MT2002: Statistical Modeling

Introduction, Course Overview, Policies

Muhammad Almas Khan
muhammad.almas@nu.nu.pk

FAST-NUCES

January 12, 2025

Course Outline

1. Course Overview
2. Why Study This Course?
3. Rules of the Game
4. Course Goals
5. Tools and Libraries
6. Why PyMC?
7. Course Summary
8. Assessment and Grading
9. Tools and Resources
10. Course Outline Image

What is Statistical Modeling?

- Understanding and analyzing data using statistical methods.
- Building models to represent real-world phenomena.
- Applications in science, engineering, and business.

Why Study This Course?

- **Purpose:**
  - Equip students with skills to analyze and interpret data effectively.
  - Prepare for careers in data science, machine learning, and research.
- **Importance:**
  - Statistical modeling is essential for data-driven decision-making.
  - Bridges the gap between theoretical concepts and real-world applications.

Rules of the Game

- **Punctuality:** Be on time and silence your phone.
- **Focus:** Engage with the lecture, not your friends.
- **Participation:** Ask questions and respect opinions.
- **Attendance:** Attendance will be recorded at the start of class. After class, ensure your presence is correctly marked on Flex and address any discrepancies promptly, don't expect any correction after class time.
- **Evaluation:** Address discrepancies within one day after updated on flex.
- **Group Work:** Disjoint groups of 3 or 4 for discussions.

What Will You Learn?

- Fundamental concepts in statistical modeling.
- Hands-on experience with tools like Python and PyMC.
- Real-world applications of statistical techniques.

Tools You Will Use

- **Python Programming Language:** The primary language for this course.
- **PyMC:** A library for probabilistic programming and Bayesian inference.
- **ArviZ:** A library for exploratory analysis of Bayesian models.
- **Plotting Libraries:** Tools like Matplotlib and Seaborn for data visualization.

Why PyMC?

- **Intuitive and Flexible:** Simplifies Bayesian modeling and inference.
- **Probabilistic Power:** Supports building complex statistical models with ease.
- **Visualization Support:** Integrates seamlessly with ArviZ for insightful visualizations.
- **Real-World Applications:** Widely used in finance, healthcare, and research for uncertainty modeling.

About This Course

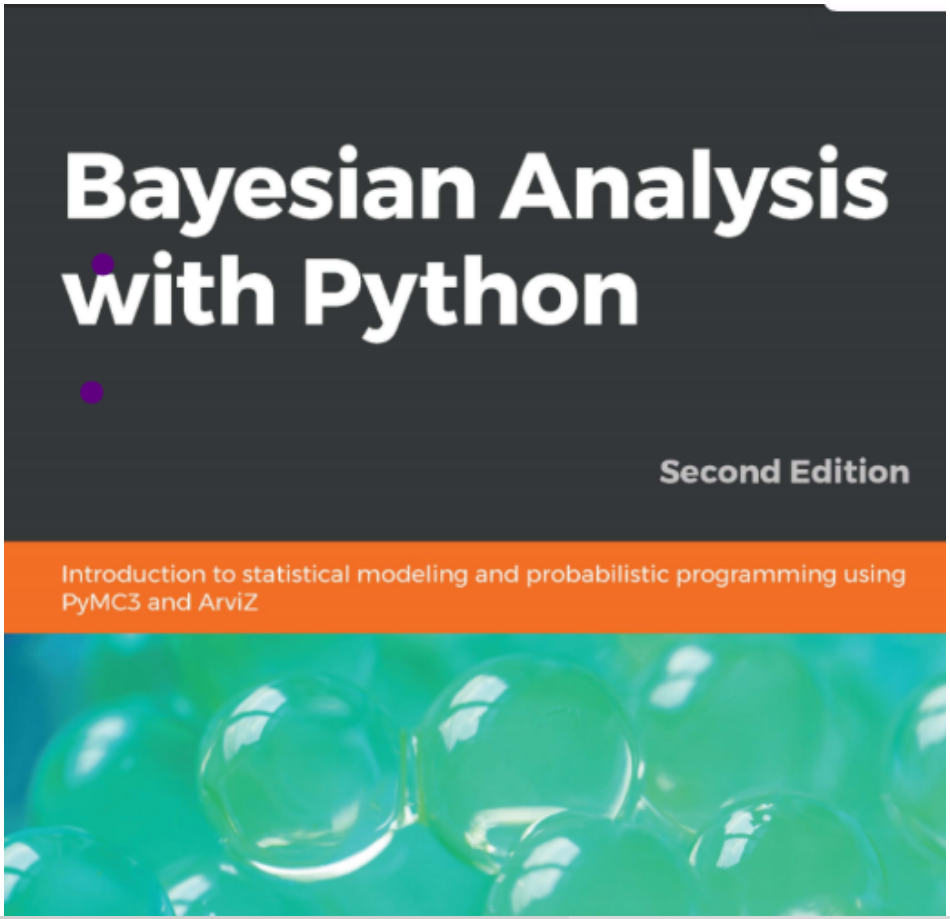
- **What is this course for?**
  - Developing statistical thinking and modeling skills.
- **What will you learn?**
  - Fundamental concepts in statistical modeling, Bayesian inference, and visualization.
- **Our Approach:**
  - Quizzes: 4 to 8.
  - Assignments: 4 to 8.
  - 2 sessional exams and 1 final exam.

How Will You Be Graded?

- Homework Assignments: 10%
- Quizzes: 10%
- Sessional: (15+15=30)%
- Final Exam: 50%

Resources for Success

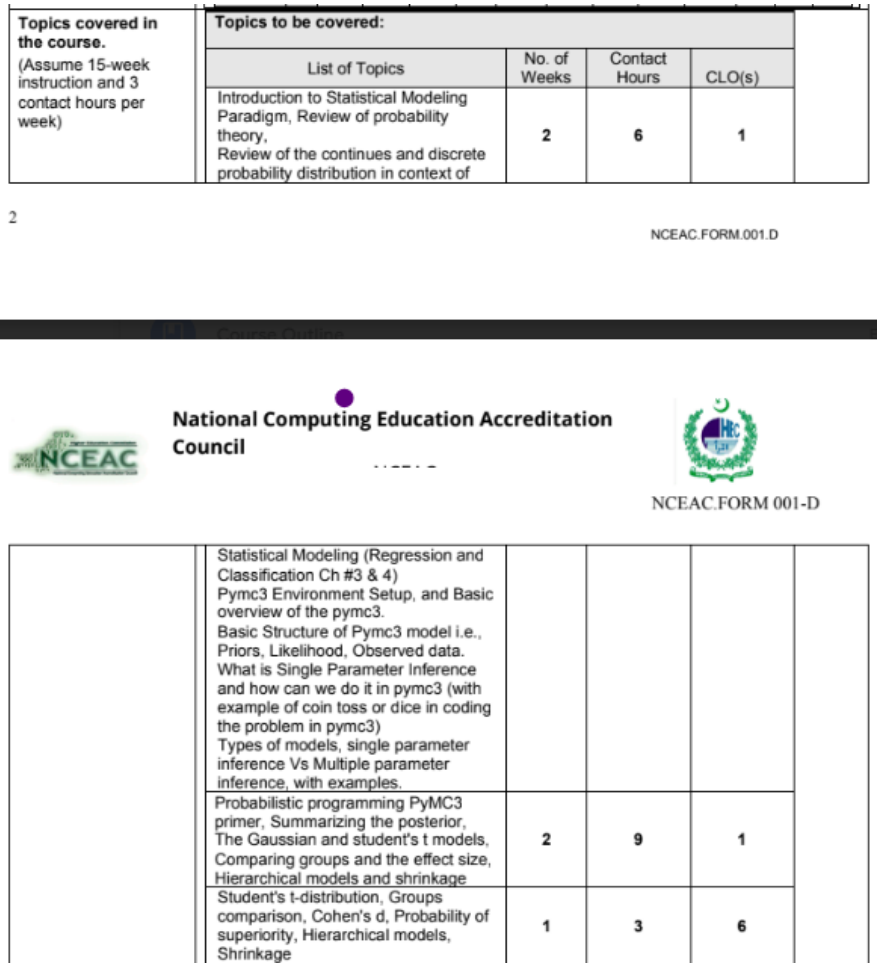
- Textbook: *Introduction to Statistical Modeling*.
- Python (with libraries for probablistic programming).
- Office Hours: Will share on GCR.



GitHub Repository:

`https://github.com/aloctavo`

Course Outline



Topics covered in the course. (Assume 15-week instruction and 3 contact hours per week)	Topics to be covered:			
	List of Topics	No. of Weeks	Contact Hours	CLO(s)
	Introduction to Statistical Modeling Paradigm, Review of probability theory, Review of the continues and discrete probability distribution in context of	2	6	1

2

NCEAC.FORM.001.D

National Computing Education Accreditation Council

NCEAC.FORM 001-D

	Statistical Modeling (Regression and Classification Ch #3 & 4) Pymc3 Environment Setup, and Basic overview of the pymc3. Basic Structure of Pymc3 model i.e., Priors, Likelihood, Observed data. What is Single Parameter Inference and how can we do it in pymc3 (with example of coin toss or dice in coding the problem in pymc3) Types of models, single parameter inference Vs Multiple parameter inference, with examples.			
	Probabilistic programming PyMC3 primer, Summarizing the posterior, The Gaussian and student's t models, Comparing groups and the effect size, Hierarchical models and shrinkage	2	9	1
	Student's t-distribution, Groups comparison, Cohen's d, Probability of superiority, Hierarchical models, Shrinkage	1	3	6

Course Outline Details:

- Week 1-5: Introduction to Statistical Modeling, Pymc, priors and likelihood, inference Vs robust inference
- Week 7-13: Bayesian Regression and classification
- Week 14-16: Model Comparison and some case studies (AB testings)

Let's Get Started with setting up environment for pymc!

Questions? Feel free to ask.

Welcome to MT2002: Statistical Modeling!