18. Mock-up of a written examination paper

NATURAL SCIENCES TRIPOS	**Part IA**
June 2019 (Sample Exam)	9 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 **not** answer all three questions from **Section 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

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

SECTION A

A1

(a) A bird of prey hunts mice. The chance of success for each attempt by the bird to catch a mouse is 0.6. The bird makes four attempts. Find the probability of each of the following outcomes.

i. Three successful attempts.
ii. No successful attempts.
iii. At most two successful attempts. *[~30% marks]*

(b) A second bird of prey hunts with a probability of success p=0.8. Two graduate students are studying the behaviour of this particular bird and devise the following game: Student A pays 50p to Student B every time the bird makes a successful attempt and gets 80p from Student B every time the bird fails to catch a mouse. They play the game over twenty attempted hunts by the bird. What is the expected value of the earnings of Student B?

[~30% marks]

(c) The same bird of prey also hunts squirrels. In its natural habitat, it can find both red and grey squirrels. The grey squirrels make up 80% of the squirrel population, while red squirrels make up 20%. The dynamics of this squirrel population are being studied and for this purpose the animals are being tagged. 92% of all the red squirrels have been tagged and 63% of the grey. A tagged squirrel is caught by the bird of prey. What is the chance that the squirrel was red?

[~40% marks]

A2

Consider the following matrix

$$A = \begin{pmatrix} 4 & 1 & -1 \\ 2 & 5 & -2 \\ 1 & 1 & 2 \end{pmatrix}$$

(a) Find $A\mathbf{v}^T$ where $\mathbf{v}$ = (1,3,-2). *[~5% marks]*

(b) Find det(A). *[~15% marks]*

(c) Find the characteristic equation of A, and hence find its eigenvalues. *[~40% marks]*

(d) Find a set of eigenvectors for A, and hence sketch its invariant lines when viewed as a linear transformation. *[~40% marks]*

SECTION B

B3 A biologist suspects that a ladybird population may be subject to infection by a male-killing bacterium, which gives rise to a female-biased sex ratio. To test this idea, he samples 10 individuals from the population and records their sex. He plans to use a one-tailed binomial test (with a significance level of α = 0.05) to evaluate whether the sex ratio is significantly skewed towards females.

(a) What is the minimum number of females (out of the 10 individuals sampled) that would cause the biologist to reject the null hypothesis of a 1:1 population sex ratio? *[~20% marks]*

(b) Suppose that the sex ratio is in fact 1:1. What is the probability of obtaining sufficient females in the sample to mistakenly reject the null hypothesis (a Type I error)? *[~10% marks]*

(c) Suppose that the sex ratio is in fact 3:1 (females:males). Calculate the probability of the biologist obtaining so few females that he mistakenly fails to reject the null hypothesis (a Type II error)? *[~20% marks]*

(d) How might the risk of both types of error be reduced? *[~10% marks]*

A statistics lecturer compares the heights (in cm) of 15 male students from three colleges: University Hall, Valance Mary Hall and Katharine Hall.

University Hall	Valance Mary Hall	Katharine Hall
177	172	172
178	179	180
180	178	182
184	180	176
182	185	187

$$\bar{y}_{University} = 180.2\,\text{cm};\ \bar{y}_{Valance\,Mary} = 178.8\,\text{cm};\ \bar{y}_{Katharine} = 179.4\,\text{cm}$$

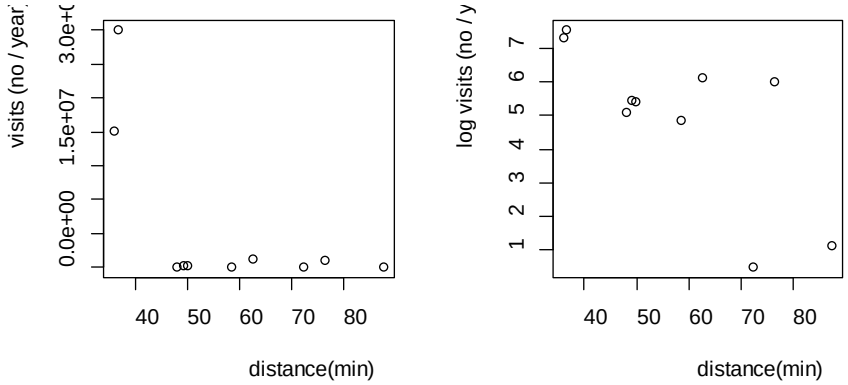
$$\sum y_i = 2692\,\text{cm};\qquad \sum (y_i - \bar{y})^2 = 255.73\,\text{cm}^2.$$

(e) Is there a significant difference between heights of male students at these Colleges? (Please show working and state null and alternative hypotheses) *[~30% marks]*

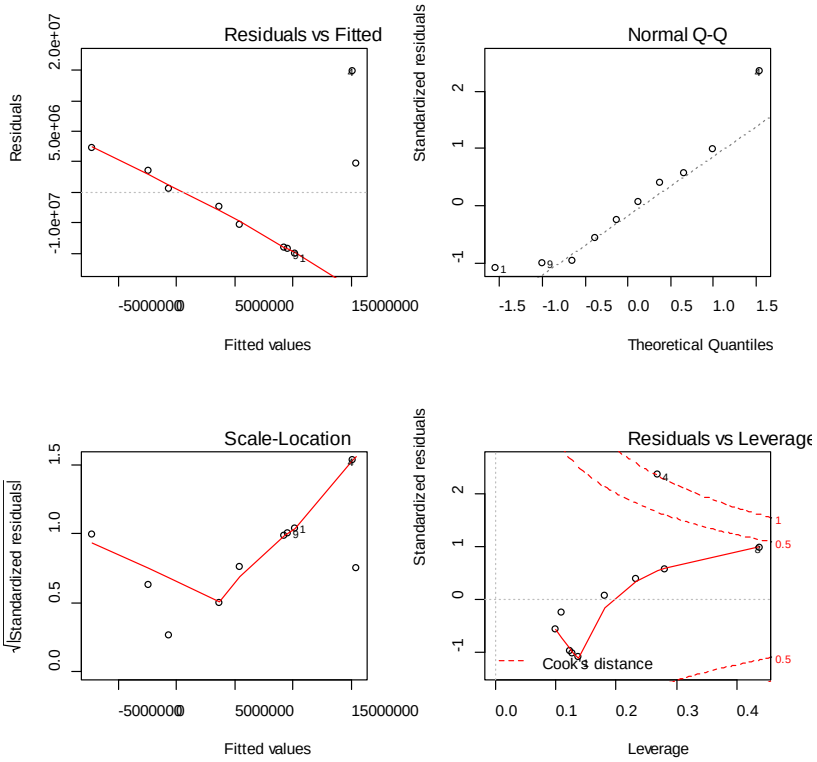
(f) Provide one or two sentences that could be used to report your analysis in a scientific paper. *[~10% marks]*

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

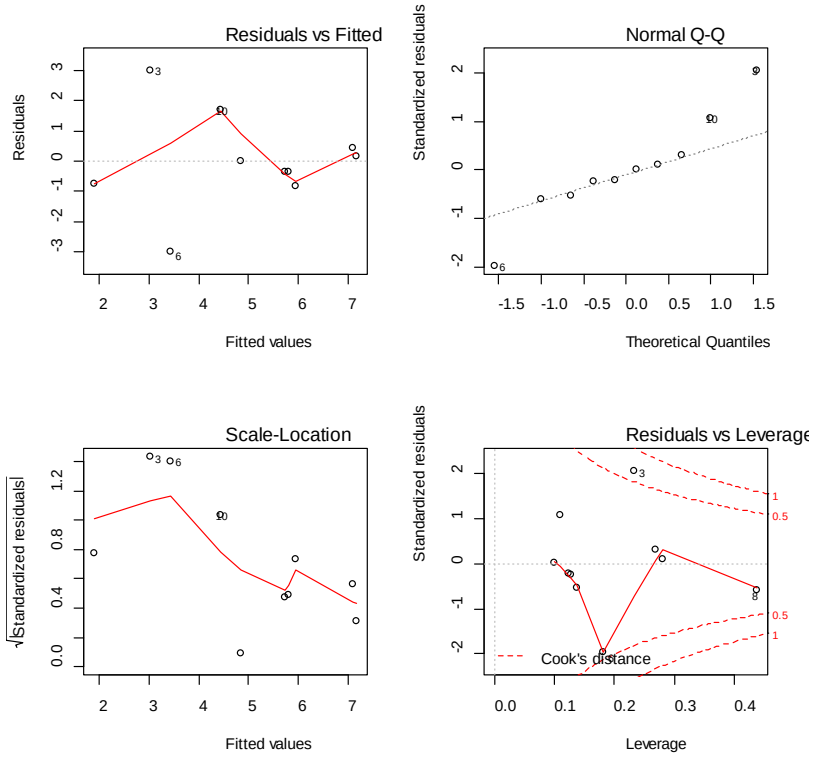
A conservation biologist is interested in testing whether accessibility affects the number of visitors to nature reserves. She collects visitor data (number of visits/year) for 10 reserves, for which she also measures driving distance (minutes) to the nearest town. She plots both the raw data as well as log transforming visit numbers:



After inspecting the diagnostic plots for a regression analysis on the raw data,



she decides to carry out her regression analysis on the log transformed data, and obtains the diagnostic plots shown in the next page.



(a) Was she justified in deciding to use the transformed data? If yes, did the transformation improve the validity of the model's assumptions?

[~15% marks]

When she carries out her regression on the log transformed data, she obtains the following ANOVA table.

	SS	df	MS	F
Model	27.904	**??**	**??**	**??**
Error	**??**	**??**	**??**	
Total	49.712	9		

(b) Complete the table (missing items are denoted by **??**), and test whether distance from the nearest town is a predictor of the number of visits.

[~20% marks]

(c) Write a sentence to summarise your results in a scientific paper.

[~15% marks]

An analysis of whether weight (in kg) and sex of patients affects their likelihood of contracting an infection while on hospital returns the following output:

```
Call:
glm(formula = infected ~ sex + weight + sex:weight,
      family = binomial(logit), data = Dataset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6236  -0.5969  -0.3051  -0.2450   2.3101

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.18270    0.92802   1.274   0.2025
sex[T.male]        -0.86963    1.29470  -0.672   0.5018
weight             -0.17608    0.09220  -1.910   0.0562 .
sex[T.male]:weight -0.04771    0.12700  -0.376   0.7072
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 83.234  on 80  degrees of freedom
Residual deviance: 63.645  on 77  degrees of freedom
AIC: 71.645
```

(d) What is the likelihood of contracting an infection for a male patient weighting 47 kg?

[~10% marks]

Dropping the interaction we get:

```
Analysis of Deviance Table

Model 1: infected ~ weight + sex
Model 2: infected ~ weight * sex
Resid. Df Resid. Dev Df Deviance
1        78     63.785
2        77     63.645  1    0.141
```

Dropping the main factors from the additive model, we get:

```
Analysis of Deviance Table

Model 1: infected ~ sex
Model 2: infected ~ weight + sex
Resid. Df Resid. Dev Df Deviance
1        79     74.956
2        78     63.785  1   11.171

Analysis of Deviance Table

Model 1: infected ~ weight
Model 2: infected ~ weight + sex
Resid. Df Resid. Dev Df Deviance
1        79     68.072
2        78     63.785  1    4.287
```

(e) What would you conclude from the output above?

[~25% marks]

(f) Provide one or two sentences that you could use in a paper to summarise your analysis.

[~15% marks]

SECTION C

C5 The giant hogweed, *Heracleum mantegazzianum*, was introduced into Great Britain by the Victorians, and has been spreading ever since. The model

$$\frac{dN}{dt} = \big(b(N) - d(N)\big)N,$$

is suggested, where $b(N)$ and $d(N)$ are functions of the size of the hogweed population at time t, $N(t)$. A particular version of the model has

$$b(N) = b_0 - b_1 N,$$
$$d(N) = d_0,$$

in which $b_0 > d_0$.

(a) Interpret the biological meaning of the model. *[~10% marks]*

(b) Demonstrate that if β and κ are chosen appropriately, the model can be written as

$$\frac{dN}{dt} = \beta N\left(1 - \frac{N}{K}\right).$$

Interpret the meaning of the expressions for β and K.

[~20% marks]

(c) Sketch the direction field associated with the model, and use it to sketch $Y(t)$, assuming that the initial hogweed population is small.

[~20% marks]

An alternative model takes a different form for $b(N)$

$$b(N) = \frac{b_0}{1 + b_1 N},$$
$$d(N) = d_0.$$

(d) Sketch $b(N)$ and interpret the biological basis of the updated model, suggesting one reason why it might be an improvement.

[~15% marks]

(e) Determine the population size at which the total number of Hogweeds is increasing most quickly according to the updated model.

[~20% marks]

(f) Suggest three features of even the updated model that may be unrealistic for an invading plant species. *[~15% marks]*

C6 A group of organisms that is living in an isolated habitat has population size at time t, $Y(t)$. The population is subject to immigration at rate α, reproduction at per capita rate β, and death at per capita rate γ, where α, β and γ are all positive constants. In parts (a)-(d) you should assume that $\beta < \gamma$.

(a) Write down a model of the form

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

to describe the evolution of $Y(t)$. *[~10% marks]*

(b) Find the equilibrium value of your model, and examine its stability using a method based on calculating dF/dY at the model's equilbrium.

[~25% marks]

(c) Solve the model to find $Y(t)$, given that the habitat was initially empty. Verify your solution is consistent with your answer to part (b).

[~30% marks]

(d) Sketch your solution and interpret the changes to your sketches as the values of the parameters are changed. *[~20% marks]*

(e) What would happen if instead parameters were such that $\beta = \gamma$?

[~15% marks]

SECTION D

D7 The dynamics of a system are described by the following pair of simultaneous first-order non-linear differential equations.

$$\frac{dx}{dt} = xy + x^2,$$
$$\frac{dy}{dt} = 6 - x^2 - y.$$

(a) Find the equilibrium point(s). *[~20% marks]*

(b) Classify the equilibrium point(s). *[~30% marks]*

(c) Sketch the null-clines for this system, and then the phase plane, showing the behaviour around the equilibrium points by adding representative trajectories. *[~30% marks]*

(d) For the initial condition (-2, 0) sketch the trajectory of the path on the phase plane, and sketch the graph of x against t, showing the main qualitative features. *[~20% marks]*

D8 Suppose that a dangerous new virus of humans has emerged and has circulated in several countries other than the United Kingdom. Last week the first cases were found in London. As far as we know the disease is not fatal and people recover with long lasting immunity. The following equations describe the dynamics of this viral infectious disease.

$$\frac{dS}{dt} = bN - bS - \beta IS,$$
$$\frac{dI}{dt} = \beta IS - \nu I - bI,$$
$$\frac{dR}{dt} = \nu I - bR.$$

(a) Briefly give biological definitions of each of the model parameters.
[~15% marks]

(b) Define R_0 and derive an expression for this quantity for the model above.
[~20% marks]

(c) Find the non-zero equilibrium. *[~20% marks]*

Fortunately a vaccine has already been developed, and this vaccine confers perfect immunity against the virus.

(d) How would R_0 and the non-zero equilibrium point have been altered if a proportion p of the population had been instantaneously vaccinated at $t = 0$?
[~20% marks]

It is discovered that newborn children are most at risk of severe disease following infection.

(e) Rewrite the model such that, instead of instantaneously vaccinating a proportion p of the population at $t = 0$, instead a proportion q of newborns is vaccinated routinely at the time of birth. What is the minimum proportion that needs to be vaccinated to prevent an outbreak?
[~25% marks]

SECTION E

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E9 Two lionesses are confronted by a potential prey of energetic value b. Each must simultaneously decide whether or not to pursue it. Pursuit entails the expenditure of c units of energy. If only one lioness pursues the prey, she captures it with probability p_1, in which case she consumes it all and thus gains b units of energy (while the other lioness gains nothing). If both lionesses pursue the prey, they capture it with probability p_2 ($> p_1$), in which case they share it equally and thus gain $b/2$ units of energy each. If neither lioness pursues the prey, neither gains anything.

(a) Write down a payoff matrix for this game (with payoffs in terms of net expected energetic gain). *[~10% marks]*

(b) Under what conditions is each of the pure strategies in this game evolutionarily stable? *[~20% marks]*

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

[~30% marks]

(d) Suppose that the two lionesses represent a mother and her daughter, and that if both pursue and capture the prey, the mother will then claim the greater share, gaining $2b/3$ units of energy (while the daughter gains only $b/3$). Write out a new payoff matrix for the game, and determine the conditions under which the strategy “pursue if mother, but not if daughter” is evolutionarily stable.

[~40% marks]

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E10

(a) For an enzyme catalysed reaction, show how the steady state approximation and other assumptions can be used to derive the Michaelis-Menten equation:

$$v_0 = \frac{V_{max}[S]}{K_M + [S]}$$

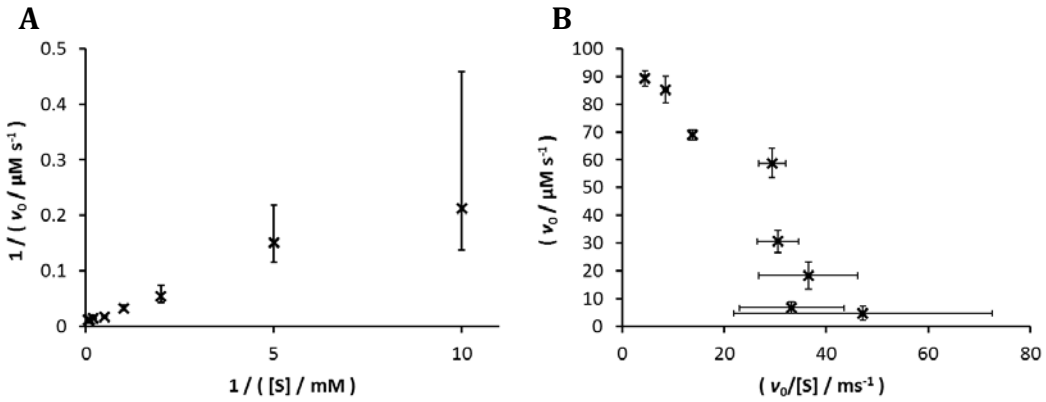
where v_0 is the initial rate of product formation, **[S]** is the initial concentration of the substrate S, and V_{max} and K_M are positive constants.

Base your answer on the following reaction scheme, where E represents the enzyme, (ES) is the enzyme-substrate complex, P is the product and k_1, k_{-1} and k_2 are rate constants for the indicated steps:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} (ES) \xrightarrow{k_2} P + E$$

[~60% marks]

(b) Panels **A** and **B** below show scatter plots for a data set of substrate concentrations **[S]** and initial rates v_0 for the same enzyme catalysed reaction after attempting to apply two different linear transformations. For each case, explain why it would not be appropriate to perform an unweighted linear regression using the horizontal axis coordinate as the explanatory variable and the vertical axis coordinate as the response variable.



[~20% marks]

(c) If the data set discussed in part (b) is expected to obey the Michaelis-Menten equation, suggest an alternative approach that would treat the experimental data more appropriately and would yield accurate values for the fitting parameters V_{max} and K_M.

[~20% marks]

ANSWER NO MORE THAN 2 QUESTIONS FROM SECTION E

E11

(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.

Hint. 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 E11 CONTINUES OVERLEAF; TURN OVER)

E11 (continued)

(b)

(i) Copy and complete the dynamic programming sequence alignment matrix given below 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.

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

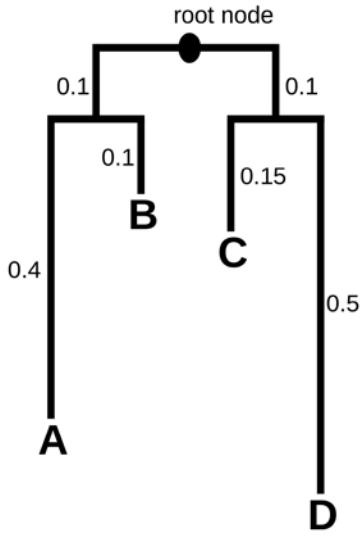
```
          C G G U A G C
          |   |   |   |
          C A G - A U C
scores:   3 2 5 4 7 6 9
```

[~10% marks]

(QUESTION E11 CONTINUES ON THE NEXT PAGE)

E11 (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