STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN DATA SCIENCE & ANALYTICS
CAT 1- Open Book
DSA 8505: Bayesian Statistics

DATE: 14th Feb 2025

Instruction

(a) Answer All Question

(b) Scan and submit your answer sheet through the Google Classroom by 23h59, 14th Feb 2025.

---

1. Consider a Bayesian estimation problem where the posterior distribution of a parameter θ follows a normal distribution:
$$\theta|x \sim \mathcal{N}(\mu, \sigma^2).$$
Construct a 90% credible interval for θ using the equal-tailed credible interval method.

2. A biased die has an unknown probability θ of landing on a six. A Bayesian statistician assumes a Beta(3,3) prior for θ. The die is rolled 15 times, and a six appears 6 times. Compute the posterior distribution of θ (***Start by deriving posterior distribution***).

3. The posterior distribution of a parameter θ is given by:
$$\pi(\theta|x) \sim N(10, 16).$$
   (a) Determine the Bayes estimator under the squared error loss function.
   (b) Determine the Bayes estimator under the absolute error loss function.

4. Suppose the posterior distribution of θ has two peaks at $\theta = 2$ and $\theta = 6$, with the highest posterior density occurring at $\theta = 6$.
   (a) Explain why the MAP estimate under the 0-1 loss function differs from the Bayes estimator under squared error loss.
   (b) If the posterior mean is 4.5, which estimate would you choose under squared error loss, and why?

5. Suppose a parameter θ follows a normal prior distribution $N(\mu_0, \sigma_0^2)$, and the likelihood is also normally distributed as $N(\mu, \sigma^2)$.

   (a) Derive the posterior distribution.

   (b) Show that the posterior mean minimizes the expected squared error loss.

   (c) If $\theta \sim N(6, 2)$ and the likelihood is $N(12, 3)$, compute the Bayesian estimator under squared error loss.

6. Bayesian logistic regression is used to model binary outcomes, such as predicting whether a patient has a disease based on their age and cholesterol level. Here is a sample dataset:

Disease (1 = Yes, 0 = No)	**Age (years)**	**Cholesterol (mg/dL)**
0	45	180
1	52	220
1	60	250
0	35	160
0	40	170
1	55	240
0	30	150
1	65	270
0	42	175
1	58	230

   (a) Explain the role of prior distributions in Bayesian logistic regression and discuss their impact on inference.

   (b) The logistic function (sigmoid) is used to transform the linear combination of predictors into a probability. Explain why it is appropriate for modeling binary outcomes.

   (c) In the Bayesian model given, the posterior estimates of beta's are:

$$\beta_0 = -0.1, \quad \beta_{\text{age}} = -0.25, \quad \text{and} \quad \beta_{\text{cholesterol}} = 0.018.$$

   Compute the probability of disease for a 50-year-old patient with a cholesterol level of 200 mg/dL. Interpret the result in the context of disease diagnosis.