Bayesian Inference for the Parameters of the Normal Distribution

December 11, 2024

Introduction

- Bayesian inference provides a framework for updating beliefs using data.
- Focus: Parameters of the Normal distribution.
- Cases to explore:
  - Unknown mean, known variance.
  - Both mean and variance unknown.

Normal Distribution Basics

- Likelihood for n observations $\boldsymbol{x} = (x_1, \ldots, x_n)$:

$$f(\boldsymbol{x} \mid \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right).$$

- Mean: μ, Variance: σ^2.
- Bayesian approach: Combine likelihood with prior beliefs.

Unknown Mean, Known Variance

- Assume σ^2 is known, μ is unknown.
- Prior for μ: Normal $\mathcal{N}(\mu_0, \tau^2)$.
- Posterior:

  $$\mu \mid \boldsymbol{x}, \sigma^2 \sim \mathcal{N}(\mu_n, \tau_n^2),$$

  where

  $$\mu_n = \frac{\frac{\mu_0}{\tau^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \quad \tau_n^2 = \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}.$$

- Notice that

  $$\mu_n = \mu_0 \frac{\frac{1}{\tau^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} + \bar{x} \frac{\frac{n}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \quad \tau_n^2 = \frac{\frac{\sigma^2}{n} \tau^2}{\tau^2 + \frac{\sigma^2}{n}}.$$

If the prior mean is very precise relative to the data mean, then we should weight it highly. Alternatively, if the data mean is more precise, then it should be assigned a larger weight. Also, posterior variance $<$ prior variance and variance of the data mean

Interpretation

- Posterior mean μ_n: Weighted average of prior mean μ_0 and sample mean $\bar{X}$.
- Posterior variance τ_n^2: Reflects reduced uncertainty after observing the data.
- As $n \to \infty$, $\mu_n \to \bar{X}$ and $\tau_n^2 \to 0$: Prior influence diminishes with large data.

Marginal likelihood calculations

- Likelihood of the data given μ:

$$p(\boldsymbol{X} \mid \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2}\right)$$

  Simplified for the sample mean $\bar{X}$:

$$p(\boldsymbol{X} \mid \mu) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{X} - \mu)^2\right)$$

- Prior on μ:

$$p(\mu) = \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\mu - \mu_0)^2}{2\tau^2}\right)$$

Marginal Likelihood Derivation

- Marginal likelihood:

$$p(\boldsymbol{X}) = \int p(\boldsymbol{X} \mid \mu) p(\mu) \, d\mu$$

- Substituting:

$$p(\boldsymbol{X}) \propto \int \exp\left(-\frac{n}{2\sigma^2}(\bar{X} - \mu)^2 - \frac{1}{2\tau^2}(\mu - \mu_0)^2\right) d\mu$$

- Combine terms inside the exponent:

$$-\frac{1}{2}\left[\frac{n}{\sigma^2}(\bar{X} - \mu)^2 + \frac{1}{\tau^2}(\mu - \mu_0)^2\right]$$

becomes:

$$-\frac{1}{2}\frac{1}{\tau_n^2}(\mu - \mu_n)^2 - \frac{1}{2}\frac{n\tau^2 + \sigma^2}{\sigma^2\tau^2}$$

where:

$$\tau_n^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}, \quad \mu_n = \tau_n^2\left(\frac{n\bar{X}}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)$$

Marginal Likelihood Result

- The integral evaluates to:

$$p(\boldsymbol{X}) = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\sigma^2 \tau^2}{n\tau^2 + \sigma^2}} \exp\left(-\frac{1}{2} \frac{n\tau^2 + \sigma^2}{\sigma^2 \tau^2} (\bar{X} - \mu_0)^2 \right)$$

- Interpretation:
  - Marginal likelihood penalizes models based on how well μ_0 aligns with $\bar{X}$.
  - Balances model complexity (τ^2) and data fit (σ^2).

Example: Numerical Calculation

- Given:
  - $n = 10$, $\bar{X} = 5$, $\sigma^2 = 4$.
  - Prior: $\mu_0 = 0$, $\tau^2 = 1$.
- Compute:

$$\tau_n^2 = \left(\frac{10}{4} + \frac{1}{1}\right)^{-1} = \frac{4}{6.5}, \quad \mu_n = \frac{4}{6.5}\left(\frac{10 \cdot 5}{4}\right) = \frac{50}{6.5}$$

- Marginal likelihood can then be numerically evaluated for these parameters.

Summary

- The marginal likelihood $p(\boldsymbol{X})$ integrates out the parameter μ using the prior.
- For a normal model with known variance and unknown mean:

$$p(\boldsymbol{X}) = \int \mathcal{N}(\bar{X} \mid \mu, \sigma^2/n)\mathcal{N}(\mu \mid \mu_0, \tau^2)\, d\mu$$

- This evaluates to a Gaussian form, penalizing models with a mismatch between prior and data.
- Used in Bayesian model selection and comparison.

Posterior Predictive Distribution

- Goal: Predict a new observation X_{new} based on the posterior distribution of μ.
- The posterior predictive distribution is:

$$p(X_{\text{new}} \mid \boldsymbol{X}) = \int p(X_{\text{new}} \mid \mu) p(\mu \mid \boldsymbol{X}) \, d\mu$$

- Substituting distributions:

$$X_{\text{new}} \mid \mu \sim \mathcal{N}(\mu, \sigma^2), \quad \mu \mid \boldsymbol{X} \sim \mathcal{N}(\mu_n, \tau_n^2)$$

Combining Distributions

- The posterior predictive distribution is a mixture of Gaussians:

$$p(X_{\text{new}} \mid \boldsymbol{X}) = \int \mathcal{N}(X_{\text{new}} \mid \mu, \sigma^2)\mathcal{N}(\mu \mid \mu_n, \tau_n^2)\, d\mu$$

- This results in:

$$X_{\text{new}} \mid \boldsymbol{X} \sim \mathcal{N}(\mu_n, \sigma^2 + \tau_n^2)$$

- Key points:
  - Mean of X_{new}: μ_n, the posterior mean of μ.
  - Variance of X_{new}: $\sigma^2 + \tau_n^2$, combining data uncertainty and parameter uncertainty.

Interpretation

- Posterior predictive variance:

$$\mathrm{Var}(X_{\text{new}} \mid \boldsymbol{X}) = \sigma^2 + \tau_n^2$$

  - σ^2: Variance of the data-generating process.
  - τ_n^2: Residual uncertainty in the mean μ, reduced as n increases.
- As $n \to \infty$, $\tau_n^2 \to 0$, and:

$$p(X_{\text{new}} \mid \boldsymbol{X}) \to \mathcal{N}(\bar{X}, \sigma^2)$$

- The posterior predictive distribution accounts for both prior beliefs and the observed data.

Example: Numerical Calculation

- Given:
  - $n = 10$, $\bar{X} = 5$, $\sigma^2 = 4$.
  - Prior: $\mu_0 = 0$, $\tau^2 = 1$.
- Compute posterior parameters:

$$\tau_n^2 = \left(\frac{10}{4} + \frac{1}{1}\right)^{-1} = \frac{4}{6.5}, \quad \mu_n = \frac{4}{6.5}\left(\frac{10 \cdot 5}{4}\right) = \frac{50}{6.5}$$

- Posterior predictive distribution:

$$X_{\text{new}} \mid \boldsymbol{X} \sim \mathcal{N}\left(\frac{50}{6.5}, 4 + \frac{4}{6.5}\right)$$

Both Mean and Variance Unknown

- Prior for μ: $\mathcal{N}(\mu_0, \tau^2)$.
- Prior for σ^2: Inverse-Gamma $\text{IG}(\alpha_0, \beta_0)$.
- Joint likelihood:

$$f(\boldsymbol{x} \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right).$$

- Full conditionals:

$$\mu \mid \boldsymbol{x}, \sigma^2 \sim \mathcal{N}\left(\frac{\frac{\mu_0}{\tau^2} + \frac{\sum x_i}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}}\right),$$

$$\sigma^2 \mid \boldsymbol{x}, \mu \sim \text{IG}\left(\alpha_0 + \frac{n}{2}, \beta_0 + \frac{\sum(x_i - \mu)^2}{2}\right).$$

Joint Prior and Posterior

- Joint prior:
$$p(\mu, \sigma^2) = p(\mu \mid \sigma^2)p(\sigma^2).$$
- Using conjugate priors:
$$p(\mu, \sigma^2 \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \mu, \sigma^2)p(\mu \mid \sigma^2)p(\sigma^2).$$
- Posterior distribution combines prior and likelihood for joint inference.

Deriving the Joint Posterior

- Assume:
  - Likelihood:

$$f(\boldsymbol{x} \mid \mu, \sigma^2) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{\sum(x_i - \mu)^2}{2\sigma^2}\right).$$

  - Priors:

$$\mu \mid \sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2/\kappa_0), \quad \sigma^2 \sim \text{IG}(\alpha_0, \beta_0).$$

- Joint prior:

$$p(\mu, \sigma^2) = p(\mu \mid \sigma^2) p(\sigma^2).$$

- Posterior:

$$p(\mu, \sigma^2 \mid \boldsymbol{x}) \propto f(\boldsymbol{x} \mid \mu, \sigma^2) p(\mu \mid \sigma^2) p(\sigma^2).$$

Posterior Derivation (continued)

- Combine terms:

$$p(\mu, \sigma^2 \mid \boldsymbol{x}) \propto (\sigma^2)^{-(n/2+\alpha_0+1)} \exp\left(-\frac{\sum(x_i-\mu)^2}{2\sigma^2}\right)$$

$$\times \exp\left(-\frac{\kappa_0(\mu-\mu_0)^2}{2\sigma^2}\right) \exp\left(-\frac{\beta_0}{\sigma^2}\right).$$

- Group terms by μ and σ^2:
  - Conditional posterior for μ:

$$\mu \mid \boldsymbol{x}, \sigma^2 \sim \mathcal{N}\left(\mu_n, \frac{\sigma^2}{\kappa_n}\right),$$

    where:

$$\mu_n = \frac{\kappa_0\mu_0 + n\bar{x}}{\kappa_0 + n}, \quad \kappa_n = \kappa_0 + n.$$

  - Marginal posterior for σ^2:

$$\sigma^2 \mid \boldsymbol{x} \sim \mathrm{IG}\left(\alpha_0 + \frac{n}{2}, \beta_0 + \frac{1}{2}\left[\sum(x_i-\bar{x})^2 + \frac{\kappa_0 n(\bar{x}-\mu_0)^2}{\kappa_0+n}\right]\right).$$

Summary and Applications

- Bayesian inference for Normal distribution allows incorporating prior knowledge.
- Applications: A/B testing, prediction intervals, hierarchical modeling.
- Extendable to other distributions and hierarchical models.