BST 5230 Bayesian Statistics - Spring 2018

Mid-term Exam

Miao Cai miao.cai@slu.edu

2018-03-22

1.(10 Points 5,5)) A disease has base rate 3%. The sensitivity of a test is 0.80 and the specificity is 0.995.

(a) Find the probability that a person selected at random tests positive for the disease.

(b) Find the positive predictive value.

Answers:

(a)

$$
\begin{aligned}
P(T=+) &= P(T=+, D=+) + P(T=+, D=-) \\
&= P(T=+|D=+) * P(D=+) + P(T=+|D=-) * P(D=-) \\
&= P(T=+|D=+) * P(D=+) + [1-P(T=-|D=-)] * [1-P(D=-)] \\
&= 0.8 * 0.03 + (1-0.995)(1-0.03) \\
&= 0.02885
\end{aligned}
$$

(b)

$$
\begin{aligned}
PPV &= P(D=+|T=+) \\
&= \frac{P(D=+)P(T=+|D=+)}{P(T=+)} \\
&= \frac{0.03 * 0.80}{0.02885} \\
&= 0.8318891
\end{aligned}
$$

2. (20 Points) Suppose that $X_1, X_2, \dots X_n \sim$ i.i.d GEOMETRIC(θ). That is, they are geometric random variables. (That is, each is the number of trials needed to get one success where the probability of success is θ.) The probability mass function for the geometric distribution is

$$f(x|\theta) = \theta(1-\theta)^{x-1}, \quad x = 1, 2, 3, \dots$$

Suppose that we assume a BETA(α, β) prior distribution for θ.

(a) Find the posterior distribution for θ.

(b) Is the BETA(α, β) prior distribution a conjugate prior? Explain.

(c) Suppose that $n = 5$ and we obtain $x_1 = 1, x_2 = 5, x_3 = 4, x_4 = 2, x_5 = 8$. Assume a BETA(2,3) prior distribution. Find the posterior distribution and plot it.

(d) Find the posterior mean and standard deviation.

Answers:

(a)

$$\begin{aligned}
P(\theta|X) &= \frac{c_1\theta^{\alpha-1}(1-\theta)^{\beta-1} * \theta(1-\theta)^{x-1}}{c_2} \\
&= c_3\theta^{\alpha}(1-\theta)^{x+\beta-1} \quad c_1, c_2, c_3 \text{are supposed to be unknow constants} \\
&\sim \text{BETA}(\alpha+1, \beta+x-1)
\end{aligned}$$

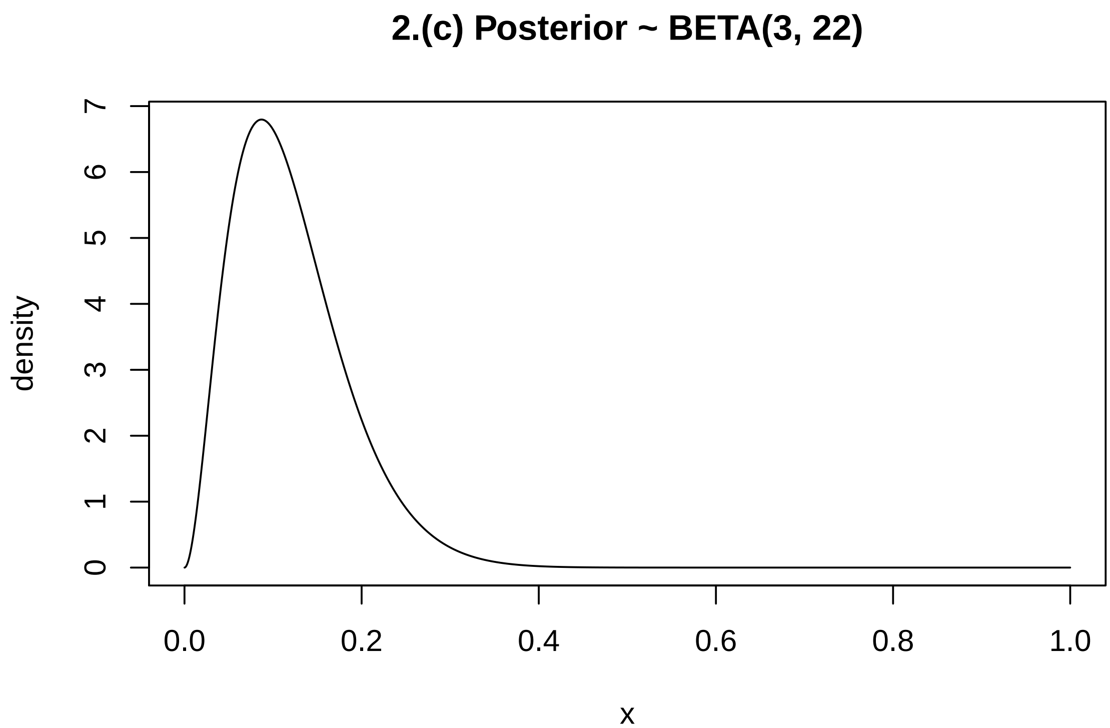
(b) Yes, BETA(α, β) is a conjugate prior for geometric data because if we have prior BETA(α, β) and geometric data, we will get posterior BETA($\alpha + 1,\ \beta + X - 1$), which is in the same family of prior distribution.

(c) According to the result from part (a), we have:

$$\begin{aligned}
P(\theta|X) &\sim \text{BETA}(\alpha+1, \beta + \sum_{i=1}^{5} x_i - 1) \\
&\sim \text{BETA}(3, 22)
\end{aligned}$$

The plot of the posterior distribution is as follows:

```
x = seq(0, 1, 0.0001)
d = dbeta(x, 3, 22)
plot(x, d, type = "l", main = "2.(c) Posterior ~ BETA(3, 22)", ylab = "density")
```



(d) According to the properties of BETA distribution:

$$Mean = \frac{\alpha}{\alpha + \beta} = \frac{3}{3 + 22} = \frac{3}{25}$$

$$s.d. = \sqrt{Var} = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}} = \sqrt{\frac{3 * 22}{(3+22)^2(3+22+1)}} = 0.0637302$$

3. (6 Points) Choose and plot a prior that describes your belief in the following circumstances. Explain your reasoning in both cases.

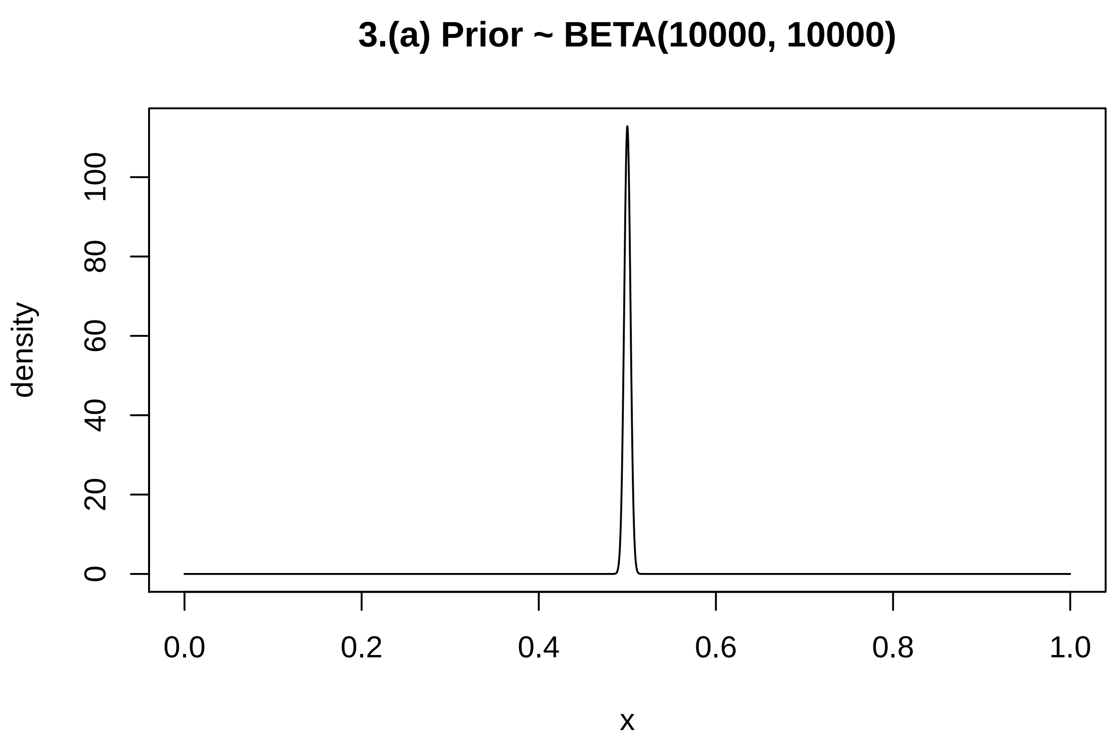
(a) Suppose you have a coin that you know is minted by the US government and has not been tampered with. Therefore you have a strong prior belief that the coin is fair.

(b) Now you have a different coin, this one made of some strange material and marked "Patent Pending: International Magic Company."

Answers:

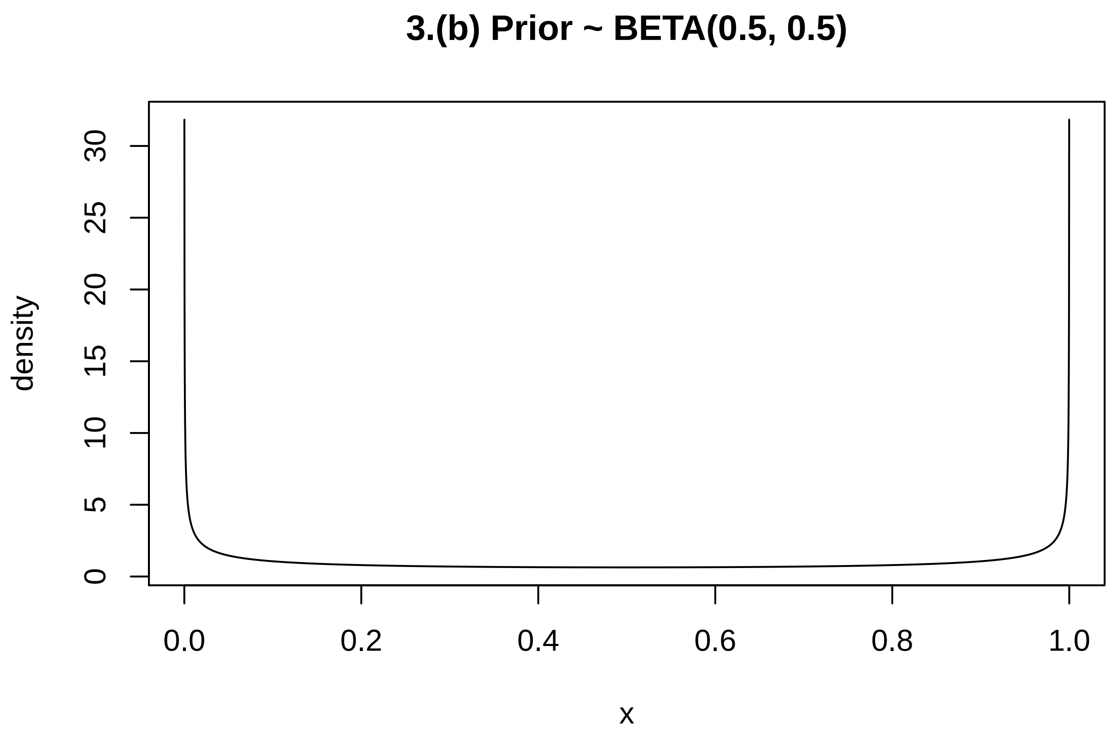
(a) Since we have a strong belief that the coin should be fair, the mean should be 0.5, the variance should be very small and the support should be within (0, 1). So I choose the prior BETA(10000, 10000). Here is the plot of the prior distribution I choose:

```r
x = seq(0, 1, 0.0001)
d = dbeta(x, 10000,10000)
plot(x, d, type = "l", main = "3.(a) Prior ~ BETA(10000, 10000)", ylab = "density")
```



(b) Since we have a strong belief that the coin should be either 0 or 1, the density should be high at either 0 or 1, and the density should be very small in between 0 and 1 and the support should be within (0, 1). So I choose the prior BETA(0.5, 0.5). Here is the plot of the prior distribution I choose:

```r
x = seq(0, 1, 0.0001)
d = dbeta(x, 0.5,0.5)
plot(x, d, type = "l", main = "3.(b) Prior ~ BETA(0.5, 0.5)", ylab = "density")
```



4. (16 Points) Suppose that $X|\theta \sim BIN(4, \theta)$, that is binomial with 4 trials and probability of success. Assume a BETA(2,3) prior distribution for . We observed $x = 1$ success out of the 4 trials.

(a) Suppose we apply the Metropolis-Hastings algorithm beginning with $\theta^{(1)} = 0.6$. Suppose we use as the proposal density normal distribution with mean $\theta^{(t)}$ and variance $\sigma^2 = 0.04$. Find the acceptance probability if the first proposal is $\theta^{(\text{prop})} = 0.56$.

(b) Suppose we apply the Metropolis-Hastings algorithm beginning with $\theta^{(1)} = 0.6$. Suppose we use as the proposal density normal distribution with mean $\theta^{(t)}$ and variance $\sigma^2 = 0.04$. Find the acceptance probability if the first proposal is $\theta^{(\text{prop})} = 0.62$.

(c) Suppose we apply the Metropolis-Hastings algorithm beginning with $\theta^{(1)} = 0.6$. Suppose we use as the proposal density a BETA(2,3) prior distribution. Find the acceptance probability if the first proposal is $\theta^{(\text{prop})} = 0.56$.

(d) Among (a) â (c) which is (are) the simpler Metropolis and which is (are) the Metropolis-Hastings algorithm.

Answers:

(a) The R codes for calculating the acceptance probability and the result are as follows:

```
# Essential part of BETA distribution
pBeta = function(p,a,b)
{
  z = 0
  if (p >= 0 && p <= 1) {  z = p^(a-1) * (1-p)^(b-1) }
  return(z)
}

# Binomial likelihood function
pLik = function(x,n,p) { p^x * (1-p)^(n-x) }

# Proposal distribution
pProp = function(p)
{
  rnorm(n=1,mean=p,sd=sqrt(0.04))
}

n = 4 # the number of trials
x = 1 # the number of success
a = 2 # alpha for beta distribution
b = 3 # beta  for beta distribution
```

```r
pp = 0.56 # current value
pc = 0.6  # proposal value

num = pBeta(pp,a,b)*pLik(x,n,pp) # numerator
den = pBeta(pc,a,b)*pLik(x,n,pc) # denominator

(alpha = num/den) # acceptance probability - alpha
```

```
## [1] 1.402933
```

(b) The R codes for calculating the acceptance probability and the result are as follows:

```r
pp = 0.62 # current value
pc = 0.6  # proposal value

num = pBeta(pp,a,b)*pLik(x,n,pp) # numerator
den = pBeta(pc,a,b)*pLik(x,n,pc) # denominator

(alpha = num/den) # acceptance probability - alpha
```

```
## [1] 0.8262261
```

(c) The R codes for calculating the acceptance probability and the result are as follows:

```r
# Proposal distribution
pProp = function(p)
{
   rbeta(n = 1, shape1 = 2, shape2 = 3)
}

pp = 0.56 # current value
pc = 0.6  # proposal value

num = pBeta(pp,a,b)*pLik(x,n,pp)*pc*(1-pc)^2 # numerator
den = pBeta(pc,a,b)*pLik(x,n,pc)*pp*(1-pp)^2 # denominator

(alpha = num/den) # acceptance probability - alpha
```

```
## [1] 1.242267
```

(d) In this question, (a) and (b) are the simpler Metropolis algorithm because the proposal density are normal distributions which are symmetric, while (c) is the Metropolis-Hastings algorithm because the proposal density is beta which is not symmetric.