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)

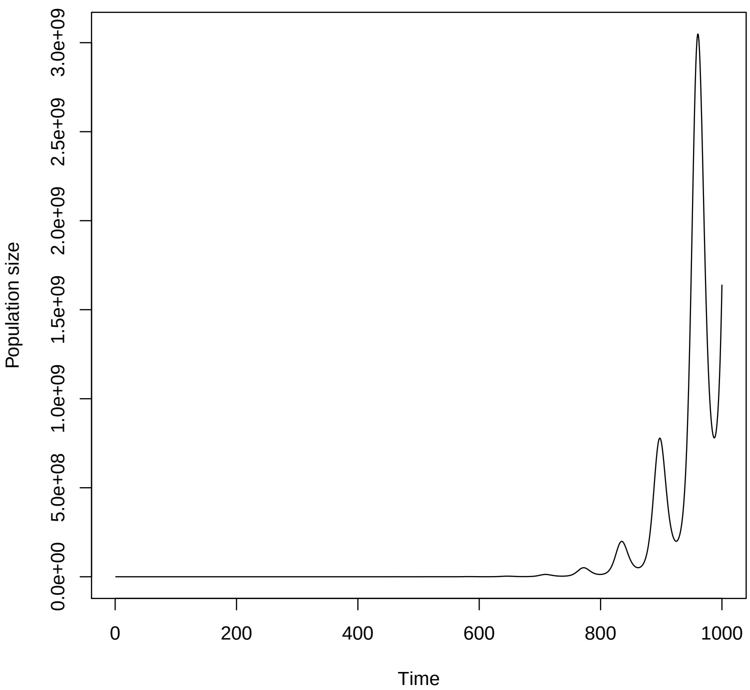
(ii) Using reasonable values for the parameters N_0, a and b, and for the units of N and t, make a sketch of the solution (40%)

(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)

Model Answer (Marker – Pawar (1st), Rosindell (2nd)):

(i) Use separation of variables to rewrite the d.e. as $\int dN/N = \int (a + b\cos(t))dt$
So $\log N = at + b\sin t + c$
and $N = N_0 exp(at + b\sin t)$

(ii) Plot of exp(0.217234t + sin(t)) — Should look something like this:



(iii) We need $N(t) \geq N_0$, so $at + b\sin t \geq= 0$ for all $t > 0$
Knowing that the slope of $\sin t$ is 1 at the origin:

Continues on next page

If $b < 0$, we need $a > |b| = -b$

If $b \geq 0$, the sine is reversed and we need a/b to be at least slightly bigger than $2/3\pi$ (0.217234 to be exact)

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:

$$\frac{dx_1}{dt} = ax_1 + bx_2$$
$$\frac{dx_2}{dt} = cx_1 + dx_2$$

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%)

Model Answer (Marker – Pawar (1st), Rosindell (2nd)):

(i) Since $x_1 + x_2 = N$, the total area, which is constant, we need $c = -a$ and $d = -b$

Some might be able to already show here that $a < 0$ and $b > 0$: bonus marks in the form of leniency towards the rest of the question

(ii) Nothing, because $det(A) = 0$, so at least one of the eigenvalues is zero; hence the general stability theory does not apply

(iii) First we eliminate x_2 using the fact that $x_2 = N - x_1$ so the first d.e. becomes $dx_1/dt = (a - b)x_1 + bN$

With a change of variables $y = x_1 + bN/(a - b)$,

this d.e. becomes $dy/dt = (a - b)y$

so $y(t) = y(0)exp((a - b)t)$, with $y(0) = N_1 + bN/(a - b)$

Finally, $x_1(t) = y(0)exp((a - b)t) - bN/(a - b)$

(iv) We need $a < b$, to avoid exponential growth, which would lead to negative solutions.

Thus $y(\infty) = 0$, so $x_1(\infty) = -bN/(a - b)$ and $x_2(\infty) = N + bN/(a - b) = aN/(a - b)$

This means we also need $b > 0$ and $a < 0$ to ensure x_1 and x_2 remain positive

These conditions supersede the previous one, $a < b$, as they imply it

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%)

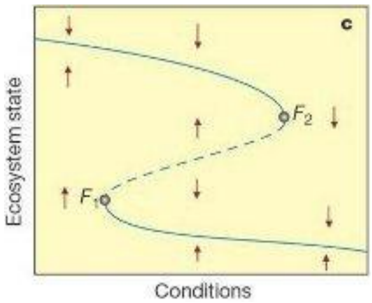
(ii) Describe and explain the dynamics in shallow fresh water lakes in terms of this bifurcation diagram. (30%)

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

Model Answer (Marker – Pawar (1st), Rosindell (or Cator?) (2nd)):

(i) A good answers will first of all recognize that this is appears to be a case of a catastrophic transition, and that the ecosystem has alternative stable states. The bifurcation diagram looks as follows

The plot should look something like:



In the diagram two fold bifurcations can be seen (labelled F). An acceptable alternative is a diagram in which one or two of the fold bifurcations are replaced by transcritical bifurcation. (This is how this was initially explained in the lectures, the folds are then replaced by sharp turns). The diagram should have the enrichment status on the abscissa, and the state of the lake (eg turbidity) on the ordinate.

(ii) To explain the dynamics it is essential to recognise that the middle state in the diagram is unstable, and the lower and upper branches stable. Therefore the dynamics lead toward the upper and lower branches. However, following a perturbation the system can switch from one stable state to the other. Under a change of conditions, a sudden change in behaviour can occur this takes you to the other side of the bifurcation. The arrows in the diagram tell the story.

(iii) In freshwater typically the floating vegetation is good in absorbing light, the submerged vegetation tends to be good at sequestering nutrients, and this fits the bill here. It is somewhat similar to the example explained in the lectures where the pond and duckweed competed. The duckweed, or submerged plants reduce the nutrient levels in the water, excluding the algae (or pondweed). Pondweed, or algae, once dominant block the submerged plants from receiving sunlight. This creates two alternative stable states, and the transition between them is catastrophic. This is dealt with in some detail in a paper on the reading list by Scheffer et al. (2001) who phrase this nicely: *"The pristine state of most shallow lakes is probably one of clear water and a rich submerged vegetation. Nutrient loading has changed this situation in many cases. Remarkably, water clarity often seems to be hardly affected by increased nutrient concentrations until a critical threshold is passed, at which the lake shifts abruptly from clear to turbid. With this increase in turbidity, submerged plants largely disappear. Associated loss of animal diversity and reduction of the high algal biomass makes this state undesired. Reduction of nutrient concentrations is often insufficient to restore the vegetated clear state. Indeed, the restoration of clear water happens at substantially lower nutrient levels than those at which the collapse of the vegetation occurred .... Experimental work suggests that plants increase water clarity, thereby enhancing their own*

Continues on next page

growing conditions. This causes the clear state to be a self-stabilizing alternative to the turbid situation .. The reduction of phytoplankton biomass and turbidity by vegetation involves a suite of mechanisms, including reduction of nutrients in the water column, protection of phytoplankton grazers such as Daphnia against fish predation, and prevention of sediment resuspension. In contrast, fish are central in maintaining the turbid state, because they control Daphnia in the absence of plants. Also, in search for benthic food they resuspend sediments, adding to turbidity. Whole-lake experiments show that a temporary strong reduction of fish biomass as 'shock therapy' can bring such lakes back into a permanent clear state if the nutrient level is not too high."

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%)

Model Answer (Marker – Pawar (1st), Rosindell (2nd)):

(i) The basic reproductive number is the number of secondary infections caused by a single infected individual in a population without the disease. (There is a slide in the lecture material saying "The number of secondary infections caused by a single infected individual in a completely susceptible population". This is a workable but more limited definition, that does not take the possibility of vaccination into account). **To control a disease the basic reproductive number needs to be reduced to less than one.** For this value the disease is certain to disappear. For Ebola, estimates for the basic reproductive number fall between 1 and 2.

(ii) Control strategies will affect the reproductive number and allow explaining the rationale underlying the intervention. The basic reproductive number is the transmission rate times the duration of the infection times the fraction of susceptibles that are present in the population. In formula, if they happen to reproduce it this is $\frac{\beta S_0/N}{\gamma+\mu+v}$, although the formula is not massively helpful. The verbal interpretation suggest that you can do one of 3 things: **(a) reduce transmission, (b) reduce the duration of the infection, (c) reduce the number of susceptibles.** I would expect to see these 3 in the answer. Examples of how this can be done (not all apply to Ebola) is through (a) quarantine, condoms, bednets, travel restrictions (b) treatment, culling (c) vaccination.

(iii) Ebola is a viral disease for which there is currently no effective treatment and no vaccine (although testing is underway). The reproductive number exceeds one. So answers should not focus on treatment or vaccines. The only thing that currently can be done to control the disease is *to reduce the transmission of the disease.* This done through the isolation of infected patients by keeping them in isolated wards and where doctors and nurses wear protective clothing. In past outbreaks, as well as this one, burial practices might have contributed to transmission: cleaning of the dead body by family members, kissing and touching the body can cause infection. Therefore another way to reduce the transmission is through persuading or forcing people to change burial practices. It appears that in previous Ebola outbreaks a change in behaviour in the population contributed to halting of the outbreak. In the outbreak in West Africa it appears this is happening to an extent. Information campaigns are therefore an important means to. But these efforts are all aimed at reducing transmission, and not changing the duration of the infection, or the number of susceptible individuals.

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%)

Model Answer (Marker – Burt (1st), Barraclough (2nd)):

(i) Homing endonucleases might be used in 2 ways:

(a) If a gene for a HEG targeting the X chromosome is put on the Y chromosome and expressed only at male meiosis, then the X chromosome may be prevented from being transmitted to the next generation, giving the Y chromosome an advantage, which would spread through the population, and produce a male-biased sex ratio as it did so. This will lead to lower numbers of mosquitoes.

(b) If a HEG targets a sequence in an essential gene, or a female fertility gene, and then is inserted in the middle of its own recognition sequence, thereby disrupting the target gene, then it may spread through a population by the homing reaction, producing gene knock-outs and either death or sterility as it does so.

(ii) If h = homing rate, then equilibrium frequency $q = h$, and fraction of individuals that die is $q^2 = e^2$

(iii) Any population model can be used; in class we did the discrete time version of the logistic. In this model the effect of the HEG is to reduce r, but has no effect on alpha.

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%)

Continues on next page

(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%]

Model Answer (Marker – Barraclough (1st), Burt (2nd)):

(i) Equation 1: inflow of substrate, outflow of substrate, metabolism of substrate by the bacterial species. Equation 2: growth of the bacterium through metabolism of substrate and outflow of bacteria.

(ii) It is the substrate concentration at which the rate of metabolism is half of $k_e E$ (i.e. V_{max}). Plot a monotonic curve and show half the asymptote and the S value for it.

(iii) Set both equations to zero and solve. Thus,

$c\frac{k_e ESN}{K+S} - DN = 0$

$c\frac{k_e ES}{K+S} = D$

$ck_e ES = D(K+S)$

$ck_e ES - DS = DK$

$S(ck_e E - D) = DK$

Therefore, $\hat{S} = \frac{DK}{ck_e E - D}$

From equation 2, at steady-state,

$\frac{k_e E\hat{S}\hat{N}}{K+\hat{S}} = \frac{D\hat{N}}{c}$

Therefore, into equation 1,

$0 = DQ - D\hat{S} - \frac{D\hat{N}}{c}$

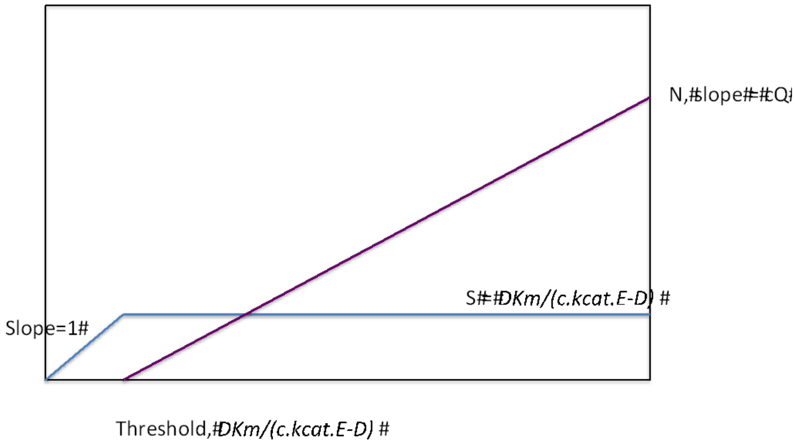
That is, $\hat{N} = c(Q - \hat{S})$
(Merit mark up to here)

Or $N = 0$ and $S = Q$ if

$Q > \frac{DK}{ck_e E - D}$ and $D < \frac{ck_e EQ}{K+Q}$
(Distinction level if they spot this as well)

The plot should look something like:

Continues on next page



y-axis= S and N at steady state, x-axis = Q

(iv) They could pick anything - adding a phage for example. Then add another equation, and a negative term reducing bacteria density (e.g. $-fNP$) and then a growth term for phage ($dP/dt = gfNP - DP$). Hypothesis could be that you would get predator prey cycles depending on the parameter values. Other features mentioned in the lectures were evolution of resource use ($dE/dt = mutationrate \times selectiongradient$, $d((1/N)dN/dT)/dE$) or competition or facilitation with another bacteria species. Looking for succinct explanation of feature, outline of how it would be added (distinction if provided with mathematical detail).

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%)

Model Answer (Marker – Tin-Yu Hui (1st), Burt (2nd)):

(i) See attachment

(ii) See attachment

(iii) See attachment

(iv) See attachment

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
```

Continues on next page

```
$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%)

Model Answer (Marker – Tin-Yu Hui (1st), Burt (2nd)):

(i) See attachment

(ii) See attachment

(iii) See attachment

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%)

Model Answer (Marker – Pawar (1st), Rosindell (2nd)):

(i) We're talking about the probability that a giraffe will breed, so this is a continuous r.v. limited to the interval [0,1]. The prior of choice is the Beta distribution, which has two parameters a and b.

(ii) These parameters can be interpreted as counts of "having seen a giraffe produce offspring in a year" and "b giraffe fail to produce offspring in a year" The expert's guess of the probability is the mean, $a/(a+b)$, and higher numbers a and b correspond to higher certainties (beliefs) in the correctness of this guess. Here, Betabuster might help in translating questions that make sense to an expert into numbers a and b.

(iii) When no expert is around, choose a reasonable value for the mean, say 0.5, but assign low values to a and b (also called an uniformative prior). Good values are $a = b = 1$, which yields the uniform distribution.

(iv) Under the present circumstances, an uniformative prior is to be preferred as the number of observations is bound to be small due to the rareness of the species, and we don't want the observations (which are objective) to be swamped by the prior (which is always somewhat subjective). It is best to profusively thank the expert, but quietly ignore his claims.

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.

Model Answer (Marker – Pawar (1st), Rosindell (2nd)):

```
# v[i] is Poissonian with mean mu[i], and we want to
# link mu[i] to a[i] and h[i] using a linear relationship

 for (i in 1:n) {
    v[i] ~ dpois(mu[i])
    mu[i] <- exp(a+b*a[i]+c*h[i]+delta[i]) # exp(): to keep mu positive
    delta[i] ~ dnorm(0,tau) # unexplained variance
 }
 a ~ dnorm(0,0.01) # uninformative priors for numbers that can be positive or ↩
    negative
 b ~ dnorm(0,0.01)
 c ~ dnorm(0,0.01)
 tau ~ dgamma(0.01,0.01) # uninformative prior for a precision, which is a positive↩
     number
```

End of paper