STAT3016/4116/7016 Tutorial 1

Dehua Tao

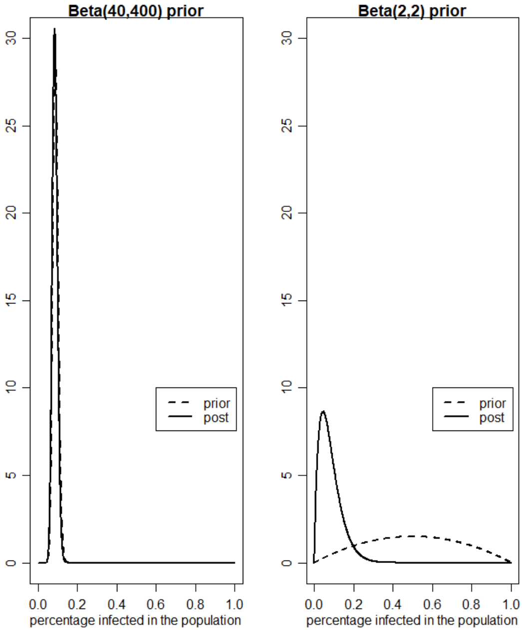
Australian National University

August 3, 2023

Revision

- Bayes theorem: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Show in another way: $P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{P(Y)}$.
- Assumption: Exchangeability.
- $P(\theta)$: Prior distribution.(noninformative, informative, weakly informative)
- $P(Y|\theta)$: Likelihood.(Contains information in the data or sample.)
- $P(\theta|Y) \propto P(Y|\theta)P(\theta)$: Posterior distribution.
- Conjugate prior: $P(\theta)$ and $P(\theta|Y)$ have the same distribution.

Question 1



Beta(40,400) prior
30
25
20
15
10
5
0
prior
post
0.0
0.2
0.4
0.6
0.8
1.0
percentage infected in the population
Beta(2,2) prior
30
25
20
15
10
5
0
prior
post
0.0
0.2
0.4
0.6
0.8
1.0
percentage infected in the population

Question 1

- Under Beta(40,400) prior: the posterior is Beta(40,420), reflects strong prior belief. Hence this is the prior dominates the posterior case.
- Under Beta(2,2) prior: the posterior is Beta(2,22). Beta(2,2) is a weak prior, data provides more information compared to the prior. Hence this is the likelihood that dominates the posterior case.
- A Uniform(0,1) prior is equivalent to a Beta(1,1) prior. The prior distribution is flat on the interval (0,1). In other words, there is no prior information and the posterior distribution of θ is completely driven by the likelihood.

Question 1

- Under the Unif $(0.15, 0.20)$ prior, then $p(\theta) = \frac{1}{0.05} \propto k$ where $\theta \in (0.15, 0.20)$, for some constant k. So

$$p(\theta \mid y) \propto \theta^y (1-\theta)^{n-y} \text{ for } \theta \in (0.15, 0.20)$$

Let's find the normalising constant c such that $\int_{0.15}^{0.20} c \cdot \theta^y (1-\theta)^{n-y} d\theta = 1$.

$$c \times B(y+1, n-y+1) \int_{0.15}^{0.20} \frac{\theta^y (1-\theta)^{n-y}}{B(y+1, n-y+1)} d\theta = 1$$
$$c \times B(y+1, n-y+1) \times [F^*(0.20) - F^*(0.15)] = 1$$

where $F^*(x)$ is the CDF of a $\mathrm{Beta}(\mathrm{y}+1, \mathrm{n}-\mathrm{y}+1)$ distribution evaluated at x.

Question 1

- Therefore

$$c = \frac{1}{B(y+1, n-y+1) \times [F^*(0.20) - F^*(0.15)]}$$

and

$$p(\theta \mid y) = \frac{\theta^y (1-\theta)^{n-y}}{B(y+1, n-y+1) \times [F^*(0.20) - F^*(0.15)]}$$

for $\theta \in (0.15, 0.20)$.

Question 2(a)

- Prior distribution for θ: $P(\theta = 1) = P(\theta = 2) = \frac{1}{2}$.
- Given $\sigma^2 = 2^2$, we have:
  $(y|\theta = 1, \sigma^2 = 2^2) \sim N(1, 2^2)$
  $(y|\theta = 2, \sigma^2 = 2^2) \sim N(2, 2^2)$
- The marginal probability density for y:

$$\begin{aligned} p\left(y|\sigma^2 = 2^2\right) &= \sum_\theta p\left(y, \theta|\sigma^2 = 2^2\right) \\ &= \sum_\theta p\left(y|\theta, \sigma^2 = 2^2\right) p(\theta) \\ &= 0.5N\left(y|\theta = 1, \sigma^2 = 2^2\right) + 0.5N\left(y|\theta = 2, \sigma^2 = 2^2\right) \end{aligned}$$

Question 2(b)

- $\mathrm{P}(\theta = 1 | y = 1, \sigma^2 = 2^2)$:

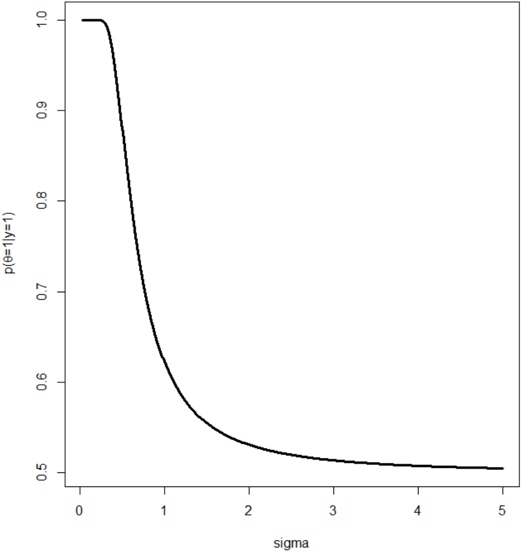
$$
\begin{aligned}
&= \frac{p(y = 1, \theta = 1, \sigma^2 = 2^2)}{p(y = 1)} \\
&= \frac{0.5N\left(y = 1 | \theta = 1, \sigma^2 = 2^2\right)}{0.5N\left(y = 1 | \theta = 1, \sigma^2 = 2^2\right) + 0.5N\left(y = 1 | \theta = 2, \sigma^2 = 2^2\right)} \\
&= 0.5312094
\end{aligned}
$$

- the posterior probability is slightly higher than 0.5 after observing a $y = 1$, this is because due to $y = 1$ has a slightly higher probability being sampled from $N(1, 4)$ distribution compared to $N(2, 4)$, θ has a slightly higher probability equal to 1.

Question 2(b)

- why $p(y=1,\theta=1,\sigma^2=2^2)=0.5N\left(y=1|\theta=1,\sigma^2=2^2\right)$?
- y follows a Normal distribution, we know that for continuous random variable y, $p(y=c)=0$.
- In Bayesian statistics, we use the value of the density to represent the joint probability, this value alone is meaningless, but the ratio is meaningful. The ratio is just the posterior odds. We can use posterior odds to do the Bayesian inference.

Question 2(c)



Question 2(c)

- when σ^2 increases, posterior odds decreases, and when σ^2 decreases, posterior odds increases.
- when $\sigma^2 = 0$, $y = 1$ is equivalent to $\theta = 1$ with no uncertainty.
- when σ^2 is large, we assume more variability in the sampling model, we have less confidence to conclude that $\theta = 1$ given $y = 1$.

Question 3(a)

- p: the probability that a passenger who is scheduled to take a particular flight, fails to show up.(the value of the parameter)
- $g(p)$: the prior probability that p equals to a certain value.
- Derive posterior probability $g(p|y)$:

$$g(p = k|y) = \frac{g(y|p = k)g(p = k)}{g(y)}$$

  for $k = 0, 0.025, 0.05, 0.075, 0.10.$
- likelihood $g(y|p = k) = \left(\begin{array}{c} 60 \\ 55 \end{array} \right) k^5(1 - k)^{55}.$

Question 3(a)

- We need to calculate $g(y)$ first because the prior distribution of p doesn't come from a known distribution with proper *pmf*. (when $p = 0$, $g(p) = 0.8$ isn't calculated from a known *pmf* formula.)
- $g(y)$:

$$g(y) = \sum_{p=k} g(y|p = k)g(p = k)$$
$$= 0.01512898$$

- $g(p = 0|y) = 0$: $g(y|p = k) = 0$.
- $g(p = 0.10|y) = \dfrac{\binom{60}{55} 0.9^{55} 0.1^{5} \times 0.015}{g(y)} = 0.1647906.$

Question 3(b)

- collect another 10 observations (11 in total), then calculate $g(p|y)$: do what we did in (a) for another 10 times.
- we choose to use the posterior of p in (a) as the prior in (b) in order to use the latest information.
- $p = 0.075$ has the largest posterior probability: 8.998135e-01 which means $p = 0.075$ is the posterior mode.
- in fact, in each iteration related to updating the posterior probability, the posterior probability in the last iteration is used as the prior probability in the next iteration.

Question 3(c)

- If the airline company maintains the same prior distribution, and bases their profit forecasts on this prior distribution, the company would be overestimating the probability that all customers show up. (will lose profit)
- The company could potentially overbook a higher proportion of seats to boost profits. (Based on posterior modes, 0.075 of passengers may not show up, overbook the seats may reduce the number of empty seats on the plane)
- The best way is to use the past data continuously update the probability of passengers showing up for particular flights allowing for more accurate predictions to boost profits or reduce losses.