

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: Wednesday, 27th March 2017, 10:00 – 13:00

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. Most sections allow you to choose between questions, answering ONE. Please read the instructions at the head of each section carefully.

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.

Section 1: Maths

A. Solve **one of the following two problems** [30%]. Please indicate clearly in your answer book which question you are answering (i.e., A(i) or A(ii)).

(i) Solve the following integral:

$$\int \frac{dx}{x^2 - 4x + 3}$$

(ii) Obtain the order three expansion of the function $f(x) = \frac{\ln x}{x}$ at $x_0 = 1$.

B. Solve **one of the following two problems** [70%]. Please indicate clearly in your answer book which question you are answering (i.e., B(i) or B(ii)).

(i) Solve the differential equation:

$$y' - y/2 = 2 \sin(3t)$$

and explain if the general solution is finite when $t \rightarrow \infty$.

(ii) Diagonalize the following matrix A , and find the matrix P such that $D = P^{-1}AP$, being D the diagonal matrix. Once you obtain D and P , demonstrate that $A = PDP^{-1}$. Finally, find the value of A^5 .

$$A = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$$

Section 2: Dynamical Models in Ecology

Please select exactly **one question** and answer it. Please indicate clearly in your answer book which question you are answering.

- A.** Alice studies competition between two species. The two species' diets overlap to a large extent. The animals she studies search for their food, mainly nuts and seeds, and once they have found food items, they spend a considerable time opening and consuming these. She wants to implement a mathematical model for the interaction between the two species, and makes the following logical arguments/assumptions:

- The more individuals there are, the fewer food items there will be.
- Animals of species 1 are on average twice the size of species 2.
- She assumes that the density of food items (F) depends linearly on the density of species 1 (N_1) and species 2 (N_2) as

$$F = F_0 - 2N_1 - N_2$$

- The intake of food items follows a Holling type II functional response, and because of the size differences, the handling times (h_1, h_2) and search rates for food items (α_1, α_2) differ between the two species. Given the size differences, the growth rate (r_1, r_2) and mortality rates (m_1, m_2) are also different.

With these considerations, she writes down the following differential equation model for the changes in the densities of species 1 and 2:

$$\frac{dN_1}{dt} = r_1 N_1 \left(\frac{\alpha_1 F}{1 + \alpha_1 h_1 F} - m_1 \right)$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(\frac{\alpha_2 F}{1 + \alpha_2 h_2 F} - m_2 \right)$$

She wants to know what the outcome of the competition between these two species

- (i) Draw the isoclines for this model in the phase plane. Based on this, predict the outcome of the competition [25%]
 - (ii) Alice needs to explain the results of her model to an ecologist. Explain her results in ecological terms [30%]
 - (iii) The ecologist asks if the outcome of competition would have been different if she would have used different parameters. When is the outcome of the competition different? [30%]
 - (iv) The ecologist asks if the result would have been qualitatively different if she would not have used a different functional response than the Holling type II. Would it? [15%]
- B.** An ordinary differential equation (ODE) to describe the logistic growth of a population of fish, where at a constant rate h fish are hunted is

$$\frac{dx}{dt} = vx \left(1 - \frac{x}{x_{\max}} \right) - hx$$

where x is the population size, v is the growth rate and x_{\max} is the maximum population or carrying capacity in the absence of harvesting. Assume that $x \geq 0$ (i.e. that population cannot be negative) and $h \geq 0$.

- (i) Put this equation in non-dimensional form,

$$\frac{dz}{d\tau} = z(1 - z) - \phi z = (1 - z - \phi)z,$$

using $\tau = vt$ and $z = x/x_{\max}$, where you should find $\phi = h/v$. Interpret what this constant ϕ means. [15%]

- (ii) Determine the fixed points of this ODE for, (a) $\phi = 0$ (normal logistic growth), (b) $0 < \phi < 1$, & (c) $\phi = 0$. [15%]
- (iii) Sketch the phase portrait of this ODE for (a), (b), & (c), by plotting the RHS of above equation as function of z and indicating the stability of each fixed point by drawing appropriate arrows on the z -axis of the plot and using filled or open circles, respectively, for stable and unstable fixed points. [35%]
- (iv) For each of (a), (b) and (c), answer whether there is always a stable population independent of the initial conditions (where $z(\tau = 0) > 0$). [15%]
- (v) Using these results write an expression for the maximum stable population x^* as a function of the parameters h, v, x_{\max} . Describe qualitatively how the maximum stable population changes as h is increased. What is the critical value of h^\dagger , for which there is no longer a stable population. Interpret this biologically. [20%]

Section 3: Population Genetics & Evolutionary Ecology

Please select exactly **one question** and answer it. Please indicate clearly in your answer book which question you are answering.

- A.** You are in charge of a chemostat facility growing single-celled freshwater algae (*Chlorella* sp.) for an animal feed company. The algae grow in classic illuminated chemostats with a flow through of nutrient medium – the growth rate of the algae is limited solely by the concentration of nitrate in the inflow, every other nutrient, light and CO₂ is in excess. The system can be modeled by the following ODE system:

$$\begin{aligned}\frac{dN}{dt} &= \frac{ckNS}{m+S} - DN \\ \frac{dS}{dt} &= D(S_i - S) - \frac{kNS}{m+S}\end{aligned}$$

- (i) Explain each term in the model, taking care to define the parameters and give indicative units for each of them, including the variables. [25%]
- (ii) Algae are harvested by filtering them out of the out-flow from the chemostats. What is the steady-state rate of production of algal biomass? What would you alter in the system in order to increase the rate of production? [25%]
- (iii) You are thinking of increasing the temperature of the chemostats in order to speed up the production rate further. Sketch out the model could be extended to predict how the algae will respond to an increase in temperature. Use diagrams and/or equations where useful for your explanations [50%]

- B.** Answer the following:

- (i) Suppose you have written a program in R to simulate the change in allele frequency over time due to genetic drift. The simulator takes three input arguments: p_0 , the initial allele frequency of a particular allele; N , the effective population size; and t , the number of generations to be simulated forward in time. The simulator returns the allele frequencies over successive generations. Explain, in as much detail as possible, how you can use this simulator to express the probability of fixation of an allele before $t = 30$ generations, as a function of p_0 and N . [40%]
- (ii) (a) Suppose there is an infinite, random-mating population with 2 alleles, A and B. The frequency for the two alleles at generation t is p_t and q_t respectively, with $p_t + q_t = 1$. The three diploid genotypes AA, AB, and BB are at Hardy-Weinberg equilibrium. Furthermore, the AB heterozygotes produce A gametes in proportion d , and B gametes with proportion $1 - d$. Show that, in the next generation, the allele frequency for A is $p_{t+1} = p_t^2 + 2p_tq_td$. [30%]
- (b) Show that, if $d > 0.5$, then the allele frequency of A is strictly increasing (i.e. $p_{t+1} > p_t$), as long as A remains polymorphic. [30%]

Section 4: Maximum Likelihood & GLMs

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

- A.** (i) Let Y be a uniform random variable with lower bound a and upper bound b . The probability density function of Y is $f(Y) = \frac{1}{b-a}$.
- (a) Calculate $E[Y]$ [20%]
- (b) Calculate $E[Y^2]$. Then show that $Var[Y] = \frac{(b-a)^2}{12}$ [25%].
- Hint: $b^3 - a^3 = (b-a)(a^2 + ab + b^2)$
- (ii) Let X_1, X_2, \dots, X_n be i.i.d. Exponential random variables with rate λ . The probability density function of an Exponential random variable X is $f(X) = \lambda e^{-\lambda x}$.
- (a) Show that the MLE for λ is $\frac{n}{\sum_{i=1}^n x_i}$. [30%]
- (b) Suppose we have five observations, 0.28, 0.66, 0.56, 0.43, and 1.07. Compute the MLE for λ and its 95% confidence interval based on approximate normality. [25%]
- B.** Risking your sanity and life, you have spent nine years gathering data on Hungarian Hornbills. You have followed these animals throughout their lives, gathering repeated measurements on individuals. These creatures are vicious and can spew extremely long flames of fire from their bills. Not much is known about Hungarian Hornbills in general, and therefore your dataset is coveted by many. The Ministry of Magical Statistics (MoMs) has tasked you to find out whether the body size of the predicts the flame reach, if that would be possible many accidents could be avoided.

Your experience with hornbills allows you to make the a priori consideration of sex – because you know that the sexes have different flame reaches. Therefore, consider an interaction between sex and body size. You run your model in R, and get the following output:

```
library(lme4)
model<-lmer(FlameReach~BodySize*Sex+(1|Individual),data=a)
summary(model)
Linear mixed model fit by REML ['lmerMod']
Formula: FlameReach ~ BodySize * Sex + (1 | Individual)
Data: a

Random effects:
   Groups      Name      Variance Std.Dev.
Individual (Intercept) 0.5494    0.7412
Residual              0.2119    0.4603
Number of obs: 2000, groups:  Individual, 100

Fixed effects:
              Estimate Std. Error t value
(Intercept)   4.3178    0.7569   -5.705
BodySize       0.9255    0.1013    9.134
Sex            -0.2222    0.2158   -1.234
BodySize:Sex   -0.3193    0.1328   -2.404
```

Flamereach is in meters, **BodySize** was standardized to a mean of 0 and a standard deviation of 1, and females are coded as 0, males as 1. Answer each of the following questions.

- (i) Calculate the repeatability (in %) of the variable **FlameReach**. You do not have to give an indication of precision for the repeatability. Write out the general equation(s) of how to calculate repeatability, and a verbal explanation of each term. [30%]
- (ii) Write out the two model equations for the fixed part for both sexes. You do not have to add the random part. Simplify both equations as much as possible to quantify the slope for each sex. [30%]

- (iii) Write a results section dealing for your report to the MoMs, as you would a combined methods and results section in a paper. [40%]

Section 5: Bayesian statistics

This section has *one compulsory question worth 60% of the total mark divided into five points (i-v). The remaining 40% will be assessed based upon your submission of the practical given to you previously in class.*

You are hiking on the Pyrenees mountain range and have discovered an ancient human bone. You extracted the DNA and sequenced its genome. You were able to obtain the genotypic information for only one locus and found that the mysterious ancient human has an AA genotype (homozygous for Adenine). You want to make some statistical inferences on whether this sample is genetically closer to modern Spanish, French or Basque individuals.

Let's assume that your parameter of interest is $\theta = \{S, F, B\}$ representing the probability that your sample comes from a Spanish (S), French (F) or Basque (B) population, respectively.

The data are $y = \{g, f_A^{(S)}, f_A^{(F)}, f_A^{(B)}\}$ where g is the ancient genotype (so that $g = AA$) and $f_A^{(S)}, f_A^{(F)}, f_A^{(B)}$ are the known population frequencies of allele A in modern Spanish, French and Basque, respectively. Under the assumption of Hardy Weinberg Equilibrium, we know that

$$p(g = AA, f_A^{(i)} | \theta = i) = (f_A^{(i)})^2$$

for a generic population i . Note the latter equation represents the likelihood function $f(y|\theta)$.

- (i) Using Bayes' law, write the equation for the posterior probability of the sample belonging to the Spanish population given the data. Assume that you have a generic prior probability $\pi(\theta)$ with known hyperparameters. Be as formal and explicit as possible. No proofs or extra calculations are required. [10%]
- (ii) Let's assume that we gather the follow population allele frequencies

$$\begin{aligned}f_A^{(S)} &= 0.7 \\f_A^{(F)} &= 0.2 \\f_A^{(B)} &= 0.1\end{aligned}$$

and that we ask for an opinion to experts regarding a prior probability of this sample belonging to any of the tested populations. Here the opinion from our 3 experts:

- (a) Dr Cobain: It is still highly debated whether ancient humans in the Pyrenees are the ancestors of modern French, Spanish or Basque. All anthropological evidences so far are not solid enough to point towards any specific population. We have no clue!
- (b) Professor Grohl FRS: It must be Spanish. I see no evidence why this sample should be the ancestor of any other modern population. I am 100% sure.
- (c) Mr Novoselic: We have collected more than 1,000 ancient samples from Pyrenees so far and we were able to assign 50% of them as Spanish, 30% as French, and 20% as Basque.

Based on this information, choose the most suitable prior distribution $\pi(\theta)$. Justify your choice. There is no right or wrong answer (although one of them is hardly acceptable) as long as it is properly justified. Formalise $\pi(\theta)$ by assigning a prior probability for each value of θ based on your choice. [10%]

- (iii) Based on your chosen prior distribution, calculate the Bayes factor for model M_1 with parameter $\{\theta = S\}$ vs. model M_2 with parameter $\{\theta \neq S\}$. Write the equation and provide the value for the Bayes factor. Approximate any calculation as much as you wish but be reasonable (e.g. $0.82/0.19 \approx 4$ is totally fine but $0.82/0.19 \approx 3$ is not). You will not be penalised for minor algebra mistakes. Provide a brief discussion on the support for M_1 or M_2 . [30%]

- (iv) Let's assume that a new prior probability on θ is now dependent on an unknown hyperparameter τ with distribution $h(\tau)$ with $\tau = \{-1, 0, 1\}$. In other words, the prior distribution is $\pi(\theta|\tau)$ and τ can only have the discrete values of -1 , 0 , or 1 . Using Bayes' law in hierarchical modelling, write the equation for the posterior distribution of the sample being Spanish. Be as formal and explicit as possible. No proofs or extra calculations are required. Note that both θ and τ are discrete distributions. [10%]