Logistic Growth Model Exercise with Paramecium Data

John M. Drake

November 4, 2024

Introduction

The purpose of this exercise is to explore a common application of statistical reasoning – *model fitting* – using our understanding of likelihood theory and different statistical distributions.

Logistic Growth Model Exercise with Paramecium Data

In this exercise, you will analyze a dataset of *Paramecium* population counts over time, applying logistic growth modeling and various fitting techniques to estimate parameters. Complete each task below, paying attention to the details in each step. The data come from an experiment by Veilleux (1976, 1979). Data file `Veilleux1.csv` contains a time series of population of *Paramecium aurelia* densities in culture as it smoothly approached an equilibrium.

Task 1: Load and Plot the Data

1. Load the provided Paramecium data. Make sure it includes columns for time (days) and population count.
2. Plot the data to visualize population growth over time.

Task 2: Write a Differential Equation Solver

1. Implement a differential equation solver in R for the logistic growth model:

   $$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)$$

   where:

   - N is the population at time t,
   - r is the intrinsic growth rate, and
   - K is the carrying capacity.

2. Test your solver with initial guesses for r and K to ensure it works as expected.

Hint: Consider using `ode()` from the `deSolve` package. There is an example with two different models in the file `demonstration.pdf`.

Task 3: Estimate Parameters r and K

1. Fit the logistic growth model to the data by estimating the parameters r and K using maximum likelihood for the Gaussian model. Recall the the maximum likelihood estimators for this error distribution are the same as the minimizing sum of squared errors.
2. Report your fitted values for r and K.

Hint: Use `optim` for numerical optimization.

Learning outcome: In this step, we interpret the solution of a differential equation as the mean of a statistical distribution of possible observations and use that interpretation to find the unknown parameter values.

Task 4: Find the Likelihood Interval for r and K (Gaussian Errors)

1. Assuming Gaussian errors, calculate the likelihood interval for r and K.
2. Summarize and plot these intervals.

Hint: Use the likelihood profile method for precise interval estimation.

Learning outcome: In this step, we learn to compute a confidence interval.

Task 5: Plot a Heatmap of the Likelihood with a 1.96 Confidence Ellipse

1. Create a heatmap of the likelihood for combinations of r and K.
2. Overlay a 1.96 confidence ellipse on this heatmap to visualize the region.

Hint: The `ggplot2` package and `geom_contour()` can be useful here.

Learning outcome: In this step, we learn about parameter correlations and how this might affect hypothesis testing.

Task 6: Repeat Tasks 3-4 for Days 3 to 6.5

1. Filter the data to include only measurements taken between days 3 and 6.5.
2. Repeat the parameter fitting and likelihood interval estimation (Tasks 3-4) on this filtered data.

Learning outcome: In this step, we learn that the information contained in the data depend on more than just the quantity of data available.

Task 7: Using a Gamma Distribution

1. Modify your model to assume a Gamma distribution of errors rather than Gaussian.
2. Repeat parameter estimation and likelihood interval estimation (Tasks 3-4) with this new assumption.

Hint: Look use the function `dgamma`, which you can incorporate into your fitting function.

Learning outcome: In this step, we generalize from our study of Gaussian distributions to arbitrary (and possibly more appropriate) distributions.

Task 8: Rounding Counts to Nearest Integer and Fitting Poisson and Negative Binomial Distributions

1. Round the observed population counts to the nearest integer.
2. Assuming these rounded counts follow either a Poisson or Negative Binomial distribution, repeat Tasks 3-4.
3. Compare the fit of each model and interpret any differences.

Learning outcome: In this step, we examine discrete distributions and the possibility of over dispersion.