BST 5230 Bayesian Statistics Homework 2

Miao Cai[*]

2018-03-22

1. (12 Points: 2 each) Suppose that we have a sample $X_1, X_2, \dots, X_n$ from the Erlang2(λ) distribution, which has PDF

$$f(x|\lambda) = \lambda^2 x \exp(-\lambda x), \quad x > 0$$

This distribution is often used to model outcomes that are necessarily positive, such as the times required to perform a task. Suppose that the prior distribution is the GAMMA(α, β) parameterized so that the mean is $\frac{\alpha}{\beta}$.

(a) Find the formula for the posterior distribution.

(b) Is the GAMMA prior a conjugate prior in this case? Explain what a conjugate prior is, and why (or why not) the GAMMA is a conjugate prior.

(c) Suppose that we run an experiment to see how long it takes medical technicians to perform a task after they have received training. In the past, such a task has taken somewhere around 20 minutes, but there is a lot of variability. Consider each of the following prior distributions for the parameter:

  (i) GAMMA(1,100)

  (ii) GAMMA(1,2)

  (iii) GAMMA(10,10)

  Discuss whether each of these priors would be appropriate given the description given above.

(d) Suppose that we test four individuals on how long it takes to do the procedure after they have received training. The four times are: 9, 14, 7, and 10 minutes. Find the maximum likelihood estimate of λ.

(e) Using a GAMMA(2,20) prior for λ and the data from part (d) find and plot the posterior.

(f) Find the posterior mean and standard deviation.

Answers:

(a)

$$p(\lambda|X) = \frac{p(\lambda)f(X|\lambda)}{p(X)} = \frac{c_1\lambda^{\alpha-1}e^{-\beta\lambda} * \prod_{i=1}^n(\lambda^2 x_i e^{-\lambda x_i})}{c_2} = c_3\lambda^{2n+\alpha-1}e^{-\lambda(\beta+\sum_{i=1}^n x_i)}$$

Assuming that c_1, c_2, c_3 are unknown constants that make $p(\theta|X)$ is a posterior distribution. Therefore, $p(\theta|X) \sim \text{GAMMA}(\alpha + 2n, \beta + \sum_{i=1}^n x_i)$

(b) Yes, GAMMA is a conjugate prior for Erlang2(λ) distribution.

Conjugate prior is when the prior distribution and posterior distribution are in the same family

[*]PhD student, Department of Epidemiology and Biostatistics, College for Public Health and Social Justice, Saint Louis University
Email address miao.cai@slu.edu

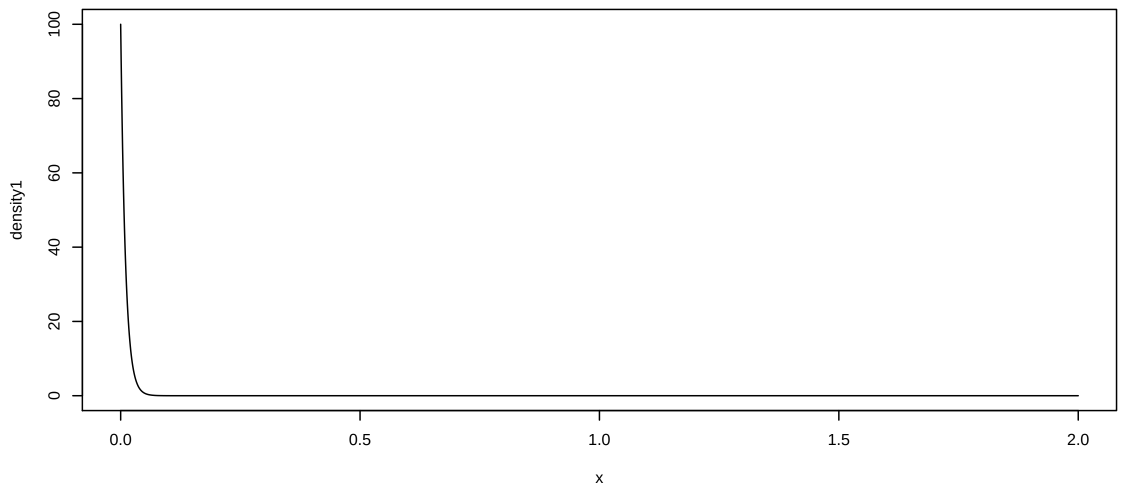
of distribution.

When we have GAMMA(α, β) prior distribution and Erlang2(λ) data, we have a GAMMA$(\alpha + n,\ \beta + n\lambda)$ posterior distribution. Therefore, we claim that GAMMA is a conjugate prior for Erlang2 data.

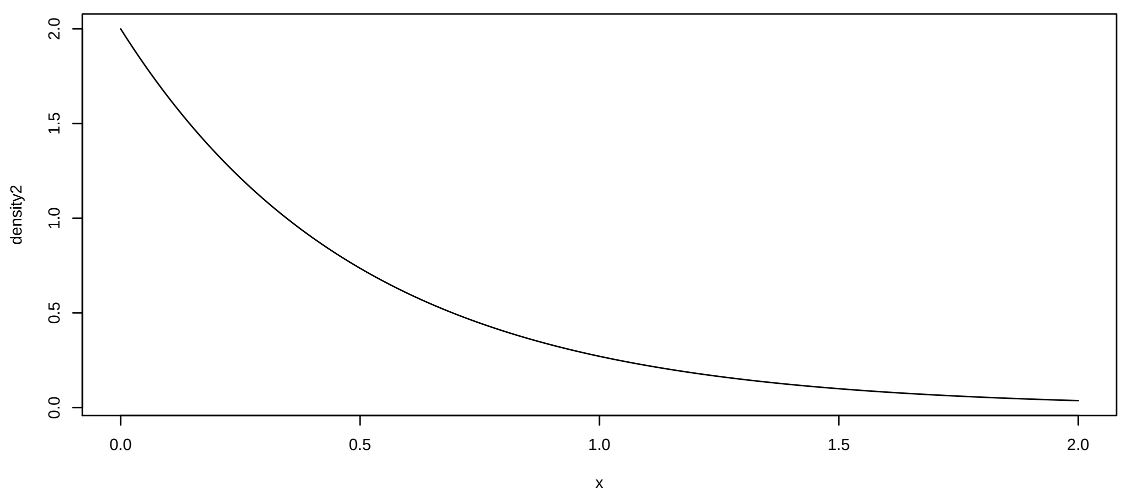
(c) The mean of Erlang2(λ) distribution is $\frac{2}{\lambda}$, and the variance of Erlang2(λ) distribution is $\frac{2}{\lambda^2}$. According to the information in the question, the mean of the task time is around 20, and which means $\frac{2}{\lambda} \approx 20\ \alpha/\beta = \lambda \approx 10$. In addition, "there is a lot of variability" implies the prior variance $\frac{\alpha}{\beta^2}$ should be fairly large. Let's plot these priors:

```
x = seq(0, 2, 0.001)
density1 = dgamma(x, 1, 100)
density2 = dgamma(x, 1, 2)
density3 = dgamma(x, 10, 10)

par(mfrow = c(3, 1))
plot(x, density1, type = "l", main = "Prior 1: GAMMA(1, 100)")
plot(x, density2, type = "l", main = "Prior 2: GAMMA(1, 2)")
plot(x, density3, type = "l", main = "Prior 3: GAMMA(10, 10)")
```

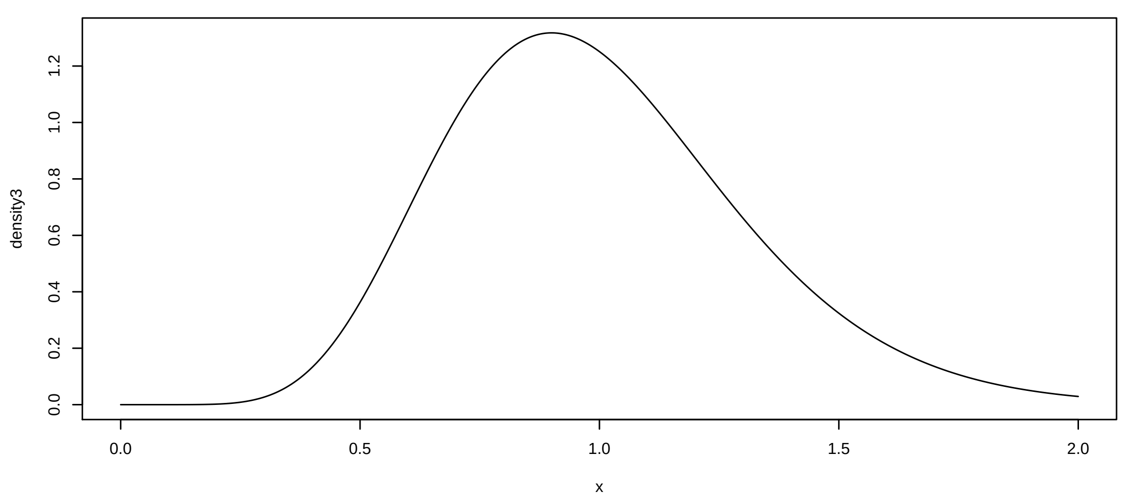
Prior 1: GAMMA(1, 100)



Prior 2: GAMMA(1, 2)



Prior 3: GAMMA(10, 10)



According to the plots above, I think that GAMMA(1, 2) is an appropriate prior given the information in the question.

(d) The likelihood function in this case is

$$
\begin{aligned}
L(\lambda) &= \prod_{i=1}^{4} \Big(\lambda^2 x_i \exp(-\lambda x_i)\Big) \\
&= \lambda^8 \exp(-\lambda \sum_{i=1}^{4} x_i) \prod_{i=1}^{4} x_i
\end{aligned}
$$

Then we can get the log likelihood function by taking the derivative of natural log of $L(\lambda)$:

$$
\begin{aligned}
l'(\lambda) &= \frac{\partial ln(L(\lambda))}{\partial \lambda} \\
&= \frac{\partial 8 ln(\lambda) + \sum_{i=1}^{4} ln x_i - \lambda \sum_{i=1}^{4} x_i}{\partial \lambda} \\
&= \frac{\partial 8 ln(\lambda) + \sum_{i=1}^{4} ln x_i - 40\lambda}{\partial \lambda} \\
&= \frac{8}{\lambda} - 40
\end{aligned}
$$

Let $l'(\lambda) = 0$ to maximize the likelihood function, we have $\lambda = \frac{1}{5}$.

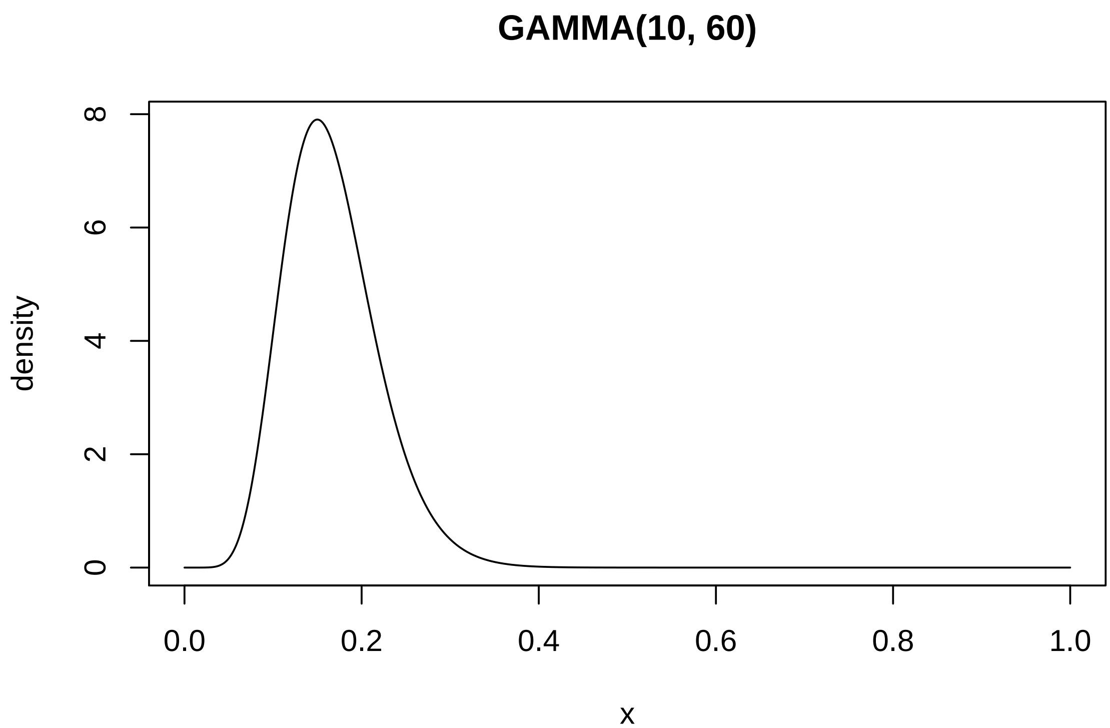
(e) According to the Bayes Theorem, we have:

$$
\begin{aligned}
p(\lambda|X) &= \frac{p(\lambda) f(X|\lambda)}{p(X)} \\
&= \frac{c_1 \lambda^{\alpha-1} e^{-\beta\lambda} * (\lambda^2)^4 x_1 x_2 x_3 x_4 e^{-\lambda(x_1+x_2+x_3+x_4)}}{c_2} \\
&= c_3 \lambda^{\alpha+7} e^{-\lambda(\beta + \sum_{i=1}^{4} x_i)}, \quad c_1, c_2, c_3 \text{ are unknown constants}
\end{aligned}
$$

Let's plug in $\alpha = 2, \beta = 20, \sum_{i=1}^{4} x_i = 40$

$$
\begin{aligned}
p(\lambda|X) &= c_3 \lambda^9 e^{-\lambda(20+40)} \\
&= c_3 \lambda^9 e^{-60\lambda} \\
&\sim \text{GAMMA}(10, 60)
\end{aligned}
$$

```r
## Here is the code for generating GAMMA(10, 60) distribution:
x = seq(0, 1, 0.001)
d = dgamma(x, 10, 60)
plot(x, d, type = "l", main = "GAMMA(10, 60)", ylab = "density")
```


GAMMA(10, 60)
density
0
2
4
6
8
0.0
0.2
0.4
0.6
0.8
1.0
x

2. (16 Points: 2 each) Suppose that $X|\theta \sim \text{BIN}(100, \theta)$. The prior distribution for θ is BETA(1, 3).

(a) Plot the prior.

(b) Compute the posterior distribution if x = 29 successes are observed in 100 trials.

(c) Is the BETA distribution a conjugate prior in this situation? Explain.

(d) Plot the posterior distribution.

(e) Apply the Metropolis-Hastings algorithm with proposal distribution $\mathcal{N}(\theta^{(t)}, \sigma^2 = 0.1)$. Here $\theta^{(t)}$ is the current value in the Markov chain. Run 10,000 simulations after an appropriate burn-in. Plot a histogram of the posterior distribution.

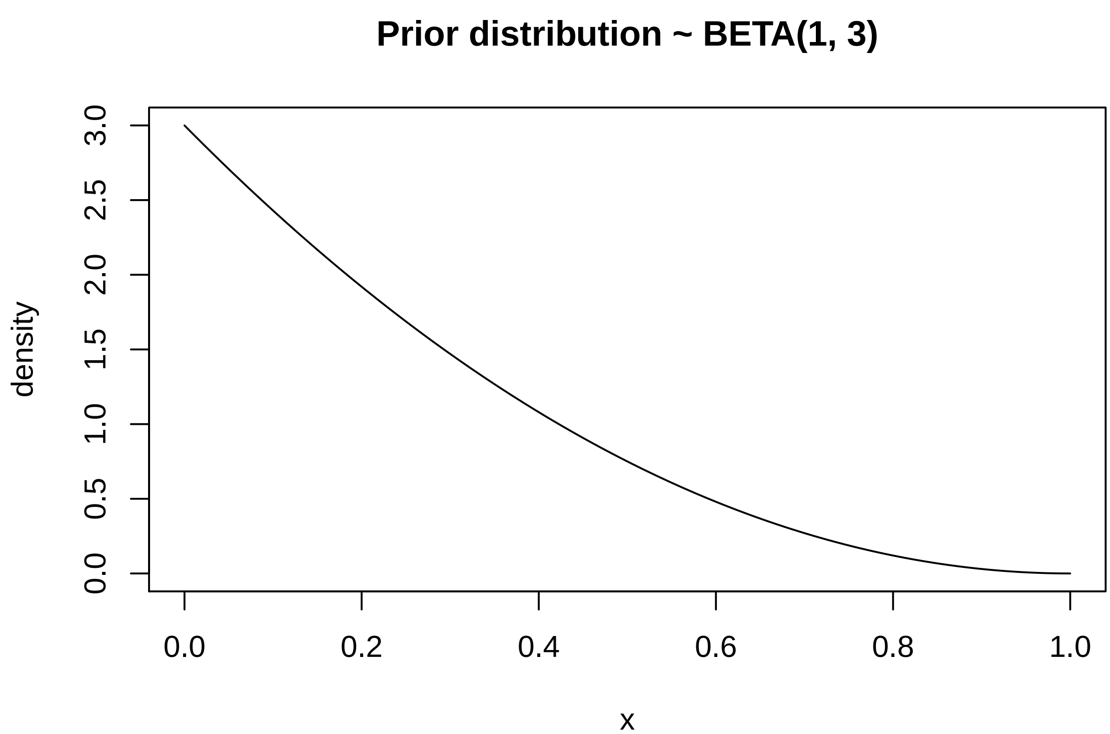
(f) Now apply the Metropolis-Hastings algorithm with proposal distribution BETA(1,3). Run 10,000 simulations after an appropriate burn-in. Plot a histogram of the posterior distribution.

(g) Are either of (e) and (f) the simpler Metropolis algorithm, rather than the full Metropolis-Hastings algorithm? Explain.

(h) Use the **plotPost** function to plot the posterior distributions and find HDI intervals from parts (e) and (f).

Answers:

(a)
```
x = seq(0, 1, 0.001)
d = dbeta(x, 1, 3)
plot(x, d, type = "l", main = "Prior distribution ~ BETA(1, 3)", ylab = "density")
```



(b)

$$
\begin{aligned}
p(\theta|X) &= \frac{p(\theta)f(X|\theta)}{p(X)} \\
&= \frac{c_1\theta^{\alpha-1}(1-\theta)^{\beta-1} * c_2\theta^{29}(1-\theta)^{100-29}}{c_3} \\
&= c_4\theta^{\alpha+28}(1-\theta)^{70+\beta} \\
&\sim \text{BETA}(\alpha+29, \beta+71)
\end{aligned}
$$

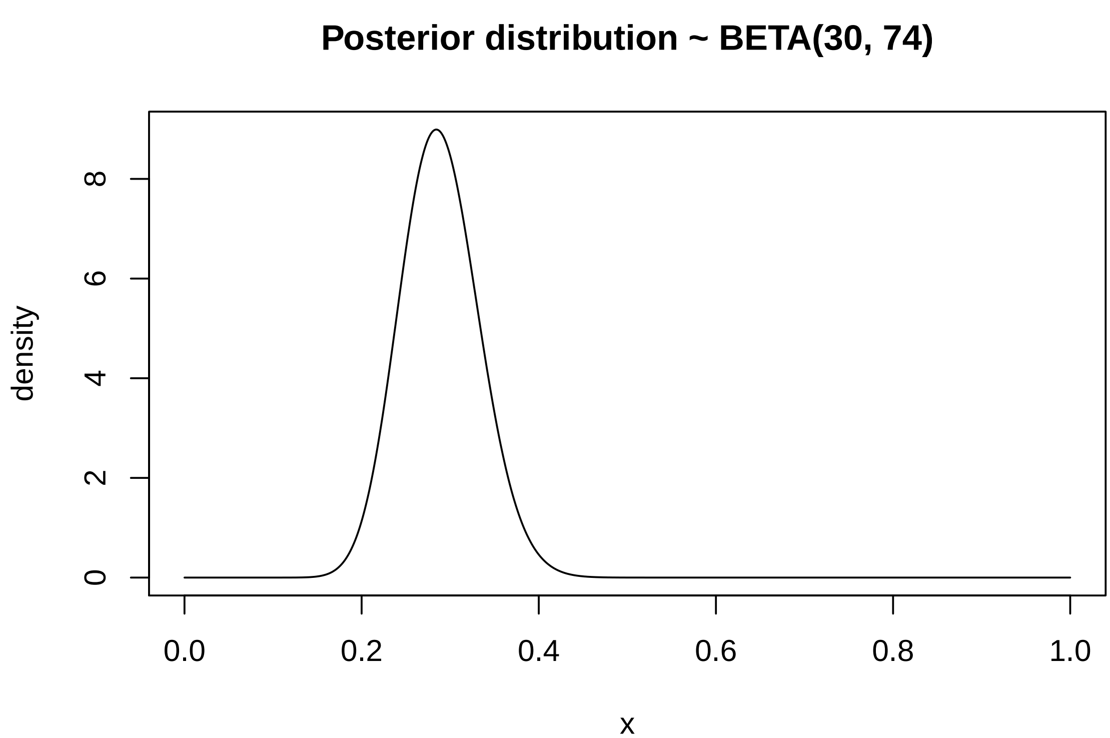
When $\alpha = 1, \beta = 3$,

$$\sim \text{BETA}(30, 74)$$

(c) Yes, it is. In this case, the prior is a BETA distribution and we have a binomial data, then we get a BETA posterior. Therefore, we can claim that BETA distribution is a conjugate prior for binomial data.

(d)

```r
x = seq(0, 1, 0.001)
d = dbeta(x, 30, 74)
plot(x, d, type = "l", main = "Posterior distribution ~ BETA(30, 74)", ylab = "density")
```

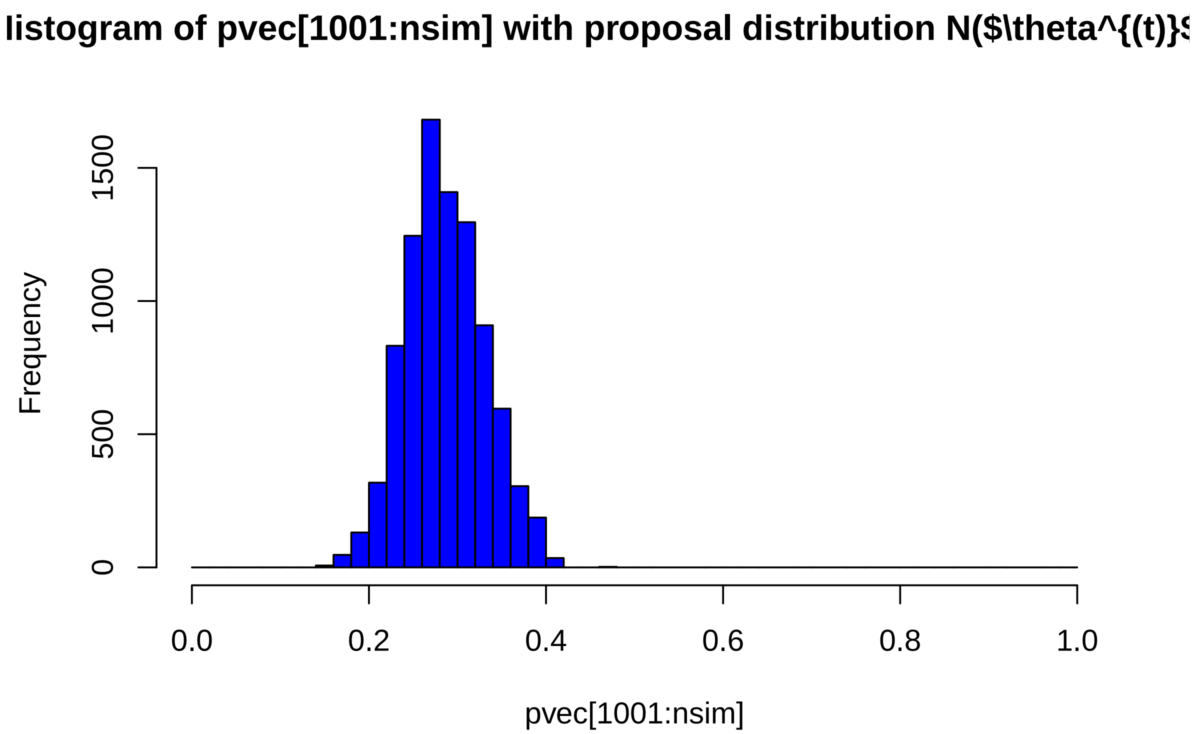


(e)

```r
# Essential part of BETA distribution
pBeta = function(p,a,b)
```

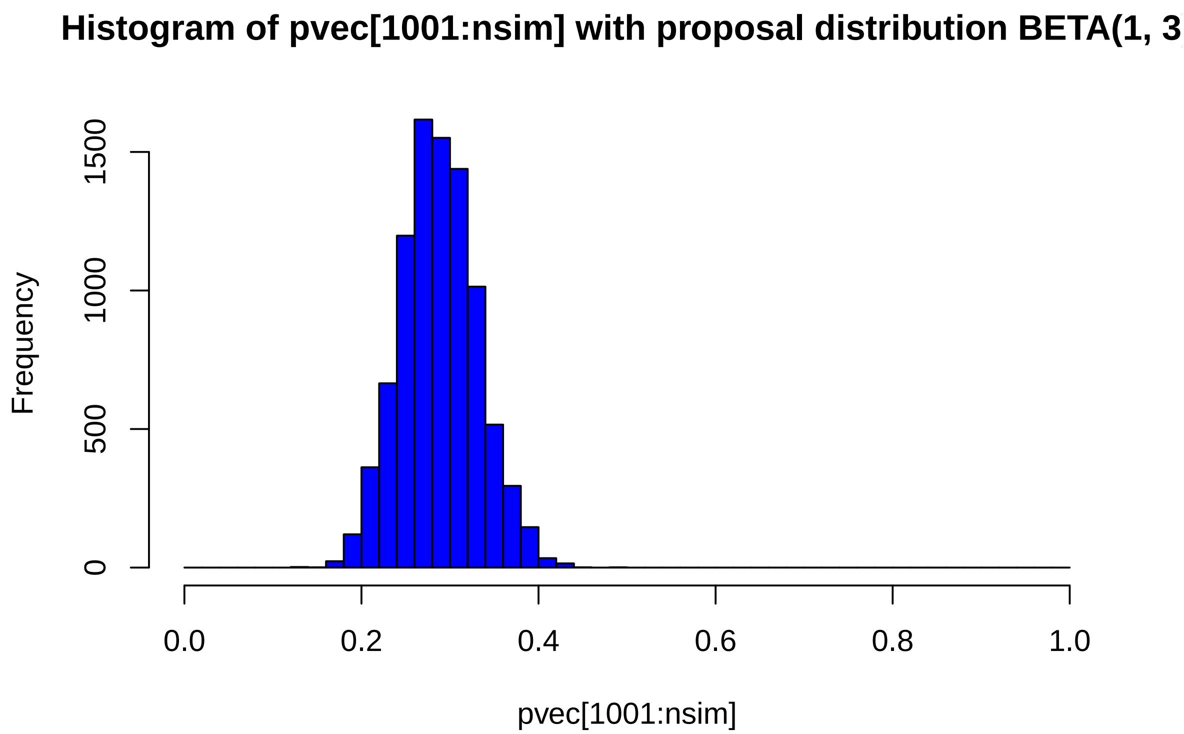
```r
{
  z = 0
  if (p >= 0 && p <= 1) {  z = p^(a-1) * (1-p)^(b-1) }
  return(z)
}

# 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(0.1))
}

n = 100
x = 29
a = 1
b = 3

set.seed(123)
nsim = 10000
pvec = rep(0,nsim)
pvec[1] = 0.50
acc = 0
for (i in 2:nsim)
{
  pp = pProp( pvec[i-1] )
  pc = pvec[i-1]
  num = pBeta(pp,a,b)*pLik(x,n,pp)
  den = pBeta(pc,a,b)*pLik(x,n,pc)
  alpha = num/den
  probAcc = min(1,alpha)
  u = runif(1)
  if ( u < probAcc ) { pvec[i] = pp; acc = acc + 1  } else pvec[i] = pc
}

result_e = pvec[1001:nsim]
```

```r
hist( pvec[1001:nsim] , xlim=c(0,1) , breaks=seq(0,1,0.02) , col="blue" ,
      main = paste("Histogram of" , "pvec[1001:nsim]",
                   "with proposal distribution N($\\theta^{(t)}$, 0.1)"))
```



(f)
```r
# 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)
{
  rbeta(1, 1, 3)
}
```

```r
n = 100
x = 29
a = 1
b = 3

set.seed(123)
nsim = 10000
pvec = rep(0,nsim)
pvec[1] = 0.50
acc = 0
for (i in 2:nsim)
{
  pp = pProp( pvec[i-1] )
  pc = pvec[i-1]
  num = pBeta(pp,a,b)*pLik(x,n,pp)*(1-pc)^2
  den = pBeta(pc,a,b)*pLik(x,n,pc)*(1-pp)^2
  alpha = num/den
  probAcc = min(1,alpha)
  u = runif(1)
  if ( u < probAcc ) { pvec[i] = pp; acc = acc + 1  } else pvec[i] = pc
}

result_f = pvec[1001:nsim]

hist( pvec[1001:nsim] , xlim=c(0,1) , breaks=seq(0,1,0.02) , col="blue" ,
      main = paste("Histogram of" , "pvec[1001:nsim]", "with proposal distribution BETA(1, 3)'
```



(g) Yes, question(e) is simpler Metropolis algorithm but question (f) is not. For question (e), the proposal distribution $\mathcal{N}(\theta^{(t)}, \sigma^2 = 0.1)$ is a symmetric distribution. For question (f), BETA(1, 3) has a pdf:

$$
\begin{aligned}
f(\theta|x) &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{\Gamma(1)\Gamma(3)}{\Gamma(1+3)} x^{1-1}(1-x)^{3-1} \\
&= \frac{1}{3}(1-x)^2
\end{aligned}
$$

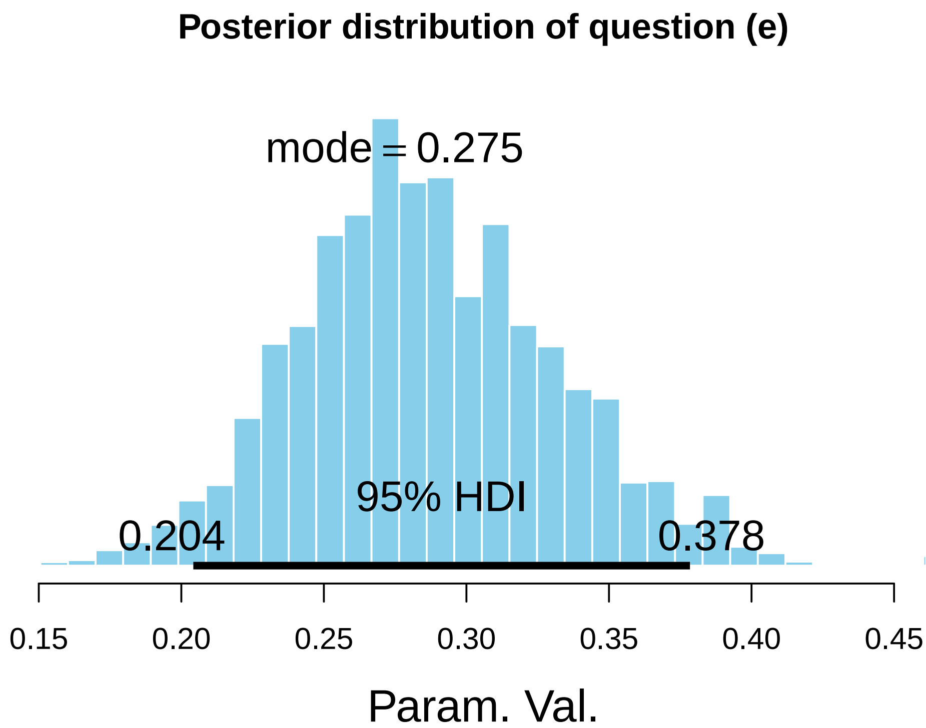
This is not symmetric in terms of "p current" and "p proposal". Therefore, only question (e) is the simpler Metropolis algorithm.

(h) For the posterior distribution of part (e):

```
source("DBDA2E-utilities.R")
```

```
##
## *******************************************************************
## Kruschke, J. K. (2015). Doing Bayesian Data Analysis, Second Edition:
## A Tutorial with R, JAGS, and Stan. Academic Press / Elsevier.
## *******************************************************************
```

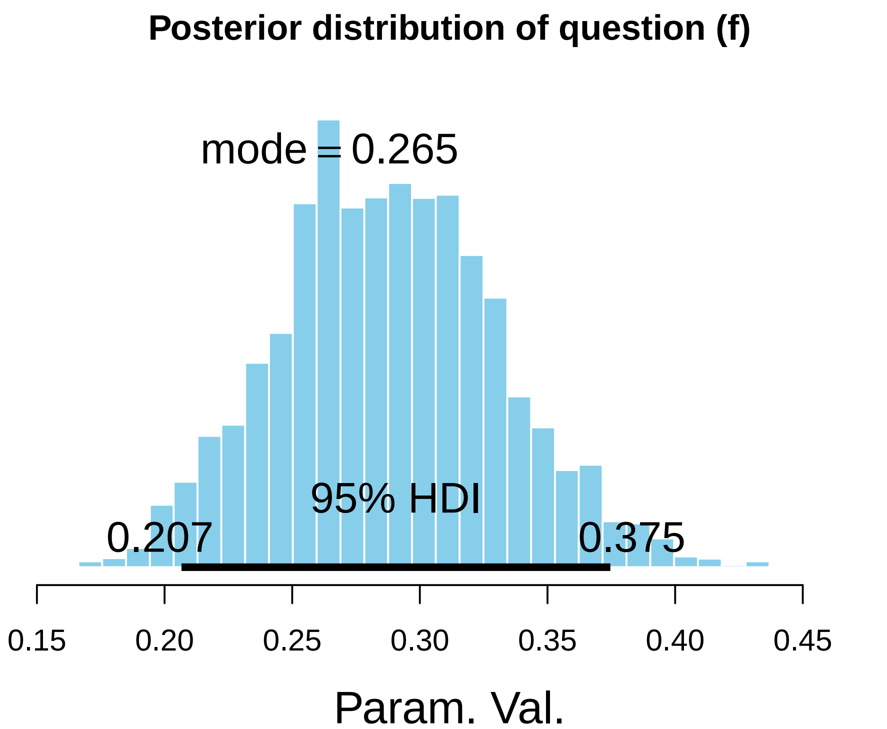
```r
plotPost(result_e, main = "Posterior distribution of question (e)")
```




```
##                   ESS      mean    median      mode hdiMass    hdiLow
## Param. Val. 1095.996 0.2862097 0.2826411 0.2747948    0.95 0.2042043
##               hdiHigh compVal pGtCompVal ROPElow ROPEhigh pLtROPE pInROPE
## Param. Val. 0.3784067      NA         NA      NA       NA      NA      NA
##             pGtROPE
## Param. Val.      NA
```

For the posterior distribution of part (f):

```r
plotPost(result_f, main = "Posterior distribution of question (f)")
```




```
##                   ESS      mean    median     mode hdiMass    hdiLow
## Param. Val. 1267.525 0.2878143 0.2869364 0.264609    0.95 0.2066278
##               hdiHigh compVal pGtCompVal ROPElow ROPEhigh pLtROPE pInROPE
## Param. Val. 0.3746417      NA         NA      NA       NA      NA      NA
##             pGtROPE
## Param. Val.      NA
```

3. (8 Points: 2 each) Suppose that $X|\theta \sim \text{BIN}(5, \theta)$. The prior distribution for θ is BETA(1, 3). We observe x = 2 successes on the n = 5 trials.

(a) Suppose we apply the Metropolis-Hastings algorithm from part (e) of Problem 2. Suppose that we begin with $\theta^{(1)} = 0.5$. If the first proposal is $\theta^{(\text{prop})} = 0.42$, compute the probability of acceptance.

(b) Same as part (a) above, but assume the first proposal is $\theta^{(\text{prop})} = 0.52$.

(c) Now suppose we apply the Metropolis-Hastings algorithm from part (f) of Problem 2. Suppose that we begin with $\theta^{(1)} = 0.5$. If the first proposal is $\theta^{(\text{prop})} = 0.42$, compute the probability of acceptance.

(d) Same as part (c) above, but assume the first proposal is $\theta^{(\text{prop})} = 0.52$.

Answers:

(a)
```
# 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(0.1))
}

n = 5
x = 2
a = 1
b = 3

pp = 0.42
pc = 0.5

num = pBeta(pp,a,b)*pLik(x,n,pp)
den = pBeta(pc,a,b)*pLik(x,n,pc)
```

```r
(alpha = num/den)
```

```
## [1] 1.482001
```

(b)
```r
pp = 0.52
pc = 0.5

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

(alpha = num/den)
```

```
## [1] 0.8819071
```

(c)
```r
# Proposal distribution
pProp = function(p)
{
  rbeta(1, 1, 3)
}

pp = 0.42
pc = 0.5

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

(alpha = num/den)
```

```
## [1] 1.101368
```

(d)
```r
pp = 0.52
pc = 0.5

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

(alpha = num/den)
```

```
## [1] 0.9569305
```