Exercise 1: Conjugate Priors for Gaussians

Question

Assume the data distribution is Gaussian:

$$p(x|\theta) \sim \mathcal{N}(x|\mu, \sigma^2)$$

where the parameters are $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$. The prior distribution on the mean is:

$$p(\mu|\mu_0, \sigma_0^2) \sim \mathcal{N}(\mu|\mu_0, \sigma_0^2).$$

Given observed data $D = (x_1, \ldots, x_n)$, show that the posterior distribution is:

$$p(\mu \mid D, \mu_0, \sigma_0^2) \sim \mathcal{N}\left(\frac{1}{n}\left(\frac{\sigma_0^2}{\sigma^2/n + \sigma_0^2}\bar{x} + \frac{\sigma^2/n}{\sigma^2/n + \sigma_0^2}\mu_0\right), \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$$

where

$$\bar{x} = \sum_{i=1}^{n} x_i.$$

Probable Solution

Given:

$$p(x|\mu, \sigma^2) \sim \mathcal{N}(x|\mu, \sigma^2), \quad p(\mu|\mu_0, \sigma_0^2) \sim \mathcal{N}(\mu|\mu_0, \sigma_0^2).$$

Likelihood for all data

The likelihood for n observations is:

$$p(D|\mu, \sigma^2) = \prod_{i=1}^{n} \mathcal{N}(x_i|\mu, \sigma^2).$$

Since each observation is independent, the joint likelihood is the product of Gaussians. Ignoring constants (which do not depend on μ):

$$p(D|\mu, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

The exponent comes from the Gaussian density formula. Only keep terms involving μ because constants will cancel in normalization.

Expand the quadratic term:

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n} x_i^2 - 2\mu\sum_{i=1}^{n} x_i + n\mu^2 = C - 2\mu\bar{x} + n\mu^2,$$

where C is constant w.r.t. μ and $\bar{x} = \sum_{i=1}^{n} x_i$.

Thus:

$$p(D|\mu, \sigma^2) \propto \exp\left(-\frac{n}{2\sigma^2}\mu^2 + \frac{\bar{x}}{\sigma^2}\mu\right).$$

Prior

The prior is:

$$p(\mu|\mu_0, \sigma_0^2) \propto \exp\left(-\frac{1}{2\sigma_0^2}(\mu - \mu_0)^2\right) = \exp\left(-\frac{1}{2\sigma_0^2}\mu^2 + \frac{\mu_0}{\sigma_0^2}\mu\right).$$

Again,ignore constants and expand the square to isolate terms involving μ.

Posterior (combine likelihood and prior)

By Bayes' rule:

$$p(\mu|D) \propto p(D|\mu)p(\mu).$$

Multiply likelihood and prior:

$$p(\mu|D) \propto \exp\left(-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)\mu^2 - 2\left(\frac{\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right)\mu\right]\right).$$

The quadratic and linear terms in μ combine. This is the kernel of a Gaussian distribution.

Identify posterior parameters

This is a Gaussian in μ with:

$$\text{Precision: } \tau = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}, \quad \text{Variance: } \sigma_*^2 = \tau^{-1} = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}.$$

Posterior mean:

$$\mu^* = \frac{\frac{\bar{x}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}.$$

The mean is the ratio of the linear coefficient to the precision. It is a weighted average of the prior mean and the data sum.

Express in required form

Rewrite using weights:

$$\mu^* = \frac{1}{n}\left(\frac{\sigma_0^2}{\sigma^2/n + \sigma_0^2}\bar{x} + \frac{\sigma^2/n}{\sigma^2/n + \sigma_0^2}\mu_0\right).$$

Thus the posterior is:

$$p(\mu \mid D, \mu_0, \sigma_0^2) \sim \mathcal{N}\left(\frac{1}{n}\left(\frac{\sigma_0^2}{\sigma^2/n + \sigma_0^2}\bar{x} + \frac{\sigma^2/n}{\sigma^2/n + \sigma_0^2}\mu_0\right), \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right).$$

Final Thought

The posterior mean is a weighted combination of:

- The sum of observed data $\bar{x}$, scaled by its precision.
- The prior mean μ_0, scaled by prior precision.

.