Bayesian Analysis Assignment

Question 1 (5 + 5 Marks)

2022 Legislative Assembly elections were held in Uttar Pradesh from 10 February to 7 March 2022 in seven phases[1]. The ABP-CVoter opinion poll survey[2] conducted a few days prior to the elections estimated that Bhartiya Janata Party (BJP+) is likely to sway 41.5% of the vote share in the state headed by Mr. Yogi Adityanath. Similarly, the Polstrat-NewsX opinion poll[3] projected that 40.9% of the voters supported BJP+.

(i) Suppose that an analyst decides to use the two opinion poll results to construct a prior for p = "percentage of voters who will support BJP+". Specifically, she decides to give equal weightage to both the opinion poll results and by matching the mean and variance, represent it as a *Beta (a, b)* distribution. Determine the parameters *a* and *b*.

(ii) Suppose that an exit poll based on a random sample of 150 found that 61 voted for BJP+. If you use the prior obtained in (i) for p, then determine the posterior distribution for p and the 20th percentile of the distribution.

Question 2 (4+3+3 Marks)

Exponential distribution is typically used to model "inter-arrival time," i.e., the time between two successive arrivals. A random variable X is said to have an exponential distribution if it has the following probability density function (pdf)

$$f(x) = \mu e^{-\mu x}, \; x \in (0, \infty), \; \mu > 0. \text{ (Same as Gamma } (1, \mu))$$

Here, μ is called the "rate" parameter and can be interpreted as average rate of arrivals per unit time. Suppose that the inter arrival time (in minutes) of customers at a bank is believed to follow an exponential distribution with unknown rate. Based on experience, the prior distribution for μ is taken to be a Gamma distribution with shape parameter $\nu = 1$ and rate parameter $\lambda = 0.2$. On a given day, the observed inter-arrival times (in minutes) for first 5 customers were as follows: 7, 5, 8, 4 and 6 minutes.

(i) Identify the known form of the posterior distribution of μ given the data, along with its parameters?

(ii) What is the classical estimate for μ obtained by maximizing the likelihood for observed data and 95% credible interval? (This procedure is called "Maximum Likelihood Estimation")

(iii) Based on the predictive distribution, compute the probability that the inter-arrival time for the 6^{th} customer to be greater than 7 minutes.

---

[1] 2022 Uttar Pradesh Legislative Assembly election - Wikipedia

[2] ABP-CVoter Opinion Poll: BJP Leads & SP Is Biggest Challenger In Latest UP Survey. Check BSP & Congress Share (abplive.com)

[3] Polstrat-NewsX Pre-Poll Survey 2 (indianewsbusiness.com)

Question 3 (4+3+3 Marks)

When a new mobile phone is launched everyone usually goes crazy over it and it is sold at such a high price just because it is new. Suppose that the number of customers at a mobile store during a week for a newly launched phone follows a Poisson distribution with mean μ. The weekly number of purchases for a newly launched mobile phone observed over a ten-week period is:

$$\{28, 15, 11, 21, 11, 17, 21, 15, 14, 18\}$$

(i) Compute the posterior distribution for μ, posterior mean, credible intervals, and predictive distributions, if Jeffry's prior is used for μ.

(ii) The mobile store also found out that their competitor experienced an average of 200 customers over a 10-week period. Construct a gamma (power) prior with this information.

(iii) Determine the posterior predictive probability distribution of the number of purchases of a newly launched mobile phone based on prior in (ii).

Question 4 (4+3+3 Marks)

In an oil factory, every week, a random sample of the weights of the oil tins are administered. Let's say you are the quality analyst responsible for quality standards at that oil factory. You want the probability that a randomly chosen tin of oil labelled "1000 g" is actually less than 1 kilogram (1,000 grams) to be 1% or less. The weight (in grams) of the oil tins produced by the machine is normal (μ, σ^2) where σ^2 = 32. The weights (in grams) of 20 oil tins from a batch are:

1003	995	999	1001	994	998	996	1002	1004	998
994	994	995	995	1001	997	1006	997	998	994

You decide to use a discrete prior distribution for μ with the following probabilities:

$$g(\mu) = \begin{cases} .05\,, & for\ \ \mu \in \{991, 992, \ldots 1010\} \\ 0\,, & otherwise \end{cases}$$

(i) Calculate the posterior probability distribution for μ. Also, compute the posterior mean and variance.

(ii) What is the probability that μ is less than 1000 under the posterior distribution?

(iii) Suppose now that the mean is fixed, i.e., μ =998 but we have a prior on σ^2. More specifically, assume an inverse gamma prior for σ^2, i.e., $1/\sigma^2$ follows Gamma distribution with shape=40 and rate=10. Using the above data of the 20 oil tins find the posterior distribution for σ^2.