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: Tuesday, 13th Jan 2015, 1300 – 1600

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: GIS and Genomics

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

A. You have been provided with a dataset showing the locations of the insectivorous Venus flytrap (*Dionaea muscipula*) plants across a wetland in North Carolina. A sticky insect trap was positioned near the location of each plant located in the wetland and used to quantify the number of potential food insects for the plant over a one month period. You also have a shapefile showing the boundary of the wetland.

Using this data (for each plant, GPS location and a count of food insects), how might you use GIS to assess the hypotheses that the density of plants is predicted by local variation in insect densities. Specifically:

(i) How might you use GIS buffers to establish the variation in plant density around individual point locations?

(ii) What options might be appropriate for quantifying insect density for each plant density estimate?

(iii) What effect does the spatial scale at which density is calculated have on your analysis? How might you assess the magnitude of this effect and how might behaviour of typical prey insects help resolve this question?

(iv) What effect might the edge of the wetland have on your analysis and how might you counteract this?

(v) What analysis might you use on your extracted data to test the hypothesis that plant density is predicted by local insect density? What assumptions of the analysis might you be particularly concerned about?

(vi) What assumptions does this analysis make about the original sample data?

Model Answer (Marker – Orme (1st), Lysenko (2nd)):

The students haven't been taught spatial statistics so we don't expect point process analysis or Ripley's K! However, they've used point data and buffering and clipping in GIS and we've talked about regression and spatial autocorrelation in stats, so all of the suggestions below should occur to them. This is a problem solving exercise rather than information recall.

(i) Each location could be buffered to a given distance. The number of plants within that buffer could be divided through by buffer area to give a point estimate of the plant density around each plant.

(ii) You could use the local point estimate for each plant or alternatively use the average insect density for all plants sampled within the buffer. These options essentially represent point and smoothed estimates of density – good answers might discuss this difference.

(iii) The size of the buffer used makes a difference to the values and the variance. At the extremes, a tiny buffer will make all of the densities the same (1/A) and a huge buffer will similarly give identical values (N/A). In the middle, you'll see a range of values. Answers should probably describe a plot of observed values for a range of buffer sizes – clever answers might talk about using the variance as a tool to choose a scale. Knowing something about the movement (flight distances, any territories, coloniality) of typical prey insects might also help support a particular scale.

(iv) Edges truncate buffers leading to inaccurate estimates of plant density close to the edge and possibly also prey density. Using the boundary to clip buffers to give a more realistic area might help alleviate the problem.

(v) A linear regression of some sort. Good answers might mention looking carefully at the variance of plant density and point out that it is truncated at zero and really good answers might talk about spatial autocorrelation.

(vi) Many possible things: the survey is complete (or reasonably unbiased), the spatial distribution of flies is stable over time, the sticky traps don't saturate with prey during the month. Any thoughtful critique.

B. Answer the following:

(i) Describe how you could generate the data required for a genome assembly project using Sanger sequencing and Next Generation Sequencing techniques (NGS). Briefly compare the differences between both methodologies.

Continues on next page

(ii) Briefly outline the steps that you would follow to obtain a eukaryote genome assembly once you got your NGS data from a sequencing facility. How would you deal with the presence of repetitive regions in the genome and its effect on the assembly? What statistic could you use to assess the quality of the resulting assembly?

Model Answer (Markers – Javi De Castro (1st), Vogler/Savolainen (2nd)):

(i) -Sanger: it was used on the Human Genome Project (2001). Genomic DNA is fragmented and cloned into bacteria, where it is amplified. Each amplicon (multiple copies of the same DNA sequence) is then sequenced. A cycle sequencing reaction is then performed where fluorescently labeled ddNTPs (terminators) are added. The resulting product contains a mixture of end-labeled extension product of all the intermediate sizes of the amplicon. It is then run through a high-resolution electrophoresis capillary system that orders them by size and reads the corresponding base with a coupled laser. -NGS: libraries are generated by initially fragmenting the genome. Addition of universal common adaptors allows us to amplify the library and attach it to a surface where the sequencing reaction will take place. This is coupled with an imaging system that sequentially detects each incorporated nucleotide. The main difference is the higher degree of parallelization offered by NGS techniques. The fact that no cloning step is required means that the whole genome can be sequenced in one go. This leads to massive differences in yield (NGS can give several orders of magnitude of more data), and has an obvious reflection on costs as well. Sanger yields, in general, longer and much more accurate reads.

(ii) This was discussed in the Genome Assembly Lecture and Practical. First, raw sequence reads have to be filtered to discard low quality reads (QC step). Then a genome assembly algorithm has to be used to obtain the assembly.

Repetitive regions can make up to 50% of the eukaryotic genomes. They make assemblies difficult, especially if the repetitive fragment is bigger than the fragment size used when building the sequencing library. To deal with this issue, several techniques are usually employed: 1) using longer fragment size libraries, typically, a genome sequencing project will include a set of different libraries with a range of fragment sizes; 2) increasing the read length. PacBio or 454 technologies are used to complement projects based on Illumina technologies, where the maximum sequence read is less than 300 bp.; 3) using paired-end reads: both ends of the fragment are sequence; and 4) using mate-pair libraries: extra larger fragment size libraries are prepared (a few kb long as opposed to hundreds of bases which is the typical size). Then the fragments are enzymatically digested and recircularised. This brings sequences that were a few kb apart in the genome in direct contact.

The final assembly can be assessed using the N50 value that Velvet outputs. N50 is the smallest scaffold/contig length above which 50% of an assembly is represented. To calculate it, all you have to do is sort the resulting contigs by size and add them starting from the largest one. The N50 is the size of the contig that makes this total sum greater or equal to 50% of the assembly.

Continues on next page

Section 2: Stats and model fitting

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

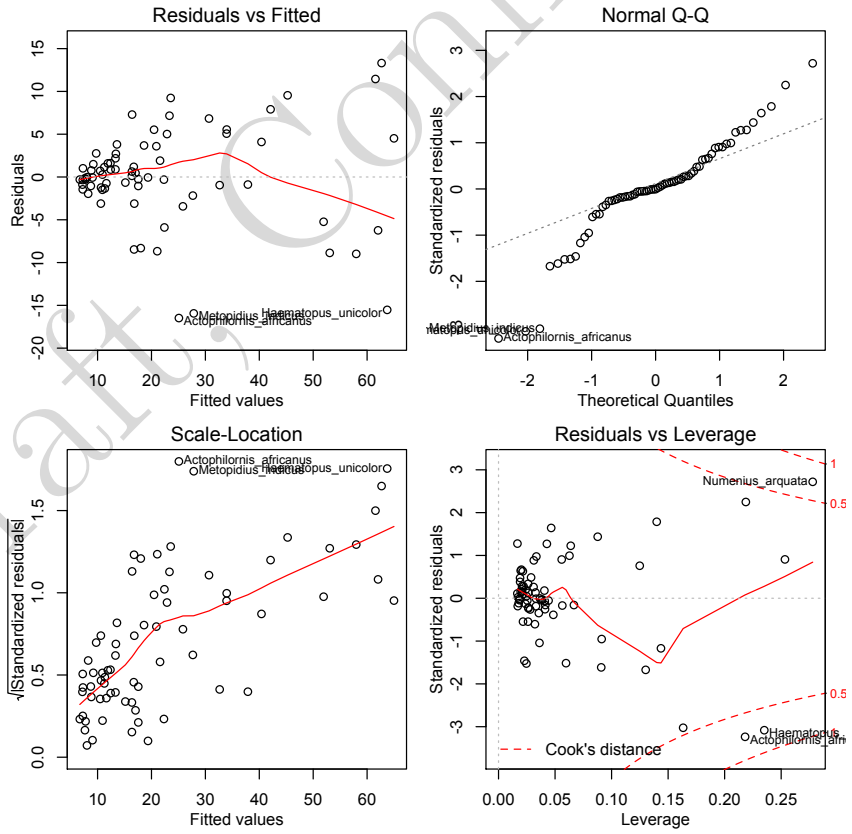
A. A student you are supervising has presented you with an initial linear model from his project looking at predicting variation in egg mass between species of shorebirds. The model summary and diagnostic plots are as follows. `M.Mass` and `F.Mass` are male and female body mass, respectively.

```
Call:
lm(formula = Egg.Mass ~ M.Mass * F.Mass, data = shorebird.data)

Residuals:
     Min       1Q   Median       3Q      Max
-16.4843  -1.3953  -0.0291   2.7322  13.3129

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    4.212e+00  1.445e+00   2.915  0.00483 **
M.Mass         7.080e-02  3.402e-02   2.081  0.04126 *
F.Mass         4.847e-02  2.686e-02   1.805  0.07560 .
M.Mass:F.Mass -5.097e-05  1.913e-05  -2.664  0.00965 **
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1          1

Residual standard error: 5.761 on 67 degrees of freedom
Multiple R-squared:  0.8915,    Adjusted R-squared:  0.8866
F-statistic: 183.5 on 3 and 67 DF,  p-value: < 2.2e-16
```



Why does your student need to go back the drawing board and what suggestions do you have for improvement?

Model Answer (Markers – Orme (first), Pawar (second)):

The diagnostic plots show all kinds of bother - variance increase is the obvious problem which lead to all sorts of warning flags. OK questions will discuss each plot in turn, really good answers will be aware that all the warning flags (long tails

Continues on next page

in QQ, flaring in RF and SL, high leverage points in RvL) stem from the same basic problem. Poorer answers might suggest that the data is wrong/mismeasured for outliers. Clearly, the variables need transforming - and the use of mass suggests a log-log transform to fit an allometric scaling model.

Clever answers will also discuss two further problems:

i) Male mass and female mass? How independent are they? In fact, their correlation is extremely high and the level of collinearity in the model is horrendous.

ii) These are species – they show complex patterns of similarity described by phylogeny. So the observations are not independent of one another and a phylogenetic model is needed.

B. In nature, predators and prey body sizes are typically correlated, and a reasonable model of the size of predator expected to take a prey item of a given size seems to be a power law relationship:

$$M_{pred} = M_0 M_{prey}^a$$

where M_{pred} is predator mass (in g), M_{prey} is prey mass (in g), M_0 is a constant, and the power-law exponent a is another constant. One argument is that this power-law relationship arises from the fact that the effects of body size on the underlying traits responsible for consumer-resource interactions (e.g., body velocity, jaw crushing power) can be captured by power-laws.

Somebody has given you a large dataset on the predator-prey body size relationship in marine organisms. When you plot predator body size against prey body size, you get the result shown in Figure 1a. The other two plots show the body size distributions of predators (Figure 1b) and prey (Figure 1c).

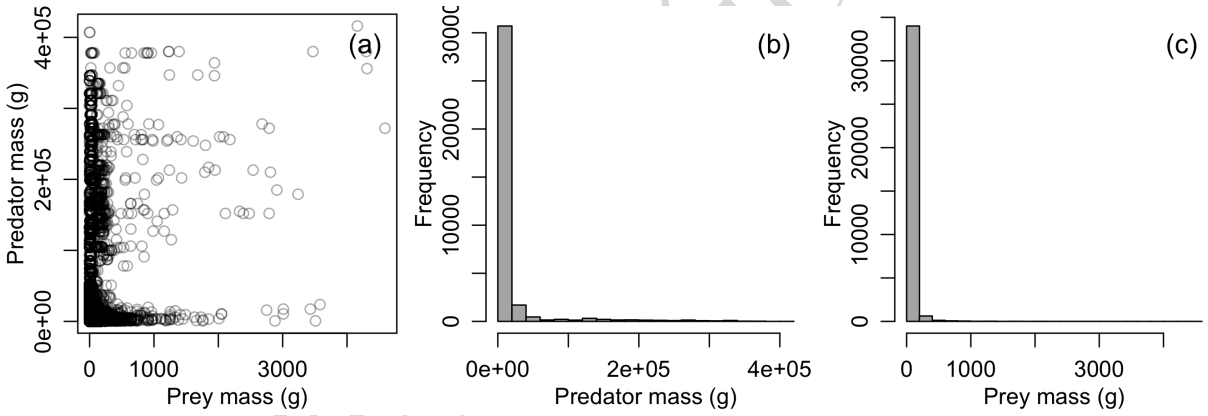


Figure 1: Predator-prey body sizes in marine organisms: (a) predator body mass as a function of prey body mass; (b) a histogram of predator body mass; and (c) a histogram of prey body mass. All masses in grams.

How would you go about testing whether the above equation is a good fit to these data? In particular, please answer the following (each question equally weighted):

(i) Why you might expect a positive relationship between predator and prey size — what biological constraints would apply, and why do you think such a relationship is not so evident in the above scatter plot?

(ii) Write out the equation of a model that should allow you to fit a linear regression model to the data and state what assumptions you would make about the data to allow you to fit the model to the data. Do you foresee any problem with using a linear model with these data?)

(iii) Sketch a graph that would best express how you think the fitted model would look like.

(iv) How you would estimate uncertainty around the parameters a and M_0?

(v) State what M_0 and a mean, in words. In particular, what physiological attribute should M_0 naturally include, especially if the organism is an ectotherm?

Continues on next page

(vi) Would you consider this model to be mechanistic? Why, or why not?

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

(i) Bigger prey will require bigger predators to eat them because physically, predators can usually only handle prey within a range of their own body size. The relationship is not so evident in the scatterplot because the predator and prey size distributions are strongly skewed. A transformation might be in order.

(ii) A linear regression model of the form below or an equivalent expression. Excellent answers might describe ϵ in more detail.

$$log(M_P) = log(M_0) + a \times log(M_f) + \epsilon$$

This model comes from taking a log of the first equation above, and is one way to look for a power law. A LM would require that the predator and prey masses are log-normally distributed, that there is homogeneity of variance over the entire prey range. Strong answers would add that the sampling error is mostly in predator mass and not prey mass. The last assumption is probably not going to be met in such datasets as both predator and prey masses are measured or estimated in the same way.

(iii) A simple scatterplot of a linear regression fitted to a scatter where x axis is $log(M_p)$ and y axis is $log(M_f)$. Extra marks for clear axis labels!

(iv) Any discussion of standard errors or confidence limits from the model

(v) The intercept (M_0) is the expected predator size for a 1 kg prey and the slope a is the increase in prey mass with predator size. Good answers might go into more detail about the power law scaling of a. Physiological attribute in M_0 would be temperature dependence. A really good answer would include a Boltzmann-Arrhenius type model for the thermal dependence.

(vi) The model is mechanistic in the sense that it captures biomechanical constraints on consumer resource size ratios. It is also a bit phenomenological in the sense that it doesn't make mathematically explicit what the constraint exactly is other than that it is a power-law relationship.

Continues on next page

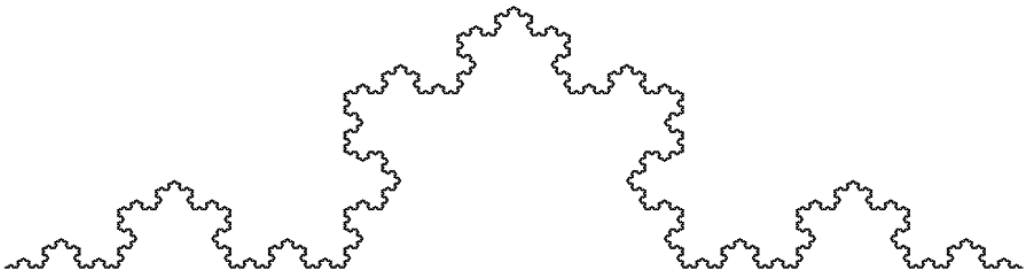
Section 3: HPC & fractals

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

A. You only need give brief bullet point style answers to these questions:

(i) Give three reasons why fractals occur in the natural world. (30%)

(ii) Calculate the dimension of the following fractal; show your workings. (20%)



(iii) Consider the function `Draw_fract` written below in pseudo code:

```
Draw_fract(x, y, r, n){
   if (n>0) {
      Draw_fract( x         ,y          ,r*1/3   ,n-1  )
      Draw_fract( x         ,y+r*2/3    ,r*1/3   ,n-1  )
      Draw_fract( x+r*2/3   ,y          ,r*1/3   ,n-1  )
      Draw_fract( x+r*2/3   ,y+r*2/3    ,r*1/3   ,n-1  )
   } else {
      Draw a square of width r with it's bottom left corner at (x,y)
  }
}
```

a. Show what `Draw_fract(0,0,27,0)` would draw (10%).
Hint: this means x = 0 , y = 0 , r = 27 and n = 0, read through the code carefully trying to think like a computer.

b. Show what `Draw_fract(0,0,27,1)` would draw so n = 1 now (10%).
Hint: you will use your answer to i) above.

c. Show what `Draw_fract(0,0,27,2)` would draw so n = 2 (20%)

d. As `n` becomes large the function `Draw_fract` described above generates a fractal, what is its fractal dimension; explain your answer? (10%)

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

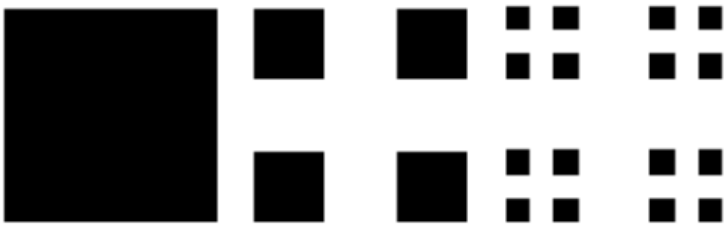
(i) Any three reasonable points (1 mark each) e.g.:

- *Fitting large surface areas into small volumes (e.g. lungs)*
- *Solving transportation problems optimally (e.g. circulatory system)*
- *As the result of a simple set of rules operating (e.g. plant growth)*
- *Same processes happening at multiple scales (e.g. coastlines)*
- *Fractal structures are heritable.*

Continues on next page

(ii) Requires 4 copies of itself to construct the same shape but 3 times larger; therefore 4 = 3D where D represents the fractal dimension (1/2 of 20%).
Log(4) = log(3D)
D = log(4)/log(3) = 1.262 (1/2 of 20%)

(iii)
a. Left-most configuration in figure below
b. Middle configuration in figure below
c. Right-most configuration in figure below



Up to half marks can be awarded for i , ii and iii if written text demonstrates good ability to work through the code and has the right idea but contains some mistakes or lacks the final diagrams

(iv) In the case where n gets large we end up with an object where four copies are required to produce something of three times the width. So the fractal dimension is the same as above (1.262). The marks are mostly for working out the pattern and seeing what happens as n gets really large.

B. You only need give brief bullet point style answers to these questions.

Consider a simple individual based neutral model containing a fixed number (J) of individual organisms.

In each time step, an individual is chosen at random to die and be replaced with the offspring of another individual in the system. With probability v the new-born individual is of an entirely new species in the system (speciation) otherwise it is of the same species as its parent.

(i) For the special case where $v = 0$. Given an initial condition where there are many species in the system, describe what will happen to the diversity of the system as the simulation progresses and why. (20%)

(ii) 2.) What happens to the system in the more general case where $v > 0$ starting from different initial conditions of both high and low diversity. (20%)

(iii) Consider two simulation tasks for this neutral model: A) simulating the system many times for a wide range of different values of v and B) simulating the system for a single value of v but with a very large value of J.

a. Can either of these simulation tasks run in parallel on multiple CPUs and why? (10%)

b. For running these tasks on a high performance computing facility you would need to write a shell script file. Your shell script contains the code `#PBS -l mem=800mb` what does this mean and how might you need to change this to perform the required simulations? (10%)

c. Your shell script also contains the code `#PBS -l walltime=12:00:00` what does this mean and how might you need to change this to perform the required simulations? (10%)

Continues on next page

(iv) Give three advantages or disadvantages to the use of simple models such as the individual based neutral model described here. (30%)

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

(i) It will decay to one species OR monodominance (1/2 of 20%) because there is no speciation to balance extinction so diversity can only decrease to the absorbing state of a single species (1/2 of 20%)

(ii) Diversity will converge to the same state from any initial condition, or some mention of dynamic equilibrium or similar (1/2 of 20%). The higher the speciation rate the higher the diversity tends to be (1/2 of 20%).

(iii) (a) Problem A is embarrassingly parallel, the simulations for different values of can be run simultaneously on difference CPUs (1/2 of 10%). Problem B is a single simulation so cannot be parallelised (1/2 of 10%)

(b) This describes how much RAM is needed to be allocated to the simulations, currently 800 megabytes (1/2 of 10%). Problem B describes a large system that might require more RAM in order for it to be stored in working memory so the number 800 may need to be larger (1/2 of 10%).

(c) This describes how much CPU time is needed for each simulation currently 12 hours (1/2 of 10%). Problem B simulations may be slow and so could require a longer walltime (1/2 of 10%).

(iv) Any three reasonable and well thought out points will do for the total of 30% mark each, e.g.,

- *ad: can use to study processes of interest in isolation from complicating factors*
- *dis: not very realistic / ignores important factors*
- *ad: computationally / analytically tractable*
- *ad: can easily explore parameter space of the model / limited number of parameters*

Continues on next page

Section 4: Maths I

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

A. Gompertz Growth

The Gompertz growth curve is given by

$$N(t) = K \exp(-a \exp(-bt)),$$

where K and b are strictly positive constants.

Now answer the following (equal weighting for all questions):

(i) Find a formula for a in terms of $N(0)$.

(ii) What is the behaviour of $N(t)$ as t goes to infinity?

(iii) Show that $N(t)$ is strictly increasing if $N(0) < K$ and strictly decreasing if $N(0) > K$.

(iv) Under what conditions on the constants b, K and $N(0)$ does $N(t)$ have an inflection point? If so, for what time t?

B. Taylor series expansion of the ascsine transform

The transform

$f(x) = \arcsin(x)$

is often used in biology as a data preprocessing step for normalising proportions and consists of taking the arcsine of the square root of a data point. Calculate the Taylor series approximation of the arcsine function at $x = 0$, up to fourth order. This will be significantly easier to calculate than the arcsine function when you don't have a computer at hand (Hint: recall that $sin^{-1}x = \frac{1}{\sqrt{1-x^2}}$)

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

A. *See attached solution notes.*

B. *Odd function, so only have odd powers:* $x + \frac{x^3}{6}$
Also see attached solution notes

Continues on next page

Section 5: Maths II

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

A. Tidal Currents in Fjords

In a simple mathematical model of the tidal currents in a fjord we model the fjord as a canal of length L and with a constant parabolic cross-section:

$$y_{can}(x) = d\left(\left(\frac{2x}{w}\right)^2 - 1\right).$$

Here, x is a horizontal coordinate perpendicular to the flow of water, $y_{can}(x)$ is the height (or depth, if negative) of the canal floor as a function of x (with respect to average sea level), and d and w are given parameters (describing the depth and width of the canal, respectively).

The diurnal change of the sea level due to the tides is modeled as

$$y_{sea}(t) = y_0 \sin(4\pi t).$$

Here, t is time measured in days, $y_{sea}(t)$ is the sea level at time t, again with respect to average sea level, and y_0 is a given parameter describing the maximal height of the tides.

You may assume that the fjord never runs dry, so $y_0 \leq d$.

Now answer the following (equal weighting for all questions):

(i) Calculate the volume of water $V(t)$ in the fjord as a function of time. You may ignore the fact that it takes time for water to enter or leave the fjord.

(ii) Calculate the strength of the tidal current by finding the instantaneous change of $V(t)$.

(iii) Calculate the sea level y_{sea} at the time when the tidal current is the strongest (I am not asking for this time t). You can use the fact that the solution is positive.

To check that your solution of part iii is correct, you can verify that for a value of y_0 that is very small compared to d, the solution is $y_{sea} \approx 0$. For $y_0 = d$ the solution is $y_{sea} = y_0/3$.

B. A Recurrence Relation

Find a formula for x_k as a function of k from the recurrence $x_k = x_{k-1} + (3/4)x_{k-2}$, with $x_1 = 0$ and $x_2 = 8/3$.

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

A. *See attached solution notes.*

B. $x_k = 2\left(\frac{3}{2}\right)^{k-2} + \frac{2}{3}\left(-\frac{1}{2}\right)^{k-2}$
Also see attached solution notes

End of paper

CMEE exam 1, 2014-2015 Maths Primer

Solutions

I.A.

1) $N(0) = K \exp(-a \exp(0))$

$= K \exp(-a)$

$\Rightarrow a = -\ln \frac{N(0)}{K} = \ln \frac{K}{N(0)}$

2) As $t \to \infty$, $\exp(-bt) \to 0$,

so $N(\infty) \to K$.

3) $\frac{dN}{dt} = abK \exp(-a \exp(-bt) - bt)$

all factors, except a, are positive.

Hence, N is strictly increasing for $a > 0$, which is for $N(0) < K$, and strictly decreasing for $a < 0$, which is for $N(0) > K$.

4) $\frac{d^2N}{dt^2} = -ab^2K \exp(-a \exp(-bt) - 2bt) \times (-a + \exp(bt))$

To have an inflection point this must be 0.

Only one factor can be 0, namely $-a + \exp bt$ and only if $a > 0$, i.e. if $N(0) < K$.

The inflection point is at t satisfying $\exp bt = a$

i.e. at $t = \frac{1}{b} \ln a$.

I.B

Taylor series of $f(x) = \arcsin x$ at $x=0$, up to fourth order.

Since $f(x)$ is odd, we already know that the series will only contain terms of odd power.

$$f(x) \cong \overbrace{f(0)}^{=0} + f'(0)\,x + \overbrace{\frac{1}{2}f''(0)\,x^2}^{=0} + \frac{1}{6}f'''(0)\,x^3 + \frac{1}{24}\underbrace{f^{(4)}(0)}_{=0}\,x^4$$

Need derivatives f' and f'''

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

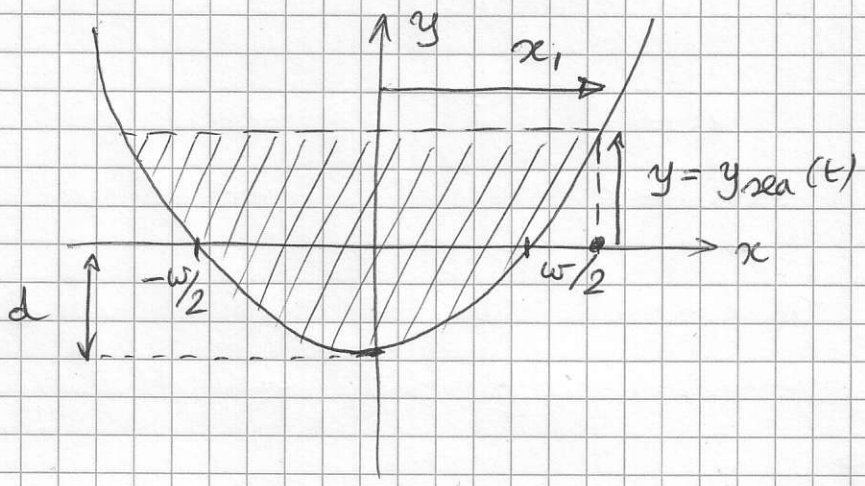
$$f''(x) = \frac{x}{(1-x^2)^{3/2}}$$

$$f'''(x) = \frac{3x^2}{(1-x^2)^{5/2}} + \frac{1}{(1-x^2)^{3/2}}$$

So $f(0) = 0$, $f'(0) = 1$ and $f'''(0) = 1$

$$\Rightarrow \quad f(x) \sim x + \frac{x^3}{6} \left(+ O[x^5]\right)$$

II. A.



1. $$V(t) = L \int_{-x_1}^{+x_1} \left(y - d\left[\left(2\frac{x}{w} \right)^2 - 1 \right] \right) dx$$

where x_1 is such that $y(t) = d\left[\left(2\frac{x}{w} \right)^2 - 1 \right]$.

Thus, $$x_1 = \sqrt{\frac{y(t)+d}{4d}}\, w$$

and $$V(t) = L\left[(y+d)x - \frac{4d}{w^2}\frac{x^3}{3} \right]_{-x_1}^{x_1}$$

$$= 2L\left[(y+d)x_1 - \frac{4d}{w^2}\cdot\frac{x_1^3}{3} \right]$$

↓ substitute x_1

$$= 2L\left[\frac{(y+d)^{3/2}}{\sqrt{4d}}\, w - \frac{4d}{w^2}\cdot\frac{1}{3}\left(\frac{y+d}{4d}\right)^{3/2} w^3 \right]$$

↓ simplify

$$= \frac{4}{3} L \underbrace{\frac{w}{\sqrt{4d}}}_{\text{Call this } a} \left(y(t)+d \right)^{3/2}$$

Finally, as $y(t) = y_0 \sin 4\pi t$,

$$V(t) = a\left(y_0 \sin 4\pi t + d \right)^{3/2}$$

2. Tidal current: $\frac{dV}{dt}$

$$= a \cdot \frac{3}{2}\left(y_0 \sin 4\pi t + d\right)^{1/2} \cdot y_0 \cdot 4\pi \cos 4\pi t$$

$$= 6\pi a y_0 \left(y_0 \sin 4\pi t + d\right)^{1/2} \cos 4\pi t$$

3. Current is strongest when $\frac{dV}{dt}$ is maximal.

Stationary point when $\frac{d^2V}{dt^2} = 0$

$$\frac{d^2V}{dt^2} = 6\pi a y_0 \left[\frac{y_0 4\pi \cos 4\pi t}{2\left(y_0 \sin 4\pi t + d\right)^{1/2}} \cdot \cos 4\pi t \right.$$

$$\left. + \left(y_0 \sin 4\pi t + d\right)^{1/2} (-4\pi) \sin 4\pi t \right]$$

$$= \frac{6\pi a y_0}{\left(y_0 \sin 4\pi t + d\right)^{1/2}} \left(2\pi y_0 \cos^2 4\pi t \right.$$

$$\left. - 4\pi \left(y_0 \sin 4\pi t + d\right) \sin 4\pi t \right)$$

So $\frac{d^2V}{dt^2} = 0 \iff$ $\underbrace{y_0 \cos^2 4\pi t}$ $= 2\left(y_0 \sin 4\pi t + d\right) \underbrace{\times \sin 4\pi t}$

$\downarrow$ y^2 $\qquad\qquad$ $\downarrow$

$$= y_0 - \frac{y^2}{y_0} \qquad = 2(y+d)\frac{y}{y_0}$$

$$\iff \quad y_0^2 - y^2 = 2(y+d)y$$

$$\iff \quad 3y^2 + 2dy - y_0^2 = 0$$

$$\iff \quad y = \frac{1}{3}\left(-d \pm \sqrt{d^2 + 3 \cdot y_0^2}\right)$$

By what is given, we take the positive solution

II.B

$$\begin{cases} x_1 = 0 \\ x_2 = \frac{8}{3} \\ x_k = x_{k-1} + \frac{3}{4} x_{k-2} \end{cases}$$

check: $x_3 = \frac{8}{3}$

$x_4 = \frac{8}{3} + \frac{3}{4} \cdot \frac{8}{3}$

$= \frac{14}{3}$

...

In matrix form:

$$\begin{pmatrix} x_k \\ x_{k-1} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & \frac{3}{4} \\ 1 & 0 \end{pmatrix}}_{A} \begin{pmatrix} x_{k-1} \\ x_{k-2} \end{pmatrix} = A^{k-2} \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$$

We need A^{k-2}, hence we need to diagonalise A.

- eigenvalues: $\det \begin{pmatrix} 1-\lambda & 3/4 \\ 1 & -\lambda \end{pmatrix} = -(1-\lambda)\lambda - 3/4$

$= \lambda^2 - \lambda - 3/4$

roots: $\lambda_\pm = \frac{1}{2}\left(1 \pm \sqrt{1+3}\right)$ → $3/2$, → $-1/2$

- eigenvectors:

for $\lambda_+ = 3/2$: $(1-3/2)x_1 + \frac{3}{4}x_2 = 0$

→ choose $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ (or any scalar multiple)

for $\lambda_- = -1/2$: $(1+1/2)x_1 + \frac{3}{4}x_2 = 0$

→ choose $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$ (or any scalar multiple)

$$\rightarrow \quad P = \begin{pmatrix} 3 & -1 \\ 2 & 2 \end{pmatrix} \qquad (\det P = 8)$$

$$\Lambda = \begin{pmatrix} 3/2 & \\ & -1/2 \end{pmatrix}$$

$$P^{-1} = \frac{1}{8}\begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix}$$

Diagonalisation of A: $\quad A = P \Lambda P^{-1}$

$$\Rightarrow \quad A^{k-2} = P \, \Lambda^{k-2} \, P^{-1}$$

$$= \begin{pmatrix} 3 & -1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \left(\frac{3}{2}\right)^{k-2} & \\ & \left(-\frac{1}{2}\right)^{k-2} \end{pmatrix} \cdot \frac{1}{8} \cdot \begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix}$$

We need x_k, which is the first element of

$$\begin{pmatrix} x_k \\ x_{k-1} \end{pmatrix} = \begin{pmatrix} 3 & -1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \left(\frac{3}{2}\right)^{k-2} & \\ & \left(-\frac{1}{2}\right)^{k-2} \end{pmatrix} \underbrace{\cdot \frac{1}{8} \begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 8/3 \\ 0 \end{pmatrix}}_{= \begin{pmatrix} 2/3 \\ -2/3 \end{pmatrix}}$$

$$= \begin{pmatrix} 3 & -1 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} \left(\frac{3}{2}\right)^{k-2} \\ -\left(-\frac{1}{2}\right)^{k-2} \end{pmatrix} \cdot \frac{2}{3}$$

$$\Rightarrow \quad \boxed{x_k = 2\left(\frac{3}{2}\right)^{k-2} + \frac{2}{3}\left(-\frac{1}{2}\right)^{k-2}}$$