Modeling weekly rat population dynamics in Chicago, Illinois during 2011-2018

10 total points

Christopher Cahill

17 November 2023

NOTE: date above is the due date

Background

Chicago's 311 service center records all service requests the city receives, including complaints for rodents and rats. When phone complaints about rodents come in, city employees are sent to inspect the reported area. Employees record things they see on these site visits including the number of premises they observe that have rats.

Data

All of these data are stored in a database and can be accessed through the Chicago data portal available at https://data.cityofchicago.org/Service-Requests/Chicago-Rodent-map/3hy6-wdqe. In the `rats.csv` dataset you have access to these data, where I summed the total number of inspected premises with rats each week in Chicago during 2011-2018.

Math behind a simple state-space population dynamics model

We want to fit a population dynamics model to the data. The simplest population dynamics model is the exponential population model. This model states that abundance N_t at time t changes as

$$N_{t+1} = N_t \lambda_t$$

for $t = 1, \ldots, T-1$, where T is the number of time steps with observed counts and λ_t are population growth rates. We can extend this and model the unobserved latent growth rates as originating from a common distribution:

$$\lambda_t \sim \text{Normal}\left(\bar{\lambda}, \sigma_\lambda^2\right).$$

where $\bar{\lambda}$ is the average population growth rate, and σ_λ^2 is the variance of the population growth process.

However, for this model to work we also need to specify the intial conditions (i.e., we need to specify $N_{t=1}$). We will specify the initial conditions by estimating $N_{t=1}$ as a parameter. We can then map the ecological process of interest to the observed rat data via:

$$\begin{aligned} y_t &= N_t + \varepsilon_t \\ \varepsilon_t &\sim \text{Normal}\left(0, \sigma_{obs}^2\right). \end{aligned}$$

where y_t is a weekly count of premises with rats originating from the Chicago service database. Note that for populations in varying environments it is better to model dynamics on the log scale and that population dynamicists call the log of λ by a different name, r.

Assignment

Your goal is to build a state-space population dynamics model described in the equaitons above that estimates an *index* of rat abundance using Chicago's 311 service request data. You should report the median +/- 90% credible intervals for your posterior predictive abundance trend (rather than the observed data trend) during 2011-2018 via a plot. The observed data points should be overlaid on the plot as well. Evaluate your model using standard diagnostics. 4 points.

This is a hard model to fit, so you should expect some headaches and warnings from the Stan samplers. In particular, you will probably run into low-ebfmi warnings. Can you get rid of these errors with more informative priors or a reparameterization? If so, do the more informative priors appear to influence your analysis and inferences? What is your interpretation of these warnings and diagnostics as they pertain to the inferences you are trying to make? You should use the vaguest priors possible for your final model. 2 points. Note that if you do this right, you can get away with weakly informative priors and the model will run without any warnings or exceptions.

Your second goal is aimed at improving planning for city workers. Upon examining your plot from part 1, city workers tell you that whenever your model shows the abundance index rising above 2100 they believe the rat levels are "bad". Thus, they want you to predict rat abundance into the future for one year beyond your final data point and to report the probability that the abundance index is > 2100 at the end of the next year. 2 points.

Lastly, there are cool ecological questions that can be examined with your state-space model. Your third goal is to explore your model fit and try to explain some of the patterns you have modeled in a post hoc manner. Also, propose explict ways in which you could improve your model and study these patterns in the future, with special focus on some of the mechanisms underlying the population fluctuations you observed. 2 points.

Submit all code, figures (I shouldn't have to wait for your models to run to see your results), and write-up necessary to support your answers.

Hints: if you are having trouble getting your model running on the real data, try starting out by simulating fake data with parameters similar to your data (or approximately so). This is how I got this model running on these data.