www.simiode.org

SIMIODE
A SYSTEMIC INITIATIVE FOR MODELING INVESTIGATIONS & OPPORTUNITIES WITH DIFFERENTIAL EQUATIONS

STUDENT VERSION

FITTING THE LOTKA-VOLTERRA MODEL
TO TIME SERIES DATA WITH gauseR PACKAGE

Iordanka N. Panayotova
Department of Mathematics
Christopher Newport University
Newport News VA USA

Maila Brucal-Hallare
Department of Mathematics
Norfolk State University
Norfolk VA USA

Abstract: This modeling scenario guides students through the process of fitting the Lotka-Volterra model of two differential equations to a real time series observational data. Students use the capabilities of R and R studio, an integrated development environment for R, and the gauseR package, a collection of tools specialized for fitting Lotka-Volterra models to time series data. Students start the modeling scenario with fitting the logistic growth model to a given set of data while they are provided with the R code. Next, they are guided through the process of fitting the Lotka-Volterra model to a time series data of predator-prey and of two competing species. Throughout this activity, students learn how to extract the model parameters and make predictions for the future behavior of the interacting species using the fitted mathematical model. This modeling scenario can be done in class or out of class or even as a combination of both.

Keywords: Lotka-Volterra model, model fitting, parameter estimation, predator-prey, competing species.

Tags: Time series data, R, gauseR, prediction, Gause, experimental data, differential equations.

SCENARIO DESCRIPTION

Introduction

In the 1920s, the Italian zoologist Umberto D'Ancona (1896–1964) was working on a set of data on the abundance of various species in the fish markets of the *Veneto* region, in Italy. He observed that in the years after WWI, when the commercial fishing had greatly decreased in the Adriatic, there was a dramatic increase in the percentage of sharks and other predator fish among the relatively few catches that were being made.

In this modeling scenario, we will use the capabilities of R, a free software environment for statistical computing and graphics, and R studio, an integrated development environment for R. to accomplish the set goals. We will also use the gauseR package, a collection of tools specialized for fitting Lotka-Volterra models to time series data (Mühlbauer *et al.*, 2020).

Fitting a Mathematical Model to Given Data
Part I: One species logistic growth

Let us start with the case of one species. Consider the data for *Paramecium caudatum* taken from Gause's experimental book (Gause, 1934b). This is a species of unicellular protists in the phylum Ciliophora, and the species is very common and widespread in marine, brackish and freshwater environments. This species was grown in a monoculture by Gause, and the data can be found in the gauseR package under the named **gause_1934_science_f02_03**. This file contains data for both species *Paramecium caudatum* and *Paramecium aurelia* grown in monoculture and in a mixture.

1. Use the given code below to load the package gauseR and plot the data in R (you can use the provided code below). What type of growth do you see for these data?

   ```
   # load package
   require(gauseR)

   # load data
   data(gause_1934_science_f02_03)

   # separate the data related to Paramecium caudatum only (they are grown in
   monoculture)
   Paramecium_caudatum = gause_1934_science_f02_03
   [gause_1934_science_f02_03$Treatment=="Pc", ]

   # data file Paramecium_ caudatum contains columns Day and Volume_Species1
   # plot Volume_Species1 (y-axis) vs Day (x-axis)
   plot (Volume_Species1 ~ Day, data = Paramecium_caudatum)
   ```

   Observe that the plotted data exhibit a logistic type of growth for *Paramecium caudatum*, i.e. first species grow very fast (exponentially), but after certain time they level out even though there is still a certain variability in the species volume. Hence, we will fit a logistic growth model to the data.

   Consider the logistic model equation in the form:

   $$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) = N(r - sN), \quad (3)$$

   where r is the intrinsic growth rate (the net rate at which new individuals are introduced to the population when the population is vanishingly sparse), s is the density dependence parameter (which reflects how the size of the population affects the overall rate), and

4. Extract the parameters r and s from the linear regression model using the command:

```
rsn_parameters = coef ( Model_Pc )
```

5. Next, we will use the extracted parameters as starting values for the nonlinear least square fitting function (**nls**). This part is generally not required as we can just use the obtained parameters to graph the logistic growth function, however we do not have a value for the initial volume. One way to approach this problem is to try different guesses for the initial volume and choose the best fitting curve. However, here we will use another, more elegant way using the command **nls** (nonlinear least squares) to choose the initial values by optimizing the fitted model.

Even though the command **nls** as shown below is sufficient and will give you a solution without knowing how it works, it might be interesting to know that it uses a version of the gradient descent method, in case you have heard of it in your numerical methods class. With the gradient descent method, as minimization algorithm, if algorithm converges, it might find a local minimum but not the (desired) global minimum. That is why we needed the extra step of finding initial values for the nonlinear least squares command to be applied.

Note also that the gradient descent type algorithms are an active area of research in computer science as it has important consequences for machine learning (AI). The better the algorithm, the faster the AI's learning rate.

Using the command **summary,** identify the carrying capacity K for this species. What can we say for the long-term species' behavior?

```
# use the code below to transform parameters into logistic growth parameters
```

```
logistic_pars = c ( r = unname (rsn_pars["(Intercept)"]),

        K=unname ( -rsn_pars["(Intercept)"]/rsn_pars[ "Volume_Species1"]))
```

#fit with **nls** command (nonlinear least squares), using linear model estimates as starting #values for parameters by running the code

```
nls_model = nls (Volume Species1 ~ get_logistic(time = Day, N0, r, K),

        data = Paramecium_caudatum,

        start = c ( N0 =
unname(Paramecium_caudatum$Volume_Species1[which.min(Paramecium_caudatum$
Day)]),   r = unname(logistic_pars["r"]), K = unname(logistic_pars["K"])))
```

```
summary ( nls_model)
```

6. Use the code below to plot the results.

```
# plot results
```

growth in the absence of predator, or in other words, a self-limitation term is included for both species:

$$\frac{dN}{dt} = N(r + aN) + bNP, \quad (5)$$

$$\frac{dP}{dt} = P(e + cP) + dNP, \quad (6)$$

where parameters $r, a, b, c, d,$ and e could be positive or negative.

Remember that, for a single species, parameter s in (1) being negative causes the population to approach a carrying capacity. The same could be expected for system (5)-(6) when parameters b and d are both negative, |one or both species approach a carrying capacity at which the population remains constant, or as constant as the external environmental conditions allow. In such a case, the species are in competition. The opposite of competition is mutualism, where each species enhances rather than inhibits the growth of the other. Both b and d are positive. The last case for these two-species equations is when one interaction parameter is positive (d) (prey species enhances the growth of the predator) and the other is negative (b) (predator species inhibit the growth of the prey) (Lehman *et al.* 2019).

Using the ideas from Part I, let us divide both sides of equations (5)-(6) by N and P respectively, to calculate the per-capita growth rate of each species. As a result, the model can be written as a system of two simple linear equations:

$$\frac{dN}{Ndt} = r + aN + bP, \quad (7)$$

$$\frac{dP}{Pdt} = e + cP + dN. \quad (8)$$

Hence, once we calculate per-capita growth rates for each species, we can use ordinary least squares regression against species abundances to identify the parameters of the mathematical model.

1. Start by exploring the data.
   - Graph the data points in R.
   - Describe the graphs of the prey and predator species – what interesting in their behavior do you observe?

2. Fit the Lotka-Volterra model in the form of equations (7)-(8) to the observational time series data of two interacting species and identify the values of the parameters $a, b, c,$ and d. Use the following outline and the provided R code as a guidance:

   - Open the provided file in R studio.
   - Load the necessary packages: *gauseR* and *deSolve*.
   - Load the data from the file **gause_1934_book_f32**
   - Create the time-lagged abundances for both prey and predator.
   - Use the time-lagged abundances to create per-capita growth rates in the form $\frac{dN}{Ndt}$ for the prey and $\frac{dP}{Pdt}$ for the predator.

column named VolumeSpecies1 and *Paramecium aurelia* data as a column named VolumeSpeceis2.

- Use the R program for the predator-prey system and modify it to fit Lotka-Volterra model to the data for the two competing species.
- Is there any difference in the behavior of the two set of species from Part II and Part III? Describe/explain the difference if there is any.
- What is the long-term behavior of the two competing species? Do they co-exist, or does one species become extinct?

If they co-exist, can you identify their carrying capacity? If so, find it for each species.

References

Gause, G. F. 1934a. Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle for Existence, *Science*, New Series, 79(2036) (Jan. 5, 1934): 16-17. Published by: American Association for the Advancement of Science Stable. http://www.jstor.org/stable/1660803.

Gause, G. F. 1934b. *The Struggle for Existence.* Baltimore: The Williams and Wilkins Company. p. 163.

Lehman, Clarence; Loberg, Shelby; Clark, Adam. 2019. *Quantitative Ecology: A New Unified Approach.* University of Minnesota Libraries Publishing. Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/204551.

Mühlbauer L.K., M.Schulze, W.S.Harpole, and A.T.Clark. 2020. gauseR: Simple methods for fitting Lotka-Volterra models describing Gause's "Struggle for Existence." *Ecology and Evolution.* 10:13275–13283. https://doi.org/10.1002/ece3.6926 .

R project. 2021. https://www.r-project.org/.