Statistics for Data Science - II
Graded Questions
Jan 2025 Term

Week-10

Question

1. A hospital tracks the daily number of emergency cases as a random variable X following a Poisson distribution with an unknown average arrival rate λ. To estimate λ, the hospital records the number of emergency cases over a period of 14 days as follows:

$$3, 6, 5, 7, 4, 6, 8, 5, 6, 7, 4, 5, 6, 7.$$

(i) The hospital models λ with a Gamma prior having parameters α and β. Derive an expression for β in terms of α, such that the prior mean is 3.5.

(a) $\beta = \frac{\alpha}{3.5}$

(b) $\alpha = 3.5\beta$

(c) $\beta = \frac{\alpha}{3}$

(d) $\frac{\alpha}{\beta} = 3.5$

Answer: a

Solution

The Gamma distribution Gamma(α, β) has a mean given by:

$$E[\lambda] = \frac{\alpha}{\beta}$$

Given that the prior mean is 3.5, we set:

$$\frac{\alpha}{\beta} = 3.5$$

Rearranging for β:

$$\beta = \frac{\alpha}{3.5}$$

Thus, the correct answer is (a).

Question

(ii) If the posterior mean of λ is computed to be 5.2 for the 14-day sample given above, determine the value of α that was used. Provide the answer correct to two decimal places.

Answer: 12.76 (Range: 12.73 to 12.79)

Solution

The posterior for λ follows a Gamma distribution:

$$\lambda | X \sim \text{Gamma}(\alpha + S, \beta + n)$$

where:

- $S = 3 + 6 + 5 + 7 + 4 + 6 + 8 + 5 + 6 + 7 + 4 + 5 + 6 + 7 = 79$ (sum of observations)
- $n = 14$ (number of days)

The posterior mean is given by:

$$E[\lambda | X] = \frac{\alpha + S}{\beta + n}$$

Setting $E[\lambda | X] = 5.2$:

$$\frac{\alpha + 79}{\beta + 14} = 5.2$$

From part (i), we substitute $\beta = \frac{\alpha}{3.5}$:

$$\frac{\alpha + 79}{\frac{\alpha}{3.5} + 14} = 5.2$$

Step-by-Step Calculation of α:

1. Rewrite the denominator using $\beta = \frac{\alpha}{3.5}$:

$$\frac{\alpha + 79}{\frac{\alpha}{3.5} + 14} = 5.2$$

2. Introduce a new variable for clarity:

$$D = \frac{\alpha}{3.5} + 14$$

Substituting this into the equation:

$$\frac{\alpha + 79}{D} = 5.2$$

Multiplying both sides by D:

$$\alpha + 79 = 5.2D$$

3. Substitute $D = \frac{\alpha}{3.5} + 14$:

$$\alpha + 79 = 5.2\left(\frac{\alpha}{3.5} + 14\right)$$

4. Expand the right-hand side:

$$\alpha + 79 = \frac{5.2\alpha}{3.5} + 72.8$$

5. Rearrange the equation:

$$\alpha - \frac{5.2\alpha}{3.5} = 72.8 - 79$$

$$\alpha - \frac{5.2\alpha}{3.5} = -6.2$$

6. Factor out α:

$$\alpha\left(1 - \frac{5.2}{3.5}\right) = -6.2$$

7. Compute the coefficient:

$$1 - \frac{5.2}{3.5} = 1 - 1.4857 = -0.4857$$

$$\alpha(-0.4857) = -6.2$$

8. Solve for α:

$$\alpha = \frac{-6.2}{-0.4857} \approx 12.76$$

Thus, the correct answer is 12.76.

Question

1. A doctor is testing patients for a rare disease. She continues testing until finding the first positive case. Let p be the probability that a patient tests negative. The prior distribution of p is $\text{Beta}(\alpha, \beta)$, with prior mean equal to 0.8.

   After testing 12 patients, she finds that the first 11 test negative and the 12th tests positive. What is the posterior mean of p in terms of β? Assume independence of test results for different patients.

   (a) $\dfrac{4\beta + 11}{5\beta + 12}$

   (b) $\dfrac{4\beta + 11}{5\beta + 1}$

   (c) $\dfrac{4\beta + 10}{5\beta + 11}$

   (d) $\dfrac{4\beta + 11}{\beta + 1}$

Solution

A doctor tests patients sequentially for a rare disease until the first positive case is found. Let p denote the probability a patient tests negative, with prior distribution $p \sim \text{Beta}(\alpha, \beta)$ and a prior mean of 0.8.

After testing 12 patients, the first 11 test negative and the 12th tests positive.

Given the prior mean is 0.8, then

$$\frac{\alpha}{\alpha + \beta} = 0.8 \implies \alpha = 4\beta$$

So the prior is $\text{Beta}(4\beta, \beta)$.

Also the observed data corresponds to 11 negatives followed by the first positive at trial 12, then the likelihood is given by:

$$P(\text{data} \mid p) = p^{11}(1 - p)$$

Using the conjugacy of Beta prior with Bernoulli likelihood we get:

$$p \mid \text{data} \sim \text{Beta}(\alpha + 11, \beta + 1) = \text{Beta}(4\beta + 11, \beta + 1)$$

then the posterior mean is given by:

$$E[p \mid \text{data}] = \frac{4\beta + 11}{4\beta + 11 + \beta + 1} = \frac{4\beta + 11}{5\beta + 12}$$

The required posterior mean is : $\boxed{\dfrac{4\beta + 11}{5\beta + 12}}$

Question

2. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim}$ Bernoulli(p) distribution. The prior distribution of p has the PMF as follows:

p	0.2	0.6
$f_P(p)$	0.3	0.7

For $n = 5$ consider the samples $\{1, 1, 0, 0, 1\}$.

(i) Find the posterior mode of p. Enter the answer correct to one decimal place.

(ii) Find the posterior mean of p. Enter the answer correct to two decimal places.

Solution

Suppose $X_1, \ldots, X_5 \overset{i.i.d.}{\sim}$ Bernoulli(p), with prior:

$$p = \begin{cases} 0.2 & \text{with probability } 0.3 \\ 0.6 & \text{with probability } 0.7 \end{cases}$$

Data observed: $\{1, 1, 0, 0, 1\}$.

Then the likelihood for the given sample is:

Number of ones = 3, zeros = 2.

For $p = 0.2$:

$$L(0.2) = 0.2^3 \times 0.8^2 = 0.008 \times 0.64 = 0.00512$$

For $p = 0.6$:

$$L(0.6) = 0.6^3 \times 0.4^2 = 0.216 \times 0.16 = 0.03456$$

Given prior probabilities and likelihoods, using Bayes' theorem, the posterior probabilities are:

$$P(p \mid \text{data}) = \frac{P(p) \times L(p)}{\sum_{p'} P(p') \times L(p')}.$$

Therefore,

$$\begin{aligned}
P(p = 0.2 \mid \text{data}) &= \frac{0.3 \times 0.00512}{0.3 \times 0.00512 + 0.7 \times 0.03456} = \frac{0.001536}{0.025728} \approx 0.0597, \\
P(p = 0.6 \mid \text{data}) &= \frac{0.7 \times 0.03456}{0.3 \times 0.00512 + 0.7 \times 0.03456} = \frac{0.024192}{0.025728} \approx 0.9403.
\end{aligned}$$

(i) **The posterior mode is the p with maximum posterior probability, which is:** $\boxed{0.6}$

(ii) **The posterior mean of p.**

$$E[p \mid \textbf{data}] = 0.2 \times 0.0597 + 0.6 \times 0.9403 = \boxed{0.58} \quad (\textbf{approx})$$

Range: 0.55 to 0.61