QUEEN MARY, UNIVERSITY OF LONDON

MTH6102: Bayesian Statistical Methods

Exercise sheet 4

2023-2024

The deadline for submission is **Monday the 30th October at 11am**. Late submissions receive zero marks. You can submit a Word document, pdf or a clearly legible image of hand-written work.

1. **25 points.** Suppose that the number of new cases of a medical condition observed each week can be modelled using a negative binomial distribution with parameters q and r, q is unknown, while r is known.

   [See the table of common distributions in the Module Content section on QMPlus for details of the negative binomial distribution.]

   We observe n weeks' worth of data, and the number of cases each week was $y_1, \ldots, y_n$.

   (a) Show that a beta distribution provides a conjugate prior distribution for q, and find the posterior distribution with such a prior.

   The column in the exercise 2 dataset labelled `y`, contains the observed data $y_1, \ldots, y_n$. Assume that r is equal to 3.

   (b) With a uniform prior distribution for q on the interval $[0, 1]$, what is the posterior distribution for q (including the numerical value of the parameters)?

   (c) What is the posterior mean?

   (d) Use R to find the posterior median and a 95% credible interval for q.

   Solution. The likelihood is

$$p(y \mid q) = \prod_{i=1}^{n} \binom{r + y_i - 1}{y_i} q^r (1-q)^{y_i} \propto q^{rn}(1-q)^S, \text{ where } S = \sum_{i=1}^{n} y_i.$$

   The pdf for a Beta(α, β) prior is

$$p(q) \propto q^{\alpha-1}(1-q)^{\beta-1}$$

   The posterior is

$$\begin{aligned} p(q \mid y) &\propto p(q)\, p(y \mid q) \\ &\propto q^{\alpha-1}(1-q)^{\beta-1} q^{rn}(1-q)^S = q^{rn+\alpha-1}(1-q)^{S+\beta-1} \end{aligned}$$

   The posterior density is proportional to a Beta(α_1, β_1) pdf with $\alpha_1 = rn + \alpha$ and $\beta_1 = S + \beta$, hence this is the posterior distribution. The prior and posterior are in the same family, so the Beta distributions are conjugate to the negative binomial likelihood.

   A uniform prior is Beta$(1, 1)$. If the data is stored in the vector y, then the following R code calculates the posterior parameters and estimates for parts (b) to (d).

```
S = sum(y)
n = length(y)
alpha = 1
beta = 1
r = 3

# posterior parameters
alpha1 = r*n+alpha
beta1 = S+beta

# posterior mean
alpha1/(alpha1+beta1)

# posterior median
qbeta(0.5, shape1=alpha1, shape2=beta1)

# equal tail 95% credible interval
qbeta(c(0.025, 0.975), shape1=alpha1, shape2=beta1)
```

2. **25 points.** For the binomial model considered in the lectures, with success probability q and observed data k successes out of n trials, assume a Beta(α, β) prior distribution for q. Show that the posterior mean for q is always in between the prior mean for q and the maximum likelihood estimator $\hat{q}$.

   Show that if the prior is uniform, then the posterior variance for q is always less than the prior variance. Find an example of α, β, k and n where the posterior variance is greater than the prior variance.

   Solution. The posterior mean for q is

$$E(q \mid k) = \frac{k+\alpha}{n+\alpha+\beta}.$$

   The MLE is $\frac{k}{n}$ and the prior mean is $\frac{\alpha}{\alpha+\beta}$.

   The posterior mean can be written as

$$\frac{k+\alpha}{n+\alpha+\beta} = w\frac{k}{n} + (1-w)\frac{\alpha}{\alpha+\beta}$$

   where

$$w = \frac{n}{n+\alpha+\beta}, \; 1-w = \frac{\alpha+\beta}{n+\alpha+\beta}.$$

   $0 \le w \le 1$, so the posterior mean is always in between the prior mean and the MLE.

   With a uniform prior distribution, the prior variance is $1/12$, using either a formula for the uniform distribution, or for the beta distribution with $\alpha = 1, \beta = 1$. The variance of a Beta(α, β) distribution is

$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

The posterior distribution is Beta$(k+\alpha, n-k+\beta)$. So the posterior variance with a uniform prior is

$$\frac{(k+1)(n-k+1)}{(n+2)^2(n+3)}.$$

$\dfrac{(k+1)}{(n+2)}$ and $\dfrac{(n-k+1)}{(n+2)}$ are two numbers that sum to 1, so their product is at most 1/4.

Also, $1/(n+3) < 1/3$.

So the product of all three is less than 1/12.

The posterior variance will be lower than the prior variance for larger n, so we should try very small n, k. Trying $k = 0, n = 1$, with various α and β gives for $\alpha = 3$ and $\beta = 1$ the prior variance

$$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.0375$$

and the posterior variance is

$$\frac{(k+\alpha)(n-k+\beta)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)} = 0.04.$$

3. **25 points.** The number of offspring X in a certain population has probability mass function

$$p(x \mid \theta, \psi) = \begin{cases} \theta & x = 0 \\ (1-\theta)\psi(1-\psi)^{x-1} & x = 1, 2, \ldots \end{cases}$$

where θ and ψ are unknown parameters in the interval $[0, 1]$.

Write down the likelihood when r zeroes and $n - r$ non-zero values $x_1, x_2, \ldots, x_{n-r}$ are observed from n independent observations on X.

Suppose θ and ψ have independent prior beta densities with parameters a, b and c, d, respectively. Show that θ and ψ have independent posterior beta distributions and identify the posterior parameters.

[*Hint:* You may use the fact that two continuous random variables X and Y are independent, if their joint density $f_{X,Y}(x, y)$ can be written as the product of their marginal densities $f_X(x)$ and $f_Y(y)$, respectively. That is, $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.]

Solution. The likelihood is

$$p(x \mid \theta, \psi) = \prod_{j=1}^{r} \theta \prod_{i=1}^{n-r} (1-\theta)\psi(1-\psi)^{x_i-1} = \theta^r(1-\theta)^{n-r}\psi^{n-r}(1-\psi)^{S-(n-r)}$$

where $S = \sum_{i=1}^{n-r} x_i$. The prior distributions are

$$p(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$$

$$p(\psi) \propto \psi^{c-1}(1-\psi)^{d-1}$$

So the posterior is

$$\begin{aligned}
p(\theta, \psi \mid x) &\propto p(x \mid \theta, \psi)\, p(\theta, \psi) \\
&= p(x \mid \theta, \psi)\, p(\theta) p(\psi) \\
&\propto \theta^r (1-\theta)^{n-r} \psi^{n-r} (1-\psi)^{S-(n-r)} \theta^{a-1} (1-\theta)^{b-1} \psi^{c-1} (1-\psi)^{d-1} \\
&= \theta^{r+a-1} (1-\theta)^{n-r+b-1} \psi^{n-r+c-1} (1-\psi)^{S-(n-r)+d-1}
\end{aligned}$$

This posterior is a product of terms involving θ and terms involving ψ, and so θ and ψ are independent in the posterior distribution. We can write

$$p(\theta, \psi \mid x) = p(\theta \mid x)\, p(\psi \mid x).$$

Picking out the terms from $p(\theta, \psi \mid x)$, we have

$$p(\theta \mid x) \propto \theta^{r+a-1} (1-\theta)^{n-r+b-1}$$

$$p(\psi \mid x) \propto \psi^{n-r+c-1} (1-\psi)^{S-(n-r)+d-1}.$$

Hence the marginal posterior for θ is $\text{Beta}(r+a, n-r+b)$, and for ψ it is $\text{Beta}(n-r+c, S-(n-r)+d)$.

4. **25 points.** Your friend transmits an unknown value θ to you over a noisy channel. The noise is normally distributed with mean 0 and a known variance 4, so the value x that you receive is modeled by $N(\theta, 4)$. Based on previous communications, your prior on θ is $N(5, 9)$.

   (a) Suppose your friend transmits a value to you that you receive as $x = 6$. Show that the posterior pdf for θ is $N(74/13, 36/13)$. For this problem, you need to derive the posterior by carrying out the calculations from scratch.

   (b) Suppose your friend transmits the same value θ to you $n = 4$ times. You receive these signals plus noise as $x_1, \ldots, x_4$ with sample mean $\bar{x} = 6$. Using the same prior and known variance σ^2 as in part (a), show that the posterior on θ is $N(5.9, 0.9)$. Plot the posterior and posterior on the same graph. Describe how the data changes your belief about the true value of θ. For this question, you may use the normal updating formulas.

   (c) How do the mean and variance of the posterior change as more data is received?

   (d) IQ in the general population follows a $N(100, 15^2)$ distribution. An IQ test is unbiased with a normal variance of 10^2; that is, if the same person is tested multiple times, their measured IQ will differ from their true IQ according to a normal distribution with 0 mean and variance 100.

      i. Tommy Vard scored an 80 on the test. What is the expected value of his true IQ?

      ii. Anna Taft scored a 150 on the test. What is the expected value of her true IQ?

Solution.

(a) The posterior density is

$$f(\theta|x) = c_1 \exp(-(\theta-5)^2/18) \times \exp(-(6-\theta)^2/8)$$

for some constant $c_1 > 0$. All we need is some algebraic manipulations of the exponent in the posterior.

$$\begin{aligned} -(\theta-5)^2/18 - (6-\theta)^2/8 &= -\frac{1}{2}\Big(\frac{\theta^2 - 12\theta + 36}{4} + \frac{\theta^2 - 10\theta + 25}{9}\Big) \\ &= -\frac{1}{2}\Big(\frac{13\theta^2 - 148\theta + 424}{36}\Big) \\ &= -\frac{1}{2}\Big(\frac{(\theta - 74/13)^2}{36/13} + C\Big), \end{aligned}$$

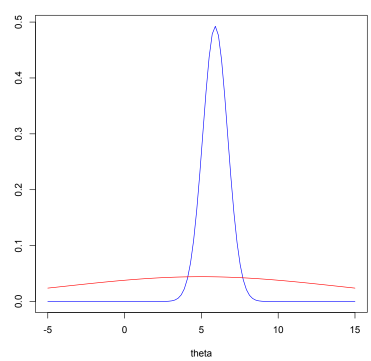
where C is a constant. Thus, the posterior is

$$f(\theta|x) \propto \exp\Big(-\frac{(\theta-74/13)^2}{2 \cdot 36/13}\Big).$$

This has the form of a pdf for $N(74/13, 36/13)$.

(b) We have $\mu_0 = 5$, $\sigma_0^2 = 9$, $\bar{x} = 9$, $\sigma^2 = 4$ and $n = 4$. So using the updating normal-normal formulas

$$a = 1/9, b = 1, a + b = 10/9 \Rightarrow \mu_1 = \frac{5/9 + 6}{10/9} = 5.9, \sigma_1^2 = 1/(10/9) = 9/10 = 0.9.$$



So the posterior is $N(5.9, 0.9)$. After observing $x_1, \ldots, x_4$, we see that the posterior mean is close to the MLE $\bar{x}$ and the posterior variance is much smaller than the prior variance. The data has made us more certain about the location of θ.

(c) As more data is received n increases, so b increases, so the mean of the posterior is closer to the data mean and the variance of the posterior decreases. Since the variance goes down, we gain more certainty about the true value of θ.

(d) With no new data we are given the prior $f(\theta) \sim N(100, 15^2)$. For data x =score on the IQ test we have the likelihood $f(x|theta) \sim N(\theta, 10^2)$. Using the updating normal-normal formulas we have $\mu_0 = 100$ $\sigma_0^2 = 15^2$, $\sigma^2 = 10^2$, $n = 1$. So $a = 1/225$ and $b = 1/100$.

  i. Tommy Vard, $x = 80$, so $\mu_1 = \frac{a100 + b80}{a+b} = 86.15$.

  ii. Anna Taft, $x = 150$, so $\mu_1 = \frac{a100 + b150}{a+b} = 134.62$.