Faculty of Sciences
Faculty of Mathematics, Statistics, and Computer Science

Stochastic Process MiniProject 2

Suppose we are working with a simple Bayesian linear model. A Bayesian model is one in which the parameters themselves are treated as random variables. Our beliefs about these parameters are updated after observing data using the posterior distribution.

Given a prior distribution $p(\theta)$ and likelihood $p(y \mid \theta)$, the posterior distribution is given by:

$$p(\theta \mid y) \propto p(\theta) \cdot p(y \mid \theta)$$

In most real-world cases, we do not have a closed-form expression for the posterior distribution. This is often because the posterior includes an intractable normalizing constant. However, for tasks such as parameter estimation, prediction, or uncertainty quantification, we still need access to the distribution of θ. To do this, we use sampling-based methods like Metropolis-Hastings to approximate the posterior and draw representative samples from it.

Introduction

Suppose we have a non-negative function $f(x)$, and we wish to construct a probability distribution from it. To do so, we define the corresponding probability density function as:

$$p(x) = \frac{f(x)}{\int f(x)\, dx}$$

Sampling from this distribution becomes challenging when the integral in the denominator is difficult or intractable to compute analytically. This is a common problem in many fields, including Bayesian statistics and statistical physics.

To overcome this, a class of algorithms known as **Markov Chain Monte Carlo (MCMC)** methods has been developed. These algorithms allow us to generate samples from complex probability distributions without requiring explicit computation of the normalizing constant. One of the foundational MCMC algorithms is the **Metropolis-Hastings algorithm**, which we will explore and implement in this project.

MCMC and the Metropolis-Hastings Algorithm

Investigate what Markov Chain Monte Carlo (MCMC) methods are and how they are used to sample from complex probability distributions.

Then, study the Metropolis-Hastings algorithm in detail — specifically how it relates to Markov chains and the concept of stationary distributions.

Implementation

This section contains three progressively challenging tasks that require implementing the Metropolis-Hastings algorithm in practical settings. All implementations must include appropriate visualizations (e.g., histograms, trace plots), empirical summaries (mean, variance, etc.), and interpretation of the results.

Task 1: Sampling from a Custom Exam Score Distribution

Suppose that the scores on a particular exam are modeled by the following unnormalized probability function:

$$f(x) = \exp\left(-\frac{(x-75)^4}{5000}\right) + 0.01\sin(0.3x), \quad \text{for } x \in [0, 100]$$

This function is not normalized and has a multi-modal shape due to the sine term. Your task:

- Implement the Metropolis-Hastings algorithm to draw samples from the corresponding distribution.
- Compute the sample mean and variance using the drawn samples.

Task 2: Estimating the Normalizing Constant

Using the same function from Task 1, estimate the normalizing constant:

$$Z = \int_0^{100} f(x)\,dx$$

Use your samples from Task 1 and explain your method clearly.

- Report your estimate for Z.
- Explain how your method approximates this integral and any assumptions involved.

Task 3: Bayesian Parameter Estimation in a Simple Linear Model

In this Bayesian model, the data is generated according to the following likelihood:

$$y_i \mid \theta \sim \mathcal{N}(\theta x_i, 1)$$

You are also given a prior belief about θ expressed as:

$$p(\theta) \propto \exp(-\theta^4 + 2\theta)$$

Your task is to construct the posterior distribution of θ given the observed data. The posterior combines the likelihood of the observed data and the prior, allowing you to update your belief about θ in light of the data.

- Derive and express the unnormalized posterior distribution of θ up to a constant.
- Implement the Metropolis-Hastings algorithm to generate samples from this posterior.
- Visualize your results: plot both a histogram of the samples (posterior approximation).
- Compute and report the posterior mean.

This task connects MCMC methods with Bayesian inference and demonstrates how Metropolis-Hastings enables sampling from posterior distributions that are not easy to sample from directly.