

Monday 4th June 20189 to 12

MATHEMATICAL BIOLOGY

*You must answer **eight** questions.*

*You must answer at least **one question** from **each of Sections A to E**.*

*You must begin each answer on a **separate** sheet.*

*Attach a **separate** cover sheet to **each** question.*

The question in Section B that is marked with an asterisk () requires knowledge of the last six lectures in the Michaelmas term.*

Indicative proportions of marks for each part of the questions are given.

*It **does not matter** whether you write on only one side of the paper or on both sides.*

STATIONERY REQUIREMENTS*Script Paper**Rough Work Pads**Blue Coversheets**Tags***SPECIAL REQUIREMENTS***Formulae Booklet**Approved Calculators Allowed*

<p>You may not start to read the questions printed on the subsequent pages of this question paper until you have been instructed that you may do so by the Invigilator</p>
--

SECTION A

A1

(a) The function $f(x)$ is defined by

$$f(x) = \begin{cases} \frac{3}{16}(4x - x^2) & \text{if } 0 \leq x \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

(i) Show that $f(x)$ is a probability density function.

[~15% marks]

(ii) Find $P(x \geq 1)$.

[~10% marks]

(b) The continuous random variable X has probability density function $f(x)$, where

$$f(x) = \frac{x^3}{4} \text{ for } 0 \leq x \leq 2.$$

(i) Find $E(X)$.

[~10% marks]

(ii) Find $E(X^2 + 4X - 2)$.

[~20% marks]

(iii) Find $\text{Var}(3X + 4)$.

[~10% marks]

(c) A scientist needs to sequence a length of DNA, and has a choice of three sequencing laboratories, A, B or C. She chooses laboratory A for 30% of the time, laboratory B for 10% of the time and laboratory C for 60% of the time. Laboratory A offers same day results on 20% of occasions, laboratory B on 6% of occasions and laboratory C on 9% of occasions.

(i) Find the probability that the laboratory chosen will not be able to provide same day results.

[~15% marks]

(ii) Given that the scientist chooses a laboratory and obtains a result on the same day, find the probability that she chose laboratory B.

[~20% marks]

A2

Suppose \mathbf{x} is an eigenvector of a matrix M with eigenvalue λ , and let

$$B = \beta I - 3M + M^2,$$

in which I is the identity matrix and β is a scalar.

(a) Show that \mathbf{x} is also an eigenvector of the matrix B . What is the eigenvalue?

[~20% marks]

The remainder of this question concerns the matrix

$$A = \begin{pmatrix} 4 & 0 & 1 \\ -2 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}.$$

(b) Find the characteristic equation for A and solve it to find the eigenvalues of A .

[~30% marks]

(c) Hence find the eigenvectors of A .

[~30% marks]

(d) The vector \mathbf{w} is given by

$$\mathbf{w} = \begin{pmatrix} 5 \\ 3 \\ -5 \end{pmatrix}.$$

Express \mathbf{w} in terms of the eigenvectors of A , and hence write an expression for $A^m \mathbf{x}$, in which m is a positive integer.

[~20% marks]

(TURN OVER)

SECTION B

B3

- (a) An ecologist wants to investigate the possible effect of pollutants on the sex ratio in the common roach, a species of fish. She collects 400 fish, 200 from pristine locations and 200 from rivers affected by pollutants. She sexes the fish, and obtains:

	Males	Females
Pristine	91	109
Polluted	78	122

Does the presence of pollutants affect the sex ratio in roach? Provide one or two sentences that you could use in a paper to summarise your analysis.

[~30% marks]

- (b) Our ecologist decides also to test whether pollutants might affect the size of roaches. She measures the total length (from snout to tip of the tail, in cm) of females collected from three different types of locations: pristine (no pollutants), medium levels of pollutants, and high levels of pollutants.

	pristine	medium	high
	23.8	24.9	21.5
	23.9	28.3	23.1
	20.7	23.2	22.1
	20.1	25.3	25.2
	20.5	27.5	23
	20.2		24.5
	21.3		
mean	21.50	25.84	23.23
n	7	5	6

$SS_{\text{Group}}=54.96$ and $SS_{\text{Total}}=98.13$

Is there evidence of an effect of pollution on roach length? Sketch the means, and provide one or two sentences that you could use in a paper to summarise your analysis.

[~30% marks]

(QUESTION B3 CONTINUES ON THE NEXT PAGE)

B3 (continued)

(c) The ecologist decides to run a post-hoc test, and obtains the following output:

Tukey multiple comparisons of means
95% family-wise confidence level

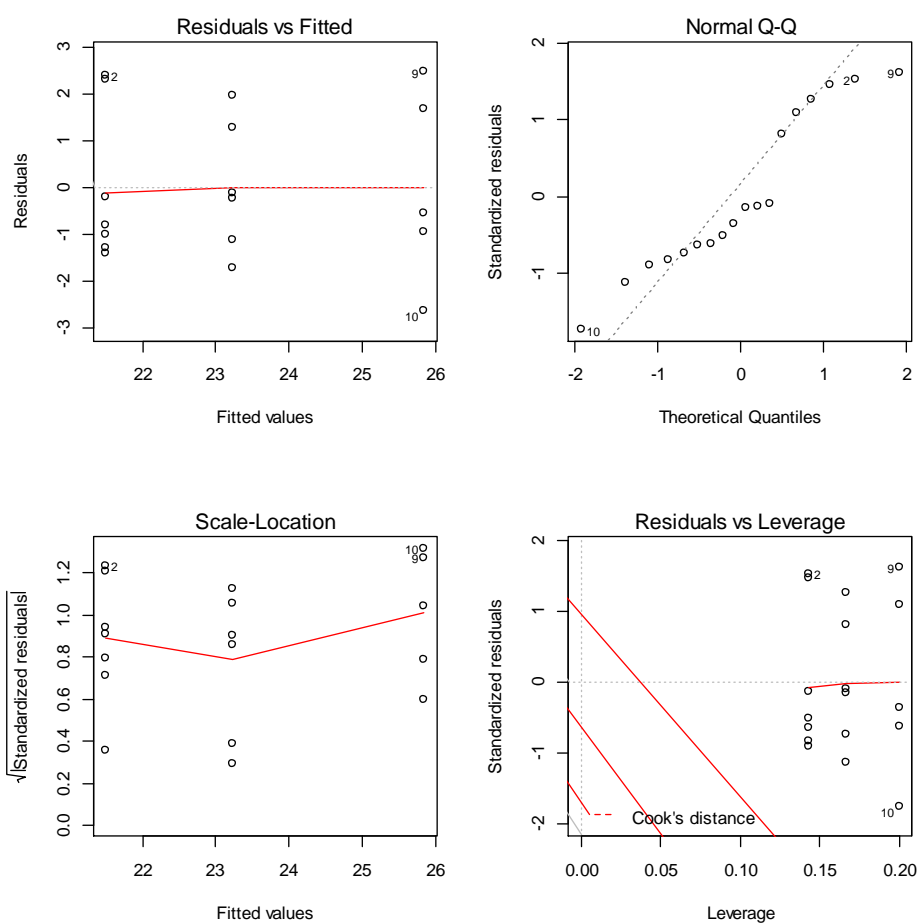
Fit: aov(formula = length ~ location, data = roach)

```
$location
      diff      lwr      upr    p adj
medium-high  2.606667 -0.06147088  5.2748042 0.0559616
pristine-high -1.733333 -4.18476352  0.7180969 0.1916554
pristine-medium -4.340000 -6.92005580 -1.7599442 0.0014994
```

Was this post-hoc test justified? Provide one or two sentences that could be used in a paper to summarise the analysis.

[~20% marks]

(d) Comment on the diagnostic plots for the analysis in (b), which are shown below:



[~20% marks]

(TURN OVER)

B4* (Requires knowledge of the last six lectures of the Michaelmas term)

A disease ecologist wants to investigate the predictors of infection load of a chytrid fungus that affects amphibians. He starts by comparing the infection load of males and females of two amphibian species (common toads and Alpine newts), sampling five individuals for every combination of levels.

The ecologist runs a linear model, and obtains the following output:

```
Call:
lm(formula = chytrid.load ~ sex + species + sex:species, data = chytrid)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.420 -1.365 -0.440  1.030  3.800
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.9800     0.9388   2.109  0.05105 .
sexmale        -0.4200     1.3277  -0.316  0.75583
speciestoad     4.7200     1.3277   3.555  0.00264 **
sexmale:speciestoad -2.5600     1.8776  -1.363  0.19162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.099 on 16 degrees of freedom
Multiple R-squared:  0.5371,    Adjusted R-squared:  0.4503
F-statistic: 6.188 on 3 and 16 DF,  p-value: 0.005395
```

(a) What is the predicted chytrid load for a male toad? And for a female newt?

[~10% marks]

(b) What can we conclude from the F statistic at the bottom of the output ($F_{3,16}=6.2$, $p=0.005$)?

[~10% marks]

(c) Our ecologist tries dropping the interaction from his model, and gets:

Analysis of Variance Table

```
Model 1: chytrid.load ~ sex + species
Model 2: chytrid.load ~ sex + species + sex:species
  Res.Df  RSS Df Sum of Sq    F
1     17  78.7
2      ??  70.5  1      ??     ??
```

Copy and complete the table above. What can be concluded from it?

[~20% marks]

(QUESTION B4 CONTINUES ON THE NEXT PAGE)

B4 (continued)

(d) He goes on to drop the main factors:

Analysis of Variance Table

```
Model 1: chytrid.load ~ species
Model 2: chytrid.load ~ sex + species
  Res.Df  RSS Df Sum of Sq    F
1      18   ??   1      14.5   ??
2      17  78.7   1      14.5   ??
```

Analysis of Variance Table

```
Model 1: chytrid.load ~ sex
Model 2: chytrid.load ~ sex + species
  Res.Df  RSS Df Sum of Sq    F
1      ?? 137.9   1      14.5   ??
2      17  78.70   1      14.5   ??
```

Was he justified to take these steps? Given this output and the output in (c), provide one or two sentences that could be used in a paper to summarise the analysis. **[~30% marks]**

(e) Finally, he decides to use information about elevation (in m) and pond area (in m²) of the sampling locations from which he collected the animals. He builds a linear model with four predictors (elevation, pond area, sex and species), and takes an information theoretic approach:

```
global model call:
lm(formula = chytrid.load ~ sex + species + elevation + pond.area,
    data = chytrid)
```

```
---
Model selection table
  (Int)      elv      pnd.area sex  spc  df  logLik  AICc
11 -0.4950      0.002112      0.001876 +   +   4  -23.705  58.1
12 -0.5939 0.0006422 0.001876      +   +   5  -22.409  59.1
15 -0.7120      0.002183      0.001865 +   +   5  -23.457  61.2
16 -0.7695 0.0006228 0.001942      +   +   6  -22.220  62.9
4  -0.4056 0.0012890 0.001865      +   +   4  -27.913  66.5
8  -0.7451 0.0012240 0.001989      +   +   5  -27.512  69.3
3  -0.0715      0.002444      0.001989 +   +   3  -31.260  70.0
7  -0.5974      0.002583      0.001989 +   +   4  -30.612  71.9
2   0.4856 0.0032890      0.001989 +   +   3  -38.859  85.2
10  0.3080 0.0026730      0.001989 +   +   4  -37.425  85.5
14  0.9526 0.0024430      0.001989 +   +   5  -36.344  87.0
6   1.0260 0.0031590      0.001989 +   +   4  -38.242  87.2
13  2.6200      0.0031590      0.001989 +   +   4  -42.078  94.8
9   1.7700      0.0031590      0.001989 +   +   3  -43.764  95.0
1   3.4900      0.0031590      0.001989 +   +   2  -48.681 102.1
5   4.3400      0.0031590      0.001989 +   +   3  -47.684 102.9
Models ranked by AICc(x)
```

What can be concluded from the output above? **[~20% marks]**

(f) Are these conclusions supported by the following estimates of variable importance?

```
Importance:      pond.area  species  elevation  sex
N containing models:      8      8      8      8
```

[~10% marks]

(TURN OVER)

SECTION C

C5

Citrus canker is a bacterial disease of citrus trees. Diseased trees produce smaller yields of fruit, but do not suffer any significant additional mortality.

Plant pathologists tracked disease spread through an experimental population of N citrus trees. The experiment was thought to be sufficiently far away from external sources of infection for its dynamics to be entirely self-contained. However, a sustained period of dry weather meant that environmental conditions slowly became less conducive to the spread of citrus canker during the epidemic.

- (a) Explain the following model for the number of infected trees, Y , as a function of time since the start of the experiment, t , and in which β and α are positive parameters (as part of the explanation, you should suggest a plausible biological meaning for each of these parameters)

$$\frac{dY}{dt} = \beta e^{-\alpha t} Y(N - Y).$$

[~15% marks]

- (b) Assuming that Y_0 trees are infected at the start of the epidemic, solve the model to find the number of trees infected at time t .

[~50% marks]

- (c) What is the maximum number of trees that ever become infected?

[~5% marks]

- (d) Verify that the maximum number of trees that ever become infected would become smaller if the parameter α were to have a larger numerical value, and explain why this is consistent with the parameter's biological meaning.

[~15% marks]

In analysing the results of the experiment, it was discovered that the model did not fit the data very well. Further investigation located a heavily infected commercial citrus grove in the vicinity of the experiment. It was sufficiently close for trees within the experimental plot to have been at risk of infection from the commercial grove.

- (e) Outline how the model could be updated to account for the presence of the nearby infected citrus grove (no further calculation is required).

[~15% marks]

C6

A model of the density of a population of fish is given by

$$\frac{dY}{dt} = \beta e^{-\alpha Y} Y,$$

in which $Y(t)$ is the density at time t , and β and α are positive parameters.

- (a) Sketch the absolute and relative growth rates of the fish population as a function of its density, Y . An expression in terms of the parameters should be given for the maximum biologically-meaningful value of each growth rate.

[~20% marks]

- (b) What does it mean to say “the model includes density-dependence”?

[~10% marks]

- (c) Suggest plausible biological interpretations of the parameters β and α .

[~10% marks]

An updated version of the model also includes fishing, and is given by

$$\frac{dY}{dt} = \beta e^{-\alpha Y} Y - \omega Y,$$

in which ω is another positive parameter.

- (d) Explain the updated model.

[~5% marks]

For the remainder of the question assume the parameters are such that $\beta > \omega$.

- (e) What are the equilibrium values of the fish population?

[~10% marks]

- (f) Examine the stability of the equilibrium values using a graphical method.

[~20% marks]

- (g) Sketch the population density of the fish as a function of time, for a range of initial population densities.

[~15% marks]

- (h) Explain briefly why the behaviour is consistent with the biological meanings of the parameters β and ω .

[~10% marks]

(TURN OVER)

SECTION D

D7

For the nonlinear system of equations given by

$$\begin{aligned}\frac{dx}{dt} &= x^2 - 4 - y, \\ \frac{dy}{dt} &= x^3 - 4x + yx.\end{aligned}$$

- (a) Find and classify the equilibria.

[~35% marks]

- (b) Sketch the phase plane, marking on clearly the null clines, direction field and sufficient trajectories to illustrate the behaviour of the system.

[~35% marks]

For the initial condition $x(0) = 0$, $y(0) = 5$.

- (c) Mark clearly on the phase plane the trajectory associated with this initial condition.

[~10% marks]

- (d) Sketch the graphs of $x(t)$ and $y(t)$ for this initial condition.

[~20% marks]

D8

Consider the dynamics of a prey-predator system consisting of hares with population density $H(t)$ and foxes with density $F(t)$, given by the following differential equations

$$\begin{aligned}\frac{dH}{dt} &= \beta H \left(1 - \frac{H}{\kappa} \right) - \omega HF, \\ \frac{dF}{dt} &= \lambda HF - \mu F.\end{aligned}$$

- (a) Give an interpretation of the five parameters in the model.

[~25% marks]

- (b) How does this model differ from the classic Lotka-Volterra predation model?

[~5% marks]

- (c) Find the equations of the null clines.

[~20% marks]

- (d) Calculate the stationary points of the model. Under what conditions is there a coexistence equilibrium featuring positive (non-zero) densities of both hares and foxes?

[~25% marks]

- (e) Sketch the phase plot (with H on the x axis and F on the y axis) in the case in which the coexistence equilibrium exists. Comment on the stability of the coexistence equilibrium (no calculation is required concerning the stability properties of the equilibrium point).

[~20% marks]

- (f) Propose and briefly explain a modification of the way predation is modelled that increases the realism of the model.

[~5% marks]

(TURN OVER)

SECTION E

E9

Two chicks in a nest must each simultaneously decide whether or not to beg for food when their parent approaches carrying a prey item. A chick that begs expends c units of energy. If one chick begs while the other does not, the parent will give the prey item, worth b units of energy (where $b > 0$), to the chick that begs. If both chicks beg, then the parent will give the prey item to one of the two offspring, chosen at random. If neither chick begs, then the parent will eat the food item itself, so that neither chick gets anything.

The payoff matrix below give the expected net energetic gain for a chick, depending on its own action (listed down the left-hand side of the matrix), and on its nestmate's action (listed along the top of the matrix)

	Beg	Do not beg
Beg	$(b/2) - c$	$b - c$
Do not beg	0	0

- (a) Under what conditions is each of the two pure strategies ('Beg' and 'Do not beg') in this game evolutionarily stable?

[~20% marks]

- (b) Under what conditions does the game yield an evolutionarily stable mixed strategy? Derive an expression for the evolutionarily stable probability of begging under these conditions.

[~25% marks]

- (c) Suppose that only one chick is hungry, while the other is satiated. If the hungry chick is given the prey item, it will consume it all (obtaining b units of energy), while if the satiated chick is given the prey item, it will only be able to consume half of it (obtaining $b/2$ units of energy). Write out a new payoff matrix for the game, and determine the conditions under which the strategy 'Beg if hungry, but not if satiated' is evolutionarily stable.

[~40% marks]

- (d) Suggest three ways in which you might modify the assumptions of the model to create a more realistic analysis of chick begging behaviour.

[~15% marks]

E10

- (a) The following dynamic programming sequence alignment matrix was completed using scores from an amino acid substitution matrix; a fixed penalty was used for gaps.

	-	T	L	W	V	N	K	C	H	V	Q
-	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	5	0	0	0	1
W	0	0	0	11	3	0	0	3	0	0	0
A	0	0	0	3	11	3	0	0	1	0	0
E	0	0	0	0	3	11	4	0	0	0	2
N	0	0	0	0	0	9	11	3	1	0	0
R	0	0	0	0	0	1	11	8	3	0	1
C	0	0	0	0	0	0	3	20	12	4	0
W	0	0	0	11	3	0	0	12	18	10	2

- (i) Was this matrix completed for finding a global or local alignment? How can you tell?
[~5% marks]
- (ii) What is the gap penalty?
[~5% marks]
- (iii) Show the highest scoring local alignment between the two sequences and the score at each position in the alignment.

Follow the format of the following example to illustrate an alignment and the score at each position:

P P E G R H
 | | | | |
 P P D - K H
 scores: 7 14 16 8 10 18

[~10% marks]

(QUESTION E10 CONTINUES OVERLEAF)

E10 (continued)

(b)

- (i) Copy and complete the dynamic programming sequence alignment matrix given on the following page for finding an optimal global alignment of the sequences ATTG and GACT.

Use the following scoring scheme: nucleotide match = +3, nucleotide mismatch = -1, gap penalty = -1. Use arrows to show the potential trace back options for every cell.

	-	A	T	T	G
-	0				
G					
A					
C					
T					

[~25% marks]

- (ii) What is the score of the optimal global alignment?

[~5% marks]

- (iii) Mark (e.g. with circles) the optimal global alignment path on the matrix.

[~5% marks]

- (iv) Show the optimal global alignment between the two sequences and the score at each position in the alignment.

Follow the format of the following example to illustrate to illustrate an alignment and the score at each position:

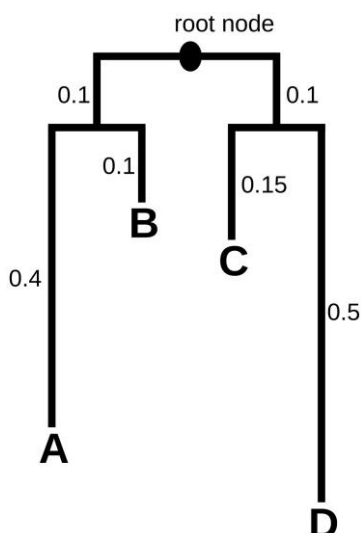
```
      C G G U A G C
      |   |   |
      C A G - A U C
scores: 3 2 5 4 7 6 9
```

[~10% marks]

(QUESTION E10 CONTINUES ON THE NEXT PAGE)

E10 (continued)

- (c) In the rooted phylogenetic tree below, vertical branch lengths (with numbers) correspond to evolutionary distance. The root represents an ancestral sequence whereas A, B, C and D represent four present-day sequences.



- (i) Does this tree have a constant molecular clock? How can you tell?
[~5% marks]
- (ii) Calculate the values of w , x , y and z in the distance matrix below, in which the distance measure is evolutionary distance.

	A	B	C	D
A	0			
B	w	0		
C	x	0.45	0	
D	y	z	0.65	0

[~10% marks]

- (iii) The UPGMA algorithm for constructing a phylogenetic tree assumes that the tree has a constant molecular clock. However, it is often still possible to apply the UPGMA algorithm even when this condition is not satisfied, although the resulting tree may be different from the true evolutionary tree. Apply the UPGMA algorithm to the distance matrix from part ii. Show the tree produced by the UPGMA algorithm and annotate the branch lengths according to the UPGMA algorithm (note that the resulting tree branch lengths may disagree with the distance matrix from part ii).

[~20% marks]

END OF PAPER