

IMPERIAL COLLEGE LONDON

MSc COURSE IN COMPUTATIONAL METHODS IN ECOLOGY AND EVOLUTION

EXAM 1

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

Exam Date: Friday, 15th Jan 2016, 1000 – 1300

Length of Exam: 3 HOURS

Instructions: All sections are weighted equally. It is a three-hour exam, and there are 4 sections, so it is a reasonable guideline to spend about 45 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.

Section 1: GIS and Genomics

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

- A.** How does deforestation detection from space work, and is global near-real time monitoring of deforestation at high-pixel resolution feasible?

Model Answer (Marker – Rob Ewers (1st), Andrew Bradley (2nd)):

Need to map where forest is first. Then create time series of maps to map and quantify area change in forest cover over time. Think about uncertainty when mapping forests using spectral data and also think about time steps: does it make sense to map annual forest change?

Mapping: Supervised methods (e.g. Maximum Likelihood or Random Forest) requires field data, i.e. collection of ground truth points with known land cover. These are then used as training and test data to link satellite data to land cover data. Mapping: Unsupervised classification (e.g. Isodata clustering). Group pixel according to their spectral similarities using clustering algorithms. You can subsequently link this to land cover data collected from the ground if you wanted to.

You can achieve better results with supervised classification, but not necessarily so. This depends on the landscape. Think about landscapes with clear habitat categories (e.g. unlogged forests versus grassland - easy to cluster as spectrally very different) and then think about landscapes with more transitional boundaries (e.g. logged vs primary forest).

Near real time monitoring: Trade-offs between spatial, spectral and temporal resolution. The spatial resolution is often the key. You can have daily monitoring of land surfaces globally at moderate to coarse pixel resolution (MODIS, AVHRR). You can have daily monitoring of selected small regions at high spatial resolution (targeted image acquisition using SPOT or similar). Then you also have clouds. Mapping in the tropics has a problem with cloud cover that obscures the land surface – optical imagery can't see through the clouds and so the maps can't be created. Clouds are obviously less prevalent in semi-arid zones and over dry forests.

Temporal frequency is particularly limited if using supervised classifications, but you could look at patterns in spectral changes only using unsupervised methods (no need to map on the ground every week) - does it become more similar to grassland (e.g. conversion forest to grassland)? Computing problem increases as you use more bands - usually there are more bands with very narrow band widths in the hyperspectral sensors (= spectral resolution). And you will also encounter problems of increasing inter-correlation between bands. However, more bands gives more chances of separating different land cover types.

- B.** You discover a population of armadillo in the Eastern Congo forest with a distinctive leopard-spotted pattern. The leopard-spotted armadillos are only found to the east of the Elila River, while those to the west are the normal tawny colour.

Be specific and provide justifications in your answers to each of the following questions:

- (i) What genetic, and comparative data would you collect to test whether the leopard-spotted armadillo was differentiated from the tawny armadillo populations? What analyses would you perform to test this? How would you determine whether the Elila River serves as a barrier to gene flow? If there has been gene flow, what would you expect to see? (40%)
- (ii) After performing these analyses you determine that the leopard-spotted armadillos are genetically structured from the tawny armadillos, however there is a tawny population with significant admixture from the leopard-spotted armadillos on the west bank of a shallow stretch of the river. Interestingly, despite the admixture, this population never seems to display the leopard-spotted phenotype. Through some additional analyses you learn that the leopard-spotted phenotype is controlled by the LSPT gene.

How would you test whether there has been selection at the LSPT locus in the leopard-spotted aardvarks? What genomic signal would you expect to see if there has been selection? How would you test whether there has been selection for the LSPT locus in the admixed population? What would you expect to see here if so? (60 %)

Model Answer (Markers – Jason Hodgson (1st), Vincent Savolainen (2nd)):

- (i) Genetic data: Genomic SNP data via RAD-seq

Comparative data: As many leopard-spotted and normal aardvark samples as possible from both sides of the river. Minimum sample sizes should be on the order of 20 animals from each putative population.

Analyses: to assess genetic structure MDS and ADMIXTURE. These analyses assume nothing about the data, and only consider allelic state and associations between markers to test for genetic substructure within the data.

If the river is a barrier to gene flow the populations samples on one side of the river will be more closely related than adjacent populations on opposite sides of the river. If there has been gene flow, you would either see no differentiation on either sides of the river, or if there has been differentiation you would expect to see evidence of admixture. For example, in the ADMIXTURE analysis you would see individuals showing ancestry from both parent populations.

- (ii) To test for selection at LSPT within the leopard-spotted aardvarks you would test for high levels of linkage disequilibrium around LSPT relative to the rest of the genome. If there has been selection LD around the selected site will be increased.

To test for selection in the admixed population you would have to test whether the frequency of leopard-spotted LSPT is increased or decreased relative to that expected given the amount of admixture measured.

Section 2: Statistics and Model Fitting

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

- A. Chris measured the abundance of one bacterial species – an *Epilithonimonas* – after 6 days in communities made up of different numbers of species: 1 (monoculture), 2, 3, 4. In order to investigate the effects of spatial structure on diversity, he grew each community in two treatments: shaken and unshaken. He hypothesized that the abundance of this species would be lower in more species-rich communities because of greater competition, but higher in unshaken cultures because they allow species to occupy distinct spatial niches.

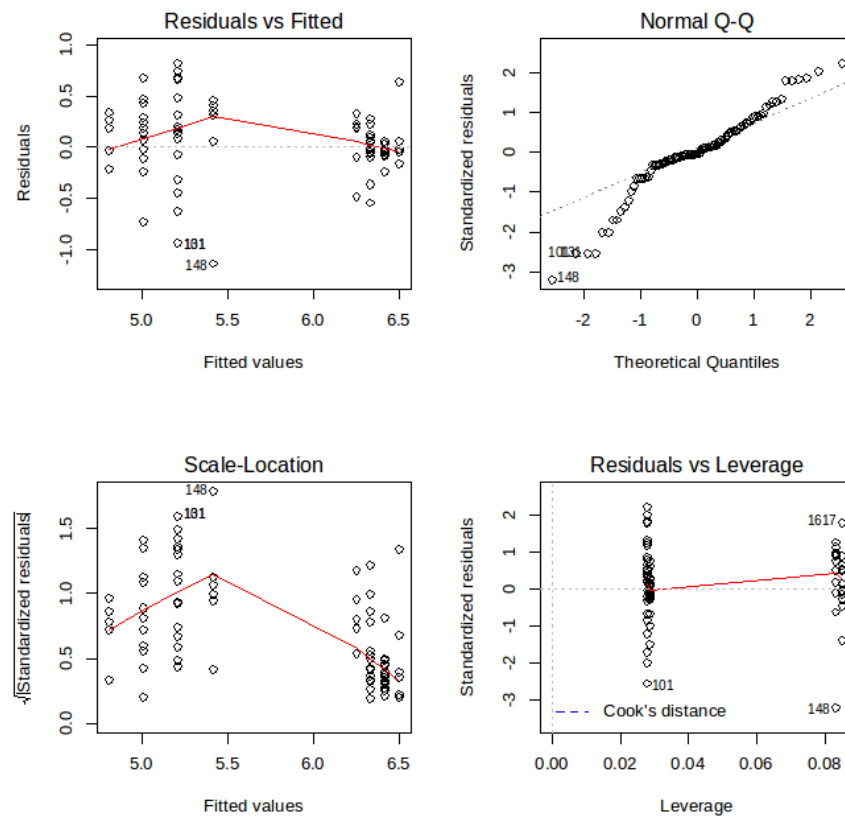
The following output shows the summary of the linear model of log abundance of *Epilithonimonas* against treatment (shaken, unshaken) and richness, and the diagnostic plots.

```
Call:
lm(formula = log.epilithonimonas ~ treatment * richness, data = Epilithonimonas,
    subset = time == 6)

Residuals:
    Min       1Q   Median       3Q      Max
-1.13513 -0.11008 -0.01354  0.19128  0.81248

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.60802    0.16366   28.156 < 2e-16 ***
treatmentunshaken  1.97433    0.23244    8.494 3.88e-13 ***
richness          0.20147    0.06186    3.257 0.00159 **
treatmentunshaken:richness -0.28397    0.08779   -3.235 0.00170 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3711 on 90 degrees of freedom
Multiple R-squared:  0.76, Adjusted R-squared:  0.752
F-statistic: 94.99 on 3 and 90 DF, p-value: < 2.2e-16
```



What do the results tell you? Were the hypotheses supported? Can you see any problems with the analysis? If so, what would you change?

Model Answer (Markers – Tim Barraclough (first), Brian Hollis (second)):

- There is a significant interaction term. The abundance of the species increases with richness in the (default) shaken treatment. The interaction term is negative. They might say this shows a negative relationship in the unshaken treatment, but Distinction level will notice that the slope in the unshaken treatment is -0.08 (the difference between the last two coefficients), but the standard error is marginally greater, so in fact – no significant relationship in the unshaken.
- They might try to interpret the intercept and treatmentunshaken terms as well. If so, tell you that at low richness (technically richness 0 which is not possible), higher abundance in unshaken treatment than in shaken treatment.
- So both hypotheses potentially supported, although richness effect only found in the shaken treatment, which (distinction level) is consistent with spatial niches protecting from competition.
- Might also say that ~75% variation explained by richness and treatment.
- Regarding problems, left-hand plots look a little wonky – higher variance at intermediate values, probably not too much to worry about. QQ plot shows some evidence of excess of extreme residuals. None of the points have undue leverage.
- Could try other transformations, but already logged response variable. Dont say remove outliers!

B. Someone gave you some data on morphometrics in Odonates (Suborder Anisoptera = Dragonflies; Suborder Zygoptera = Damselflies). Having eagerly imported the data into R, you use `str(MyData)` to find:

```
'data.frame': 100 obs. of 16 variables:
 $ Suborder : Factor w/ 2 levels "Anisoptera","Zygoptera": 1 1 1 1 1 1 1 1 1 1↵
 ...
```

```

$ BodyWeight      : num  0.159 0.228 0.312 0.218 0.207 0.22 0.344 0.128 0.392 0.029 ↵
...
$ TotalLength     : num  67.6 72 78.8 72.4 73 ...
$ HeadLength      : num  6.83 6.84 6.27 6.62 4.92 6.48 7.53 5.74 8.05 5.28 ...
$ ThoraxLength    : num  11.8 10.7 16.2 12.5 11.1 ...
$ AbdomenLength   : num  48.9 54.4 56.3 53.3 57 ...
$ ForewingLength  : num  45.5 46 51.2 49.8 46.5 ...
$ HindwingLength  : num  45.4 45.5 49.5 48.8 46 ...
$ ForewingArea    : num  370 411 461 469 382 ...
$ HindwingArea    : num  484 517 574 591 481 ...

```

Now, you would like to build a statistical model to predict Odonate body weights from their body lengths. Figure 1 shows what relevant subsets of the data look like.

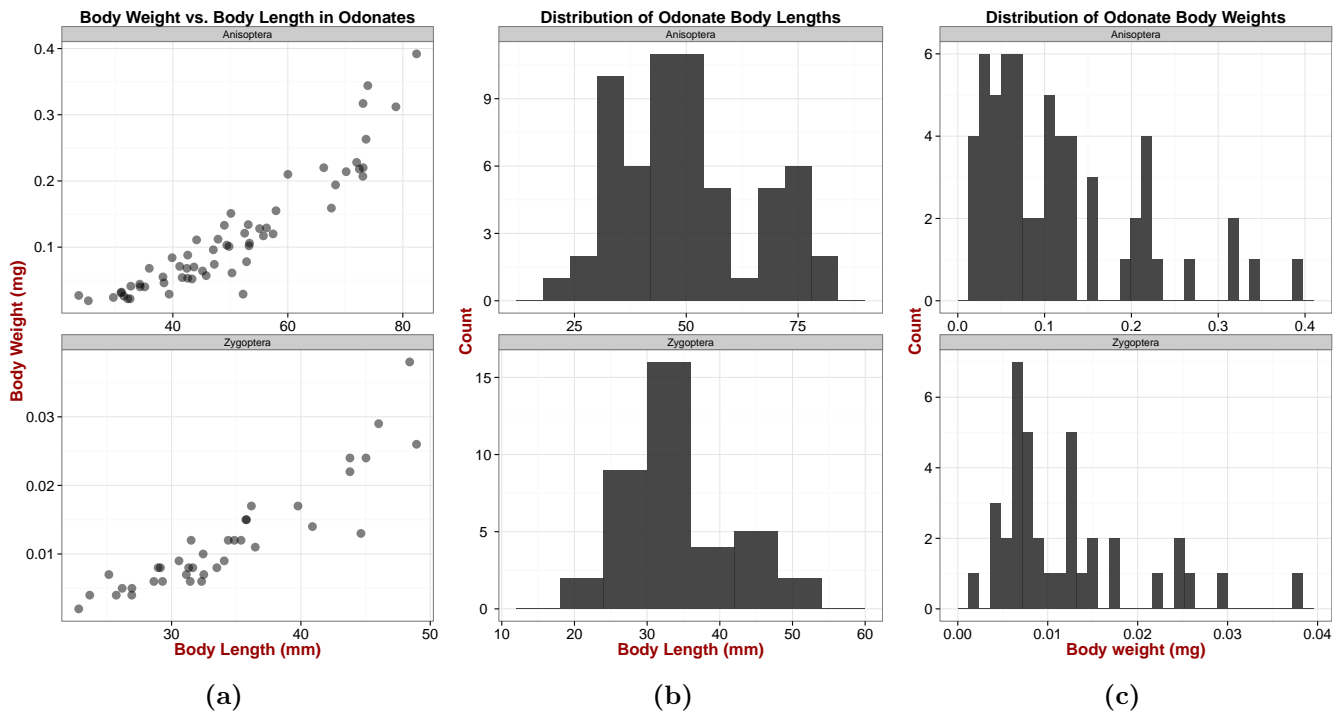


Figure 1: Body weight vs. length in Odonates: (a) Scatterplot; (b) Histograms of body lengths; and (c) Histograms of body weights.

You then perform some analyses in R, and obtain the following output with `summary(LmDragonLin)`:

```

Call:
lm(formula = BodyWeight ~ TotalLength, data = subset(MyData,
  Suborder == "Anisoptera"))

Residuals:
    Min       1Q   Median       3Q      Max
-0.098639 -0.022063 -0.003622  0.020156  0.098628

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.1592206   0.0160166  -9.941 3.86e-14 ***
TotalLength  0.0054933   0.0003081  17.829 < 2e-16 ***
---
Signif. codes:  0   ***    0.001   **    0.01   *    0.05   .    0.1    1

Residual standard error: 0.03496 on 58 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared:  0.8457,    Adjusted R-squared:  0.843
F-statistic: 317.9 on 1 and 58 DF,  p-value: < 2.2e-16

```

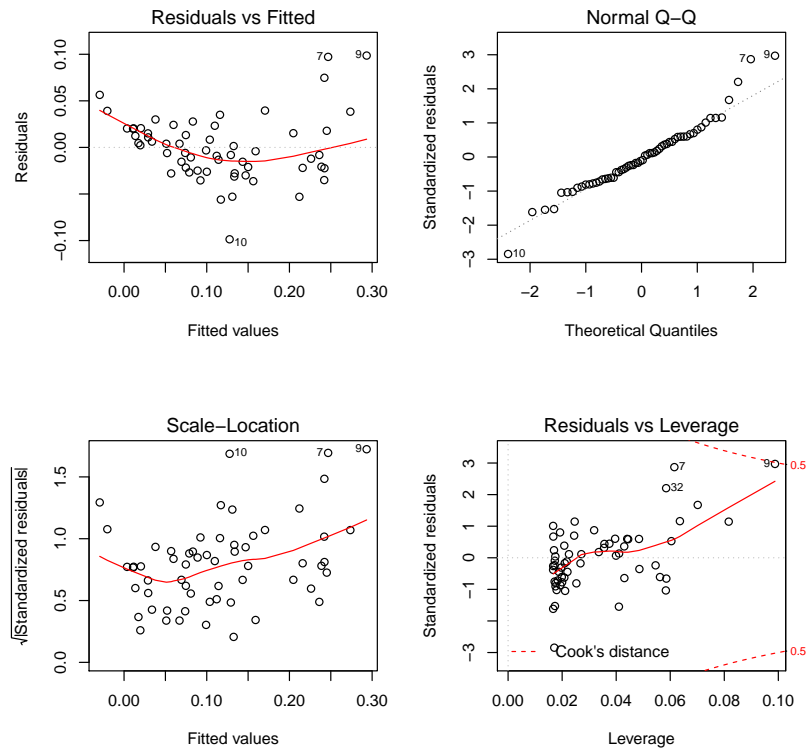


Figure 2: The result of `plot(LmDragonLin)`

Similarly, you get the following output with `summary(LmDamsellin)`:

```
Call:
lm(formula = BodyWeight ~ TotalLength, data = subset(MyData,
  Suborder == "Zygoptera"))

Residuals:
    Min       1Q   Median       3Q      Max
-0.0097445 -0.0012688 -0.0001252  0.0014485  0.0113548

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.344e-02  2.777e-03  -8.443 4.65e-10 ***
TotalLength  1.035e-03  7.981e-05  12.964 3.96e-15 ***
---
Signif. codes:  0   ***    0.001   **    0.01   *    0.05   .    0.1    1

Residual standard error: 0.003381 on 36 degrees of freedom
Multiple R-squared:  0.8236,    Adjusted R-squared:  0.8187
F-statistic: 168.1 on 1 and 36 DF,  p-value: 3.965e-15
```

Furthermore, `par(mfrow=c(2,2)); plot(LmDragonLin)` gives you the plots shown in Figure 2 for the dragonfly subset of the data.

Given this information, please answer the following (each question equally weighted):

- Why might you expect a relationship between body weight and length (see Figure 1)? Describe the statistical analysis that was performed. Make sure you state what model was fitted, and write out the mathematical equation for the model, explaining what each of the parameters in the equation/model is. Also write out the representation of the model's equation in R syntax.
- State the assumptions that the data need to satisfy to allow you to fit the model to these data, and which of assumptions do and do not the data satisfy (how can you tell?). Thus, do you see

any problems with using the model that was chosen with these data?

- (iii) Based upon the R outputs above, can you conclude that weight increases more rapidly with length in Dragonflies vs. Damselflies? Why or why not?
- (iv) What would be a more appropriate model than the one that was fitted above? Show it's equation and state what its parameters are.

Model Answer (Marker – Samraat Pawar (1st), Brian Hollis (2nd)):

- (i) For full points, the student should be able to point out that:
 - Volume increases as a cube power of length, so one would expect a positive relationship between body weight and length: $\text{Body Weight} \propto (\text{Body Length})^3$
 - The analysis was an ordinary least squares regression of body weight vs. Body length applied on dragonfly and damselfly subsets of the data. The model is a linear regression model.
 - The models equation is $\text{BodyWeight} = c + m * \text{TotalLength} + \epsilon$ (or an equivalent expression), where c is the intercept, m is the slope, and ϵ are normally distributed errors.
 - R representation of the model is `BodyWeight ~ TotalLength`
- (ii) For full marks, the student should point out that:
 - Assumptions are: (i) Body length and body weight are both normally distributed, (ii) there is homogeneity of variance in body weight over the entire body length data range, and (iii) the sampling error is mostly in body weight (dependent variable) and not body length (independent variable).
 - Assumption (i) is not strictly met, as both Figure 1 and 2 show (student should point out what elements of the figures exactly).
 - Assumption (ii) too is questionable in these data, variance in body weight seems to increase at upper body lengths (student should point out what elements of figure 1 show this exactly).
 - Assumption (iii) is probably not going to be met in such datasets as both length and weight are subject to measurement error.
 - So lots of problems with using this model. One problem, arguably the most important one, is that the data deviate systematically from a straight line equation, as indicated by the residuals. Some exceptional answers might also point out that the model diagnostics plots indicate inordinately high leverage for certain data points at upper ends of Body length...
- (iii) For full marks, the student should point out that:
 - It is hard to make this conclusion definitively without a proper statistical analysis of differences in coefficients.
 - For example, the confidence intervals of the slope coefficient should not overlap (equivalent to doing a t-test for differences in slopes).
 - Nevertheless, the R output does suggest that dragonflies have a steeper increase in weight with length (slope coefficient in R output is almost five-fold higher: 0.0054933 vs 0.001035)
- (iv) A linear regression model of the form below or an equivalent expression. Excellent answers might describe ϵ in more detail (log-normally distributed).

$$\log(\text{BodyWeight}) = \log(c) + m \times \log(\text{TotalLength}) + \epsilon$$

This model comes from taking a log of the equation $\text{Body Weight} \propto (\text{Body Length})^3$, and is one way to look for a power law (allometry).

An outstanding answer might also mention NonLinear Least Squares Regression as an alternative to LM.

Section 3: HPC & fractals

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

A. Answer the following questions. Please be brief in your answers to these questions – bullet points are OK as you are being marked on content:

- (i) What is an embarrassingly parallel problem? (10%)
- (ii) Give an example of a simulation task that is embarrassingly parallel. (10%)
- (iii) Give an example of a simulation task that is NOT embarrassingly parallel. (10%)
- (iv) What is a pseudo random number generator and a random seed? (20%)
- (v) The following question concerns an individual based simulation of organisms living in a community. An R function `stochastic_sim` accepts the community as an input and returns the community one generation later according to some unknown stochastic model. You expect that the species richness in your community will converge to a dynamic equilibrium if you repeatedly apply the function `stochastic_sim` and you wish to find out the mean value of the species richness at this equilibrium. Describe briefly how would you go about solving this problem. Please discuss random number generation, burn in and use of high performance computing in your answer. (50%)

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

- (i) A simulation problem that can easily be split up into a number of tasks that can be run in parallel
OR

A simulation problem that can be run as a number of independent parallel processes that do not need to exchange information.

- (ii) Any sensible answer, e.g., one of the following:
 - Computer graphics problems.
 - Stochastic simulations where large numbers of replicates are needed.
 - Numerical simulations where a wide area of parameter space needs to be explored.
- (iii) Any sensible answer e.g. one of the following
 - Fluid dynamics simulations.
 - Simulations on a large landscape with interaction between different habitat patches.
- (iv) A pseudo random number generator is the outcome of a deterministic process that produces a set of apparently random numbers. (half of the full credit)

The random seed is a value used to initialise a pseudo random number generator. Using the same seed again will give you exactly the same sequence of pseudo random numbers. (other half of the full credit)
- (v) Points that should be put forward in the answer:
 - Running a large number of simulations (with different initial conditions) on HPC in parallel will help get a better idea of the expected mean at equilibrium (1 mark)
 - Parallel simulations should not be conducted with the same random seed. (1 mark)
 - Burn in is the time it takes for the system to reach equilibrium from an initial condition. (1 mark)

- Data collected during burn in should be discarded. Data collected after burn in should be averaged to produce the desired answer. (1 mark)
- It can be troublesome to determine how long the burn in should be. Time-series data should be outputted from each simulation and analysed later, plotting a graph of species richness vs. time to suggest when the burn in period is over. (1 mark)

If the total score from the above is less than (5 marks), one or two bonus points can be given at the marker's discretion (up to a maximum score of 5) if the candidate demonstrates a sound knowledge of HPC in his/her answer bringing up valid and relevant points about the practicalities of HPC use that are not mentioned above.

B. Answer the following:

- (i) You have collected some fern leaves on a field trip and you wish to calculate their fractal dimensions. Describe briefly how you could do this using the box counting method. (40%)



- (ii) What is the fractal dimension of the object constructed in stages below (please explain your workings)? (30%)



- (iii) The following R code function generates a fractal along a line of a certain **width** centred on position **x** provided that the **threshold** is small.

```
linedraw <- function(x,width,threshold){
  if (width>threshold){
    linedraw(x-width/3,width/3,threshold)
    linedraw(x+width/3,width/3,threshold)
  } else {
    lines(c(x-width/2,x+width/2),c(0,0))
  }
}
```

- (a) What will happen if you call `linedraw(x=0,width=27,threshold=27)`? (10%)
- (b) Describe the fractal object that appears as you reduce the threshold, e.g. by calling `linedraw(x=0,width=27,threshold=1)`. You can do this with a series of drawings corresponding to decreasing thresholds. (20%)

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

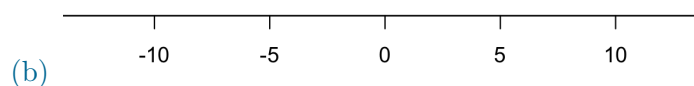
- (i) The answer should explain the following steps:
- Place a grid of a certain width over the fern and count the number of boxes that intersect the fern. (1 mark)

- Repeat the process for grids with a variety of smaller widths. (1 mark)
- Plot a graph showing the log of the number of squares needed to cover the fern on the y axis against the log of the length of each square on the x axis. (1 mark)
- A fractal fern should produce a straight line graph with a gradient equal to $-D$ where D is the fractal dimension. (1 mark)

(ii) The answer should comprise:

- To triple the width requires 5 copies of the original (1 mark)
- So $5 = 3^D$ where D is the dimension of the object (1 mark)
- Therefore $\log(5) = D * \log(3)$ and $D = \log(5)/\log(3) = 1.465$ (1 mark)

(iii) (a) You would just get a straight line of length 27 centred around 0 (1 mark)



1 mark for one correct iteration, 2 marks for the limiting cases

OR

To triple the width requires 2 copies of the original (1 mark).

So $2 = 3^D$ where D is the dimension of the object

Therefore $\log(2) = D \log(3)$ and $D = \log(2)/\log(3) = 0.631$ (1 mark)

Section 4: Maths I

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

A. Matrices:

Consider the matrix

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

- (i) Find the eigenvalues and eigenvectors of A . Find matrices P and P^{-1} such that $D = P^{-1}AP$ is diagonal. Thereby compute the matrix e^{tA} . (50%)
- (ii) Compute the matrix e^{tA} again, but now using the Taylor series of the exponential. Express the answer in terms of $\cos t$ and $\sin t$. (50%)

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

See attached.

B. Answer the following questions:

- (i) The following model is used in fisheries to describe the recruitment of fish as a function of the size of the parent stock. If we denote the number of recruits by R and the size of the parent stock by P , then

$$R(P) = \alpha P \exp(-\beta P), \text{ for } P \geq 0,$$

where α and β are positive constants.

Sketch the graph of the function $R(P)$ when $\alpha = 2$ and $\beta = 1$ and find all points on the curve that have a horizontal tangent. (20%)

- (ii) Calculate the derivative of the *inverse function* of the function $f(x) = \tan(x)$. (10%)
- (iii) Calculate the Taylor series of the arctangent function around the point $x = 0$ up to order 3. (20%)
- (iv) Spruce budworms are a major pest that defoliate balsam fir. They are preyed upon by birds. A model for the per capita predation rate is given by

$$f(N) = \frac{aN}{k^2 + N^2},$$

where N denotes the density of spruce budworm and a and k are positive constants.

For which density of spruce budworms is the per capita predation rate maximal? (50%)

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

See attached.

CMEF Exam 2015-16
Mathematics Primer, Week 1
Solutions

A. Question 1: Matrices

1) Diagonalisation of $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$.

• Eigenvalues of A : $0 = \det(A - \lambda I_2) = \det \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix}$
 $= \lambda^2 + 1$
 $\Leftrightarrow \lambda^2 = -1 \Leftrightarrow \lambda = \pm i$ where $i = \sqrt{-1}$

• Eigenvectors of A :

• for $\lambda = +i$: $\begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$
 $\Leftrightarrow -ix_1 + x_2 = 0$

so, one choice would be the vector $\begin{pmatrix} -i \\ 1 \end{pmatrix}$

• for $\lambda = -i$: similarly, $\begin{pmatrix} +i \\ 1 \end{pmatrix}$

• Construction of matrices P and D :

$$D = \begin{pmatrix} +i & \\ & -i \end{pmatrix}$$

$$P = \begin{pmatrix} -i & +i \\ 1 & 1 \end{pmatrix}$$

$$\det P = -2i$$

$$P^{-1} = \frac{1}{-2i} \begin{pmatrix} 1 & -i \\ -1 & -i \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} i & 1 \\ -i & 1 \end{pmatrix}$$

(One may now check that $A = PDP^{-1}$)

• Computation of e^{tA} from this:

$$e^{tA} = P e^{tD} P^{-1}$$

$$= P \begin{pmatrix} e^{t \cdot i} & \\ & e^{t(-i)} \end{pmatrix} P^{-1}$$

$$= P \begin{pmatrix} \cos t + i \sin t & \\ & \cos t - i \sin t \end{pmatrix} P^{-1}$$

$$= \cos t \cdot \underbrace{PP^{-1}}_{=I_2} + i \sin t \underbrace{P \begin{pmatrix} 1 & \\ & -1 \end{pmatrix} P^{-1}}_{= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}}$$

$$= \cos t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin t \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

other ways
to perform
the algebra
are of course
possible

2) Computation of e^{tA} from the Taylor series

$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k$$

$$\text{Thus } e^{tA} = \sum_{k=0}^{\infty} \frac{1}{k!} t^k A^k$$

By direct calculation:

$$A^0 = I, A^1 = A, A^2 = -I, A^3 = -A, A^4 = I \dots$$

Thus we have a pattern that repeats itself every 4 steps

$$\Rightarrow e^{tA} = I + tA - \frac{t^2}{2!} I - \frac{t^3}{3!} A + \frac{t^4}{4!} I + \frac{t^5}{5!} A - \frac{t^6}{6!} I - \frac{t^7}{7!} A + \dots$$

$$= \left(I - \frac{t^2}{2!} I + \frac{t^4}{4!} I - \frac{t^6}{6!} I + \dots \right) + \left(tA - \frac{t^3}{3!} A + \frac{t^5}{5!} A - \frac{t^7}{7!} A + \dots \right)$$

even v. odd terms

Here we recognise the Taylor series of \sin , \cos

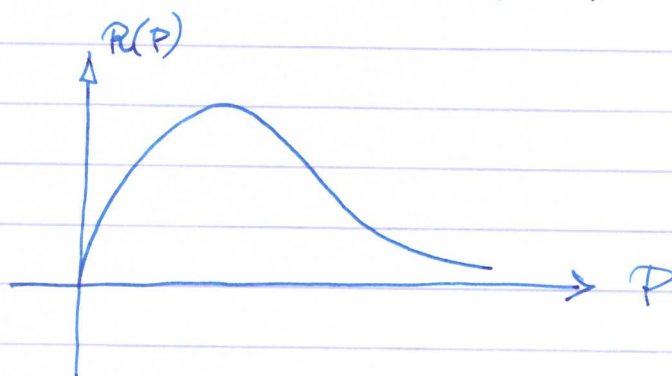
$$= \cos t \cdot I + \sin t \cdot A$$

$$= \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

B. Question 2: Differentiation

1) Fish model

$$\text{Sketch } R(P) = 2P \exp(-P)$$



$$\text{Points with horizontal tangent: } \frac{dR}{dP} = 0$$

$$\begin{aligned} \frac{dR}{dP} &= 2 \exp(-P) + 2P \exp(-P) \cdot (-1) \\ &= 2(1-P) \exp(-P) \end{aligned}$$

Then: 2 such points

at $P=1$ (where $1-P=0$), $R(1)=2/e$
 at $P=+\infty$ (where $\exp(-P)=0$), $R(+\infty)=0$

2) Derivative of inverse function of $\tan x$
 ($= \arctan x$)

Using the formula $\frac{d}{dx} f^{-1} = \frac{1}{f'(f^{-1}(x))}$

$$f(x) = \tan x = \frac{\sin x}{\cos x}$$

$$f'(x) = \frac{\cos x (\cos x) - \sin x (-\sin x)}{\cos^2 x} = \frac{1}{\cos^2 x}$$

$$\text{but also} = 1 + \frac{\sin^2 x}{\cos^2 x} = 1 + \tan^2 x$$

no need to
derive this
from memory
is fine

$$\begin{aligned} \text{Then } \frac{d}{dx} \arctan x &= \frac{1}{1 + \tan^2(\arctan x)} \\ &= \frac{1}{1 + x^2} \end{aligned}$$

3) Taylor series of $\arctan x$ around $x=0$.

We also need 2nd and 3rd derivatives, hence

$$\frac{d}{dx} \left(\frac{1}{1+x^2} \right) = -\frac{2x}{(1+x^2)^2} \quad \text{in } x=0 \text{ this is } 0$$

$$\frac{d^2}{dx^2} \left(\frac{1}{1+x^2} \right) = \frac{(1+x^2)^2 (-2) - (-2x) \cdot 2(1+x^2) \cdot 2x}{(1+x^2)^4}$$

which, in $x=0$, is -2 .

$$\begin{aligned} \text{Thus } \arctan x &\approx \arctan(0) + \left. \frac{1}{1+x^2} \right|_{x=0} \cdot x \\ &\quad + 0 \cdot \frac{x^2}{2!} - 2 \cdot \frac{x^3}{3!} + \underbrace{O(x^4)}_{\text{not required}} \\ &= x - \frac{x^3}{3} \end{aligned}$$

4) Maximum of $f(N) = \frac{aN}{k^2 + N^2}$

Stationary points: $f' = 0$

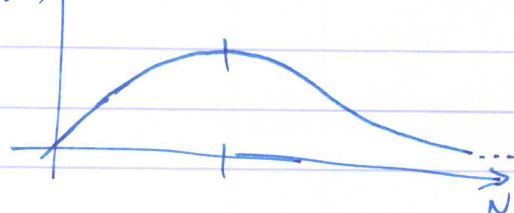
$$\begin{aligned} f' = \frac{df}{dN} &= \frac{(k^2 + N^2) \cdot a - aN \cdot 2N}{(k^2 + N^2)^2} \\ &= a \frac{k^2 - N^2}{(k^2 + N^2)^2} \end{aligned}$$

thus: $f' = 0$ for $N = \pm k$ (but $N \geq 0$ by virtue of its being a population)
and for $N = +\infty$

$$f(k) = \frac{ak}{2k^2} = \frac{a}{2k}$$

$$f(+\infty) = 0$$

$f(N)$ sketch:



From the sketch: maximum at $N=k$