Statistics for Data Science-2

Week 9 Solve with us

Table of contents

Question: 1

Suppose that the number of customers reaching a particular shop in an one-hour time period follows the Poisson distribution with an unknown parameter λ. Previous records suggest that the prior probabilities of λ are $P(\lambda = 5) = 0.4$ and $P(\lambda = 7) = 0.6$. If in a particular one-hour time period nine customers reach the the shop, find the posterior mode of λ.

(a) 5

(b) 7

(c) 6

(d) 9

Solution:

Prior probabilities of λ are $P(\lambda = 5) = 0.4$ and $P(\lambda = 7) = 0.6$.
The posterior probabilities of λ will be

$$P(\lambda = 5|X = 9) = \frac{P(X = 9|\lambda = 5).P(\lambda = 5)}{P(X = 9)}$$
$$= \frac{e^{-5}(5)^9(0.4)}{9!P(X = 9)} \qquad ...(1)$$

$$P(\lambda = 7|X = 9) = \frac{P(X = 9|\lambda = 7).P(\lambda = 7)}{P(X = 9)}$$
$$= \frac{e^{-7}(7)^9(0.6)}{9!P(X = 9)} \qquad ...(2)$$

Dividing equation (2) by (1), we get

$$\frac{P(\lambda = 7|X = 9)}{P(\lambda = 5|X = 9)} = \frac{e^{-7}(7)^9(0.6)}{e^{-5}(5)^9(0.4)}$$
$$= \frac{7^9(6)(e^{-2})}{(4)(5^9)} > 1$$

It implies that $P(\lambda = 7|X = 9) > P(\lambda = 5|X = 9)$
So, posterior mode will be 7.

Question: 2

Call duration of daily stand up meetings of employees of a certain company follows the exponential distribution with an unknown parameter λ. Duration (in minutes) of last twelve meetings are 35, 30, 25, 15, 25, 20, 20, 28, 34, 30, 15, 23. Find the Bayesian estimate (posterior mean) of λ using the prior distribution of Exp($\frac{1}{20}$) for λ.

(a) 0.4

(b) 25

(c) 0.04

(d) 2.5

Solution:

Let Λ be the prior distribution of λ.

From the given information, $f_{\Lambda}(\lambda) \sim \text{Exp}(\frac{1}{20})$.

It implies that $f_{\Lambda}(\lambda) = \frac{1}{20} e^{-\lambda/20}$.

Now, posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \Lambda = \lambda) f_{\Lambda}(\lambda)$

$\Rightarrow$ posterior density $\propto \lambda^n e^{-\lambda(X_1+X_2+\ldots+X_n)} (\frac{1}{20} e^{-\lambda/20})$

$\Rightarrow$ posterior density $\propto \lambda^n e^{-\lambda(X_1+X_2+\ldots+X_n+\frac{1}{20})}$

$\Rightarrow$ posterior density $=$ Gamma$(n+1, X_1 + X_2 + \ldots + X_n + \frac{1}{20})$

$$\Rightarrow \text{posterior mean} = \frac{n+1}{X_1 + X_2 + \ldots + X_n + \frac{1}{20}} = \frac{13}{300 + \frac{1}{20}}$$

$$\Rightarrow \text{posterior mean} = \frac{13 \times 20}{20 \times 300 + 1} = 0.04$$

Question: 3

Marks of tenth class students of a school follow the normal distribution with an unknown mean μ and variance 36. Marks of 15 students of the tenth class are 50, 60, 92, 42, 75, 85, 45, 32, 73, 75, 50, 95, 38, 55, 33. Find the Bayesian estimate (posterior mean) of μ assuming the Normal$(70, 25)$ prior distribution.

(a) 72.13

(b) 63.29

(c) 62.13

(d) 60.87

Solution:

We know that normal distribution is conjugate to the normal distribution. That is if prior distribution of μ is normal(μ_0, σ_0^2) and sample is taken from Normal(μ, σ^2), then posterior distribution of the μ will be Normal with mean $\overline{X}\dfrac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} + \dfrac{\mu_0\sigma^2}{n\sigma_0^2 + \sigma^2}$
here $\overline{X} = 60$, $n = 15$, $\mu_0 = 70$, $\sigma_0^2 = 25$, $\sigma^2 = 36$
Therefore,

$$\begin{aligned}\text{Posterior mean} &= \frac{60 \times 15 \times 25}{15(25) + 36} + \frac{70 \times 36}{15(25) + 36} \\ &= \frac{22500}{411} + \frac{2520}{411} \\ &= 54.74 + 6.13 = 60.87\end{aligned}$$

Question: 4

Three out of the last ten candidates wins a treasure hunt game. Previous record shows fraction of winners follows the Beta$(a, 6)$ distribution with an average of 40%. Estimate the long-term fraction of winners of the treasure hunt game.

(a) $\frac{7}{13}$

(b) $\frac{7}{20}$

(c) $\frac{13}{20}$

(d) $\frac{6}{18}$

Solution:

Let the long-term fraction of winners (probability of winning) be denoted by $\boldsymbol{p}$.
Previous data shows that fraction of winners follows the Beta$(a, 6)$ distribution with an average of 40%.
It implies that $E[\text{Beta}(a, 6)] = 0.4$

$$\Rightarrow \frac{a}{a+6} = 0.4$$

$$\Rightarrow a = 4$$

Therefore, prior distribution of $\boldsymbol{p}$ is Beta$(4, 6)$.
It implies that $f_{\boldsymbol{p}}(p) \propto p^3(1-p)^5$

Now, posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \boldsymbol{p} = p) f_{\boldsymbol{p}}(p)$

$\Rightarrow$ posterior density $\propto p^3(1-p)^7(p^3(1-p)^5)$

$\Rightarrow$ posterior density $\propto p^6(1-p)^{12}$

$\Rightarrow$ posterior density $=$ Beta$(7, 13)$

$\Rightarrow$ posterior mean $= \dfrac{7}{7+13} = \dfrac{7}{20} = 0.35$

Question: 5(a)

Following frequency data shows the number of patients (n) arriving in an emergency room between 12:00 AM and 6:00 AM.

n	frequency	n	frequency
0	2	4	10
1	7	5	2
2	17	6	4
3	8	7+	0

Fit the data into Poisson distribution (Find the parameter).

(a) 2.78

(b) 3.12

(c) 2.98

(d) 3

Solution:

We know that $\hat{\lambda} = \overline{X}$ is an estimate of λ.

$$\text{Sample mean, } \overline{X} = \frac{\sum_i f_i n_i}{\sum_i f_i}$$
$$= \frac{0 + 7 + 34 + 24 + 40 + 10 + 24}{2 + 7 + 17 + 8 + 10 + 2 + 4}$$
$$= \frac{139}{50} = 2.78$$

Therefore, $\hat{\lambda} = 2.78$

Question: 5(b)

Find an approximate 95% confidence interval using a normal approximation for the error distribution.
(Use the following information:
sample variance $S^2 = 2.34$ and
$P(-0.42 < \text{N}(0, 0.0468) < 0.42) = 0.95$)

(a) $[2.15, 3.40]$
(b) $[1.95, 3.20]$
(c) $[1.95, 3.55]$
(d) $[2.36, 3.20]$

Solution:

Error, e is given by
$e = \hat{\lambda} - \lambda$
Now, $E[\hat{\lambda} - \lambda] = E[\hat{\lambda}] - \lambda = \lambda - \lambda = 0$
$\text{Var}(\hat{\lambda} - \lambda) = \text{Var}(\hat{\lambda}) = \frac{\sigma^2}{n} \approx \frac{s^2}{n} = \frac{2.34}{50} = 0.0468$
It implies that error follows Normal$(0, 0.0468)$
Let 95% confidence interval be $[\hat{\lambda} - \delta, \hat{\lambda} + \delta]$ Now,

$$P(|\text{error}| < \delta) = 0.95$$
$$\Rightarrow P(|\text{Normal}(0, 0.0468)| < \delta) = 0.95)$$

It is given that $P(-0.42 < \text{N}(0, 0.0468) < 0.42) = 0.95$
Therefore $\delta = 0.42$
So, 95% confidence interval will be $[2.36, 3.20]$

Question: 6

The outcomes on tossing a coin fifteen times are: H T H H T H H T T H T H T T H. Let p be the probability of heads. Previous records show that heads appear on an average 70% of the time. Find the posterior mean of p using the Beta$(7, \beta)$ prior.

(a) $\frac{4}{5}$

(b) $\frac{2}{5}$

(c) $\frac{3}{5}$

(d) $\frac{6}{7}$

Solution:

Let $\boldsymbol{p}$ denote the probability of heads.
Given that prior of $\boldsymbol{p}$ is Beta$(7, \beta)$ with an average of 0.7.
It implies that $E[\text{Beta}(7, \beta)] = 0.7$

$$\Rightarrow \frac{7}{7+\beta} = 0.7$$

$$\Rightarrow \beta = 3$$

Therefore, prior distribution of $\boldsymbol{p}$ is Beta$(7, 3)$
It implies that $f_{\boldsymbol{p}}(p) \propto p^6(1-p)^2$
Now, posterior density $\propto P(X_1 = x_1, \ldots, X_n = x_n | \boldsymbol{p} = p) f_{\boldsymbol{p}}(p)$

$$\Rightarrow \text{posterior density} \propto p^8(1-p)^7(p^6(1-p)^2)$$

$$\Rightarrow \text{posterior density} \propto p^{14}(1-p)^9$$

$$\Rightarrow \text{posterior density} = \text{Beta}(15, 10)$$

$$\Rightarrow \text{posterior mean} = \frac{15}{15+10} = \frac{15}{25} = 0.6$$