DSA 8505 Bayesian Statistics

CAT 2

Kevin Obote - 190696

SU

Field of study

Strathmore Institute of Mathematical Science
Strathmore University

February 17, 2026

Project Repository: `https://github.com/Kevinobote/Bayesian-Statistics/blob/main/CATS/CAT%202/cat2.ipynb`

Question 1: Bayesian Inference in Security Systems

A security agency uses an automated facial recognition system to identify individuals on a watchlist at a busy airport. Based on historical records, it is known that only 0.5% of travelers are actually on the watchlist. The facial recognition system has the following performance characteristics:

- If a traveler is on the watchlist, the system correctly flags them with probability 95% (true positive rate).
- If a traveler is not on the watchlist, the system incorrectly flags them with probability 3% (false positive rate).

A randomly selected traveler is scanned by the system and is flagged as a match.

Required

(a) Define the prior probability that a randomly selected traveler is on the watchlist.

(b) Define the likelihood of the system flagging a traveler, conditional on whether the traveler is on the watchlist or not.

(c) Compute the marginal probability that a randomly selected traveler is flagged by the system.

(d) Using Bayes' Theorem, compute the posterior probability that the traveler is actually on the watchlist given that they were flagged by the system.

(e) Interpret the result from a policy perspective.

Solution

(a) Prior Probability

The prior probability represents the initial belief or prevalence of the condition before observing new data.

Let W be the event that a traveler is on the watchlist. Based on historical records:

The probability that a traveler is on the watchlist (W) is:

$$P(W) = 0.005$$

The probability that a traveler is not on the watchlist (W^c) is:

$$P(W^c) = 1 - P(W) = 0.995$$

(b) Likelihood

The likelihood is the probability of the data (being flagged, F) given the state of the parameter (θ).

- **True Positive Likelihood:** $P(F|W) = 0.95$
- **False Positive Likelihood:** $P(F|W^c) = 0.03$

These describe the measurement model of the facial recognition system.

(c) Marginal Probability

The marginal probability (evidence) is the total probability of observing the data under all possible hypotheses.
We use the Law of Total Probability:

$$\begin{aligned}
P(F) &= P(F|W)P(W) + P(F|W^c)P(W^c) \\
P(F) &= (0.95 \times 0.005) + (0.03 \times 0.995) \\
P(F) &= 0.00475 + 0.02985 \\
P(F) &= 0.0346
\end{aligned}$$

So about 3.46% of travelers will be flagged.

(d) Posterior Probability

Using Bayes' Theorem, we compute the updated probability of a hypothesis given the observed evidence:

$$P(W \mid F) = \frac{P(F \mid W)\, P(W)}{P(F)}$$

Where:

- **Prior** $P(W)$: Represents the initial background knowledge or population prevalence before observing the evidence.
- **Likelihood** $P(F \mid W)$: Measures how well the hypothesis explains the observed outcome.
- **Marginal Likelihood** $P(F)$: The normalization constant in Bayes' Theorem that ensures the posterior distribution sums to 1.
- **Posterior** $P(W \mid F)$: The updated degree of belief in the hypothesis after combining the prior and the likelihood.

Substituting the values:

$$P(W|F) = \frac{0.95 \times 0.005}{0.0346}$$
$$P(W|F) = \frac{0.00475}{0.0346}$$
$$P(W|F) \approx 0.1373$$

The posterior probability that the traveler is on the watchlist is **13.73%**.

(e) Policy Interpretation

Based on the computed posterior probability of **13.73%**, there is an 86.27% chance that the flag is a false positive. This occurs because the watchlist is extremely rare in the general population (0.5%), a phenomenon known as the **Base Rate Fallacy**.
Recommendation: Airport security should not immediately detain the traveler. Instead, they should apply **additional screening procedures**, such as manual ID verification or secondary interviewing, to confirm the identity while minimizing the disruption to innocent travelers.

Question 2: Bayesian Model Comparison and Averaging

Consider the problem of modelling monthly electricity consumption y (in kWh) using the average monthly temperature x (in degrees Celsius). Two competing regression models are proposed:

- **Model 1 (M_1):** $y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$
- **Model 2 (M_2):** $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$

Table 1: Observed electricity consumption data

Average Temperature x	**Electricity Use** y
10	220
15	240
20	280
25	350
30	460

The observed dataset is shown in Table 1:

Assume equal prior probabilities for the models: $P(M_1) = P(M_2) = 0.5$. The priors for the parameters are $\beta_j \sim N(0, 10^2)$ and $\sigma^2 \sim$ Inverse-Gamma$(1, 1)$.

Solution

(a) Ordinary Least Squares (OLS) Estimation for M_1

To estimate the regression coefficients $\hat{\beta}_1$ (slope) and $\hat{\beta}_0$ (intercept) for the linear model $y = \beta_0 + \beta_1 x + \epsilon$, we utilize the following OLS formulas:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Step 1: Data Summation from Table 1

From the provided dataset $n = 5$, we calculate the required sums:

- $\sum x = 10 + 15 + 20 + 25 + 30 = 100$
- $\sum y = 220 + 240 + 280 + 350 + 460 = 1550$
- $\bar{x} = \frac{100}{5} = 20, \quad \bar{y} = \frac{1550}{5} = 310$
- $\sum x^2 = 10^2 + 15^2 + 20^2 + 25^2 + 30^2 = 100 + 225 + 400 + 625 + 900 = 2250$
- $\sum xy = (10 \times 220) + (15 \times 240) + (20 \times 280) + (25 \times 350) + (30 \times 460)$
- $\sum xy = 2200 + 3600 + 5600 + 8750 + 13800 = 33950$

Step 2: Calculating the Slope ($\hat{\beta}_1$)

Using the computational formula:

$$\hat{\beta}_1 = \frac{5(33950) - (100)(1550)}{5(2250) - (100)^2}$$

$$\hat{\beta}_1 = \frac{169750 - 155000}{11250 - 10000}$$

$$\hat{\beta}_1 = \frac{14750}{1250} = 11.8$$

Step 3: Calculating the Intercept ($\hat{\beta}_0$)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_0 = 310 - (11.8 \times 20)$$

$$\hat{\beta}_0 = 310 - 236 = 74$$

Final OLS Estimates

For Model M_1, the estimated regression coefficients are:

$$\hat{\beta}_0 = 74, \quad \hat{\beta}_1 = 11.8$$

The estimated linear regression equation is:

$$\hat{y} = 74 + 11.8x$$

(b) Likelihood Function for Model M_1

The likelihood function $L(\beta_0, \beta_1, \sigma^2 | D)$ represents the joint probability of the observed electricity consumption data y given the temperature x and the model parameters β_0, β_1, and σ^2.

Assuming the errors ϵ_i are independent and identically distributed (i.i.d.) such that $\epsilon_i \sim N(0, \sigma^2)$, the likelihood for n observations is given by the product of individual normal probability density functions:

$$L(\beta_0, \beta_1, \sigma^2 | D) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)$$

Equivalently, this can be expressed in its simplified form:

$$L(\beta_0, \beta_1, \sigma^2|D) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\right)$$

Where:

- n is the number of observations ($n = 5$).
- y_i is the observed electricity use.
- $\beta_0 + \beta_1 x_i$ is the mean predicted by the linear model.

(c) Posterior Distribution for Model M_2

Using Bayes' Theorem, the joint posterior distribution for the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ and the variance σ^2 under the quadratic model M_2 is proportional to the product of the likelihood and the prior distributions:

$$p(\beta_0, \beta_1, \beta_2, \sigma^2|D, M_2) \propto L(D|\beta_0, \beta_1, \beta_2, \sigma^2) \times p(\beta_0)p(\beta_1)p(\beta_2) \times p(\sigma^2)$$

Substituting the specific distributions:

$$\begin{aligned} p(\beta, \sigma^2|D, M_2) \propto & \left[(\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2\right)\right] \\ & \times \left[\prod_{j=0}^{2} \exp\left(-\frac{\beta_j^2}{2(10^2)}\right)\right] \\ & \times \left[(\sigma^2)^{-(1+1)} \exp\left(-\frac{1}{\sigma^2}\right)\right] \end{aligned}$$

Where:

- The first bracketed term is the **Likelihood** for the quadratic model.
- The second bracketed term is the **Prior** for the regression coefficients β_j.
- The third bracketed term is the **Prior** for the variance σ^2.

(d) Need for MCMC Methods

Markov Chain Monte Carlo (MCMC) methods are required for this analysis because the joint posterior distribution $p(\beta_0, \beta_1, \beta_2, \sigma^2|D, M_2)$ is **analytically intractable**.

Specifically, the integration required to calculate the normalizing constant (marginal likelihood) in the denominator of Bayes' Theorem:

$$P(D|M_2) = \int \cdots \int L(D|\beta, \sigma^2)p(\beta)p(\sigma^2)d\beta_0 d\beta_1 d\beta_2 d\sigma^2$$

cannot be solved in closed form due to the complexity of the likelihood and the non-conjugate nature of the parameter space.

MCMC algorithms, such as Gibbs Sampling or the Metropolis-Hastings algorithm, circumvent this problem by generating a sequence of dependent samples (a Markov Chain) whose stationary distribution is the target posterior. This allows us to estimate parameters like the posterior mean by taking the average of these simulated samples.

(e) Model Comparison

(i) **Bayes Factor (B_{21}):**

$$B_{21} = \frac{P(D|M_2)}{P(D|M_1)} = \frac{0.045}{0.015} = 3$$

(ii) **Posterior Model Probabilities:** Since $P(M_1) = P(M_2) = 0.5$:

$$P(M_2|D) = \frac{P(D|M_2)P(M_2)}{P(D|M_1)P(M_1) + P(D|M_2)P(M_2)} = \frac{0.045}{0.015 + 0.045} = 0.75$$

$$P(M_1|D) = 1 - 0.75 = 0.25$$

(iii) **Interpretation:** A Bayes Factor of 3 indicates substantial evidence in favour of Model M_2 over M_1 according to Jeffreys' scale. The Quadratic model is 3 times more likely to have generated the data than the Linear model.

(f) Bayesian Model Averaging (BMA)

(i) General BMA Formula

The Bayesian Model Averaging (BMA) estimate for a parameter (or a prediction) Δ is the weighted average of the individual model-specific posterior means, where the weights are the posterior model probabilities. The general formula is:

$$\hat{\Delta}_{BMA} = \sum_{k=1}^{K} \hat{\Delta}_k P(M_k|D)$$

Where:

- $\hat{\Delta}_{BMA}$ is the final averaged estimate.
- K is the total number of competing models ($K = 2$ in this case).
- $\hat{\Delta}_k = E[\Delta|D, M_k]$ is the posterior mean of the parameter under model M_k.
- $P(M_k|D)$ is the posterior probability of model M_k given the data D.

(ii) Computation of BMA Estimates

Using the posterior model probabilities $P(M_1|D) = 0.25$ and $P(M_2|D) = 0.75$ as weights, we compute the BMA estimates for each coefficient:

1. **BMA Estimate for β_0:**

$$\hat{\beta}_{0,BMA} = \hat{\beta}_{0,M_1} P(M_1|D) + \hat{\beta}_{0,M_2} P(M_2|D)$$
$$\hat{\beta}_{0,BMA} = (150 \times 0.25) + (120 \times 0.75)$$
$$\hat{\beta}_{0,BMA} = 37.5 + 90 = \mathbf{127.5}$$

2. **BMA Estimate for β_1:**

$$\hat{\beta}_{1,BMA} = \hat{\beta}_{1,M_1} P(M_1|D) + \hat{\beta}_{1,M_2} P(M_2|D)$$
$$\hat{\beta}_{1,BMA} = (8.5 \times 0.25) + (6.2 \times 0.75)$$
$$\hat{\beta}_{1,BMA} = 2.125 + 4.65 = \mathbf{6.775}$$

3. **BMA Estimate for β_2:** Since β_2 is only present in M_2, its estimate in M_1 is 0:

$$\hat{\beta}_{2,BMA} = \hat{\beta}_{2,M_1} P(M_1|D) + \hat{\beta}_{2,M_2} P(M_2|D)$$
$$\hat{\beta}_{2,BMA} = (0 \times 0.25) + (0.45 \times 0.75)$$
$$\hat{\beta}_{2,BMA} = 0 + 0.3375 = \mathbf{0.3375}$$

(iii) Final BMA Regression Equation

The final Bayesian Model Averaging (BMA) regression equation is formed by substituting the weighted average coefficients into the quadratic model structure. Given $\hat{\beta}_{0,BMA} = 127.5$, $\hat{\beta}_{1,BMA} = 6.775$, and $\hat{\beta}_{2,BMA} = 0.3375$:

$$y = 127.5 + 6.775x + 0.3375x^2 + \epsilon$$

Where:

- y is the predicted monthly electricity consumption (kWh).
- x is the average monthly temperature (°C).
- ϵ is the error term representing remaining uncertainty.

This equation provides a robust forecast by incorporating evidence from both the linear and quadratic model hypotheses.