Prior and Posterior Distributions

Antonio Aguirre

Introduction

Statistical modeling aims to describe the behavior of real-world phenomena through mathematical functions. Such functions are assumed to be parametric, meaning they depend on unknown parameters. These parameters determine the underlying structure of data, and **statistical inference** aims to estimate them from observations.

There are two main philosophical approaches to inference, which differ in their treatment of randomness:

- **Classical (Frequentist) Statistics:** Treats parameters as fixed but unknown quantities. Randomness arises solely from the sampling procedure, meaning probability is interpreted as a measure of long-run relative frequency.
- **Bayesian Statistics:** Treats parameters as unknown but also as random variables. Here, randomness arises both from the sampling process and from the uncertainty in the parameter itself. Bayesian probability is a measure of **credibility** or **belief**, not just relative frequency.

Prior Distribution

A **prior distribution** represents our beliefs about an unknown parameter θ before observing any data. It encodes any previous knowledge or assumptions about θ. The prior is denoted as:

$$\pi(\theta) \tag{1}$$

where $\pi(\theta)$ is a probability distribution over possible values of θ.

Key Intuition

Unlike the frequentist approach, Bayesian statistics considers probability as a measure of belief or credibility rather than a long-run frequency of repeated experiments.

Example

Suppose we are estimating the proportion p of defective items in a factory's production line. A classical approach would use a sample proportion to estimate p, assuming it is fixed but unknown. In Bayesian analysis, we start with a prior distribution, such as $p \sim \text{Beta}(2, 8)$, reflecting prior knowledge that defects are typically rare.

Posterior Distribution

The **posterior distribution** represents our updated belief about θ after observing data D. It combines the prior information with the likelihood of the observed data through Bayes' Theorem:

$$\pi(\theta \mid D) = \frac{L(D \mid \theta)\pi(\theta)}{\int L(D \mid \theta)\pi(\theta)d\theta} \tag{2}$$

where:

- $\pi(\theta \mid D)$ is the posterior distribution,
- $L(D \mid \theta)$ is the likelihood of the data given θ,
- $\pi(\theta)$ is the prior,
- The denominator ensures the posterior integrates to 1.

Key Insight

The posterior updates our prior beliefs using observed data, giving us a refined estimate of the parameter.

Comparison to the Classical Approach

In classical (frequentist) statistics, inference is typically based only on the likelihood function:

$$L(D \mid \theta) \tag{3}$$

A parameter estimate, such as the maximum likelihood estimate (MLE), is obtained by maximizing this likelihood. However, this approach does not incorporate prior information.

Main Difference

Frequentist statistics treat parameters as unknown but fixed, whereas Bayesian statistics treat parameters as unknown and random. This broader Bayesian perspective allows for probability to quantify belief, rather than just long-run frequency.

Final Discussion

The Bayesian approach provides a coherent way to update beliefs as new data becomes available, making it especially useful in situations with prior knowledge or limited data. The choice of prior can strongly influence results, making sensitivity analysis an important part of Bayesian modeling.

However, Bayesian inference also has some downsides:

- **Computational Complexity:** Many Bayesian models do not have closed-form solutions, requiring numerical methods such as Markov Chain Monte Carlo (MCMC) or Variational Inference. These methods can be computationally expensive and slow, especially for high-dimensional models.
- **Prior Sensitivity:** The choice of prior can significantly impact the results, which can be problematic when prior knowledge is weak or incorrect.

Despite these challenges, Bayesian methods remain powerful and widely used in statistical modeling, particularly in fields where incorporating prior knowledge is essential.