Programming Probabilistically

Chapter 2 Overview

Muhammad Almas Khan

FAST-NUCES

January 13, 2025

What is Probabilistic Programming?

- Combines programming and probability to handle uncertainty.
- Enables building complex statistical models using simple code.
- Key library: **PyMC3**, a powerful tool for Bayesian modeling.

PyMC3 Primer

Steps in PyMC3 Modeling:

- **Model Specification:** Define priors, likelihoods, and observed data.
- **Inference:** Use PyMC3's sampling algorithms to compute the posterior.
- **Example:** Flipping coins the PyMC3 way.

Summarizing the Posterior

- **Posterior-based Decisions:** Make decisions based on the posterior.
- **ROPE (Region of Practical Equivalence):** Helps determine if an effect is practically zero.
- **Loss Functions:** Assess costs of incorrect decisions.

Gaussian Inferences

- **Gaussian Priors and Likelihoods:** Ideal for normally distributed data.
- **Robust Inferences:** Handle outliers using distributions like Student-t.

Groups Comparison

- **Cohen's d:** Quantify the effect size between two groups.
- **Probability of Superiority:** Estimate the likelihood that one group is superior to another.
- **Example Dataset:** *The Tips Dataset* for real-world analysis.

Hierarchical Models

- **Shrinkage:** Stabilize parameter estimates by pooling data across groups.
- **Applications:** Handle multi-level data with group dependencies.
- **Example:** A detailed demonstration to illustrate the benefits.

Summary

Key Takeaways:

- Probabilistic programming bridges programming and uncertainty modeling.
- PyMC3 makes Bayesian inference accessible and powerful.
- Posterior summarization enables practical decision-making (e.g., ROPE, loss functions).
- Gaussian and robust models address different data characteristics.
- Hierarchical models leverage shared information for better estimates.

Exercises

Try the following:

- Build a PyMC3 model for coin flips using Beta priors.
- Compare groups using Cohen's d on the Tips dataset.
- Apply hierarchical modeling to multi-level data.