Hierarchical_AIML_Book

William M. Murrah

2025-02-23

Table of contents

Preface

This is a Quarto book.

To learn more about Quarto books visit https://quarto.org/docs/books.

1 Introduction

2 Summary

In summary, this book has no content whatsoever.

3 Logistic, Gompertz, and Richards Growth Models: Understanding Theoretical Parameters

3.1 Overview

This chapter provides a hands-on approach to modeling growth curves using logistic, Gompertz, and Richards functions, focusing on the interpretability of theoretical parameters such as the asymptote, inflection point, growth rate, and shape parameter. The chapter is structured in parallel for R (using the brms package) and Python (using PyMC3 and TensorFlow Probability) to facilitate learning across both programming environments.

3.2 Objectives

1. Understand the theoretical background and functional form equations for logistic, Gompertz, and Richards curves.
2. Learn how to specify, fit, and interpret nonlinear growth models in R and Python.
3. Explore differences in growth parameters by demographic groups using hierarchical modeling.
4. Compare models using Bayesian model comparison techniques.
5. Extend analyses to Gaussian Processes and Neural Networks while retaining interpretability.

3.3 Structure of the Chapter

1. **Introduction and Theoretical Background**
2. **Data Preparation and Exploration**
3. **Logistic Growth Modeling**
4. **Gompertz Growth Modeling**
5. **Richards Curve Modeling**
6. **Comparison and Interpretation**
7. **Extensions: Nonparametric and Neural Networks**

8. **Conclusions and Insights**
9. **References and Further Reading**

3.4 Requirements

3.4.1 R Packages

- `tidyverse`, `data.table`, `ggplot2`, `EdSurvey`, `mice`
- `brms`, `loo`, `bayesplot`, `cmdstanr`

3.4.2 Python Packages

- `pandas`, `numpy`, `matplotlib`, `seaborn`
- `pymc3`, `arviz`, `tensorflow-probability`
- Optional for Advanced Modeling: `pyro-ppl`, `gpytorch`

3.5 Part 1: Introduction and Theoretical Background

3.5.1 Motivating Example: Educational Disparities and the ECLS-K Dataset

Educational disparities in academic achievement are a persistent issue with far-reaching implications for economic and societal outcomes. Students from lower socioeconomic backgrounds and marginalized racial groups often face systemic disadvantages, resulting in achievement gaps that influence lifelong opportunities, including access to higher education, employment prospects, and income potential. Understanding and intervening in educational systems to alleviate these disparities requires a nuanced examination of how academic growth unfolds across different demographic groups.

The Early Childhood Longitudinal Study Kindergarten Class (ECLS-K) dataset provides a rich source of longitudinal data to investigate these issues. By modeling growth trajectories in academic achievement, researchers can identify critical developmental periods and examine how factors such as socioeconomic status (SES) and race influence growth parameters. In this context, SES and race serve as proxies for access to resources and exposure to societal factors, including school quality, neighborhood environment, and family support systems.

3.5.2 Research Questions

This chapter is motivated by the following research questions:
- How do academic achievement trajectories differ by SES and race?
- Are there gender differences in academic achievement trajectories?
- How do academic achievement trajectories differ by SES and race?
- When do the most rapid changes in academic achievement occur for different demographic groups?
- Do students from different SES or racial backgrounds reach different levels of achievement asymptotically?

3.5.3 Why Use Functional Form Models?

Functional form models such as logistic, Gompertz, and Richards curves provide a powerful framework for examining these research questions. Unlike more flexible machine learning models, functional form models are interpretable, with parameters directly linked to theoretical constructs:
- **Inflection Point**: The timing of the most rapid change, indicating critical periods for intervention.
- **Asymptote**: The maximum achievement level, highlighting disparities in long-term educational outcomes.
- **Growth Rate**: The speed of academic progress, providing insights into the effectiveness of educational environments.
- **Shape Parameter (Richards Curve)**: Captures asymmetry in growth, potentially revealing differences in early vs. late acceleration of achievement.

3.5.4 Theoretical and Practical Implications

Using functional form models not only enhances interpretability but also supports theoretical development by linking empirical findings to educational and cognitive development theories. For example, differences in inflection points may indicate developmental milestones influenced by resource access or educational quality, while variations in asymptotes may reflect systemic inequalities. Additionally, the parametric nature of these models facilitates cross-study comparisons, enabling cumulative knowledge building.

By focusing on functional form models, this chapter aims to provide researchers with tools that offer both flexibility in modeling nonlinear growth and interpretability essential for informing educational policy and practice.

3.5.5 Overview of Nonlinear Modeling Methods

Nonlinear modeling is a powerful approach to understanding complex growth patterns and developmental trajectories. Traditional linear models are often inadequate for capturing the dynamic, non-constant rates of change observed in natural phenomena, such as cognitive development and academic achievement. To address this, various nonlinear modeling techniques have been developed, ranging from spline models and polynomial regression to more flexible machine learning approaches, such as neural networks and Gaussian processes.

Spline Models and Polynomial Regression: Spline models and polynomial regressions provide flexible tools for modeling nonlinearity by segmenting data into pieces or by fitting higher-order polynomials. However, these approaches often suffer from overfitting and lack clear interpretability of parameters related to growth processes.

Machine Learning Approaches: Modern machine learning methods, such as Random Forests, Gradient Boosting Machines, and Deep Neural Networks, can capture complex nonlinear relationships and interactions. Although these models offer high predictive accuracy, they are often criticized for their lack of interpretability and theoretical grounding, particularly in educational and cognitive development research.

3.5.6 Why Focus on Functional Form Models?

Unlike purely data-driven approaches, functional form models like the logistic, Gompertz, and Richards curves offer a balance between flexibility and interpretability. These models are grounded in theoretical equations that describe growth processes with interpretable parameters, such as the asymptote (maximum level of achievement), the inflection point (timing of most rapid growth), and the growth rate (speed of development). Importantly, the parameters of these models are directly related to theoretical constructs, making them highly relevant for hypothesis testing and theory development.

3.5.7 Interoperability and Implications for Theory Development

A key advantage of using functional form models is the interoperability of their parameters across different studies and populations. For example, comparing the inflection point across demographic groups provides insights into differences in the timing of developmental milestones. This interpretability fosters theoretical advancements, enabling researchers to link empirical findings to cognitive and educational theories. Additionally, the use of parametric functional forms facilitates replication and comparison across studies, contributing to a cumulative body of knowledge.

By focusing on logistic, Gompertz, and Richards curves, this chapter aims to equip researchers with tools that offer both flexibility in modeling nonlinear growth and interpretability crucial for theory development.

3.5.8 The Logistic Curve

- Equation:

$$y(t) = \frac{\alpha}{1 + \exp(-\beta(t - \gamma))}$$

- Parameters:
  - α: Asymptote (Maximum achievement level)
  - β: Growth rate
  - γ: Inflection point (Timing of most rapid development)

3.5.9 The Gompertz Curve

- Equation:

$$y(t) = \alpha \exp(-\exp(-\beta(t - \gamma)))$$

- Parameters:
  - α: Asymptote
  - β: Growth rate
  - γ: Inflection point

3.5.10 The Richards Curve

- Equation:

$$y(t) = \frac{\alpha}{(1 + \delta \exp(-\beta(t - \gamma)))^{\frac{1}{\delta}}}$$

- Parameters:
  - α: Asymptote
  - β: Growth rate
  - γ: Inflection point
  - δ: Shape parameter (Asymmetry)

3.6 Part 2: Data Preparation and Exploration

3.6.1 Data Source

- Using the 2011 ECLS-K dataset for educational achievement.
- Data will be prepared and explored using:
  - **R**: `tidyverse`, `EdSurvey`
  - **Python**: `pandas`, `seaborn`

3.6.2 Preprocessing Steps

- Loading data and initial exploration.
- Handling missing data using multiple imputation.
- Visualizing growth trajectories by demographic groups.

3.7 Next Steps

The next section will cover detailed coding examples for Logistic Growth Modeling in both R and Python. Stay tuned!

4 Logistic Growth Modeling: Understanding Theoretical Parameters

4.1 Overview

This tutorial focuses on modeling growth curves using the Logistic function, emphasizing the interpretability of theoretical parameters such as the asymptote, inflection point, and growth rate. It builds on the foundational knowledge from the introductory tutorial on nonlinear growth models, with a hands-on approach for implementation in both R (using the brms package) and Python (using PyMC3).

4.2 Objectives

1. Understand the functional form and theoretical parameters of the Logistic curve.
2. Specify, fit, and interpret Logistic growth models in R and Python.
3. Investigate differences in growth parameters by demographic groups, including SES, race, and gender.
4. Compare models using Bayesian model comparison techniques and evaluate model fit with posterior predictive checks.

4.3 Structure of the Tutorial

1. **Introduction to the Logistic Curve**
2. **Model Specification**
3. **Implementation in R using brms**
4. **Implementation in Python using PyMC3**
5. **Parameter Interpretation and Visualization**
6. **Model Comparison and Diagnostics**
7. **Extensions and Advanced Topics**
8. **Conclusions and Insights**
9. **References and Further Reading**