Bayesian Analysis and Inference

Introduction

Bayesian analysis is a powerful statistical methodology that updates the probability for a hypothesis as more evidence becomes available. This report explores various Bayesian modeling techniques, including Gaussian distributions, Student's t-distributions, hierarchical models, and Bayesian linear regression using PyMC3.

Gaussian Distribution Modeling

Overview

The Gaussian (Normal) distribution is commonly used to approximate many real-world data distributions due to the Central Limit Theorem and its mathematical tractability.

Probability Density Function

The probability density function (PDF) of a normal distribution is:

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ mu is the mean and σ sigma is the standard deviation.

PyMC3 Implementation

The dataset of chemical shifts was analyzed with a Gaussian distribution model. The mean and standard deviation were estimated using uniform and half-normal priors, respectively.

Robust Models with a Student's t-Distribution

Motivation

Gaussian models may not handle outliers well. The Student's t-distribution, which has heavier tails, provides a more robust alternative.

Model Implementation

The model was updated with an additional ν\nu parameter representing degrees of freedom, allowing better adaptability to outliers.

Hierarchical Models

Introduction

Hierarchical models allow partial pooling of data, balancing individual group estimates with global information.

Problem Statement

The analysis involved water quality measurement across three districts, where contamination levels were modeled using a Bernoulli distribution with hierarchical priors.

Sampling Model

A beta distribution was used to model the prior probability of acceptable contamination levels, with hyperpriors governing the group-level parameters.

Bayesian Linear Regression

Data Generation

Synthetic data was generated with a known linear relationship and added noise to simulate real-world variability.

Model Specification

A Bayesian linear regression model was implemented with the following priors:

- Intercept (α\alpha): Normal(0, 10)
- Slope (β\beta): Normal(0, 1)
- Error term (ϵ\epsilon): Half-Cauchy(5)

Inference and Posterior Predictions

Using Markov Chain Monte Carlo (MCMC) sampling, posterior distributions were estimated. Pair plots and posterior predictive checks validated model assumptions.

Conclusion

Bayesian analysis provides a flexible and interpretable approach to statistical modeling. The models presented demonstrated its application to normal distributions, robust inference, hierarchical analysis, and regression. Future work may explore Bayesian nonparametrics and advanced hierarchical structures.