QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Practical 9

2023-2024

SOLUTIONS

1 Review

In a factory producing computer chips, on the production line chips are inspected one at a time until a defective chip is found. The probability of each chip being defective is q, where q is unknown. The number of chips inspected before the first defective one is found can be modelled using the geometric distribution with parameter q, and pmf

$$p(x|q) = (1-q)^x q, \quad x = 1, 2, \ldots$$

This process is repeated n times, and the number of chips inspected before the first defective one was $x_1, \ldots, x_n$. Let that the observed data for $n = 5$ be $x_1, \ldots, x_5 = 6, 2, 7, 0, 4$.

1. What is the maximum likelihood estimate (MLE), $\hat{q}$, of q?
2. A beta(α, β) prior distribution is chosen for q. Before seeing the data, we would like the prior mean to be 0.2 and the standard deviation 0.1. Find the parameters of a beta distribution that satisfy this.
3. In R, generate a random sample of size $10,000$ from the beta distribution derived as a prior in question (b), and check that the mean and standard deviation of the sample are approximately 0.2 and 0.1, respectively.
4. Find the posterior distribution for q under the geometric likelihood and the prior computed from (b). What is the posterior mean?
5. Show that the posterior mean for q is always in between the prior mean and the MLE for this example. As $n \to \infty$ show that the posterior mean tends to the MLE of q.
6. Use R to find the posterior median and a 95% credible interval for q.

   Let $\psi = E(q^2)$, be the second moment of q.

7. Use Monte Carlo integration to approximate ψ and plot the histogram of an IID sample from ψ.

1. The likelihood is

$$p(x \mid q) = \prod_{i=1}^{n} q(1-q)^{x_i} = q^n(1-q)^{(S)}, \text{ where } S = \sum_{i=1}^{n} x_i.$$

   Taking the derivative with respect to q, set it to 0, gives that the MLE for q, is

$$\hat{q} = \frac{n}{n+S} = 0.208$$

2. If we take a Beta(α, β) prior distribution for q, we derived the posterior distribution as Beta(α_1, β_1) with $\alpha_1 = n + \alpha$ and $\beta_1 = S + \beta$.

   For choosing the prior parameters, the mean and variance of a beta distribution are

$$m = \frac{\alpha}{\alpha+\beta}, \quad v = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{m(1-m)}{(\alpha+\beta+1)}$$

   Rearranging gives

$$\begin{aligned}
(\alpha+\beta+1)v &= m(1-m) \\
\alpha+\beta &= m(1-m)/v - 1 \\
\alpha &= m(\alpha+\beta) = m\left(m(1-m)/v - 1\right) \\
\beta &= m(1-m)/v - 1 - \alpha
\end{aligned}$$

   Substituting $m = 0.2, v = 0.1^2$ gives $\alpha = 3$, $\beta = 12$ for the prior parameters.

3.
```
prior_sample=rbeta(10000,shape1=alpha_prior,shape2=beta_prior)
mean(prior_sample)
sd(prior_sample)
```

4. The posterior density of $p(q \mid x)$ is beta$(n+\alpha, S+\beta)$. We also have $n = 5$ and $S = 19$, so the posterior distribution for q is Beta$(8, 31)$, with posterior mean $\dfrac{8}{39} = 0.205$.

5. The posterior mean is $\hat{q}_B = \frac{\alpha+n}{\alpha+\beta+n+S}$. Hence, we have

$$\begin{aligned}
\hat{q}_B = \frac{\alpha+n}{\alpha+\beta+n+S} &= \frac{\alpha+\beta}{\alpha+\beta+n+S}\frac{\alpha}{\alpha+\beta} + \frac{n+S}{\alpha+\beta+n+S}\frac{n}{n+S} \\
&= w\frac{\alpha}{\alpha+\beta} + (1-w)\frac{n}{n+S}
\end{aligned}$$

   with $w = \frac{\alpha+\beta}{\alpha+\beta+n+S}$. Hence, we observe as $n \to \infty$, $w \to 0$ and $1-w \to 1$, and $\hat{q}_B \to \frac{n}{n+S}$

6. The R commands for the posterior median and a 95% credible interval are

```
> alpha_prior=3
> beta_prior=12
> #posterior parameters
> alpha_posterior=alpha_prior+n
> beta_posterior=beta_prior+S
```

```
> qbeta(0.5,shape1=alpha_posterior,shape2=beta_posterior)
[1] 0.2000546
> qbeta(c(0.025, 0.975),shape1=alpha_posterior,shape2=beta_posterior)
[1] 0.09554112 0.34326199
>
```

which gives 0.2 and $(0.096, 0.343)$.

ψ is the second moment of q.

7.
```
> q_post_sample=rbeta(10000,shape1=alpha_posterior,shape2=beta_posterior
  )
> mean(q_post_sample^2)
[1] 0.04606217
> hist(q_post_sample^2,freq=FALSE)
```