STAD91: Quiz 2 Practice (Bayesian Calculus)

Instructor: Thibault Randrianarisoa

Questions

Question 1 You are modeling a single data point X using a Normal likelihood $X|\theta \sim \mathcal{N}(\theta, 1)$. You choose a standard Normal prior for the mean: $\theta \sim \mathcal{N}(0, 1)$. If you observe a single value $X = 4$, what is the mean of the posterior distribution $\pi(\theta|X = 4)$?

a) 4

b) 2

c) 0

d) 1

Question 2 In the context of linear regression, computing the Maximum A Posteriori (MAP) estimator with a Laplace prior, $\pi(\theta) \propto \exp(-\lambda||\theta||_1)$, is equivalent to which frequentist method?

a) Ridge Regression

b) Ordinary Least Squares (OLS)

c) Lasso Regression

d) Elastic Net

Question 3 For a parameter θ in a model with Fisher Information $I(\theta)$, select all that apply to the Jeffreys prior:

a) The resulting posterior gives different results depending on the model parameterization.

b) It is a subjective prior designed to encode expert beliefs.

c) It puts more prior mass in regions where the Fisher Information is large.

d) It is always a proper prior.

Question 4 Suppose you have a Binomial likelihood $X|\theta \sim \text{Bin}(n, \theta)$. If you choose a Beta prior $\theta \sim \text{Beta}(\alpha, \beta)$, what family of distributions does the posterior belong to?

a) Beta

b) Binomial

c) Bernoulli

d) Dirichlet

Question 5 Which of the following best describes the Empirical Bayes approach, where the prior itself depends on a parameter γ?

a) We treat the hyperparameters γ as fixed and known quantities chosen by an expert.

b) We place a "hyperprior" on the hyperparameters γ and integrate them out.

c) We use an improper prior to ensure the posterior is invariant to this parameter choice.

d) We estimate the hyperparameters γ from the data (e.g., by maximizing marginal likelihood) and then proceed as if they were fixed.

Question 6 You are estimating the parameter describing the size of a square. You assign a Uniform prior to the *side length* L, such that $L \sim \mathcal{U}[0,1]$. Your colleague prefers to parametrize the problem using the *area* $A = L^2$. Which of the following describes the induced prior density $\pi_A(a)$ on the area?

a) It puts more mass on small areas: $\pi_A(a) = \frac{1}{2\sqrt{a}}$.

b) It is also Uniform on $[0,1]$: $\pi_A(a) = 1$.

c) It is proportional to the square root: $\pi_A(a) \propto \sqrt{a}$.

d) It puts more mass on large areas: $\pi_A(a) = 2a$.

Solutions

1. Correct Answer: b) 2
Reasoning: For a Normal likelihood $\mathcal{N}(\theta, 1)$ and prior $\mathcal{N}(0, 1)$, the posterior for a single observation x is $\mathcal{N}(x/2, 1/2)$. Here $x = 4$, so the mean is $4/2 = 2$.

2. Correct Answer: c) Lasso Regression
Reasoning: As shown in the slides, Ridge regression corresponds to a Gaussian prior, while Lasso (L_1 penalty) corresponds to a Laplace prior.

3. Correct Answer: c) $\sqrt{I(\theta)}$
Reasoning: The Jeffreys prior is defined as $\pi(\theta) \propto \sqrt{I(\theta)}$ in 1D (or $\sqrt{\det I(\theta)}$ in higher dimensions) to ensure invariance under reparameterization.

4. Correct Answer: c) Beta
Reasoning: The Beta distribution is the conjugate prior for the Binomial likelihood. The posterior updates to $\text{Beta}(\alpha + x, \beta + n - x)$.

5. Correct Answer: d) We estimate the hyperparameters γ from the data...
Reasoning: Empirical Bayes differs from Hierarchical Bayes (answer b) by estimating hyperparameters directly from the data (e.g., using marginal likelihood) rather than placing a prior on them (See Lecture 2, Slide 39-40).

Solution to Q6: c) It puts more mass on small areas...
Reasoning: Using the change of variable formula with $A = L^2$ (so $L = \sqrt{A}$), we have $\frac{dL}{dA} = \frac{1}{2\sqrt{A}}$. The prior on A is $\pi_A(a) = \pi_L(\sqrt{a}) \left|\frac{dL}{dA}\right| = 1 \cdot \frac{1}{2\sqrt{a}}$. This density explodes near 0, meaning a "flat" belief on length implies a strong belief that the area is small.