IMPERIAL COLLEGE LONDON

MSc COURSE IN COMPUTATIONAL METHODS IN ECOLOGY AND EVOLUTION

EXAM 2

For Internal Students of Imperial College of Science, Technology and Medicine

Exam Date: Tuesday, 24th March 2015, 1300 – 1600 hrs

Length of Exam: 3 HOURS

Instructions: All sections are weighted equally. It is a three-hour exam, and there are 5 sections, so it is a reasonable guideline to spend about 35 minutes on each section. All sections allow you to choose between two questions, answering one. Read instructions carefully at the head of each section.

PLEASE PUT ANSWERS TO EACH SECTION IN A SEPARATE EXAM BOOK.

WE REALLY MEAN IT. PLEASE PUT ANSWERS TO EACH SECTION IN A SEPARATE EXAM BOOK. THE REASON FOR THIS IS THEN WE CAN PARALLELIZE MARKING AMONG THE DIFFERENT LECTURERS AND YOU GET THE MARKS BACK SOONER.

Continues on next page

Section 1: Maths III

Please select exactly **one question** and answer it.

A. Exponential growth with oscillations

The following differential equation models a modified version of the exponential growth model, where the per capita growth rate exhibits oscillations (e.g. for species whose reproductive cycle is linked to the seasons):

$$\frac{dN}{dt} = (a + b\cos t)N(t), \quad N(0) = N_0$$

Now answer the following:

(i) Solve this initial value problem (60%) (Hint: use separation of variables)

(ii) Using reasonable values for the parameters N_0, a and b, and for the units of N and t, make a sketch of the solution (40%)

(iii) Extra-credit: Determine conditions on a and b that guarantee that the population never gets below its initial value N_0 (for a max of additional 30%) (Hint: Note that slope of $\sin t$ is 1 at the origin)

B. Forest gap dynamics

A very simple model for gap dynamics in a forest assumes that gaps are created by disturbances (e.g. fire, but not logging) and that gaps revert to forest as trees grow in gaps. In a forest of area N, let $x_1(t)$ denote the area occupied by gaps and $x_2(t)$ the area occupied by adult trees.

Here is the model:

$$\begin{aligned} \frac{dx_1}{dt} &= ax_1 + bx_2 \\ \frac{dx_2}{dt} &= cx_1 + dx_2 \end{aligned}$$

Now answer the following:

(i) Determine conditions on the parameters a, b, c and d so that these equations correctly and realistically model the given situation.

For the rest of the entire question, assume that these conditions hold. Think carefully here, because if you incorrectly answer this part, the remaining parts will be much harder to solve. (40%)

(ii) What do the trace and determinant of the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ tell you about the stability of the model? (30%)

(iii) Solve the system of differential equations. (30%)

(iv) Extra credit: Determine further conditions on the parameters so that this model has a stable solution. Hence, determine the asymptotic value of $x_1(t)$ and $x_2(t)$, as $t \to \infty$ (for a max of additional 30%)

Continues on next page

Section 2: Ecological theory and modelling

A. Gradual enrichment of shallow fresh water lakes with nutrients can lead to a sudden change from a transparent state, in which submerged plants dominate, to a turbid state dominated by floating algae. This change is not reversible by a slight decrease in the nutrient level.

(i) Draw and label an appropriate bifurcation diagram. (30%)

(ii) Describe and explain the dynamics in shallow fresh water lakes in terms of this bifurcation diagram. (30%)

(iii) Explain the possible role of competition in this ecosystem. (40%)

B. In March 2014 Ebola was first reported in West Africa. Since, well over 20,000 cases have been reported, over 10,000 of which fatal. There has been a substantial effort to control the spread of this disease.

(i) What is the basic reproductive number? For what value of the basic reproductive number can a disease be controlled? Put your answer in the context of Ebola to the extent possible. (30%)

(ii) List different categories of control measures and explain how they affect the basic reproductive number. (30%)

(iii) Using this to explain the logic behind efforts to control the 2014/15 Ebola outbreak in West Africa. (40%)

Continues on next page

Section 3: Evolutionary theory and modelling

A. Answer the following:

(i) Explain how homing endonuclease might be used to control mosquito vectors of malaria (30%)

(ii) For a homing endonuclease that targets an essential gene for which knock-out mutations are recessive lethal, derive an expression for the equilibrium frequency (g) of the HEG and the equilibrium fraction of individuals that die because of the HEG as a function of the homing rate (h) (30%)

(iii) Describe a model of density-dependent population regulation, and derive an expression within that model of what effect the HEG will have on equilibrium population size (40%)

B. A single bacterial species grows on a single input resource in a flow-through chemostat according to the following differential equations:

$$\frac{dS}{dt} = DQ - DS - \frac{k_e ESN}{K+S} \tag{1}$$

$$\frac{dN}{dt} = c\frac{k_e ESN}{K+S} - DN \tag{2}$$

Where,
D is the dilution rate
Q is the concentration of the substrate in the inflow
S is the concentration of the substrate in the chemostat
k_e is the rate parameter of the enzyme
E the amount of enzyme per cell ($k_e E$ is the maximum rate of the enzyme in excess substrate)
N is the density of the bacteria
K is the Michaelis-Menten constant of the enzyme
c is the number of cells of bacteria produced per unit of substrate metabolised

(i) Explain each component in the equations (10%) (extra credit for stating units of each parameter or variable)

(ii) Explain what the Michaelis-Menten constant signifies in words and/or graphically. (10%)

(iii) Solve the equations to work out the steady-state concentrations of substrate (S) and density of bacteria (N). Plot how they change as the concentration of substrate in the inflow (Q) increases. (40%)

(iv) Pick any other biological feature currently absent from the model. Explain how you would add it into the model and generate a hypothesis for how you expect it to effect the dynamics of the system. [40%]

Continues on next page

Section 4: Maximum Likelihood

Please select exactly **one question** and answer it. Calculator may be required in some questions.

You may use the chi-square table below for critical values:

Degrees of freedom	$\chi^2_{0.95}$
1	3.84
2	5.99
3	7.81
4	9.49

A. Answer the following:

(i) Please explain the concept of Maximum Likelihood Estimation, and also state the three items required to perform MLE. (30%)

(ii) In 20 independent coin tosses, 14 of them were head. Let p be the probability of getting a head, and the likelihood function of p is

$$L(p) = \binom{n}{y} p^y (1-p)^{n-y}$$

where n is the number of independent trials and y is the number of heads observed. Perform a likelihood ratio test for $H_0 : p = 0.5$ vs. $H_1 : p \neq 0.5$ at 5% significance level. (35%)

(iii) Describe, as precisely as possible, that how you can obtain 95% confidence interval (and confidence region for multiple parameters) from the log-likelihood function. You may use equations or graphs as examples in your explanation. (35%)

B. Answer the following:

(i) Please state the four properties of Maximum Likelihood Estimators, and also give a brief explanation to each of them. (30%)

(ii) Let $X_1, X_2, \ldots, X_3$ be independent and identically distributed samples from $Exponential(\lambda)$. Given the probability density function of an exponential random variable is $f_X(x) = \lambda e^{-\lambda x}$, show that the MLE for λ is $\frac{n}{\sum_{i=1}^{n} x_i}$. (30%)

(iii) A CMEE student is trying to implement MLE in R. She writes her own log-likelihood function `log.like` which contains two parameters, plus an input dataset `dat`. She then uses `optim()` to maximise the log-likelihood function. She types the following command into R:

```
optim(par=c(100,0.1), fn=log.like, dat=dat,
control=list(fnscale=-1), hessian=T)
```

And this is the (partial) output from R:

```
$par
[1] 401.6857134   0.1048165

$value
[1] -3507.493

$hessian
               [,1]         [,2]
[1,] -0.0007751169    -1.64175
[2,] -1.6417499182 -8259.37872
```

Please describe, as precisely as possible, each component of the input and output screen. (40%)

Continues on next page

Section 5: Bayesian statistics

Please select exactly **one question** and answer it.

A. Consider the problem of Bayesian inference for a random variable describing whether the relatively rare species of West African giraffe is able to produce offspring in a given year when held in captivity. The gestation period of a giraffe can be between 400 and 460 days. Explain how you choose an appropriate prior probability density:

(i) What type of prior distribution do you choose? (20%)

(ii) How do you sensibly choose its hyper-parameters if expert information is available? (30%)

(iii) How do you do this if no such information is around? (20%)

(iv) What do you do if the only available expert makes very strong claims? (30%)

B. Write a program in pseudo-code, BUGS, or R syntax to analyse the probability of occurrence of a certain woodland bird species in terms of habitat.

Data is given as a vectors v, a and h of length n, where $v[i]$ is the number of birds counted in one day for survey site i, $a[i]$ is the area of the patch in which the survey site is located, and $h[i]$ is a binary variable indicating the type of vegetation of the site (coniferous versus deciduous). Use linear regression with uninformative, but reasonable priors.

End of paper