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 23606 USA
iordanka.panayotova@cnu.edu

Maila Brucal-Hallare
Department of Mathematics
Norfolk State University
Norfolk VA 23504 USA
mcbrucal-hallare@nsu.edu

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.

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.

D'Ancona hypothesized that this increase in predatory fish was because many Italian fishermen had enlisted in the army and the fishing activity had diminished, allowing the fish to reproduce in peace and long-lived fish to reach maturity. Hence, predatory fish had more prey than before and their population had increased, too.

To justify this hypothesis, D'Ancona asked his father-in-law, the mathematician Vito Volterra, if a mathematical explanation is possible for such puzzling data. He asked Volterra if he could explain mathematically why a decrease in the overall level of fishing would benefit predator fish to a greater extent than it benefited their prey.

Volterra applied differential equations to this interesting biological problem, and his work is now fundamental in every textbook on theoretical ecology. The differential equations model Volterra proposed, which is known as the Lotka-Volterra equations (due to related work of a physical chemist/statistician named Alfred Lotka), investigates two interacting populations that are in a predator-prey relationship and explains D'Ancona's biological phenomenon.

To formulate the problem in the form of differential equations, Volterra made the following assumptions:

(I) Plankton, the food source for the prey fish, is abundant in the Adriatic, so there is negligible competition within the prey population for food. In the absence of predators, the prey population would thus grow, at least in the short term, according to a Malthusian model.

(II) The predator fish are dependent upon the prey fish for their food supply; in the absence of prey, predator population would decline at an exponential rate.

Can you suggest what form will the two differential equations have? Explain the terms included in each differential equation.

Volterra proposed the following system of two differential equations for the population N(t) (the prey fish) and $P(t)$ (sharks and other predator fish) in the Adriatic Sea:

$$\frac{dN}{dt} = aN - bNP, \quad (1)$$

$$\frac{dP}{dt} = -cP + dNP, \quad (2)$$

where a, b, c, and d are positive constants.

These two assumptions explain the "aN" and "$-cP$" terms in these equations. The interaction terms "$-bNP$" and "dNP" occur with a positive sign for the predator P (which benefits by the interaction) and a negative sign for the prey N as the number of "encounters" between predator and prey will be proportional to the product of the two populations; this explains the form of the interaction terms.

One of the most key features of any mathematical model is its predictive power, i.e. we can use the model to predict future behavior of the system under consideration. In the case of the Lotka-Volterra model, it can be used to predict the future prey-predator dynamics of the two interacting species. In this project, we will illustrate this process using data obtained by the Gause's classic competition experiment between two unicellular organisms *Paramecium aurelia* and *Didinium nasutum*, where both species were grown in a mixture (Gause, 1934a).

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:

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

   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 $K = \frac{r}{s}$ is the carrying capacity (the maximum population size) (Lehman *et al*., 2019). Per-capita growth rate then has the following form:

$$\frac{dN}{Ndt} = (r - sN), \qquad (4)$$

where the right-hand side is a linear equation. This form is particularly useful when analyzing empirical data as the parameters can be estimated using ordinary least squares regression of per-capita growth rates against species abundances. Hence, if we calculate the per-capita growth rate using the data, we can fit a regression line to those data. To find the per-capita growth rate data, we will use the difference equations $dN = N_i - N_{i-1}$ and $dt = t_i - t_{i-1}$.

If we fit the regression line to the data, which parameter will the y-intercept represent? How will you find the parameter s using properties of a line? What will the x-intercept represent?

All calculations can be done using R.

2. Run the code below to calculate per-capita growth rate (dNNdt) using the lagged differences and then plot it versus abundances N.

```
# calculate time-lagged abundance using the gauseR command get_lag

lagged_data = get_lag ( x = Paramecium_caudatum$Volume_Species1, time =
Paramecium_caudatum $Day)

# calculate per-capita growth rate (dNNdt) and save it as a column in the data file
#Paramecium_caudatum

Paramecium_caudatum$dNNdt = percap_growth ( x = lagged_data$x, laggedx =
lagged_data$laggedx, dt = lagged_data$dt)

# plot relationship

plot ( Paramecium_caudatum$dNNdt ~ Paramecium_caudatum $Volume_Species1,
xlab="Lagged Abundance (N)", ylab="Per-capita Growth Rate (dN/Ndt)", xlim = c ( 0,
250 ), ylim = c ( 0, 1) )
```

3. Run the code to fit a linear regression model (a line) to the newly obtained per-capita growth rate dNNdt.

```
# fit model to relationship to get dNNdt ~ r + s*N
Model_Pc = lm ( dNNdt ~ Volume_Species1, data = Paramecium_caudatum)

#plot the model as a line in red
abline(Model_Pc, lwd = 2, col = 2)
```

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
   plot (Volume_Species1 ~ Day, data = Paramecium_caudatum, type = "b", ylab = "P.
   caudatum Volume")
   time_seq = seq (0, 30, length=100)
   Ntest = get_logistic(time = timesq, N0=coef(nls_mod)["N0"], r=coef(nls_mod)["r"],
                  K=coef(nls_mod)["K"])
   ```

```
lines ( time_seq, Ntest, col="red")
```

7. We want to be able to quantify how good is the fit of the found model to the given data, rather than just looking at a figure and testing the fit by eye. To do that, we can use the **test_goodness_of_fit**, function from the gauseR package that generates an R-squared-like value, based on the difference between observed and predicted values. Values close to 1 indicate a good fit, whereas values at or below zero indicate a poor fit.
   Calculate the goodness of fit using the code below:

   ```
   #finding goodness of fit
   prediction_short = get_logistic(time = Paramecium_caudatum$Day, N0 =
   coef(nls_mod)["N0"], r = coef(nls_mod)["r"], K=coef(nls_mod)["K"])

   goodness_of_fit = test_goodness_of_fit(observed =
   Paramecium_caudatum$Volume_Species1, predicted = prediction_short)
   ```

   What does the found value tell us for this model?

8. Repeat the analysis for another species *Paramecium aurelia* when grown in monoculture. You can extract data from the same file **gause_1934_science_f02_03** using the command below. Plot the data (please note that the corresponding column has the name Volume_Species2.

   ```
   # separate the data related to Paramecium aurelia only (they are grown in monoculture)
   logistic_data = gause_1934_book_f22[gause_1934_book_f22$Treatment=="Pa",]
   ```

Part II: Two species predator-prey system

Next, we will study the interactions of two species *Paramecium aurelia* as the prey and *Didinium nasutum* as its predator when grown in a mixture. The file that contains the data from the corresponding experiment conducted by Gause is named **gause_1934_book_f32** (from the gauseR package).

When self-limitation is absent for both species, Lotka-Volterra predator-prey models display "neutrally stable" oscillations, meaning that oscillations maintain a fixed amplitude and period. When self-limitation exists for either the prey species or the predator, or both, then oscillations become damped, meaning that over time they decrease in amplitude and period until they reach stable equilibrium values (Lehman *et al*. 2019).

We want to fit the Lotka-Volterra model to the given data. If we plot the data, we will observe that the oscillations for both species are dumped. Hence, to fit the appropriate model, we will consider the generalized Lotka-Volterra model where the prey grows according to the logistic growth in the absence of predator, or in other words, a self-limitation term is included for both species:

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

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

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:

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

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

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.
    - Fit a linear model to $\frac{dN}{Ndt}$ and $\frac{dP}{Pdt}$ using the linear regression command, *lm.*

    If you graph the fitted model versus the given data points, you will observe that this model is not a good fit. The reason for this is because the method that we are using

for estimating parameters is subject to high error and Lotka_Volterra is highly sensitive to even smallest parameter changes (small changes in the parameters values cause large change to the model's predictions).

To get around this problem, we will use an optimizer to directly fit the predicted dynamics to the observed data. This can be done using the **optim** command in R using **lv_optim** function which is part of the gauseR package. Note that this optimizer works in the log space, so we need to set the sign of the parameters (i.e. positive or negative) before we conduct the analysis.

- Extract parameters from both regression models. These parameters will serve as initial values (more precisely, they will provide us the signs of the parameters) for the optimizer function.
- Optimize parameters so that the sum of the squared errors is smallest using the function **optim.** Please note that the optimizer works in log space and parameters' signs must be specified separately.
- Use the found parameters to create the Lotka-Volterra system of two differential equations and solve numerically the system using the *deSolve* package and *ode* command.
- Plot both the given data and the fitted model.

Remark: In general, the difference in fit between these two approaches is driven by the sensitivity of the Lotka-Volterra equations to noisy parameter estimates—even very small changes in parameter values can lead to large changes in model predictions.

3. Use the obtained model to predict future behavior of the interactions between the two species for another 10 time steps.

4. Reflect on your findings.

Part III: Two competing species system.

Finally, let us track the competitive interactions between *Paramecium aurelia* and *Paramecium caudatum*. These two unicellular species were grown in a mixture by Gause (Gause, 1934b) and the observational data are given in the same file that we used in Part I: **gause_1934_science_f02_03.** To extract the data, use the command:

```
PA_PAC_data =  gause_1934_science_f02_03
[gause_1934_science_f02_03$Treatment=="Mixture",  ]
```

This will give you a data file PA_PC_data which will contain *Paramecium caudatum* data as a 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/.