MATH35600 Assessed Practical 2020

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.

To compensate for the effects of strike action, this practical has been simplified and made more prescriptive than usual.

The data loaded with the command

```
cov <- read.table("https://people.maths.bris.ac.uk/~sw15190/TOI/covid19.csv",
                  sep=",",header=TRUE)
```

give data on UK COVID-19 cases up until 11/3/2020. Here we will concentrate on the second column, giving the number of new cases per day.

Suppose that you are part of the team analysing these data for government in order to offer advice. Early in an epidemic in which a disease is spreading unchecked a simple model is that its increase will be exponential. That is the rate of generation of new cases, N, will follow the model

$$N(t) = N(0)e^{rt} = e^{n_0+rt}$$

where r is a parameter relating to the number of new infections created by each existing infection and $n_0 = \log N(0)$ the log of the new case rate at time 0. *The government would like to know if there is any indication that the rate of transmission is slowing over time, perhaps indicating that the measures taken to slow the disease spread are working.*

A simple model that allows for departure from unbounded exponential growth is

$$N(t) = e^{n_0+\sum_i^K \beta_i t^i}$$

where K has to be selected and the β_i and n_0 to be estimated. A reasonable model for the observed data is that the number of cases at time t is $y_t \sim \text{Poi}(N(t))$, with the y_t being independent.

You should do the following, obviously checking results as you go to make sure the answers are sensible.

1. Write R code to evaluate the negative log likelihood of the model given above suitable for optimization using `optim`. Make your code sufficiently general that you can control K simply by changing the dimension of the parameter vector that you supply to your function. Notice that the model is invariant to the units that time is measured in: all linear rescalings of time are equivalent. When creating your time vector, it is best to have it run from -1 to 1 in equal steps (see R function `seq`, for example): this will ensure good numerical performance of `optim`.

2. The simplest way to select K is to use AIC covered in section 8.9 of the notes. Fit models with values of K from 1 to 5 and select the best one by AIC. Check that the selected model could not be further simplified using a GLRT.

3. By making use of the `hessian=TRUE` argument to `optim` and the `solve` function, find the approximate covariance matrix for the parameter estimators of your model.

4. By making use of standard results on transformation of covariance matrices, find approximate 95 confidence intervals for your estimates of $\log N(t)$ and hence plot intervals for $N(t)$ for your selected model overlaid on the raw case data.

What to hand in.

1. A 4 page report (A4, normal margins, at least 10pt font, PDF file) consisting of 2 sections.

(a) A one page summary for ministers, addressing the key question of whether there is evidence for the rate of spread slowing, or not and how reliable the evidence is.

(b) A three page (maximum) report explaining what you did statistically for other statisticians (e.g. a civil service statistician). This should include no, or minimal, R code. The idea is that you explain what you did and why and what you concluded so that a competent statistician could replicate it for themselves, and could judge how well founded your conclusions are based on the results presented.

Your report should not assume that the reader has seen this assignment sheet.

2. A text file containing the R code you used, carefully structured and commented. Again a statistician should be able to take your code and use it to replicate your analysis, while understanding what it is doing.

One report (as a pdf file) and one R code file (plain text is best) per group should be emailed to `simon.wood@bristol.ac.uk` with the subject `MA35600 COVID19` followed by your surnames, by **12 noon, on Monday 30th March**. As a result of the early university closure, this date is provisional. It will not be moved to an earlier date, but may be moved later, in which case the revised date will be posted on the course webpage. You should not assume that the deadline will be extended.

Mark scheme guidance

First class marks will be awarded for work that could be passed on to the government 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 government, 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.