STAT 404 Midterm Answers

2018150420 정해원

2022-10-20

1

(a)

```
prior.no.k0 <- function(x) {(20*x*((1-x)^3)+1)}
integrate(prior.no.k0, 0, 1)

## 2 with absolute error < 2.2e-14
```

So $k_0 = \frac{1}{2}$

The prior mean is

$$E(\theta) = \int_0^1 \theta \frac{1}{2}(20\theta(1-\theta)^3 + 1)d\theta$$

```
prior.mean <- function(x) {(1/2)*x*(20*x*((1-x)^3)+1)}
integrate(prior.mean, 0, 1)$value

## [1] 0.4166667
```

The prior variance can be calculated by

$$Var(\theta) = E(\theta^2) - \left(E(\theta)\right)^2 = \int_0^1 \theta^2 \frac{1}{2}(20\theta(1-\theta)^3 + 1)d\theta \ \ - \left(E(\theta)\right)^2$$

```
prior.sum.sq <- function(x) {(1/2)*(x^2)*(20*x*((1-x)^3)+1)}
integrate(prior.sum.sq, 0, 1)$value-(integrate(prior.mean, 0, 1)$value)^2

## [1] 0.06448413
```

Since when $n = 22$, $\sum x_i = 15$, then the likelihood is $L(\theta|x) = \binom{22}{15}\theta^{15}(1-\theta)^7$, so the posterior distribution is proportional to $p(\theta|x) \propto \theta^{15}(1-\theta)^7(\theta(1-\theta)^3 + 1)$

The integral coefficient C can be calculated by

```
post.con <- function(x) {(x^15)*((1-x)^7)*(20*x*((1-x)^3)+1)}
const <- 1/(integrate(post.con, 0, 1)$value)
const

## [1] 2535546
```

The posterior mean is

$$E(\theta) = \int_0^1 C \cdot \theta \cdot \theta^{15}(1-\theta)^7(20\theta(1-\theta)^3 + 1)d\theta$$

```
post.mean <- function(x) {x*const*(x^15)*((1-x)^7)*(20*x*((1-x)^3)+1)}
integrate(post.mean, 0, 1)

## 0.6456196 with absolute error < 9.8e-10
```

The posterior variance is

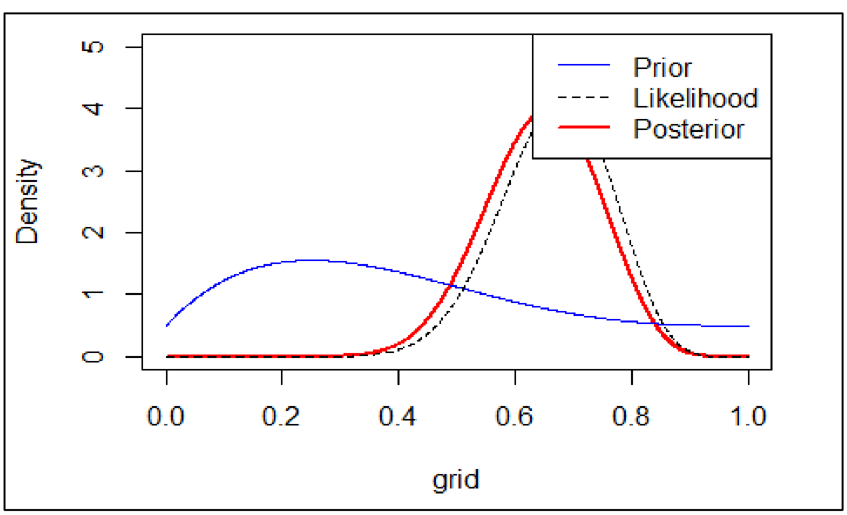
$$Var(\theta) = E(\theta^2) - \left(E(\theta)\right)^2 = \int_0^1 C \cdot \theta^2 \cdot \theta^{15}(1-\theta)^7(20\theta(1-\theta)^3 + 1)d\theta \ \ - \left(E(\theta)\right)^2$$

```
post.sum.sq <- function(x) {(x^2)*const*(x^15)*((1-x)^7)*(20*x*((1-x)^3)+1)}
integrate(post.sum.sq, 0, 1)$value-(integrate(post.mean, 0, 1)$value)^2

## [1] 0.009463911
```

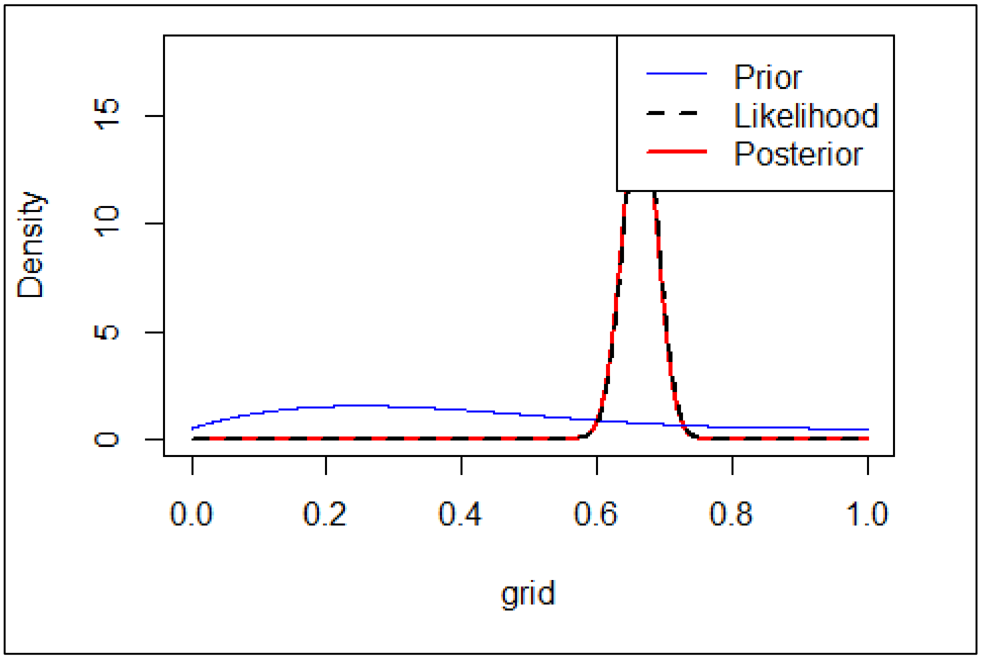
Since θ from posterior has minimum variance, it estimates the true value strongly.

(b)

```
grid <- seq(0, 1, 0.001)
prior <- 0.5*dbeta(grid, 2, 4) + 0.5*dbeta(grid, 1, 1); prior <- prior/sum(pr
ior)
like <- dbinom(15, 22, grid); like <- like/sum(like)
posterior <- prior*like; posterior <- posterior/sum(posterior)
plot(grid, posterior*length(grid), type='l', lwd=2, col="red", ylim=c(0,5), y
lab="Density")
lines(grid, prior*length(grid), col="blue")
lines(grid, like*length(grid), lty=2, col="black")
legend("topright", c("Prior", "Likelihood", "Posterior"), lwd=c(1,1,2), col=c
("blue", "black","red"), lty=c(1,2,1))
```



```r
like <- dbinom(220, 330, grid); like <- like/sum(like)
posterior <- prior*like; posterior <- posterior/sum(posterior)
plot(grid, posterior*length(grid), type='l', lwd=2, col="red", ylim=c(0,18),
ylab="Density")
lines(grid, prior*length(grid), col="blue")
lines(grid, like*length(grid), lty=2, col="black", lwd=2)
legend("topright", c("Prior", "Likelihood", "Posterior"), lwd=c(1,2,2), col=c
("blue", "black","red"), lty=c(1,2,1))
```



(c)

The predictive PMF is

$$p(\tilde{x}|x) = \int_0^1 p\,(\tilde{x}|\theta)p(\theta|x)d\theta = \int_0^1 \binom{15}{\tilde{x}} \theta^{\tilde{x}}(1-\theta)^{15-\tilde{x}} \cdot C\theta^{15}(1-\theta)^7(20\theta(1-\theta)^3+1)d\theta$$

$$= \binom{15}{\tilde{x}} \int_0^1 20\,C \cdot \theta^{16+\tilde{x}}(1-\theta)^{25-\tilde{x}} + C\theta^{15+\tilde{x}}(1-\theta)^{22-\tilde{x}}d\theta$$

$$= \binom{15}{\tilde{x}} \cdot 20C\frac{\Gamma(17+\tilde{x})\Gamma(26-\tilde{x})}{\Gamma(43)}\int_0^1 \frac{\Gamma(43)}{\Gamma(17+\tilde{x})\Gamma(26-\tilde{x})}\theta^{16+\tilde{x}}(1-\theta)^{25-\tilde{x}}d\theta + \binom{15}{\tilde{x}}$$

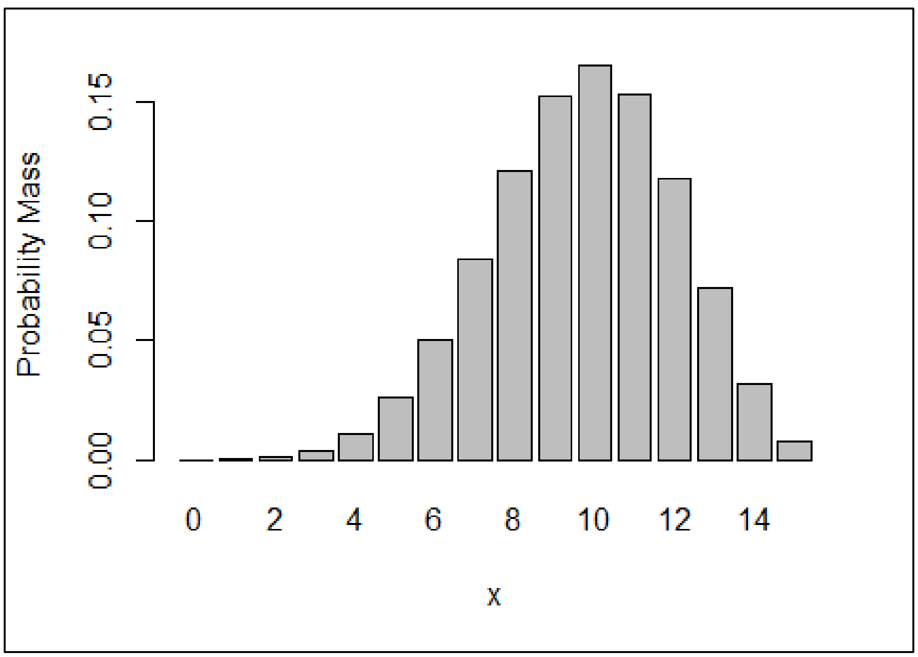
$$\cdot\, C\frac{\Gamma(16+\tilde{x})\Gamma(23-\tilde{x})}{\Gamma(39)}\int_0^1 \frac{\Gamma(39)}{\Gamma(16+\tilde{x})\Gamma(23-\tilde{x})}\theta^{15+\tilde{x}}(1-\theta)^{22-\tilde{x}}d\theta$$

$$p(\tilde{x}|x) = \binom{15}{\tilde{x}} C \cdot (\frac{20\Gamma(17+\tilde{x})\Gamma(26-\tilde{x})}{\Gamma(43)} + \frac{\Gamma(16+\tilde{x})\Gamma(23-\tilde{x})}{\Gamma(39)})$$

```
x<-0:15
pmf <- function(x){
  return(choose(15,x)*const*(20*(1/(choose(41,(16+x))*42))+1/(choose(37,(15+
x))*38)))
}
pmf(x)

##  [1] 1.883908e-05 1.972113e-04 1.067748e-03 3.963313e-03 1.127755e-02
##  [6] 2.607041e-02 5.063646e-02 8.429492e-02 1.215638e-01 1.523698e-01
## [11] 1.653262e-01 1.532945e-01 1.183368e-01 7.236811e-02 3.165630e-02
## [16] 7.557977e-03

sum(pmf(x))

## [1] 1
```

Then, our *pmf* satisfies the condition to be a PMF.

```
barplot(pmf(x), names=c(0:15), xlab="x", ylab="Probability Mass")
```



(d)

1.

Let $p(\theta) = \frac{1}{2}\left(20(\theta(1-\theta)^3) + \frac{1}{2}\right)$ and $c_1 = \frac{1}{B(17,11)}$, $c_2 = \frac{20}{B(16,8)}$. Then,

$$p(\theta)p(x|\theta) \propto \frac{1}{2}\theta^{15}(1-\theta)^7\left(20(\theta(1-\theta)^3) + \frac{1}{2}\right) \propto (20\theta^{16}(1-\theta)^{10} + \theta^{15}(1-\theta)^7$$

$$\propto \frac{c_2}{c_1+c_2} Beta(17,11) + \frac{c_1}{c_1+c_2} Beta(16,8)$$

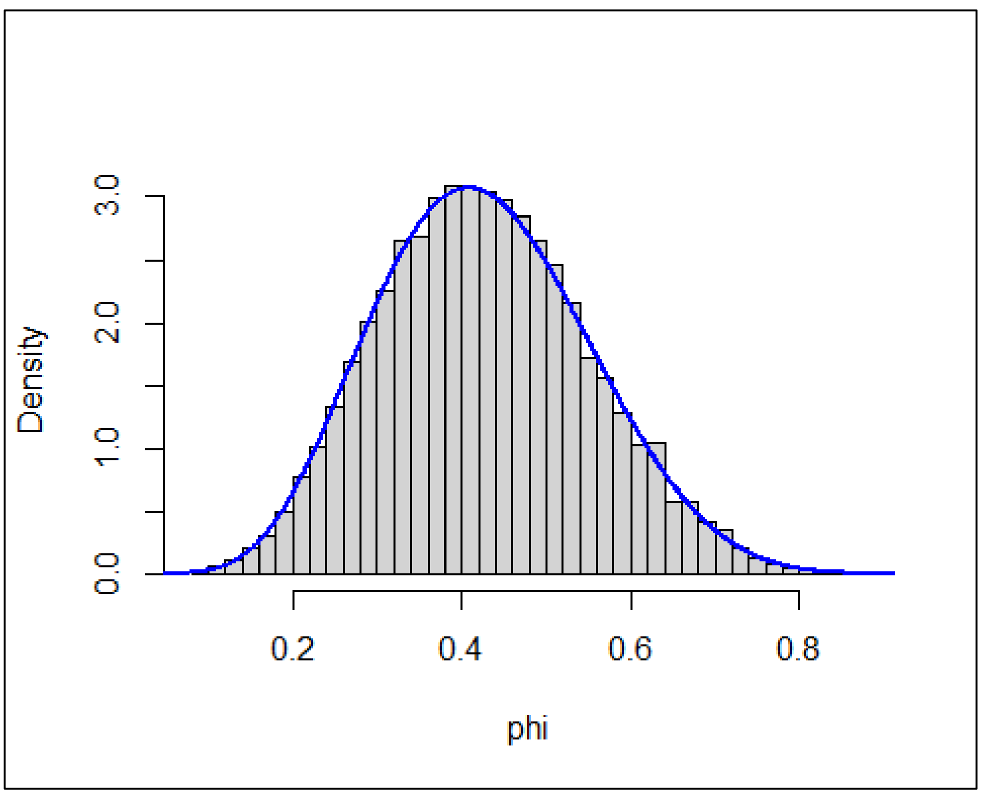
where $W_1 = \frac{c_2}{c_1+c_2}$ and $W_2 = \frac{c_1}{c_1+c_2}$

```
W1 <- 20*23*choose(22,7)/(27*choose(26,10)+20*23*choose(22,7))
W2 <- 27*choose(26,10)/(27*choose(26,10)+20*23*choose(22,7))
W1; W2

## [1] 0.3535912

## [1] 0.6464088
```

The posterior distribution is a mixture of $Beta(17,11)$, $Beta(16,8)$ with weights (0.3536, 0.6464).

The Monte-Carlo approximation of the mixture posterior of $\phi = \theta^2$is as following

```
n <- 10^4
theta <- c(rbeta(n*round(W1,4), 17, 11), rbeta(n*round(W2, 4), 16, 8))
phi <- theta^2
hist(phi, freq=F, breaks = 30, xlab="phi", main="")
grid <- seq(0, 1, 0.001)
mixture <- function(x) {(W1*27*choose(26,10)*(x^8)*((1-sqrt(x))^10) + W2*23*c
hoose(22,7)*(x^(15/2))*((1-sqrt(x))^7))/(2*sqrt(x)) }
lines(grid, mixture(grid), col="blue", lwd=2)
```

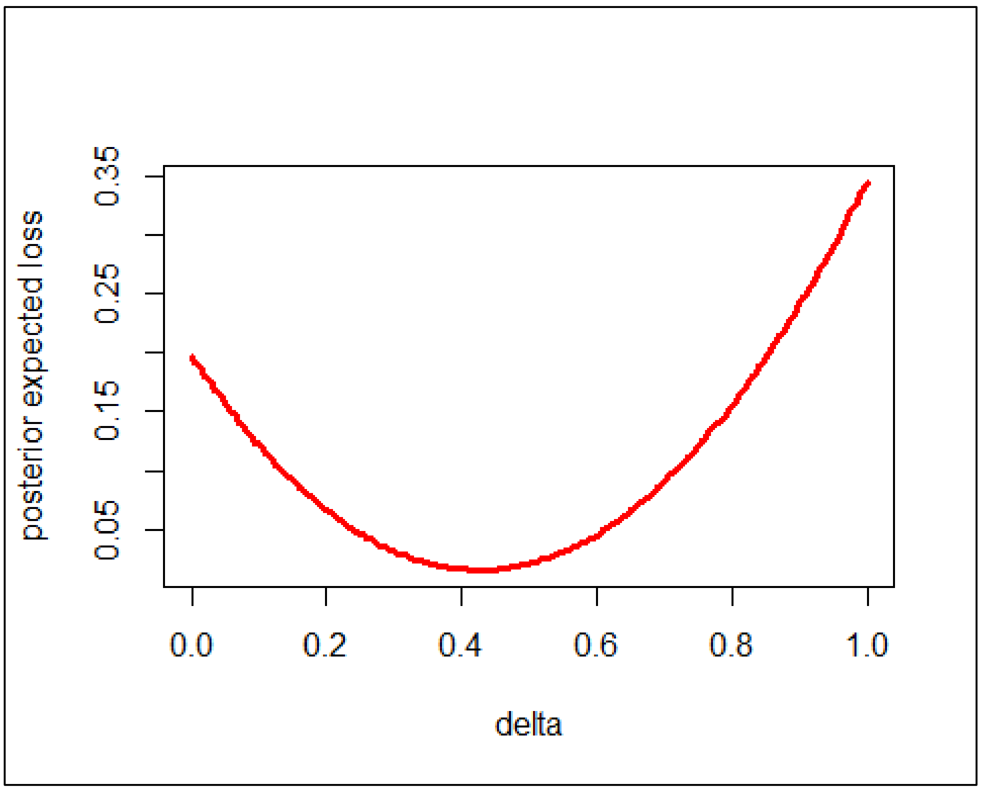


and it is similar as the exact posterior density of phi, where $p_\phi(\phi|x) = p_\theta(\theta|x)\left|\frac{1}{2\sqrt{\phi}}\right|$

2.

By estimating $\hat{\phi}$,

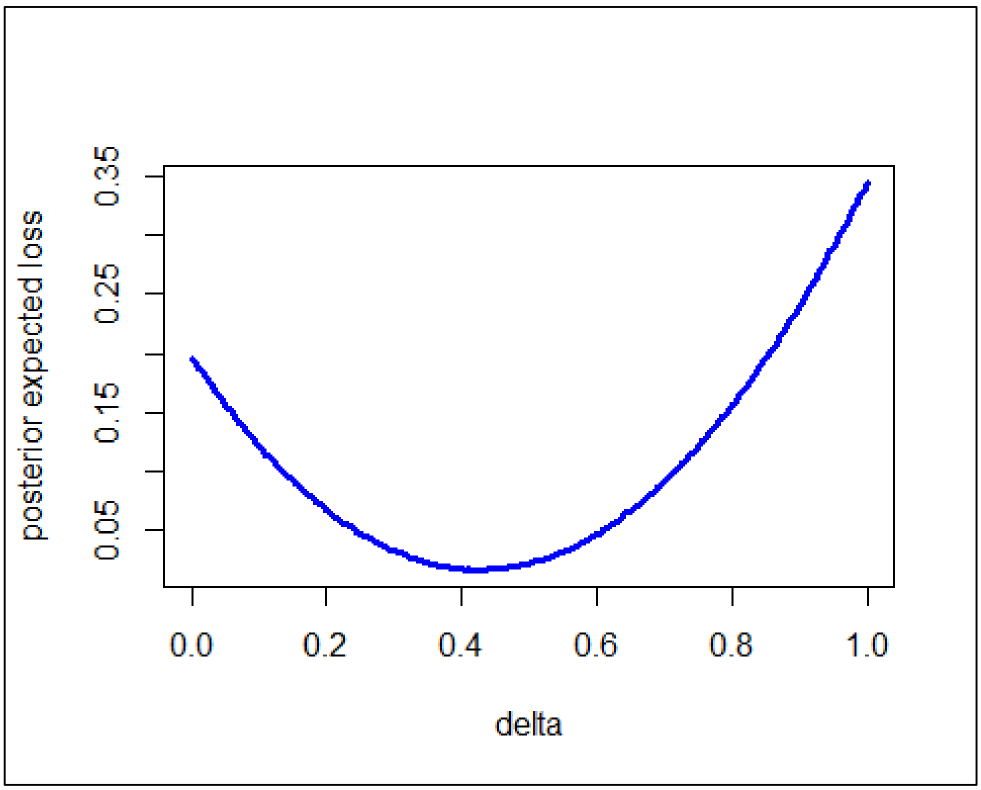
```
loss_function1 <- function(phi,a){
  return((phi-a)^2)
}
posterior_exploss_mc1 = function(delta, S = 10000){
  grid <- seq(0.005, 0.995, by=0.005)
  phi <- sample(grid, size=10^4, prob=mixture(grid), replace=T)
  loss <- apply(as.matrix(phi),1,loss_function1,delta)
  risk=mean(loss)}
delta <- seq(0, 1, by = 0.005)
post_exploss1 <- apply(as.matrix(delta),1,posterior_exploss_mc1)

par(mfrow=c(1,1))
plot(delta, post_exploss1, type = 'l', col='red',
     lwd = 3, ylab ='posterior expected loss')
```




```
delta[which.min(post_exploss1)]

## [1] 0.43
```

By estimating $\hat{\theta}^2$

```
loss_function2 <- function(theta,a){
  return((theta-a)^2)
}
posterior_exploss_mc2 = function(delta, S = 10000){
  theta <- c(rbeta(n*round(W1,4), 17, 11), rbeta(n*round(W2, 4), 16, 8))
  theta2 <- theta^2
  loss <- apply(as.matrix(theta2),1,loss_function2,delta)
  risk=mean(loss)}
delta <- seq(0, 1, by = 0.005)
post_exploss2 <- apply(as.matrix(delta),1,posterior_exploss_mc2)

par(mfrow=c(1,1))
plot(delta, post_exploss2, type = 'l', col='blue',
     lwd = 3, ylab ='posterior expected loss')
```




```
delta[which.min(post_exploss2)]

## [1] 0.425
```

2

(a)

1.

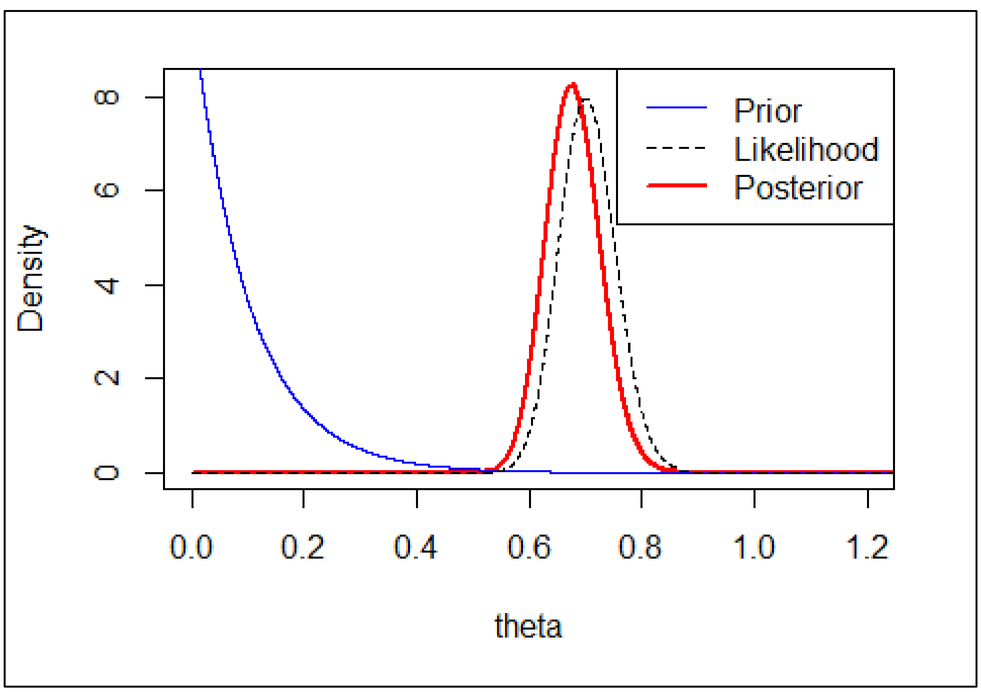
```
grid <- seq(0, 1.5, by=0.0001)
sample <- c(rep(0, 144), rep(1, 91), rep(2, 32), rep(3, 11), rep(4, 2))
a <- 1; b <- 10; n <- 280; sum.x <- sum(sample)
prior <- dgamma(grid, shape=a, rate=b); prior<-prior/sum(prior)
like <- dpois(sum.x, n*grid); like<-like/sum(like)
posterior <- prior*like; posterior <- posterior/sum(posterior)
```

The 95% Bayesian CI of theta is

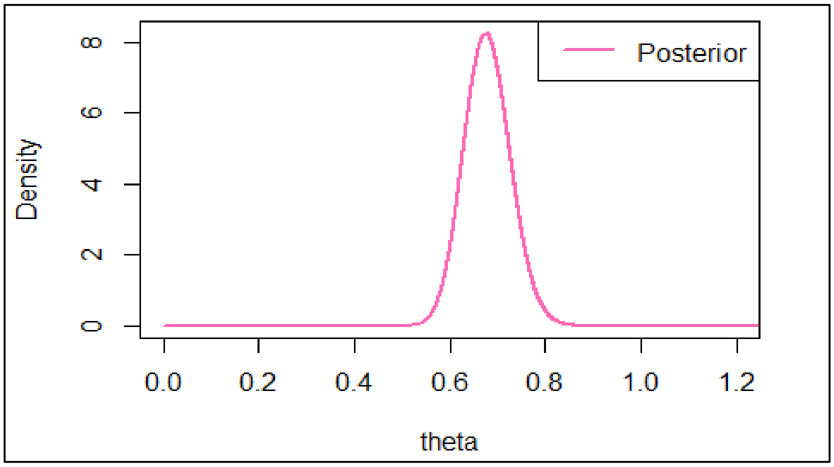
```
qgamma(c(0.025, 0.975), shape=a+sum.x, rate=b+n)

## [1] 0.5877574 0.7773930
```

(My option) The posterior distribution can be easily plotted by **dgamma()** function, and it is the same with grid approximation result. The plot of prior and posterior is as following

```
plot(grid, posterior*length(grid)/1.5, type='l', col='red', lwd=2, xlab="thet
a", ylab="Density", xlim=c(0, 1.2))
lines(grid, prior*length(grid)/1.5, type="l", col="blue")
lines(grid, like*length(grid)/1.5, type="l", col="black", lty=2)
legend("topright", c("Prior", "Likelihood", "Posterior"), lwd=c(1,1,2), col=c
("blue","black","red"), lty=c(1,2,1))
```

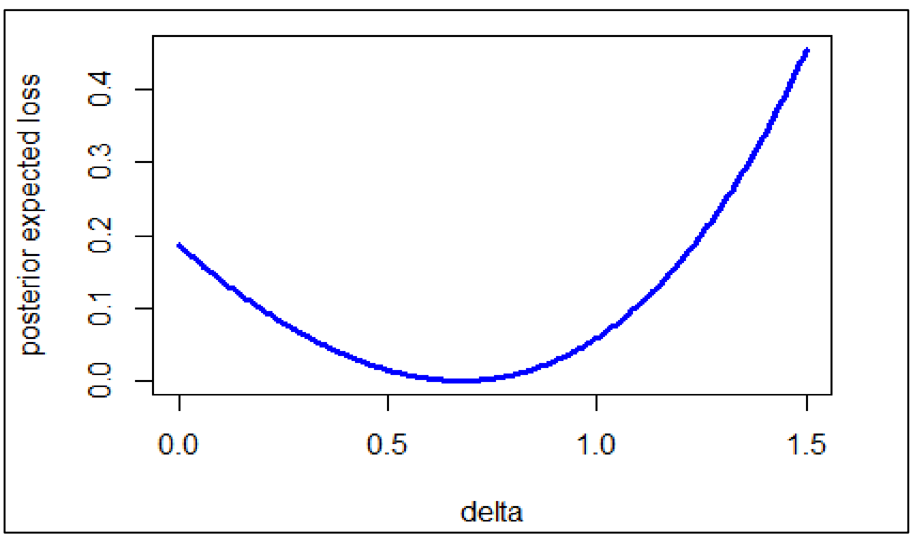


```
posterior2 <- dgamma(grid, shape=a+sum.x, rate=b+n);
plot(grid, posterior2, type='l', col='hotpink', lwd=2, xlab="theta", ylab="De
nsity", xlim=c(0, 1.2))
legend("topright", c("Posterior"), lwd=2, col=c("hotpink"), lty=1)
```



2.

```
loss_function <- function(theta,a){
  return(exp(a-theta)-(a-theta)-1)
}
posterior_exploss_mc = function(delta, S = 10000){
  theta <- rgamma(S, shape=a+sum.x, rate=b+n)
  loss <- apply(as.matrix(theta),1,loss_function,delta)
  risk=mean(loss)}
delta <- seq(0, 1.5, by = 0.005)
post_exploss <- apply(as.matrix(delta),1,posterior_exploss_mc)

plot(delta, post_exploss, type = 'l', col='blue',
     lwd = 3, ylab ='posterior expected loss')
```



```
delta[which.min(post_exploss)]

## [1] 0.68
```

(b)

1.

```
grid <- seq(0.0001, 10, by=0.0001)
prior <- 1/(grid^2)
like <- dpois(sum.x, n*grid); like<-like/sum(like)
posterior <- prior*like; posterior <- posterior/sum(posterior)
sum(posterior)

## [1] 1
```

The integral of posterior density converges to 1.

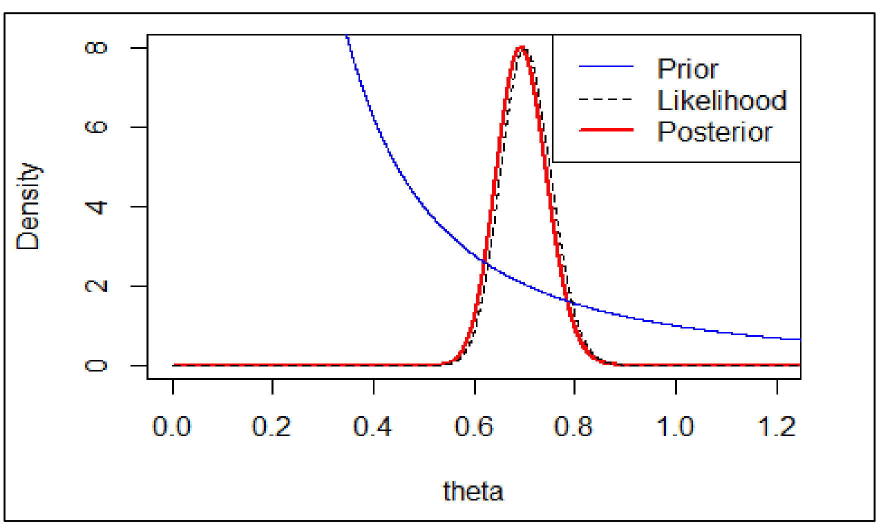
Since $p(\theta) \propto \theta^{-2}$ and $L(\theta|x) \propto \theta^{\Sigma x_i} e^{-n\theta}$, so

$$p(\theta|x) \propto \theta^{-2+196} e^{-280\theta} = \theta^{194} e^{-280\theta}$$

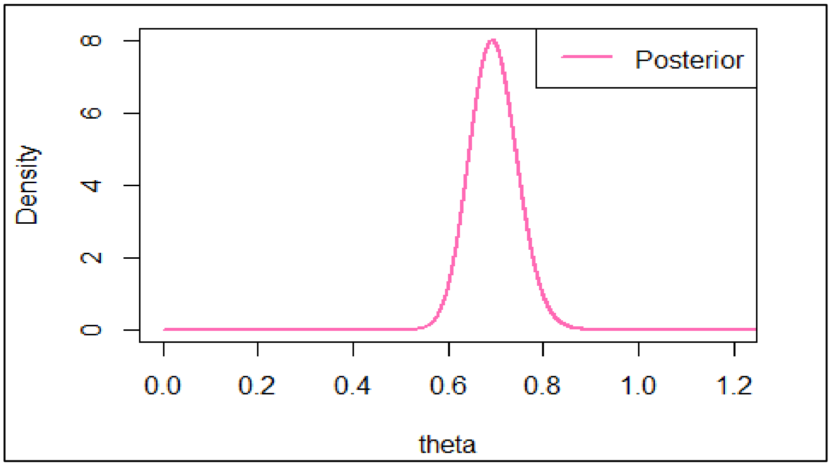
$$p(\theta|x) \sim \Gamma(195{,}280)$$

In this case as prior $p(\theta) \propto \frac{1}{\theta^2}$ given, the integral of prior distribution does not converge to 1, and it diverges to infinity. We call this **improper prior density function**. However, even though the improper prior density is given, the **posterior distribution is a proper posterior density function**.

```
plot(grid, posterior*length(grid)/10, type='l', col='red', lwd=2, xlab="theta
", ylab="Density", xlim=c(0,1.2))
lines(grid, prior, type="l", col="blue")
lines(grid, like*length(grid)/10, type="l", col="black", lty=2)
legend("topright", c("Prior", "Likelihood", "Posterior"), lwd=c(1,1,2), col=c
("blue","black","red"), lty=c(1,2,1))
```



```
posterior2 <- dgamma(grid, shape=195, rate=280)
plot(grid, posterior2, type='l', col='hotpink', lwd=2, xlab="theta", ylab="De
nsity", xlim=c(0,1.2))
legend("topright", c("Posterior"), lwd=2, col=c("hotpink"), lty=1)
```



The mean of Gamma distribution is $\frac{\alpha}{\beta}$, so the mean is

```
195/280
## [1] 0.6964286
```

2.

Let $A = \alpha + \sum x_i$ and $B = \beta + n$ whereas α, β from the prior distribution.

The posterior predictive can be calculated by the following

$$p(\tilde{x}|\mathrm{x}) = \int p(\tilde{x}|\theta)p(\theta|\mathrm{x})d\theta = \int \frac{e^{-\theta}\theta^{\tilde{x}}}{x!}\frac{B^A}{\Gamma(A)}\theta^{A-1}e^{-B\theta}d\theta = \int \frac{B^A}{x!\,\Gamma(A)}\theta^{A+\tilde{x}-1}e^{-(B+1)\theta}d\theta$$

$$p(\tilde{x}|\mathrm{x}) = \frac{B^A}{x!\,\Gamma(A)}\frac{\Gamma(A+\tilde{x})}{(B+1)^{A+\tilde{x}}}\int \frac{(B+1)^{A+\tilde{x}}}{\Gamma(A+\tilde{x})}\theta^{A+\tilde{x}-1}e^{-(B+1)\theta}d\theta = \frac{\Gamma(A+\tilde{x})}{\Gamma(A)x!}(\frac{B}{B+1})^A(\frac{1}{B+1})^{\tilde{x}}$$
$$= \binom{A+\tilde{x}-1}{\tilde{x}}(\frac{B}{B+1})^A(\frac{1}{B+1})^{\tilde{x}}$$

This is the form of negative binomial distribution when the parameters are A, $\frac{1}{B+1}$.

$$\tilde{x} \sim NegBin\left(195, \frac{1}{281}\right)$$

The probability that the new prediction $\tilde{x}$ is greater than 0, which is $P(\tilde{x} > 0|x_1, \cdots, x_{280})$ is

```
library(extraDistr)
1-dnbinom(0, 195, 1-1/281)
## [1] 0.5010193
```

3

(a)

Since $p(\theta|x) = p(\theta)L(\theta|x) \propto \theta^{24}(1-\theta)^{16}$

```
posterior.no.c <- function(x) {(x^24)*(1-x)^16}
const <- 1/integrate(posterior.no.c, 0.35, 0.6)$value
posterior.mean <- function(x) {const*x*(x^24)*(1-x)^16}
integrate(posterior.mean, 0.35, 0.6)$value

## [1] 0.5375141
```

(b)

```
prior1.no.c <- function(x) {(1+x)*exp(-x)}
k1 <- 1/integrate(prior1.no.c, 0, Inf)$value
prior2.no.c <- function(x) {1/(1+64*x^2)}
k2 <- 1/integrate(prior2.no.c, 0, Inf)$value
k1; k2

## [1] 0.5

## [1] 5.092958
```

So, the two priors are $p_1(\theta) = \frac{1}{2}(1+\theta)e^{-\theta}$, $\theta > 0$ and $p_2(\theta) = 5.093(1+(8\theta)^2)^{-1}$, $\theta > 0$

```
post1.no.c <- function(x) {(1+x)*(x^196)*exp(-281*x)}
c1 <- 1/integrate(post1.no.c, 0, 1)$value
post1.mean <- function(x) {c1*x*(1+x)*(x^196)*exp(-281*x)}
post1.ss <- function(x) {c1*(x^2)*(1+x)*(x^196)*exp(-281*x)}
integrate(post1.ss, 0, 1)$value - (integrate(post1.mean, 0, 1)$value)^2

## [1] 0.002503185
```

When $p_1(\theta)$, $\sqrt{Var(\theta|x_1,...,x_n)} = 0.002503$ When $p_2(\theta)$, $\sqrt{Var(\theta|x_1,...,x_n)} = 0.002486$