Homework 8

36-467/667

Due at 6 pm on Thursday, 29 October 2020

The file lv.R, on the class website, defies a function, `lv.sim()`, which simulates a famous ecological model of how predators and prey animals interact, leading to cycles in their populations. (You'll need to install the package `deSolve`.) We will revisit models like this later in the course, when we look at what are called "state-space" models. For now, you can treat that function as a black box whch just simulates the model on command. I am deliberately saying so little about how the model works in order to emphasize how much it's possible to learn about a model from simulation alone.

The out put of `lv.sim()` is a time series, i.e., a series of values $X(t_1), X(t_2), \ldots X(t_n)$. There are two arguments to `lv.sim()`: the first, `theta.vec`, is a vector of parameters for the model, and the second, `times`, is the vector giving the desired times $t_1, \ldots t_n$. By default, `times` is set to match up with the `lynx` data set we've looked at in Lecture 2 and Lecture 16; you shouldn't need to alter it for this problem set. The code also provides a "reference" value of the parameters, `theta.ref`; some parts of the assignment will ask you to investigate what happens when you change this.

1. *Exploring the real data.* Plot the `lynx` series. (Feel free to refer to the class notes for code.)
    a. (5) Describe the shape of the curve over time.
    b. (5) What, roughly, is the typical amount of time that passes between peaks of the lynx population? Call this the **inter-peak interval**. (You don't need to write code to do this, but you can if you want.)
    c. (5) Plot the autocovariance function. Is there a peak in the ACF at the inter-peak interval? Should there be?
    d. (5) What's the mean of the series? The variance? The autocovariance at a lag corresponding to the inter-peak interval?
    e. (2) Fit an AR(2) to the data and report the estimated intercept and slopes. *Hint*: Lecture 16.
2. *Exploring the simulated data.* Run the model *once*, with the default parameter values, and store the results.
    a. (5) Describe the shape of the curve over time. How does it resemble the curve of the data, and how is it different?
    b. (5) What, roughly, is the inter-peak interval?
    c. (5) Plot the autocovariance function. Is there a peak in the ACF at the inter-peak interval?
    d. (5) What's the mean of this series? The variance? The autocovariance at a lag corresponding to its inter-peak interval? The autocovariance at a lag corresponding to the data's inter-peak interval (if they differ)?
    e. (2) Fit an AR(2) to the simulation and report the estimated intercept and slopes. (This illustrates how any statistic can be used as a summary, descriptive statistic, even one based on some model.)

3. *Repeating the same action and expecting a different result* Re-run the model five more times, saving the results. (*Hint:* `replicate()`.)

   a. (5) Plot the additional simulation runs. (Ideally, you should have all the runs in *one* plot, with some sort of visual distinction between them.) What aspects of the shape stay the same from run to run, and which change?

   b. (5) Find the mean of each series. What's the standard deviation of these means? What standard error would you report for the expected value of the process?

   c. (5) Find the variance of each series. What's the standard deviation of these variances? What standard error would you report for the variance of the process?

   d. (5) Find the autocovariance at the lag corresponding to the inter-peak interval. What standard error would you report for this autocovariance?

4. *Extracting distributions*

   a. (5) Using at least 1000 replications of the model, find the distribution of $X(1871)$, i.e., the number of lynxes in the year 1871. Plot the histogram and describe the shape of this distribution.

   b. (5) Using at least 1000 replications of the model, find the means and variances for the years 1821, 1831, ... 1931. That is, calculate estimates for both $\mathbb{E}[X(t)]$ and $\mathrm{Var}[X(t)]$ for each of these years t. (If you think it's valid to use the same simulations to get the means and variances for all years, do so, but explain why.) Plot the mean against the variance. What relationship do you see between these quantities?

   c. (5) Explain why the relationship you found in the previous plot would be a difficulty for fitting any kind of linear AR model here.

5. *"Try it, and see what happens"*

   a. (5) Change the last parameter from 550 to 1000 and run the model 5 times. What changes, and what stays the same? What happens if you change the parameter to 100? Can you characterize what the parameter does in words?

   b. (5) Revert back to the original parameters, and double the third parameter to 0.66. What changes, and what stays the same? If you also double the fourth parameter, to 0.24, what changes, and what stays the same?

6. (1) How much time did you spend on this assignment?

Presentation rubric (10): The text is laid out cleanly, with clear divisions between problems and sub-problems. The writing itself is well-organized, free of grammatical and other mechanical errors, and easy to follow. All plots, tables, etc., are generated automatically by code embedded in the R Markdown file. Plots are carefully labeled, with informative and legible titles, axis labels, and (if called for) sub-titles and legends; they are placed near the text of the corresponding problem. All quantitative and mathematical claims are supported by appropriate derivations, included in the text, or calculations in code. Numerical results are reported to appropriate precision. Code is properly integrated with a tool like R Markdown or knitr, and both the knitted file and the source file are submitted. The code is indented, commented, and uses meaningful names. All code is relevant, without dangling or useless commands. All parts of all problems are answered with coherent sentences, and raw computer code or output are only shown when explicitly asked for.