} \underset{k_{-3}}{\overset{k_{+3}}{\rightleftharpoons}} \mathrm{A} + \mathrm{BC} \underset{k_{-4}}{\overset{k_{+4}}{\rightleftharpoons}} \mathrm{ABC}$$

(a) Assuming that this reaction system has reached an equilibrium, determine two expressions, one in terms of species concentrations and the other in terms of rate constants, for each of the following dissociation constants:

i. $K_D(\mathrm{AB} \rightarrow \mathrm{A} + \mathrm{B})$
ii. $K_D(\mathrm{BC} \rightarrow \mathrm{B} + \mathrm{C})$
iii. $K_D(\mathrm{ABC} \rightarrow \mathrm{A} + \mathrm{BC})$
iv. $K_D(\mathrm{ABC} \rightarrow \mathrm{AB} + \mathrm{C})$

[8 marks]

The effect that protein C has on the interaction between A and B can be described by a coefficient α, defined by:

$$\alpha = \frac{K_D(\mathrm{ABC} \rightarrow \mathrm{A} + \mathrm{BC})}{K_D(\mathrm{AB} \rightarrow \mathrm{A} + \mathrm{B})}$$

Similarly, the effect that protein A has on the interaction between B and C can be described by a coefficient β, defined by:

$$\beta = \frac{K_D(\mathrm{ABC} \rightarrow \mathrm{AB} + \mathrm{C})}{K_D(\mathrm{BC} \rightarrow \mathrm{B} + \mathrm{C})}$$

(b) Explain what each set of assumptions below means for the binding affinity of each protein, considering whether or not the other protein is already bound to the DNA:

i. $0 < \alpha < 1$ and $0 < \beta < 1$.
ii. $\alpha = 1$ and $\beta = 1$.
iii. $\alpha > 1$ and $\beta > 1$.

[6 marks]

The question continues on the next page.

(c) Use your results from part (a) to derive a relationship between α and β.

[2 marks]

(d) Assuming that the rate constants (k_+) for all four binding steps are fixed at a value of 10^6 L mol^{-1} s^{-1}, that rate constant $k_{-1} = 10^2$ s^{-1} and that $\alpha = 0.02$, calculate the values of $K_D(AB \rightarrow A + B)$ and k_{-4}.

[4 marks]

Section E – Question E10

A parent bird returns to its nest with an indivisible food item, of value $v > 0$, and must decide to which of its two chicks it will give the item. Each chick chooses a (non-negative) level of expenditure to invest in begging for the food. We suppose that the probability of a chick receiving the food item is then equal to its own begging expenditure relative to the total expenditure of both chicks. Given these assumptions, the expected net payoff to a chick that expends x_{self} on begging while its nestmate expends x_{mate} is given by:

$$W(x_{\text{self}}, x_{\text{mate}}) = \left(\frac{x_{\text{self}}}{x_{\text{self}} + x_{\text{mate}}}\right) v - x_{\text{self}}$$

(a) Write down an expression for $\frac{\partial W(x_{\text{self}},\ x_{\text{mate}})}{\partial x_{\text{self}}}$, the partial derivative of a chick's payoff with respect to its own begging expenditure x_{self}.

[2 marks]

(b) Using your answer from part (a), derive a formula for a chick's optimal (or 'best response') begging expenditure, $x_{\text{opt}}(x_{\text{mate}})$, which maximises its payoff given that its rival expends x_{mate}.

[4 marks]

(c) Using your answer to part (b), find a stable level of begging expenditure x^* (> 0) that satisfies the condition $x_{\text{opt}}(x^*) = x^*$, i.e. a level of begging that is optimal for a chick when its nestmate expends the same level of effort.

[2 marks]

Next, we suppose that the value of the food to chick 1, denoted v_1, may differ from its value to chick 2, denoted v_2. The respective expected payoffs to chick 1 and chick 2, assuming that the former expends x_1 on begging and the latter x_2, are given by:

$$W_1(x_1, x_2) = \left(\frac{x_1}{x_1 + x_2}\right) v_1 - x_1, \text{ and } W_2(x_1, x_2) = \left(\frac{x_2}{x_1 + x_2}\right) v_2 - x_2$$

(d) Using a similar approach as in parts (a) and (b), derive a formula for chick 1's optimal begging expenditure, $x_{1\text{opt}}(x_2)$, which maximises its payoff W_1 given that its rival expends x_2. Likewise, derive a formula for chick 2's optimal begging expenditure, $x_{2\text{opt}}(x_1)$, which maximises its payoff W_2 given that its rival expends x_1.

[6 marks]

(e) Using your answer to part (d), find stable levels of begging x_1^* and x_2^* (> 0) that satisfy the conditions $x_1^* = x_{1\text{opt}}(x_2^*)$ and $x_2^* = x_{2\text{opt}}(x_1^*)$, i.e. levels of begging each of which is a best response to the other.

[6 marks]

— END OF PAPER —