M0019 Assessed Practical 2019

You should work in groups of 3 for this assignment, handing in one report at the end. It is worth 20% of the marks for this unit.

The data loaded with the command

```
bf <- read.table("https://people.maths.bris.ac.uk/~sw15190/TOI/blowfly.dat")
```

give the counts[1] of adult sheep blowfly in a laboratory experiment against day of the experiment. The population counts are every 2 days. The data were collected in the 1950s in the lab of the Australian ecologist Nicholson, and have been the source of a large amount of investigation by theoretical ecologists and mathematical biologists every since, because of the remarkable cycles that they show.

A possible dynamic model for the population data, N_t, simply adjusts the population by the births B_t and deaths D_t at each two day time step, as follows:

$$N_{t+1} = N_t + B_t - D_t$$

where $B_t \sim \text{Poi}(PN_{t-\tau}e^{-N_{t-\tau}/N_0})$ and $D_t \sim \text{binom}(N_t, \delta)$. P, N_0 and δ are positive parameters and also $\delta \leq 1$. τ is an integer parameter (likely to be somewhere in the range 4 to 9).

Your tasks are as follows:

1. Estimate the parameter values using maximum likelihood estimation and obtain interval estimates for all but τ.
2. Critically assess the adequacy of the model – could it really have produced the observed dynamics?
3. Explore whether there are model modifications that might give more plausible models.

For part 1, you will need to develop an appropriate likelihood. Given that the N_t are observed without error, you can use the differences (see `?diff` in R) $\Delta N_t = N_{t+1} - N_t$ as the data for the purposes of constructing a likelihood. The idea is that the model gives the distribution of ΔN_t, given the parameters and the preceding sequence of N_t values. Because of the model's dependence on lagged values of N_t you will have to discard the first few ΔN_t values, as they can not be predicted.

You will need to do some work to be able to compute a log likelihood. The key is to use the normal approximations to the Poisson and binomial (not forgetting to think about how well justified this is here) to obtain an approximate likelihood. Do think a bit about plausible initial parameter values before optimizing your log likelihood. Also think about how to handle τ: you can't optimize it using `optim` or `nlm` as it is discrete, but it is easy to do something sensible to optimize it.

When critically assessing your model you should try using the best fit parameter to iterate the model forward in time to generate a full run of blowfly populations, starting from the observed first few N_t data, to get started. You should look at a deterministic run where B_t and D_t are set to their expected value at each timestep (do the dynamics look plausible, given that all randomness has been neglected?), and at some stochastic simulations from the model in which B_t and D_t are Poisson and binomial random variables. Could the model plausibly have generated the data? If necessary you might try modifying it to see if a more plausible model can be found.

What to hand in: You should write, as a group, a concise report of no more than 5 sides of A4 (normal margins $\geq$ 10pt font). The report should be accompanied by an appendix containing well structured, clearly commented R code for performing the analysis. Both the report and the appendix should start with a title and the names of all the group participants.

[1] the experimenters did not count every adult fly every 2 days, instead they counted dead flies and discarded pupal cases every 2 days, from which the number of live adults can be worked out: the raw data then are in fact the changes in adult fly population every 2 days, and these are indeed what you will model.

The report should be suitable for a statistician, explaining the analysis, allowing them to understand what you have done, what you have concluded and why. You should not assume that they have seen this sheet of instructions, so make sure the work is introduced sensibly. There should be enough detail for a statistical reader to judge the appropriateness of the approach. The report should include appropriate plots. The main body of the report should ideally include no R code (technicalities should be explained with maths, if necessary).

One report (as a pdf file) and one appendix (plain text file is best) per group should be emailed to `simon.wood@bristol.ac.uk` with the subject `M0019 blowfly` followed by your surnames, by **12 noon, on Thursday 21 March 2019**.

Mark scheme guidance

First class marks will be awarded for work that could be passed on to statistically literate scientists interested in these data, essentially without modification. That is to say the statistics is appropriate and clearly explained, the conclusions appropriately drawn and any limitations are discussed fairly.

Upper second class marks will be awarded for work that could be passed on to the scientists after a round of revision correcting some errors of presentation, interpretation or statistics that are relatively minor.

Lower second class marks will be awarded to work that has some more substantial flaws of presentation, interpretation or statistical reasoning which would require some more work to correct.

Third class marks will be awarded for work that contains some indication of substantive understanding and engagement, but contains more serious errors and misunderstandings.