Why do time series analysis? When is fitting a mechanistic model useful?

Edward Ionides

Department of Statistics, University of Michigan

ionides@umich.edu

Eisenberg MIDAS group meeting

Tuesday 24th January, 2017

Slides for this talk are at

`http://dept.stat.lsa.umich.edu/~ionides/talks/midas17.pdf`

Overview

1. Observational time series data on epidemiological systems.
2. Classical time series analysis: what it can do, and what it can't.
3. How to fit mechanistic models to time series data.

Observing population-level dynamics

- We're interested in understanding epidemiological mechanisms:
  - Who gets infected by whom? How can we reduce disease transmission?
- Mechanisms operate at the individual level, and can be studied at the individual level.
- For large, complex systems (such as human society) it is hard to use knowledge about one scale (individual behavior) to understand another scale (population behavior).
  - Example: macroeconomics should in principle be consistent with microeconomics, but doing this is an open challenge.
  - We can study how humans and mosquitoes and malaria parasites respond to temperature changes. What will be the consequences of climate change for malaria prevalence?
- Ultimately, public health is concerned with population-level outcomes.
- Representative randomized controlled studies are occasionally possible. Failing that, we can test our theories on observational population-level data.

Basic questions: Is there a trend? Is there a change-point?

- Sometimes we don't ask the data to directly confirm whether it is consistent with a mechanistic model. We just want to check whether there is any trend or pattern that could possibly be interpretable.
- Looking for statistical evidence for trends (and linear associations) is a classical topic of time series analysis.
- Case study: Malaria in the African highlands.
  - Hay et al (*Nature*, 2002) used time series analysis to argue that there was no evidence for African highland malaria trends being driven by temperature trends.
  - This finding was controversial, and turned out to be sensitive to the particular choice of dataset and time period analyzed.
  - Siraj et al (*Science*, 2014) settled the controversy using spatially resolved time series, showing how malaria prevalence changes with altitude and temperature.

Limitations of classical time series analysis

- Classical time series analysis extends standard regression analysis to allow for correlated data.
- These classical time series methods have been around for 50 years.
- The usual statistical challenges remain: obtaining and understanding the data, and applying the statistical methodology thoughtfully.

But, not all epidemiological questions one might want to address using time series data are of this type...

A perspective from 2001

The following six issues identified by Bjørnstad and Grenfell (*Science*, 2001) are not solved by classical time series methodology. They require consideration of **nonlinear mechanistic models** as statistical tools for biological systems:

1. Combining measurement noise and process noise.
2. Including covariates in mechanistically plausible ways.
3. Continuous time models.
4. Modeling and estimating interactions in coupled systems.
5. Dealing with unobserved variables.
6. Modeling spatial-temporal dynamics.

Fast forward to 2014

Dobson (*Science*, 2014)

"Powerful new inferential fitting methods (Ionides, Bretó and King, 2006) considerably increase the accuracy of outbreak predictions while also allowing models whose structure reflects different underlying assumptions to be compared. These approaches move well beyond time series and statistical regression analyses as they include mechanistic details as mathematical functions that define rates of loss of immunity and the response of vector abundance to climate."

Fitting mechanistic models to data

- Process noise is often important to get a good fit to the statistical properties of data. For example, stochastic differential equation models may be preferred to deterministic differential equations.
- Simulation-based methods can evaluate and maximize the likelihood for general partially observed stochastic dynamic models:
  - `http://kingaa.github.io/sbied/`
  - `http://ionides.github.io/531w16/`
- This opens the possibility to use all the usual tools of likelihood-based inference: likelihood ratio tests, AIC, profile likelihood confidence intervals, Fisher information.

References


Dobson, A. (2014).
Mathematical models for emerging disease.
Science, 346(6215):1294–1295.


Hay, S. I., Cox, J., Rogers, D. J., Randolph, S. E., Stern, D. I., Shanks, G. D., Myers, M. F., and Snow, R. W. (2002).
Climate change and the resurgence of malaria in the East African highlands.
Nature, 415:906–909.


Ionides, E. L., Bretó, C., and King, A. A. (2006).
Inference for nonlinear dynamical systems.
Proc. Natl. Acad. Sci. USA, 103:18438–18443.


Siraj, A. S., Santos-Vega, M., Bouma, M. J., Yadeta, D., Carrascal, D. R., and Pascual, M. (2014).
Altitudinal changes in malaria incidence in highlands of Ethiopia and Colombia.
Science, 343(6175):1154–1158.