<div align="center">

IMPERIAL COLLEGE LONDON

MSc COURSE IN COMPUTATIONAL METHODS IN ECOLOGY AND EVOLUTION

EXAM 2

</div>

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

Exam Date: Wednesday, 23rd March 2016, 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. All sections allow you to choose between two questions, answering one. Read instructions carefully at the head of each section.

<div align="center">

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

</div>

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 II

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

A. Preston (1962) and MacArthur and Wilson (1963) investigated the effect of area on species diversity in oceanic islands. It is assumed that species can immigrate to an island from a species pool of size P and that species on the island can go extinct. We denote the immigration rate by $I(S)$ and the extinction rate by $E(S)$, where S is the number of species on the island. Then the change in species diversity over time is

$$\frac{dS}{dt} = I(S) - E(S) \tag{1}$$

For a fixed island, the simplest functional forms for $I(S)$ and $E(S)$ are

$$I(S) = c(1 - \frac{S}{P}) \tag{2}$$

$$E(S) = \frac{mS}{P} \tag{3}$$

where c, m and P are positive constants.

  (i) Find the equilibrium species diversity $\hat{S}$ of eqn 1 with $I(S)$ and $E(S)$ given in 2 and 3. [20%]

 (ii) It is reasonable to assume that the extinction rate is a decreasing function of island size. That is, we assume that if A denotes the area of the island, then m is a function of A with $\frac{dm}{dA} < 0$. Furthermore, we assume that the immigration rate I does not depend on A. Use these assumptions to investigate how the equilibrium species diversity changes with island size. [20%]

(iii) Assume that $S(0) = S_0$. Solve eqn 1 with $I(S)$ and $E(S)$ given in 2 and 3. [40%]

 (iv) Assume that $S(0) = 0$. That is, the island is initially devoid of species. One then defines the time constant T as the time required for S to attain 63.2% of its equilibrium value i.e. it is the solution of

$$S(T) = (1 - e^{-1})\hat{S}$$

Find an explicit expression for T in terms of P , c, and m. [20%]

B. You are planning to conduct a field study to determine the relative abundance of a certain variety of a given species, i.e. the proportion with which this variety is present among the total population. How large a sample size would you need to estimate this proportion within 0.01 of the true value with probability at least 0.95?

Hint: Use the central limit theorem. You may need some of the following Φ values (the cumulative distribution function of the standard normal distribution):

$\Phi(1.28) = 0.9$
$\Phi(1.44) = 0.925$
$\Phi(1.65) = 0.95$
$\Phi(1.96) = 0.975$

Section 2: Dynamical Models in Ecology and Evolution

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

A. Sheep were first introduced in Tasmania in 1810. Over the period 1814–1864 their numbers increased as follows:

Year	No. of sheep (thousands)
1814	125
1824	275
1834	830
1844	1200
1854	1750
1864	1650

Based upon these data, answer the following:

(i) How could you evaluate whether or not the growth of the sheep population is density dependent? Outline/explain the possible methods. [40%]

(ii) Based on the information in the table, discuss whether the growth of the sheep population is likely to be density dependent or not. [30%]

(iii) Discuss what mechanism could have caused the pattern of population growth. [30%]

B. Historical measles data often show a biannual pattern, with measles outbreaks in every second year. Between 1928 and 1935 triennial cycles were observed in the incidence of measles in Baltimore.

(i) Explain how triennial cycles can occur in measles epidemics. [40%]

(ii) Discuss how the periodicity of measles varies with the reproductive number (R_0) and the strength of forcing. [40%]

(iii) Based on the occurrence of triennial cycles, what can you say about the value of the reproductive number (R_0) and the strength of forcing in Baltimore in the 1930s? [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. Answer the following:

   (i) Drive is one of five processes affecting allele frequencies within a population. What are the other four? [10%]

   (ii) Construct a model of allele frequency changes due to drive along the following lines. Suppose there is an infinite, random-mating population with 2 alleles, A and B, with frequencies p and q, respectively, with the diploid genotypes at Hardy-Weinberg frequencies. Suppose in AB heterozygotes the B allele is transmitted to a proportion d of gametes ($d = 0.5$ being Mendelian inheritance with no drive). What is the expected frequency of the alleles in the next generation? If $p = 0.8$, $q = 0.2$, and $d = 0.95$, what are the expected frequencies in the next generation? What is the long-term equilibrium frequency of the two alleles (i.e., frequencies to which the population will tend many generations in the future)? [35%]

   (iii) Suppose, in addition, BB homozygotes are embryonic lethal. Find an expression for the change in the frequency of the B allele from one generation to the next. Find the equilibrium frequency. What fraction of the population are BB homozygotes and die at this equilibrium? [35%]

   (iv) Explain how the process of drive might be used for public health. [20%]

B. Messrs Barraclough & Co, the famous red wine manufacturers, have invented a new process for producing extra strong claret. By genetically engineering a super' strain of yeast, they pump grape juice through a continuous flow-through system, and produce an extremely high alcohol wine for the mass market.

   Assume that the dynamics of conversion from glucose to ethanol and of the growth of the yeast in the system are specified by the following equations:

$$\frac{dS_1}{dt} = D\left(Q_1 - S_1\right) - \frac{aS_1}{(b + S_1)}N$$
$$\frac{dS_2}{dt} = \frac{aS_1}{(b + S_1)}N - DS_2$$
$$\frac{dN}{dt} = c\frac{aS1}{(b + S_1)}N - DN$$

where,
S_1 is the concentration of glucose (moles per litre) in the flow-through system
Q_1 is the concentration of glucose in the input grape juice
S_2 is the concentration of ethanol (moles per litre)
N is the density of yeast in cells per litre
D is the rate of flow through of juice (litres per second)
a is a rate parameter (sometimes called V_{max} in the microbial literature)
b is the Michaelis-Menten constant
c is the number of cells produced per mole of glucose metabolised per litre

   (i) Explain each component term in the three equations in words and/or graphically. [10%]

   (ii) What are the units of the constants a and b? [10%]

   (iii) What is the concentration of alcohol that is produced at steady-state? How would you tweak the parameters to increase the concentration? [40%]

   (iv) Imagine that parameters a, b, and c are now variables that can evolve in the yeast. Describe how you would approach including evolution into the model, and make verbal predictions for how the yeast population would evolve. [40%]

Section 4: Maximum Likelihood

Please select exactly **one question** and answer it. Calculator may be required in some questions. 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) Let X be a binomial random variable with probability mass function

$$f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

    Show that $\sum_{k=0}^{n} f_X(k) = 1$. [30%]

    (ii) In 30 independent coin tosses, 21 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.

    Please perform a likelihood ratio test for $H_0 : p = 0.5$ vs $H_1 : p \neq 0.5$ at 5% significance level. [40%]

    (iii) Describe, as precisely as possible, that how you can obtain 95% confidence interval (and joint confidence region for **two parameters**) from the log-likelihood function. You may use appropriate equations or graphs to support your answer. [30%]

B. Answer the following:

    (i) Let $X \sim N(\mu, \sigma^2)$ and the associated moment generating function is $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Show that $E[X] = \mu$ and $E[X^2] = \mu^2 + \sigma^2$. [40%]

    (ii) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed samples from $Poisson(\lambda)$. Given the probability mass function of an Poisson random variable is $f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, please find the MLE for λ. [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 her R console:

```
> optim(par=c(100,0.5), fn=log.like, method= L-BFGS-B , lower=c(1,0),
upper=c(9999,1), dat=dat, control=list(fnscale=-1), hessian=T
```

    Please describe, as precisely as possible, *each component of the input and output screen.* [30%]

Continues on next page

Section 5: GLMs & Bayesian stats

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

A. A colleague has conducted an observational study focused on a species of beetle present on all twenty islands of an archipelago. On each island, she counted the number of individuals she saw in one hour of observation, and repeated this at ten different spots on the island. She also measured the distance to the nearest food resource, in meters, for each spot where she had counted beetles. She then determined whether a major predator of the beetle was present or absent on each island.

```
> str(beetle_counts)
'data.frame':   200 obs. of  4 variables:
 $ predator   : Factor w/ 2 levels "NO","YES": 1 1 1 1 1 1 1 1 1 1 ...
 $ island     : Factor w/ 20 levels "A","B","C","D",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ distance   : num  14.26 6.49 7.97 5.14 7.41 ...
 $ individuals: int  27 28 30 29 25 17 22 23 25 28 ...
```

She hypothesizes that there will be fewer beetles farther away from food resources in the absence of predators. She also hypothesizes that this effect of distance from food in reducing the number of beetles observed will be even stronger when predators are present on the island. She has fit a generalized linear model in order to test this:

```
> mymodel<-glm(individuals ~ predator * distance, family=poisson(link=log), data=beetle_counts)
> summary(mymodel)

Call:
glm(formula = individuals ~ predator * distance, family = poisson(link = log),
    data = beetle_counts)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-7.3047  -2.0802  -0.2908   2.1894   6.7238

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         4.655983   0.030484 152.737  < 2e-16 ***
predatorYES        -0.142953   0.040329  -3.545 0.000393 ***
distance           -0.109915   0.004181 -26.291  < 2e-16 ***
predatorYES:distance 0.053042  0.004683  11.327  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 3221.6  on 199  degrees of freedom
Residual deviance: 1633.3  on 196  degrees of freedom
AIC: 2746.4
```
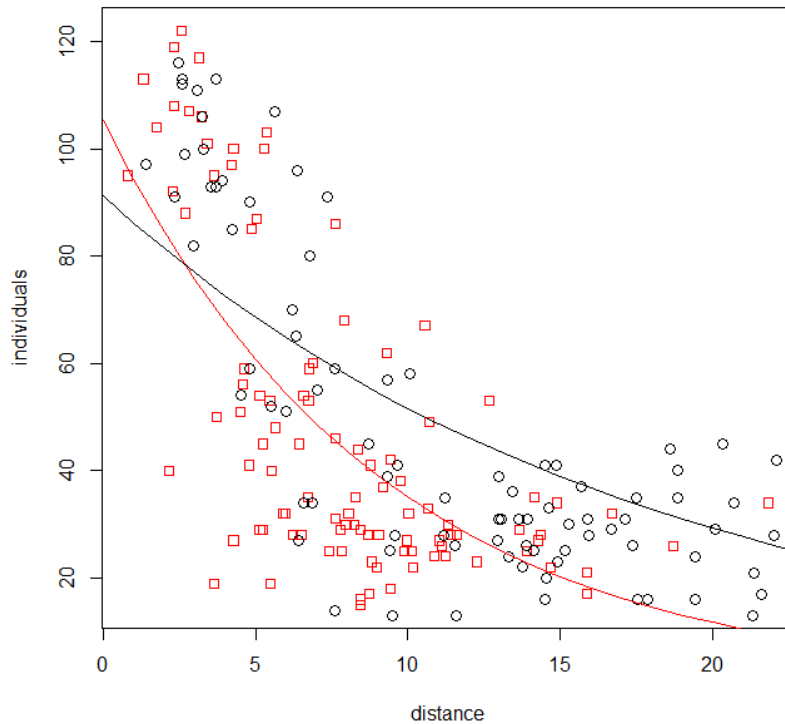
The plot below shows the raw data along with predicted values from the model across a range of distance from food source:

(i) Based on the model, show mathematically how you would calculate the predicted value for an island with predators, at a site 5 meters from a food resource. [30%]

(ii) Which color of line represents predictions for "predator present" and which color represents "predator absent"? [10%]

(iii) If we decided to trust this model, what can you say about your colleague's hypotheses? (30%)

(iv) The model shows symptoms of overdispersionhow can you tell this? Instead of using a "quasipoisson" error distribution to deal with the issue, what other (and better) approach could you take that incorporates an aspect of the sampling neglected by your colleague? Explain in detail why you would use this approach and also how you would do this, including either R code or pseudocode. [30%]

B. Answer the following

(i) State Bayes' theorem, naming all the symbols you use, and describe, in general terms, how Bayes' theorem can be used to perform Bayesian inference. [30%]

(ii) Explain the difference between an informative and a non-informative prior. Give an example for each. [20%]

(iii) In 2010 Ott and Rogers found that "Gregarious desert locusts have substantially larger brains compared with the solitarious phase". Given a set of brain size data for each of the two types of locusts, how would you use Bayesian inference to substantiate the authors' claim? [50%]

Your answer should include:

i. A statistical model suitable for coding in JAGS, including suitable priors; motivate your choice!

ii. The variable(s) supplied as data

iii. The variable(s) you would monitor in JAGS, and

iv. How you would interpret the results so as to validate the claim (or not). Don't worry too much about JAGS syntax, nor the actual values of the parameters of the priors you use.

End of paper