

# Homework 6, due 11:59pm Monday 3/21

STATS/DATASCI 531, Winter 2022

Please submit your homework report to Canvas, including both the Rmarkdown (.Rmd) source file and an HTML file compiled from it. If necessary, you can submit other files needed for your Rmd file to compile, but please do not submit a copy of the data. Your Rmd file can read in the Consett measles data from the internet, via

```
read.csv("https://kingaa.github.io/sbied/stochsim/Measles_Consett_1948.csv")
```

Your report should contain a reference section listing sources. The grader should be able to clearly identify where the sources were used, for example using reference numbers in the text. Anything and anyone consulted while you are working on the homework counts as a source and should be credited. The homework will be graded following the grading scheme in the syllabus (./syllabus.html).

This homework is conceptually quite simple, but involves overcoming various technical hurdles. The hurdles may be overcome quite quickly, or could turn into a longer battle. To make progress on statistical inference for POMP models, we have to solve these underlying computational issues. If technical difficulties arise, do not wait long before asking your colleagues, coming to office hours, or posting on Piazza.

## Installing the **pomp** package

- Computation time is an unavoidable consideration when working with simulation-based inference, for all but small datasets and simple models.
- The **pomp** package therefore allows you to specify the most computationally intensive steps—usually, simulation of the stochastic dynamic system, and evaluation of the measurement density—as snippets of C code.
- Consequently, to use **pomp**, your R program must have access to a C compiler. In addition, **pomp** takes advantage of some Fortran code and therefore requires a Fortran compiler.
- Installing the necessary compilers should be fairly routine, but does involve an extra step beyond the usual installation of an R package, unless you are running the Linux operating system for which they are usually installed by default. Given how fundamental C and Fortran are to scientific computing, it is unfortunate that Mac and Windows do not provide these compilers by default.
- Detailed instructions for installing **pomp** and other software that we will use with it are provided in the following places:
  - The pomp website installation page (<https://kingaa.github.io/pomp/install.html>)
  - Additional instructions on our course website ([https://ionides.github.io/531w22/pomp\\_prep/index.html](https://ionides.github.io/531w22/pomp_prep/index.html))

## Homework questions

### Question 6.1. Exploring behavior of a POMP model: simulating an SIR process.

Write a solution to Exercise 2.3 from Chapter 12 (Simulation of stochastic dynamic models). Note the following:

- We are working toward formal inference for POMP models. Nevertheless, playing with your model by plotting simulations at various parameter values is a useful exercise for getting to understand how your model behaves. It is not enough to know just what parameter value maximizes the likelihood, we also want to understand enough about the model to be able to interpret this MLE. What types of behavior can the model exhibit? How could we describe the behaviors that are consistent with the data?
- Your solution will have to build a copy of the measles model so that you can experiment with it. The Chapter 12 R script (<https://kingaa.github.io/sbied/stochsim/main.R>) may be useful. The script uses Hadley Wickham's `tidyverse` and `ggplot` approach to R. This is a widely used approach, and well worth learning if you have not seen it before, but you may also stick with basic R. To read the script, you will need to know that `x %>% myfunc(y)` is equivalent to `myfunc(x, y)`, so `%>%` is simply a convenient way to chain together functions, where the output of one function is piped into the next. Check that you understand this syntax for the code

```
library(tidyverse)
read_csv("https://kingaa.github.io/sbied/stochsim/Measles_Consett_1948.csv") %>%
  select(week, reports=cases) -> meas
```

- Here, the `tidyverse` version `read_csv` is used in place of the basic R function `read.csv`.
- Worked solutions are linked from the notes, if you get stuck. Ideally, you may like to look at them after solving the homework independently. Your solution is welcome to discuss the relationship between your investigation of the model and the posted solutions.
- Another example of building a pomp model is the Ricker model (<https://kingaa.github.io/sbied/intro/ricker.html>), originally developed to model fish populations and used in this example to model a bird population.
- Various other tutorials and resources are available on the pomp package web site (<https://kingaa.github.io/pomp/docs.html>).

---

## Question 6.2. Modifying a POMP model: Adding a latent period to the SIR model

Write a solution to Exercise 2.4 from Chapter 12 (Simulation of stochastic dynamic models).

You should use Csnippets for this. It should not require techniques beyond those developed in Chapter 12. However, if you are interested in learning more about writing compiled C code for R, you can look at the R extensions manual (<https://cran.r-project.org/doc/manuals/r-release/R-exts.html>). The section on distribution functions (<https://cran.r-project.org/doc/manuals/r-release/R-exts.html#Distribution-functions>) is particularly relevant.

---

## Acknowledgements

The questions derive from material in a short course on Simulation-based Inference for Epidemiological Dynamics (<http://kingaa.github.io/sbied/index.html>)

---