Contents

Introduction to Bayesian Estimation

- Bayesian estimation we assume parameters are not fixed, they have a probability distribution.

- Distribution maybe discrete or continuous.

- Before estimation we assume a prior distribution for the parameter.

- This is our knowledge about parameter before observing sample data.

- We may obtain this from literature review, recent researches etc.

- After observing sample data, we use this data to update our prior distribution.

- This updated Distribution is called Posterior distribution.

- Using this Posterior distribution we make probabilistic inferences about Population parameter.

We know the bayes therem is

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Here,

- θ is parameter, y is observed data from sample.

- P(θ) is Prior distribution of parameter θ.

- P(θ | y) is Posterior Distribution.

- P(y | θ) is Joint density of sample = Likelihood function.

- The denominator is a constant after observing data y and for a given value of θ. We don't need to worry about it, it doesn't matter much in our calculations.

So we can say

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

So if we know Likelihood and Prior, we can find posterior distribution of parameter θ.

θ could be Population proportion, Population mean etc.

Motivation

In real life parameters are never random. They are indeed fixed for a population at a given time.

But also in real life we may have some prior beliefs about our parameter.

Lets suppose we want to estimate proportion of women in the population. Definitely it's gonna be between 0 to 1.

But our common sense says its somewhere near 0.50. So how can we use this prior belief in our estimate ?

This use of prior belief is what bayesian estimation is all about.

Now, lets assume Actual proportion of women is somewhere between 0.48 to 0.52. And in bayesian we say parameters follow a probability distribution.

So lets suppose our proportion follows a Beta(a,b) distribution. Beta is a good choice here because its range is between 0 to 1. And also we can actually choose values of a & b to generate our prior distribution.

Our prior belief was proportion of women is around 0.48 to 0.52. We may be wrong, but we are pretty sure we won't be far off.

Now if we pick a = 300, b = 300, then the corresponding beta distribution has a 95% CI of [0.48 0.52]. If a=b then we get a symmetric distribution.

Now lets compare what happens to our estimates when we do a simple srs and bayesian estimation.

Lets suppose we take 50 samples from our population. We will apply both simple SRS method and Bayesian method to estimate Population Proportion.

Lets suppose in SRS

$$\hat{p} = \frac{y}{n}$$

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Where y = total number of women out of 50, n = 50.

And sampling distribution of $\hat{p}$ is Normal($\hat{p}$, SE($\hat{p}$)).

And In bayesian
We have the posterior distribution is Beta(a+y, n+b-y). How we got this Posterior is discussed later.

Lets see 2 graphs

Posterior = Beta (a+y, n+b−y)

Sampling Distribution of Sample Proportion from SRS
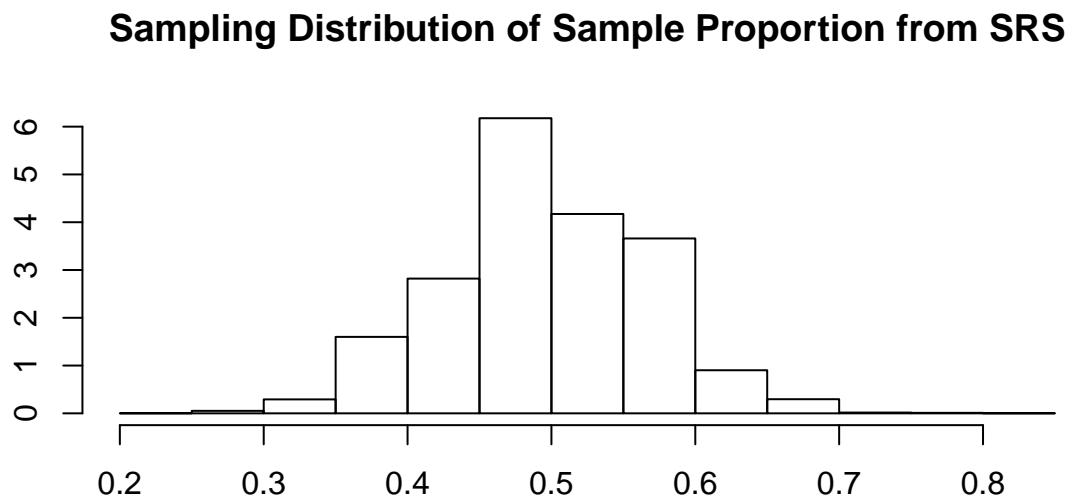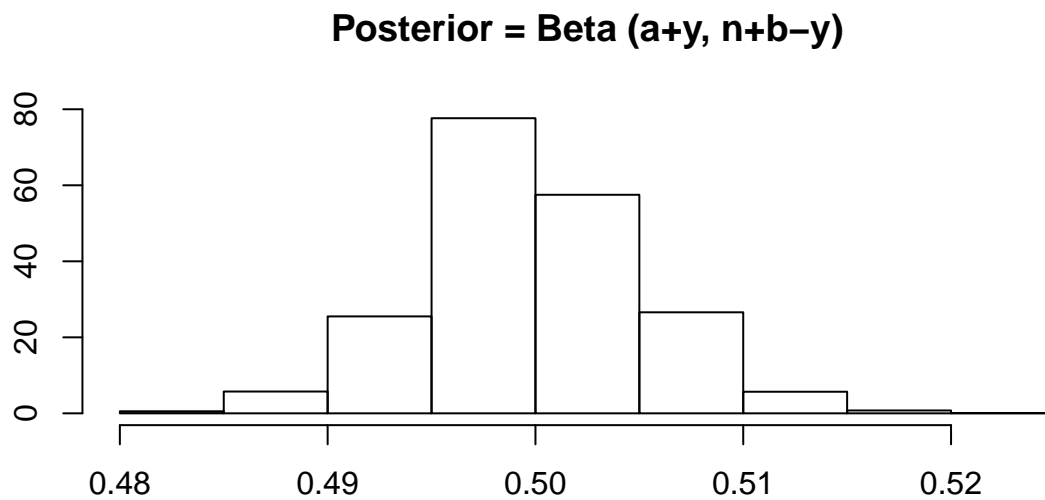
We see the posterior has very small range, so CI will be Narrow. Where from SRS sampling distribution has larger range, so CI from SRS will be Wider. As we used our prior belief to get the posterior, we got much better estimate of $\hat{p}$.

So bayesian is not about parameters being random, it's more about using our belief to make better estimates.

Now lets see how we can do all these in R.

Use set.seed(21) to match results.

Some pages are intentionally left blank.

Binomial Bayesian with Continuous Uniform Prior

This is the problem from day 06, 27th March'18.

We want an estimate of Population Proportion **P** .

Prior

$Prior = uniform(0, 1)$, so all prior values of P has equal probability.

$$f(P)_{prior} = 1, 0 < P < 1$$

Likelihood

$Likelihood = P^{29}(1 - P)^{121}$ is given in the question.

Posterior

$$Posterior \propto likelihood * Prior$$

$$=> Posterior \propto P^{29}(1 - P)^{121} * 1$$

$$=> Posterior \propto P^{29}(1 - P)^{121}$$

So the posterior is nothing but a *KerneL* of beta distribution with a = 30 , b = 122.

$$f(x)_{beta} = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1 - x)^{b-1}, 0 < x < 1$$

$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$ is the constant part called **Normalizer**, and the rest $x^{a-1}(1 - x)^{b-1}$ is called **kernel**.

As we have got our posterior distribution, now we can make any **probabilistic conclusion** about out parameter θ using this distribution. We can find

- Mean, Variance of our Parameter θ

- Credible Interval

- Hypothesis Testing

Graph of posterior from the given likelihood function

```r
lfun <- function(p){  p^29 * (1-p)^121 }
x <- seq(from=0, to=1, length.out = 5000)
y <- lfun(x)
plot(x, y,
     main = "Plot of likelihood function",
     xlab = "P")
```

Plot of likelihood function
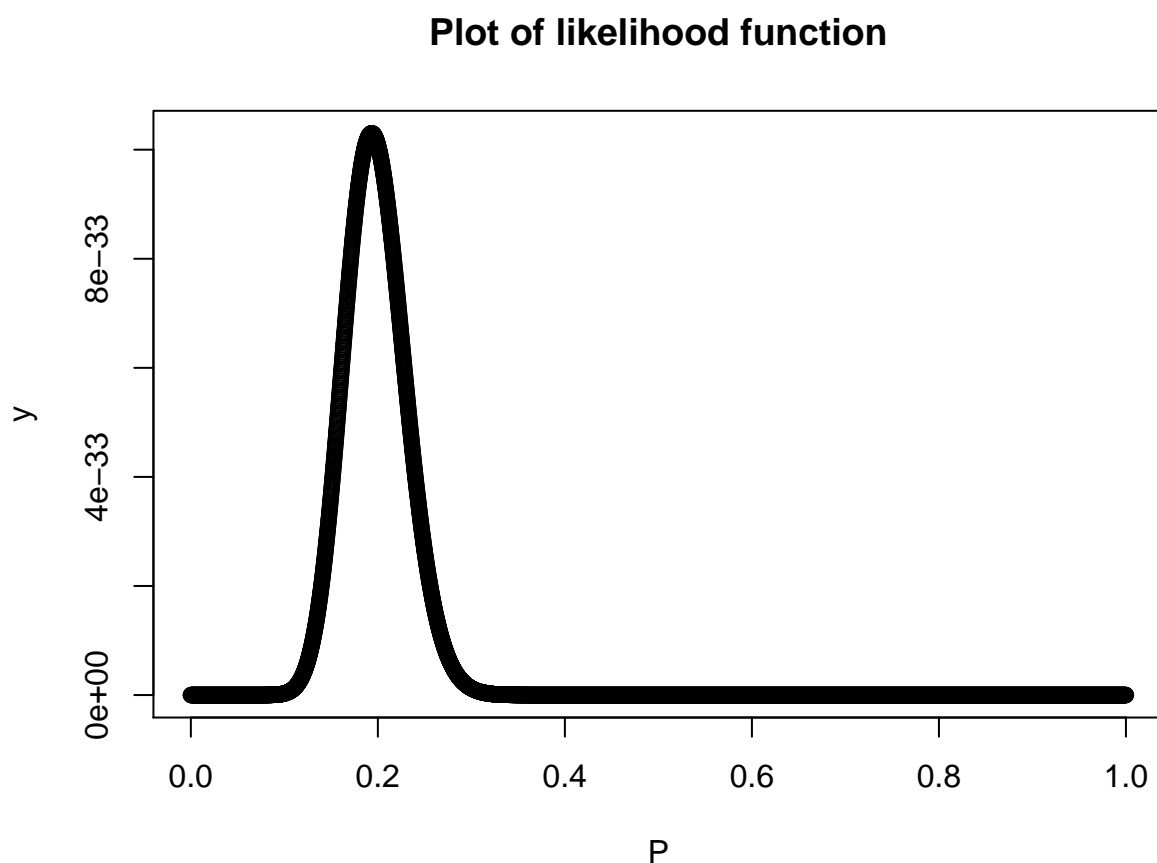
Kernel determines shape of curve, and normalizer is multiplied with kernel to make area under curve 1.

So our goal in bayesian is to match the posterior with a kernel of a known pdf.

Graph of posterior from Beta pdf

```r
fbeta <- function(x){  dbeta(x, shape1 = 30, shape2 = 122) }
x <- seq(from=0, to=1, length.out = 5000)
y <- fbeta(x)
plot(x, y,
     main = "beta(30,122)",
     xlab="P",
     ylab="Density")
```


beta(30,122)
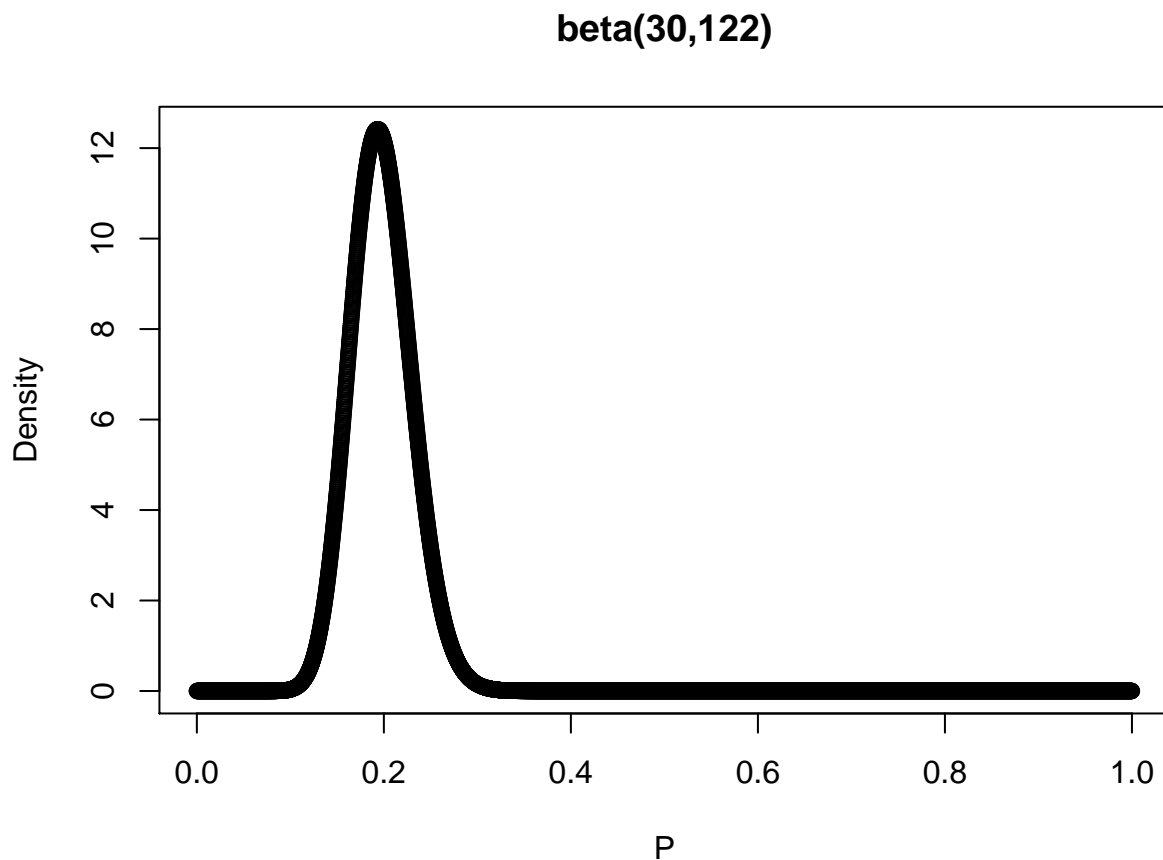
Both graphs look similar in *Shape*, but the difference is in the scaling of y-axis. As we didn't include the *constant* part in the first graph, the total area under curve is not equal 1.

But in the 2nd graph the area is 1 under curve.

However both graphs maximize at same P value.

Mean

If P ~ Beta(a,b) then

$$E[P] = \frac{a}{a+b}$$

```
a<-30
b<-122
a/(a+b) # Mean
```

```
## [1] 0.1973684
```

Variance

If P ~ Beta(a,b) then

$$V[P] = \frac{ab}{(a+b)^2(a+b+1)}$$

```
a*b/( (a+b)^2 * (a+b+1) ) # Variance
```

```
## [1] 0.001035386
```

Maximum

```
x[which.max(y)]
```

```
## [1] 0.1932386
```

Mean and maximum is not same here, because the beta distribution is not symmetric like Normal distribution.

Credible interval

We may find any 2 points amid which the area under curve is 0.95. But usual practise is to use the middle interval like done here below. Here 0.975 - 0.025 = 0.95. And both tails have area 0.025.

```
qbeta(shape1 = 30, shape2 = 122, p = c(0.025, 0.975))
```

```
## [1] 0.1382419 0.2640322
```

Hypothesis

$$H_o : P <= 0.25$$

$$H_a : P > 0.25$$

```
pbeta(q = 0.25 , shape1 = 30, shape2 = 122, lower.tail = FALSE)
```

```
## [1] 0.05751862
```

```
# Fail to reject NULL as pvalue < 0.95
```

Posterior

- Cut down the x-axis to make the graph look big, no big deal !
- Area Shaded in **red** equals 0.057
- **IN Bayesian, if pvalue > 0.95 then reject Null**

We see probability that P is greater than 0.25 is only 5.7 %. So we fail to reject NULL.

We will reject NULL if this probability was more than 0.95.

12

Another Hypothesis

$$H_o : P <= 0.14$$
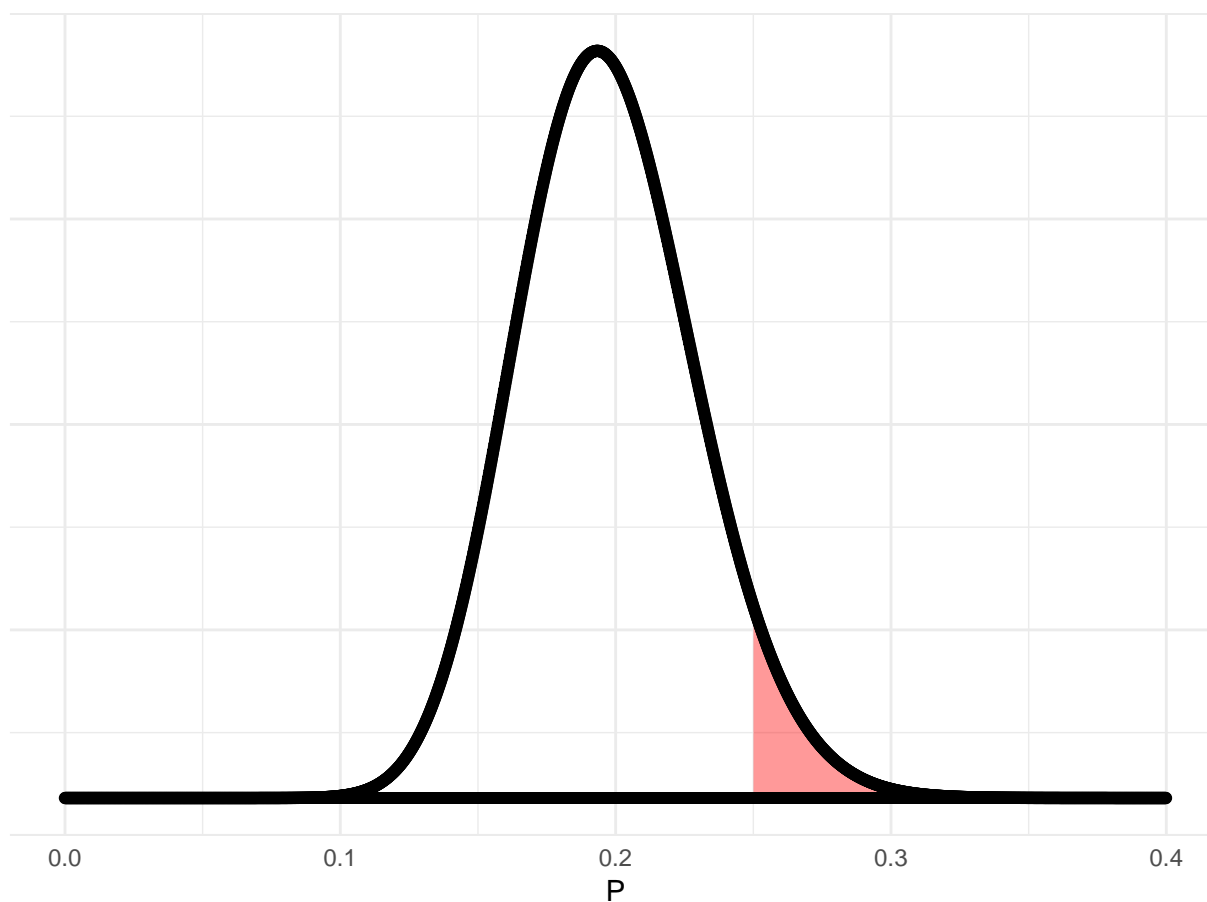
$$H_a : P > 0.14$$

```
pbeta(q = 0.14 , shape1 = 30, shape2 = 122, lower.tail = FALSE)
```
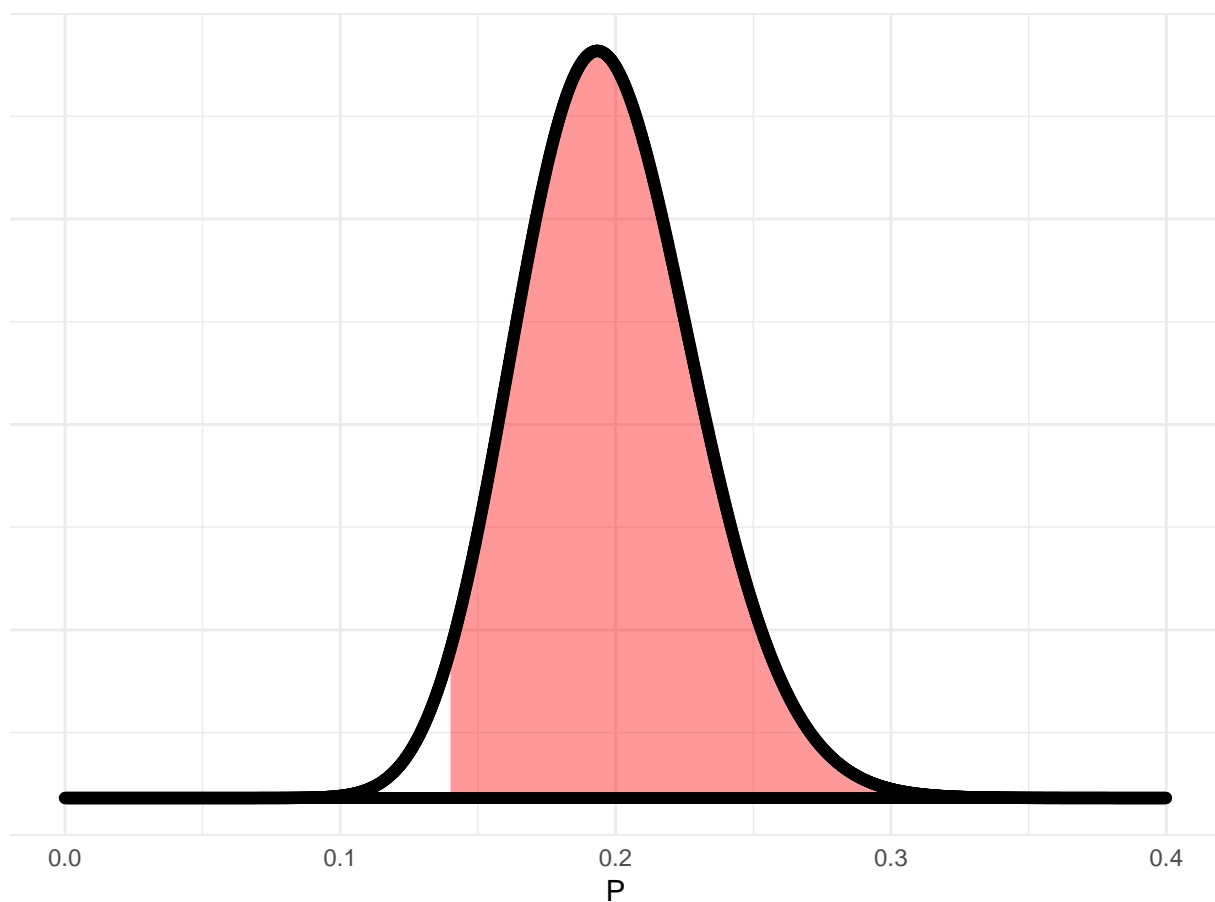
```
## [1] 0.9709004
```

```
# Reject NULL as pvalue > 0.95
```

Posterior

- **IN Bayesian, if pvalue > 0.95 then reject Null**
- We tested for P > 0.14
- From the graph we see p is greater than 0.14, its around 0.20
- Area shaded in Red is 0.97

We see probability that P > 0.14 is 97 % , so we can say P is infact greater than 0.14.

Binomial Bayesian with Continuous Beta Prior

Interest

- We are interested in finding a Posterior Disribution of our parameter θ.
- First we assume a prior distribution.
- Then we collect data and calculate likelihood
- Then we find Posterior distribution of θ

Prior

Lets assume our prior distribution is Beta(5,8) for parameter θ = Population Proportion.

$$P(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

$$P(\theta) = \frac{\Gamma(5+8)}{\Gamma(5)\Gamma(8)}\theta^{5-1}(1-\theta)^{8-1}$$

```
a=5
b=8
prior <- function(theta){
  (gamma(a+b) / (gamma(a)*gamma(b))) * theta^(a-1) * (1-theta)^(b-1)
}
```

Likelihood

Lets Assume We have drawn our data. We got 12 success out of 20 trials.

That is we run 20 bernoulli trials here.

Equivalently we run a Binomial(n =20, p=θ) trial once.

Let Y = Number of success in 20 trials.

So $Y \sim Binomial(n,\ \theta)$. The likelihood function is

$$P(y|\theta) =_n C_y\theta^y(1-\theta)^{n-y}$$
$$P(y|\theta) =_{20} C_{12}\theta^{12}(1-\theta)^{20-12}$$

```
n=20 ; y=12
likelihood <- function(theta){
  choose(n ,y) *  theta^(y)*(1-theta)^(n-y) }
```

Posterior

We have our Prior, and we have our data, now we can calculate posterior distribution.

$$Posterior \propto likelihood * prior$$

$$P(\theta|y) \propto P(y|\theta)P(\theta)$$

$$P(\theta|y) \propto [{}_nC_y\theta^y(1-\theta)^{n-y}][\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}]$$

$$P(\theta|y) \propto [\theta^y(1-\theta)^{n-y}][\theta^{a-1}(1-\theta)^{b-1}]$$

$$P(\theta|y) \propto \theta^{a+y-1}(1-\theta)^{n-y+b-1}$$

- You can drop any constant constant term from these equations.

- Also you can multiply by any constant term

- The last equation is *Kernel* of Beta(a+y, n-y+b)

- We may multiply it by $\frac{\Gamma(a+y+n-y+b)}{\Gamma(a+y)\Gamma(n-y+b)}$ to get the Exact Graph of Posterior Distribution.

- The shape of posterior doesn't change by multiplying by constant, but after multiplying the area under curve equals 1.

```
posterior <- function(theta){
  likelihood(theta) * prior(theta)
  }
```

ggplot2

```r
x <- seq(0,1, length.out = 1000)
pri <- prior(x)
lik <- likelihood(x)
pos <- posterior(x)
bet <- dbeta(x, a+y, n-y+b)  #Given in page 15
dat <- data.frame(x, pri, lik, pos, bet)

require(ggplot2)
g1=ggplot(dat, aes(x=x, y=pri)) + geom_point() + labs(title="Prior", x="Theta" )
g2=ggplot(dat, aes(x=x, y=lik)) + geom_point() + labs(title="Likelihood", x="Theta" )
g3=ggplot(dat, aes(x=x, y=pos)) + geom_point() + labs(title="Posterior", x="Theta" )
g4=ggplot(dat, aes(x=x, y=bet)) + geom_point() + labs(title="Beta(a+y, n-y+b)", x="Theta

require(gridExtra)
grid.arrange(g1, g2, g3, g4, ncol=2)
```

Scaling the posterior

```r
prior <- function(theta){
  theta^(a-1) * (1-theta)^(b-1) # NO Constant
}


n=20
y=12
likelihood <- function(theta){
  theta^(y)*(1-theta)^(n-y)    # NO Constant
}

posterior <- function(theta){
  (gamma(a+y+n-y+b) / (gamma(a+y)*gamma(n-y+b))) * likelihood(theta) * prior(theta)
}
# Multiplyed by a constant




#ggplot2
x <- seq(0,1, length.out = 1000)
pri <- prior(x)
lik <- likelihood(x)
pos <- posterior(x)
bet <- dbeta(x, a+y, n-y+b)    #Given in page 15
dat <- data.frame(x, pri, lik, pos, bet)

require(ggplot2)
g1=ggplot(dat, aes(x=x, y=pri)) + geom_point() + labs(title="Prior" , x="Theta" )
g2=ggplot(dat, aes(x=x, y=lik)) + geom_point() + labs(title="Likelihood", x="Theta")
g3=ggplot(dat, aes(x=x, y=pos)) + geom_point() + labs(title="Posterior", x="Theta")
g4=ggplot(dat, aes(x=x, y=bet)) + geom_point() + labs(title="Beta(a+y, n-y+b)", x="Theta

require(gridExtra)
grid.arrange(g1, g2, g3, g4, ncol=2)
```
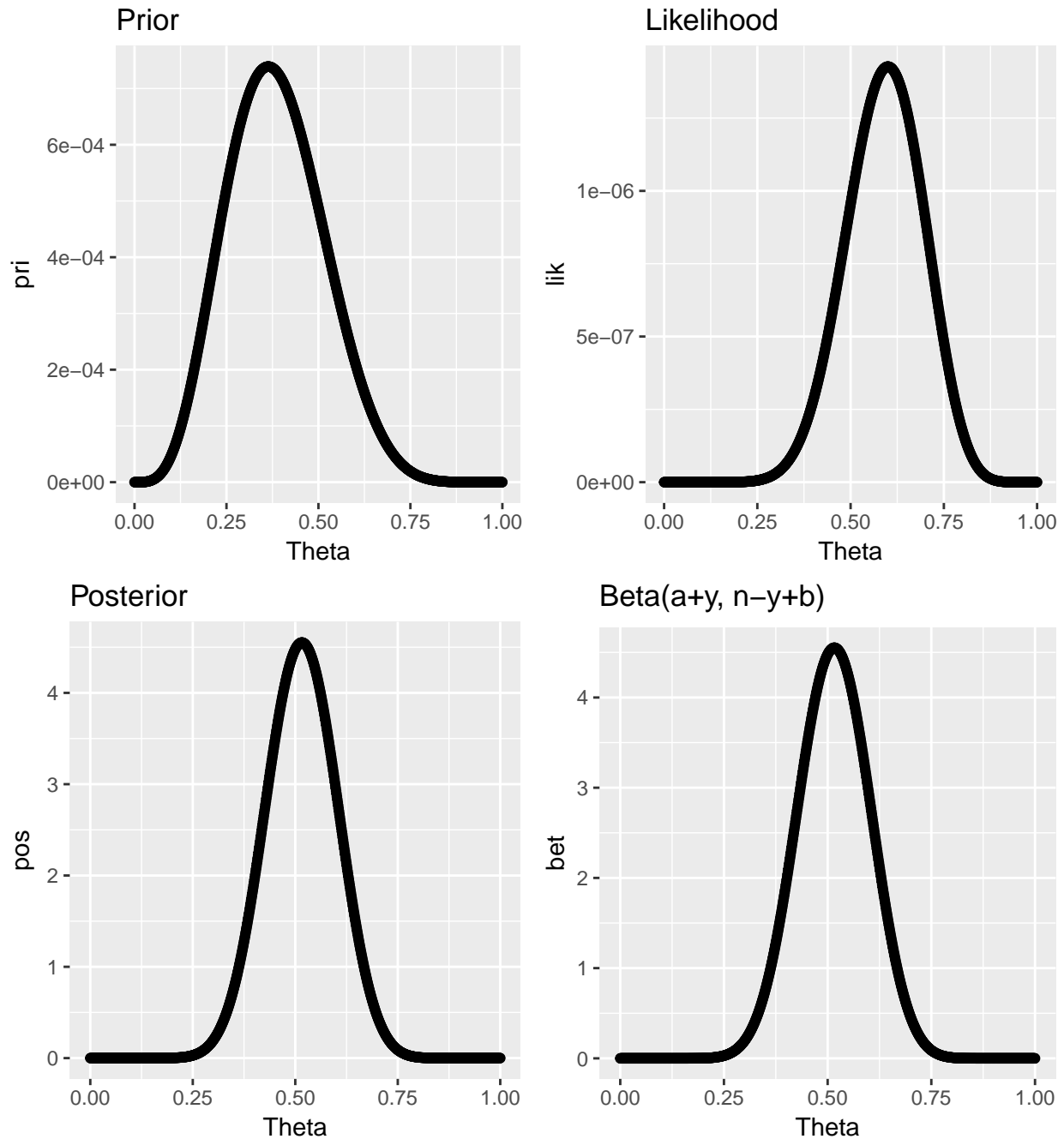
Comment

Now the posterior from posterior function and dbeta function matches exactly on y-axis scaling. Previously y-axis scaling was different for 2 graphs.

The prior was spread, posterior is clustered around near center, so posterior gives better estimates.

We dont need to scale in exam, just use dbeta, qbeta, pbeta for everything.

Credible Interval

```r
qbeta(c(0.025, 0.975), shape1 = a+y, shape2 = n-y+b)
```

```
## [1] 0.3474368 0.6811250
```

Interpretation of Credible Interval

Given the data, we have a posterior distribution that tells us that θ lies in this range with probability 0.95

Mean

```r
a <- a+y      # Makes writing formula easy
b <- n-y+b

a
```

```
## [1] 17
```

```r
b
```

```
## [1] 16
```

```r
a / (a+b)
```

```
## [1] 0.5151515
```

Variance

```r
a*b/( (a+b)^2 * (a+b+1) )
```

```
## [1] 0.007346189
```

Hypothesis

$$H_o : \theta <= 0.50$$

$$H_a : \theta > 0.50$$

```r
pbeta(0.50, shape1 = a, shape2 = b , lower.tail = FALSE)
```

```
## [1] 0.569975
# a, b are from page 20
```

Posterior
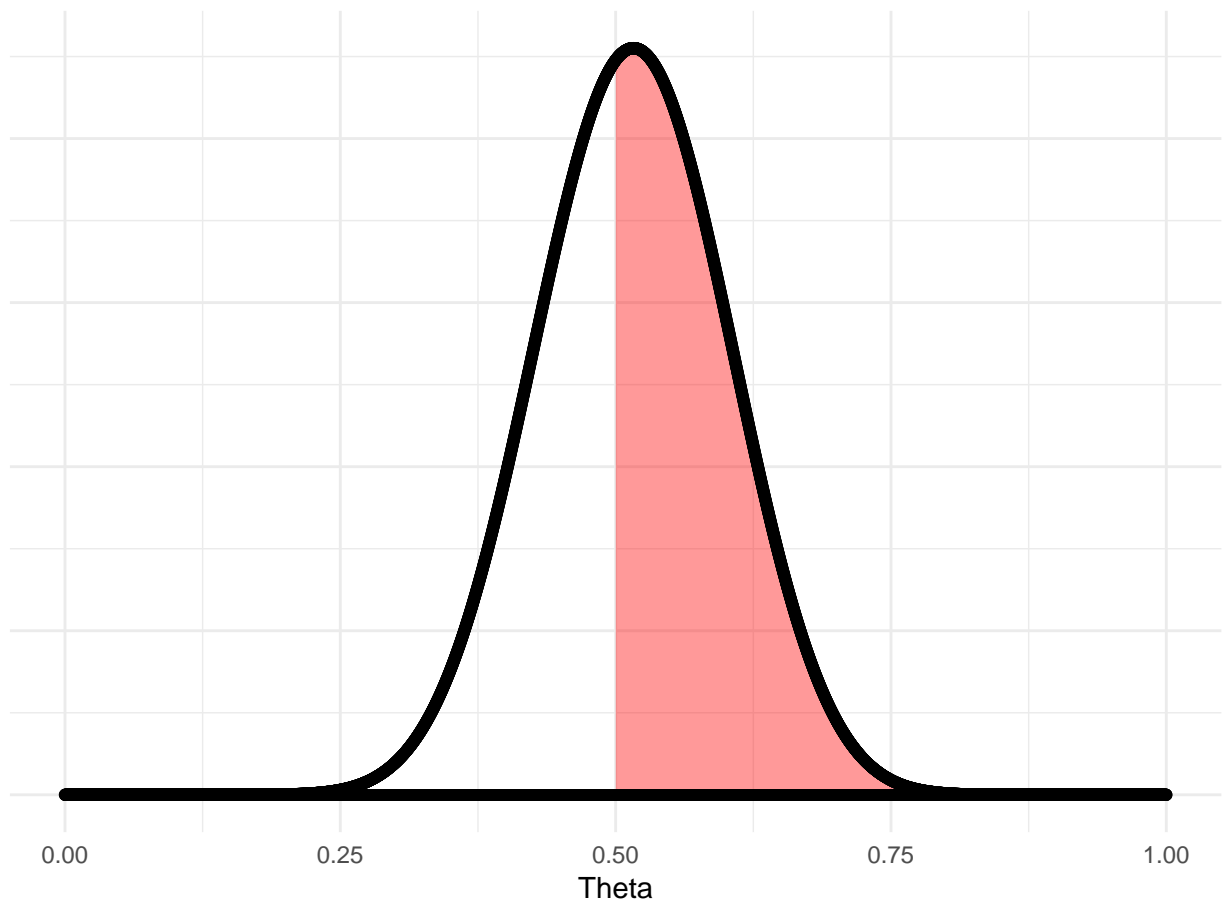
So

Probability that $\theta > 0.50$ is only 0.57. So we cannot say surely that θ is greater than 0.50. So we fail to reject null.

Another Hypothesis

$$H_a : \theta < 0.70$$

```
pbeta(0.70 , shape1 = a, shape2 = b , lower.tail = TRUE )
```

```
## [1] 0.9861601
```

Posterior
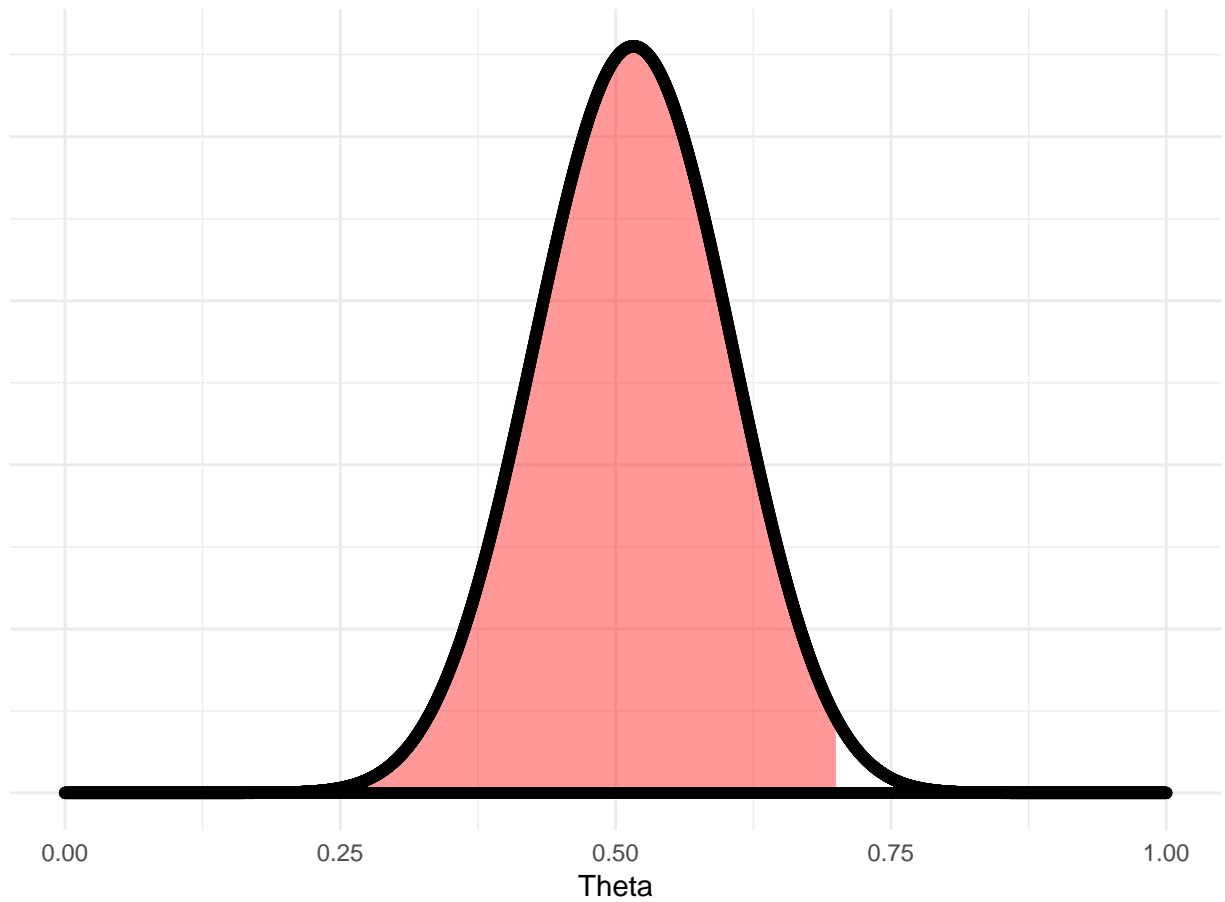
This time we calculate probability that $\theta < 0.70$.

So we calculate area to the left of $\theta = 0.70$.

We see probability that $\theta < 0.70$ is 0.98. So we can Reject NUll and say θ is indeed less than 0.70 .

Conjugate Prior

- If Prior and Posterior comes from same family of distribution, like both from Beta Distribution in this case, then Beta Distribution is said to be Conjugate prior for Binomial Distribution.

- Also Gamma pdf is conjugate prior for Poisson Distribution.

- And Normal is conjugate for Normal Distribution.

- Parameters maybe same or different for prior and posterior, but density function must be same.

Binomial Bayesian with Discrete Flat Prior

Intro

Flat prior means all prior values of parameter have equal probability. This is not same as uniform(0,1).

Flat prior maybe discrete or continuous.

Here the Prior is discrete, so we will get the Posterior as a Tabular form, rather than a specific distribution.

Procedure is same if prior is not flat, but discrete.

Prior

Lets assume we have picked 3 values of Parameter θ = Population Proportion. We give each value equal probability, thus this is a flat prior.

```r
priorprob <- rep( 1/3 , times=3)

m <- c(0.40, 0.60, 0.70)

dat <- data.frame(m , priorprob)
colnames(dat) <- c("Prior Values of theta", "Prior Probability")
dat
```

```
##   Prior Values of theta Prior Probability
## 1                   0.4         0.3333333
## 2                   0.6         0.3333333
## 3                   0.7         0.3333333
```

Likelihood

$$P(y|\theta) = \prod_{i=1}^{n} {}_nC_{y_i} \theta^{y_i} (1-\theta)^{n-y_i}$$

y are data from a Binomial(n, P= θ) distribution. We want to estimate this distribution's proportion θ using bayesian method.

We want the likelihood function to be a function of proportion, as the data y is constant after observing in bayesian method.

Lets generate 20 data from Binomial(n=10, p=0.40).

Then we will see how close our bayesian estimate is to TRUE parameter value = 0.40.

```
# size = 10 = trial numbers in each binomial experiment = n = 10

y <- rbinom(n=20, size = 10, prob=0.40)
y
```

```
##  [1] 5 3 5 3 7 6 2 3 7 6 5 6 2 4 3 1 4 3 4 5
```

```
likelihood <- function(x){
  prod( dbinom(y, size = 10, prob = x  ))
}
```

Finding Posterior

We need to find likelihood function's value for the 3 prior proportion values.

So we will have 3 values of likelihood function for 3 prior proportions.

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{\sum P(y|\theta)P(\theta)}$$

```
lik <- NULL

for(i in 1:3){
  lik[i] <- likelihood(m[i])
  }

pos <- (lik*priorprob) / sum(lik*priorprob)

dat <- data.frame( m, round(pos, 6)  )
names(dat) <- c("Value", "Probability")
```

Posterior Distribution
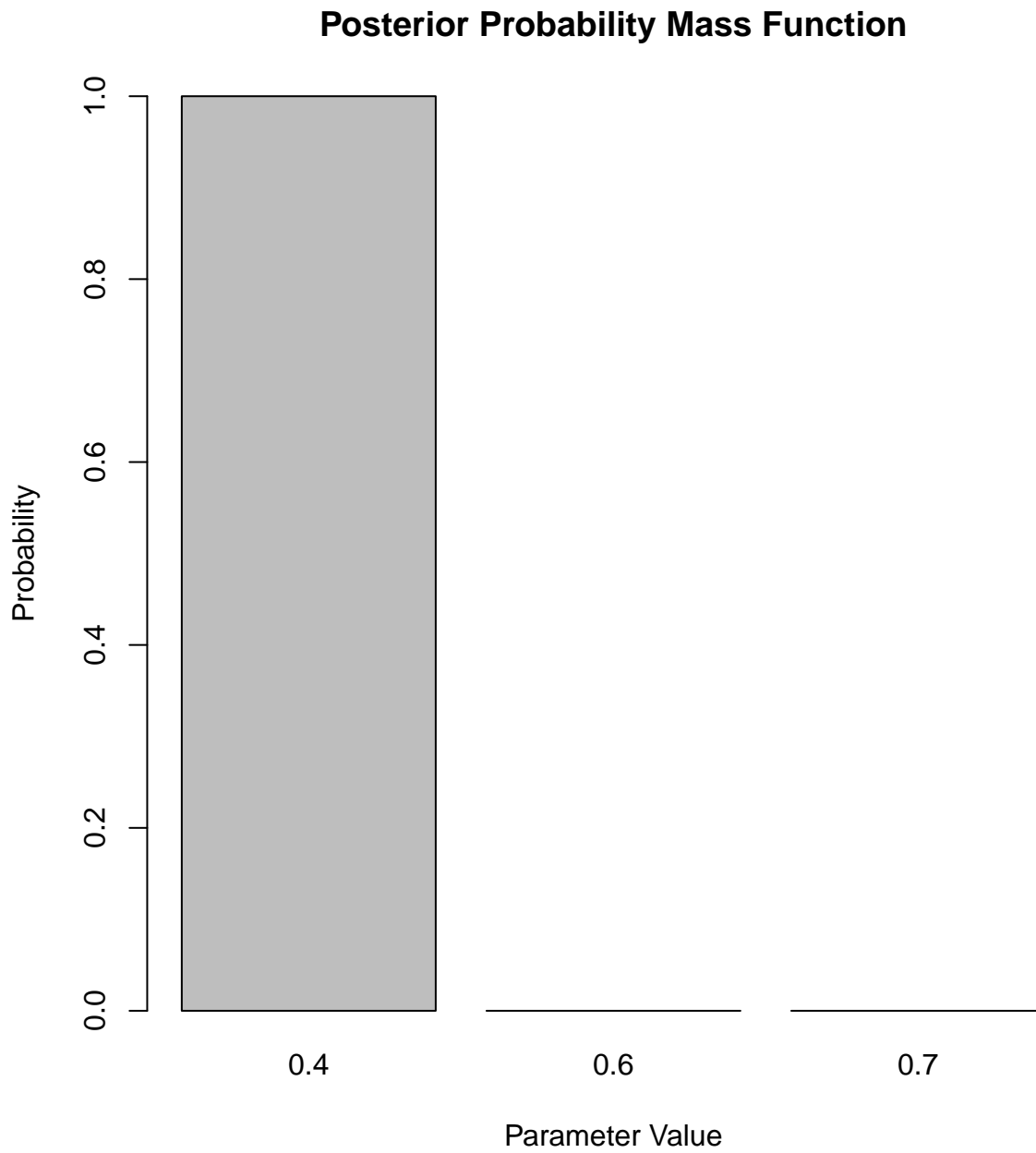
```
dat
```

```
##    Value Probability
## 1    0.4    0.999998
## 2    0.6    0.000002
## 3    0.7    0.000000
```

Graph of Posterior

```r
barplot(dat$Probability,
        names.arg = dat$Value,
        ylim = c(0,1),
        xlab = "Parameter Value",
        ylab = "Probability",
        main = "Posterior Probability Mass Function")
```

Posterior Probability Mass Function

Normal Bayesian with Continuous Normal Prior

Prior

Let's Assume Prior distribution for μ is Normal(mean=10, sd=2).

Let m = 10 = mean of prior, s = 2 = sd of prior.

$$f(\mu)_{Prior} = \frac{1}{\sqrt{(2\pi s^2)}} e^{-\frac{1}{2}\frac{(\mu-m)^2}{s^2}}$$

```r
m <- 10
sdprior <- 2

prior <- function(meu){

  dnorm(x = meu, mean = m, sd = sdprior)

}
```

Likelihood

Let we are drawing data from a normal distribution, where $\sigma = 2$ is known.

$$L(\mu|y) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{1}{2}\frac{(y_i-\mu)^2}{\sigma^2}}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\frac{\sum(y_i-\mu)^2}{\sigma^2}}$$

Lets take 200 samples from Normal(μ=15, σ=2), then see how close our bayesian estimate is to true parameter $\mu = 15$.

```r
y <- rnorm(200, mean=15, sd=2)

likelihood <- function(meu){
  prod(exp( - 0.5 * (y - meu)^2  / 2^2 ))
  }
```

Posterior

```r
posterior <- function(meu){
  likelihood(meu) * prior(meu)
}
```

```r
mu <- seq(0, 20, length.out = 10000 )
lik <- pri <- pos <- NULL

for(i in 1:10000){
  lik[i] <- likelihood(mu[i])
  pri[i] <-        prior(mu[i])
  pos[i] <-  posterior(mu[i])
}

par(mfrow=c(1,2))
plot(mu, pri)
plot(mu, pos, xlim = c(14,16))
```
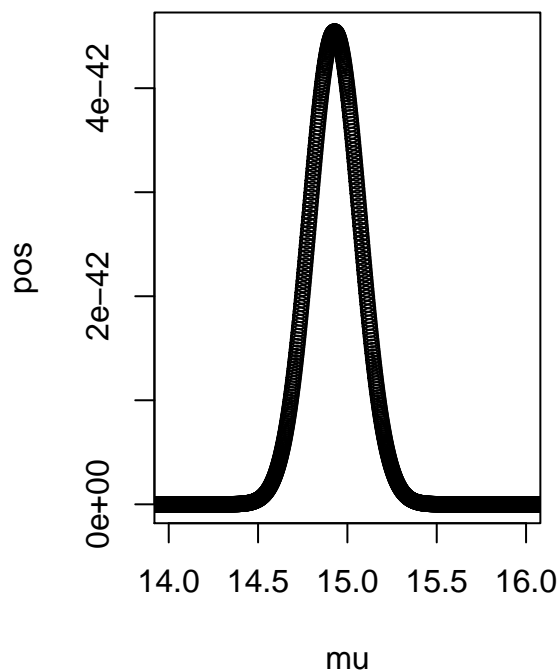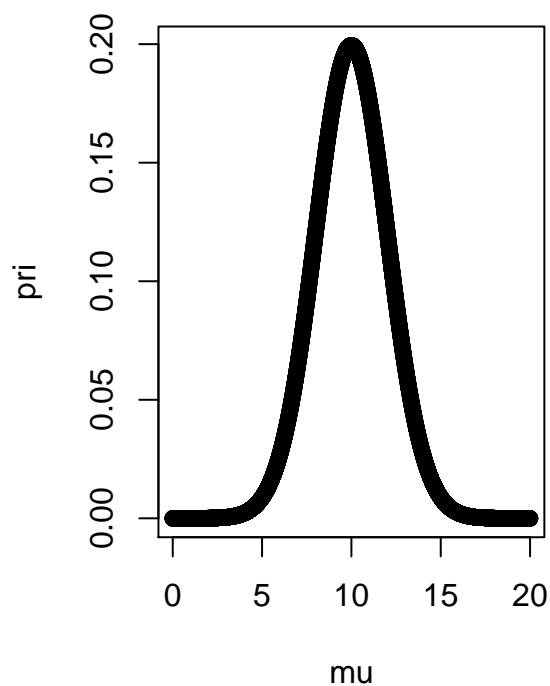
From Formula

Posterior follows Normal (mean , variance), where

$$mean = \frac{1/s^2}{n/\sigma^2 + 1/s^2} m + \frac{n/\sigma^2}{n/\sigma^2 + 1/s^2} \bar{y}$$

$$variance = \frac{\sigma^2 s^2}{\sigma^2 + ns^2}$$

```
s2      <-   4
sigma2 <-   4
n       <- 200
m       <- 10
ybar    <- mean(y)



# Need to remember these formulas is exam
# otherwise wont be able to find pdf and test hypothesis
# Infact need to remember formula for Normal, Binomial, Poisson distributions
# Need to remember for continuous posteriors
```

Mean

```
meanpos <-     ( (1/s2)  / ( ( n/sigma2 ) + (1/s2) ) ) * m +
          ( (n/sigma2) / ( ( n/sigma2 ) + (1/s2) ) ) * ybar


meanpos
```

```
## [1] 14.9283
```

Variance

```
varpos  <- ( (sigma2 * s2)  / ( ( sigma2 ) + (n * s2) ) )


varpos
```

```
## [1] 0.0199005
```

Graph of Posterior

```r
curve(dnorm(x, mean=meanpos,
           sd=sqrt(varpos)),
      from= 14, to= 16,
      xlab = "MEU", ylab = "Density")
```
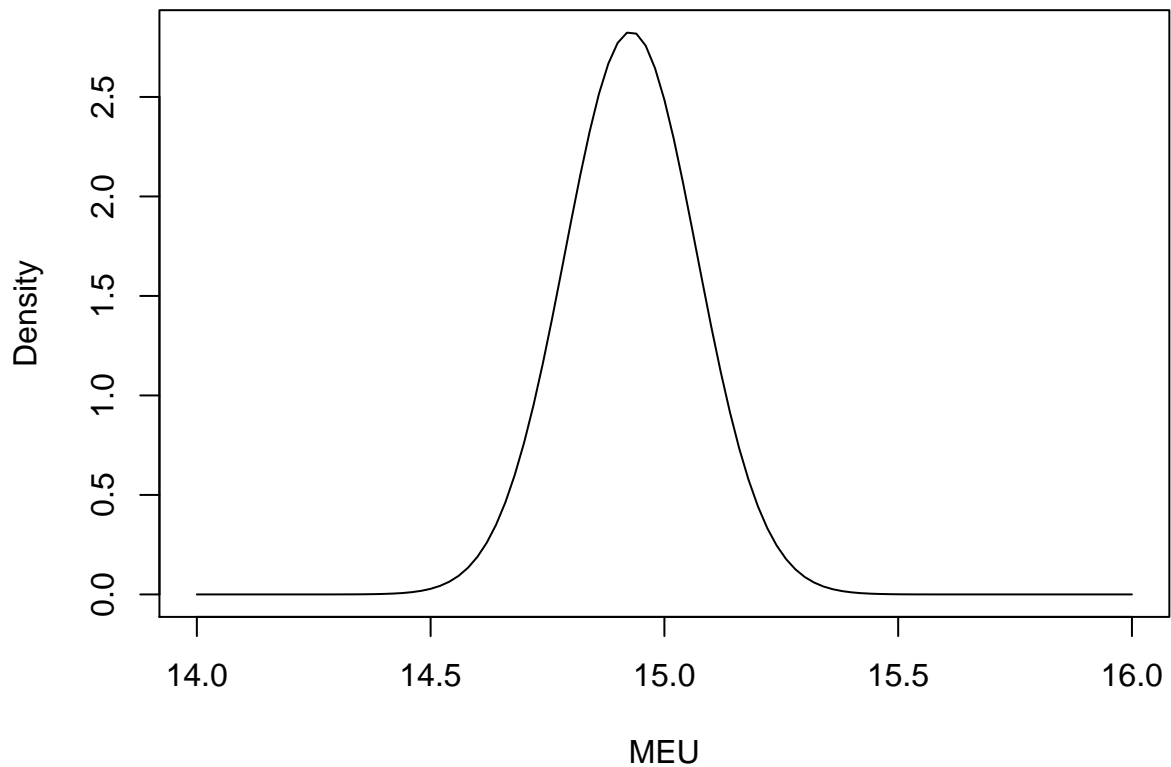

95 % Credible Interval

```r
qnorm( c(0.025, 0.975), mean=meanpos, sd=sqrt(varpos) )
```

```
## [1] 14.65181 15.20479
```

Hypothesis

$$H_a : \mu > 15$$

```
pval = pnorm( 15 , mean = meanpos, sd = sqrt(varpos), lower.tail = FALSE )

pval
```

```
## [1] 0.3056364
```

```
# IF pval > 0.95 then reject null
```

Posterior

MEU MEU MEAOW

Fail to reject NULL, as probability that $\mu > 15$ is only 0.30.

We would have rejected NULL if this probability was > 0.95.

Normal Bayesian with Discrete Flat Prior

Intro

Flat prior means all prior values of parameter have equal probability. This is not same as uniform(0,1).
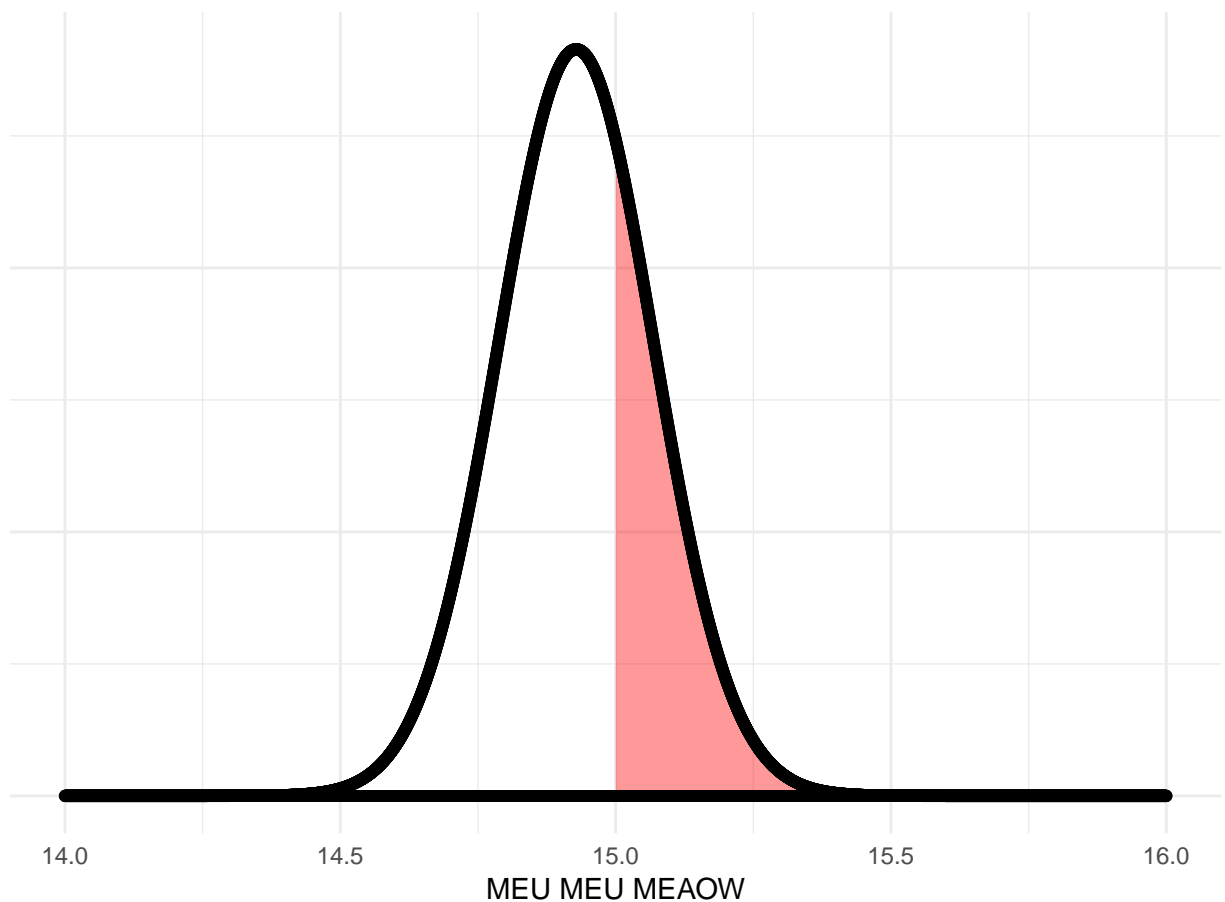
Flat prior maybe discrete or continuous.

Here the Prior is discrete, so we will get the Posterior as a Tabular form, rather than a specific distribution.

Procedure is same if prior is not flat, but discrete.

Prior

Lets assume we have picked 8 values of Parameter μ = Population Mean. We give each value equal probability, thus this is a flat prior.

```
priorprob <- rep(0.125, times=8)

m <- c(160, 170, 180, 190,
       200, 210, 220, 230)

dat <- data.frame(m , priorprob)
colnames(dat) <- c("Prior Mean", "Prior Probability")
dat
```

```
##   Prior Mean Prior Probability
## 1        160             0.125
## 2        170             0.125
## 3        180             0.125
## 4        190             0.125
## 5        200             0.125
## 6        210             0.125
## 7        220             0.125
## 8        230             0.125
```

Likelihood

$$P(y|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi\sigma^2)}} e^{-\frac{1}{2}\frac{(y_i-\mu)^2}{\sigma^2}}$$

y are data from a Normal distribution. We want to estimate this distribution's mean using bayesian method. sd=15 is known.

We want the likelihood function to be a function of mean, as the data y is constant after observing in bayesian method.

```
y <- c(175, 190, 215, 198, 184,207, 210, 193, 196, 180)

likelihood <- function(x){
  prod(dnorm(y, mean=x, sd=15)) }
```

Finding Posterior

We need to find likelihood function's value for the 8 prior mean values.

So we will have 8 values of likelihood function for 8 prior means.

$$P(\mu|y) = \frac{P(y|\mu)P(\mu)}{\sum P(y|\mu)P(\mu)}$$

```
lik <- NULL

for(i in 1:8){
  lik[i] <- likelihood(m[i])
  }

pos <- (lik*priorprob) / sum(lik*priorprob)

dat <- data.frame( m, round(pos, 6)  )
names(dat) <- c("Value", "Probability")
```

Posterior Distribution

```
dat
```

```
##    Value Probability
## 1    160    0.000000
## 2    170    0.000001
## 3    180    0.006624
## 4    190    0.516098
## 5    200    0.472202
## 6    210    0.005074
## 7    220    0.000001
## 8    230    0.000000
```

Graph of Posterior

```r
barplot(dat$Probability,
        names.arg = dat$Value,
        ylim = c(0,1),
        xlab = "Parameter Value",
        ylab = "Probability",
        main = "Posterior Probability Mass Function")
```

Posterior Probability Mass Function
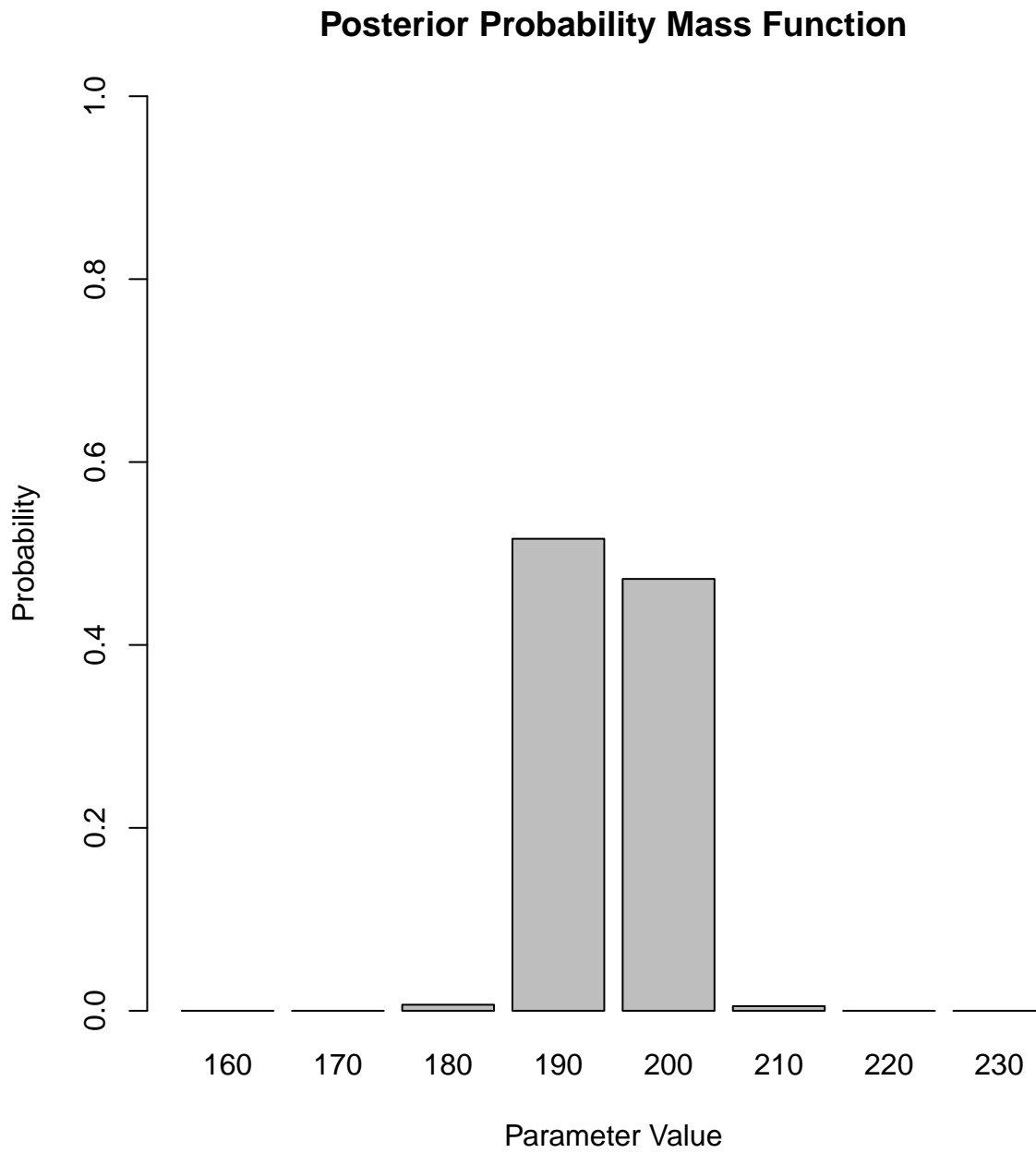
Poisson Bayesian with Continuous Gamma Prior

Prior

Lets assume our prior distribution is Gamma(α,V). And also α=2, v=3. Here parameter λ is the Random variable.

$$f(\lambda)_{prior} = \frac{V^{\alpha}}{\Gamma(\alpha)}\lambda^{\alpha-1}e^{-V\lambda}$$

We can ignore the constant $\Gamma(\alpha)$ in the denominator while deriving posterior.

```
V <- 3
a <- 2

prior <- function(lambda){

  (V^a) * lambda^(a-1)  * exp(-V*lambda)

}
```

Likelihood

We have the prior as Gamma(α,V). Now we want to estimate λ of Poisson(λ). So we need to collect data from a Poisson Distribution. Then we can write the likelihood as

$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

$$L(\lambda|y) = \prod_{i=1}^{n}\frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

$$= \frac{e^{-n\lambda}\lambda^{\sum y_i}}{\prod(y_i!)}$$

We can ignore the constant term in denominator.

```
likelihood <- function(lambda){

  lambda^sum(y) * exp(-n*lambda)

}
```

Posterior

$$f(\lambda)_{posterior} \propto likelihood * Prior$$

$$\propto [e^{-n\lambda}\lambda^{\sum y_i}][V^{\alpha}\lambda^{\alpha-1}e^{-V\lambda}]$$

$$\propto e^{-(V+n)\lambda}\lambda^{\alpha+\sum y_i-1}$$

Ignored the constant V^{α}. It is Kernel of Gamma Distribution.

So Posterior is a Gamma(a, l), where a = $\alpha + \sum y_i$ = Shape Parameter , l = V+n = Rate parameter . And λ is the random variable.

```
posterior <- function(lambda){

  likelihood(lambda) * prior(lambda)

}
```

Data

Let's generate data from poisson ($\lambda = 5$) and lets see how close our bayesian estimate is to True parameter value = 5.

```
y <- rpois(lambda = 5, n = 20)

y
```

```
##  [1]  7  4  5  4 12  2  4  8  5  4  3  5  4 10  5  6  4  8  5  9
```

```
lambdamle <- mean(y)
lambdamle
```

```
## [1] 5.7
```

ggplot2

You Don't need ggplot2 in exam.

```r
n <- 20 #Sample Size
a <- 2  #prior
V <- 3  #prior
lam <- seq(0, 10, length.out = 1000) # values of lambda to draw graph
pri <- prior(lam)
lik <- likelihood(lam)
pos <-  posterior(lam)
gam <- dgamma(x = lam, shape = a+sum(y) , rate = V+n  ) # Exact posterior with area=1
dat <- data.frame(lam, pri, lik, pos, gam)
# ggplot requires a data frame, you don't need all these in exam
require(ggplot2)
g1 <- ggplot(dat, aes(x=lam, y=pos)) + geom_point() + labs(title="Posterior")
g2 <- ggplot(dat, aes(x=lam, y=gam)) + geom_point() + labs(title="Gamma(a+sum(y), V+n)"
require(gridExtra)
grid.arrange(g1, g2, ncol=2)
```
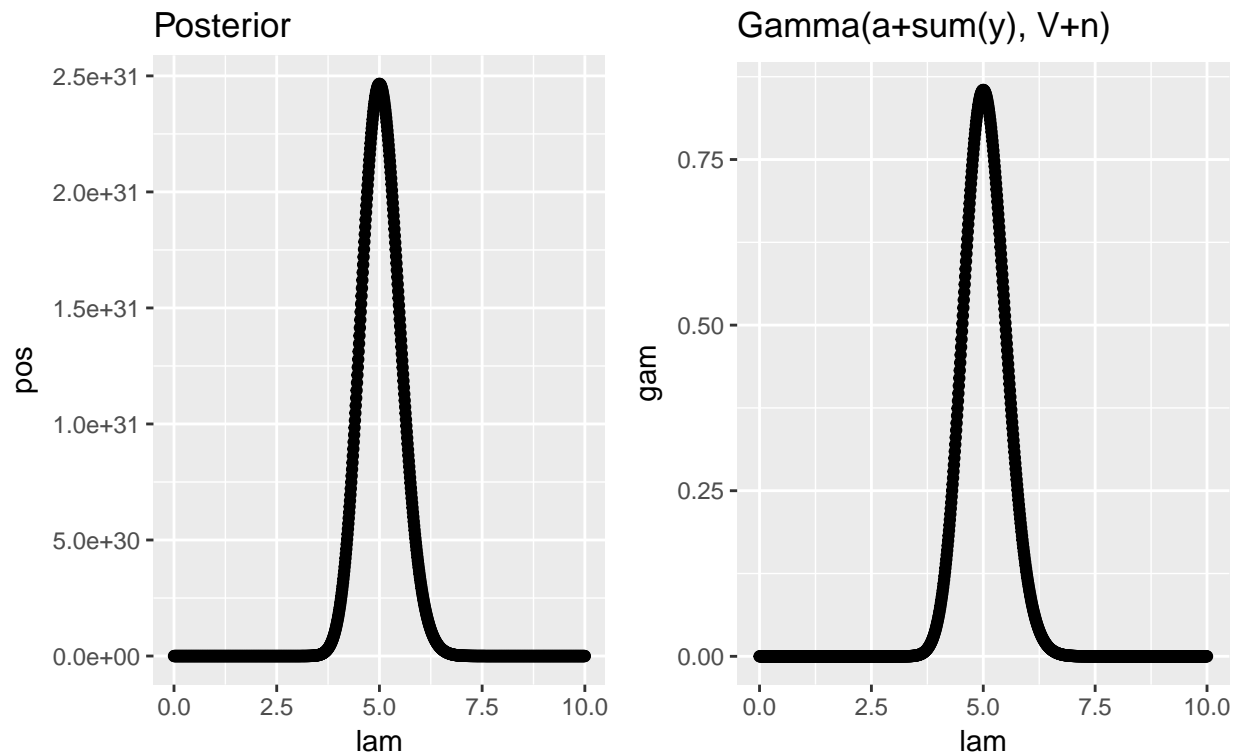

The left graph just indicates the shape of posterior distribution, and the right graph is exact Posterior Distribution with area=1 under curve.

We can use the distribution Gamma(a, l), where $a = \alpha + \sum y_i$, $l =$ V+n ,to draw probabilistic conclusion about parameter λ .

Credible Interval

```
a = a + sum(y)
l = V + n

qgamma(c(0.025,0.975), shape = a, rate = l)
```

```
## [1] 4.167524 6.001743
```

Mean

If X ~ Gamma(Shape = α , Rate= λ), then

$$E[X] = \frac{\alpha}{\lambda}$$

```
a / l
```

```
## [1] 5.043478
```

$$Var[X] = \frac{\alpha}{\lambda^2}$$

Variance

```
a / l^2
```

```
## [1] 0.2192817
```

Cheat Formula

- In exam remember formulas of **Continuous** posterior distributions for poisson, binomial, normal and find Posterior distribution, its graph, mean, variance, hypothesis tests etc.

- You don't need to write functions and draw their graphs for Continuous Posteriors, just remember formulas and find posterior.
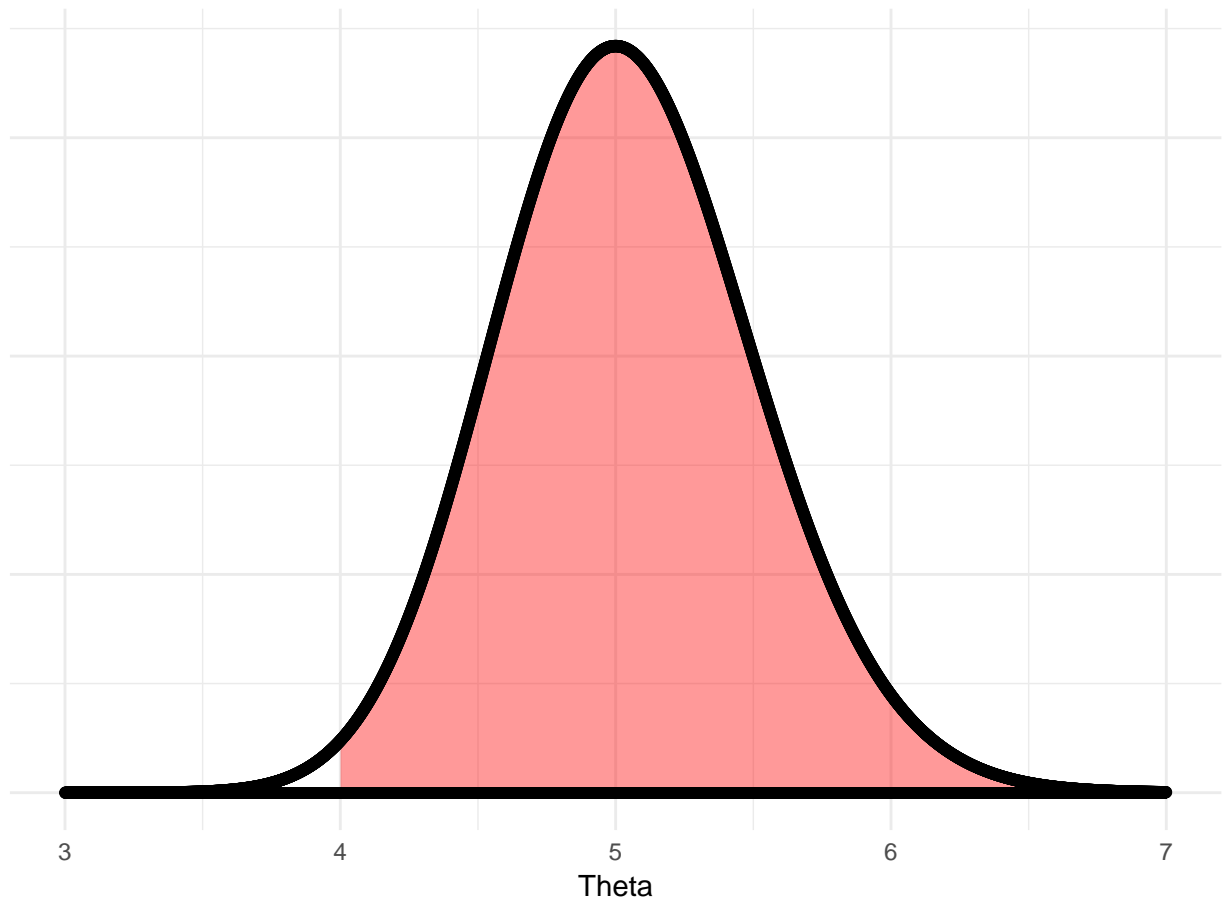
Hypothesis

$$H_o : \lambda = 4$$
$$H_a : \lambda > 4$$

```
pgamma(4, shape = a, rate=l, lower.tail = FALSE)
```

```
## [1] 0.9911591
```

Posterior

Theta

Probability that λ is greater than 4 is 99%.

So we can surely say λ is greater than 4, and Alternative hypothesis is true.

We can varify that by checking that the 95% credible interval for λ DO NOT contain 4.
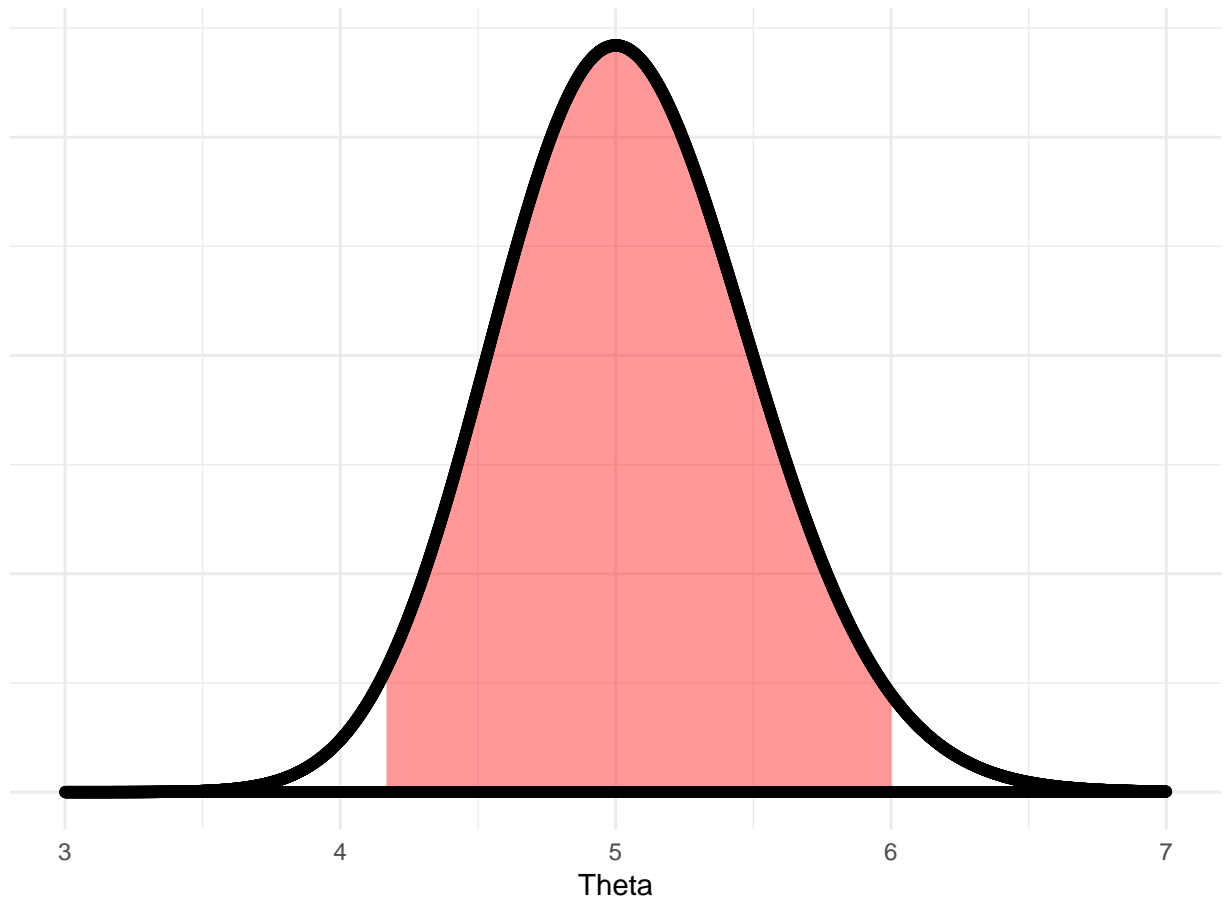
Two Sided Hypothesis

$$H_o : \lambda = 4$$
$$H_a : \lambda \neq 4$$

In this case we use 95% Credible Interval to Reject NULL hypothesis. If Credible Interval contains the NULL value then we fail to reject. If Credible Interval Do Not contain NULL Value, then we Reject NULL.

Posterior

Credible Interval

```
qgamma( c(0.025, 0.975), shape = a, rate = l )
```

[1] 4.167524 6.001743

Credible Interval Do Not contain the Value 4, so we reject NuLL here and conclude λ is greater than 4.

Poisson Bayesian with Disrete Flat Prior

Intro

Flat prior means all prior values of parameter have equal probability. This is not same as uniform(0,1).

Flat prior maybe discrete or continuous.

Here the Prior is discrete, so we will get the Posterior as a Tabular form, rather than a specific distribution.

Procedure is same if prior is not flat, but discrete.

Prior

We pick 3 rates as our prior values for λ. We give each of them equal probability, thus this is a flat prior.

```r
priorprob <- rep( 1/3 , times=3)

m <- c(4 , 6 , 7)

dat <- data.frame(m , priorprob)
colnames(dat) <- c("Prior Values of Lambda", "Prior Probability")
dat
```

```
##   Prior Values of Lambda Prior Probability
## 1                      4         0.3333333
## 2                      6         0.3333333
## 3                      7         0.3333333
```

Likelihood

$$P(y|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{y_i}}{y_i!}$$

y are data from a Poisson(rate = θ) distribution. We want to estimate this distribution's rate θ using bayesian method.

We want the likelihood function to be a function of rate = λ, as the data y is constant after observing in bayesian method.

Lets generate 20 data from Poisson($\lambda = 4.5$).

Then we will see how close our bayesian estimate is to TRUE parameter value $= 4.5$.

```
y <- rpois(n=20, lambda = 4.5)
y
```

```
## [1]  5  4  4  3  5  4  1  1  5  6  9 10  2  6  3  6  5  3  1  6
```

```
lambdamle <- mean(y)
lambdamle
```

```
## [1] 4.45
```

```
likelihood <- function(x){
  prod( dpois(y, lambda = x  ))
}
```

Finding Posterior

We need to find likelihood function's value for the 3 prior λ values.

So we will have 3 values of likelihood function for 3 prior λ.

$$P(\lambda|y) = \frac{P(y|\lambda)P(\lambda)}{\sum P(y|\lambda)P(\lambda)}$$

```
lik <- NULL
for(i in 1:3){
  lik[i] <- likelihood(m[i])
  }
pos <- (lik*priorprob) / sum(lik*priorprob) # Posterior Probability

dat <- data.frame( m, round(priorprob, 3), lik, lik*priorprob, round(pos, 6)  )
names(dat) <- c("Value", "Prior Probability", "Likelihood","Likelihood x Prior", "Poster
dat
```

```
  Value Prior Probability   Likelihood Likelihood x Prior
1     4             0.333 1.577910e-20       5.259699e-21
2     6             0.333 3.150830e-22       1.050277e-22
3     7             0.333 5.899288e-25       1.966429e-25
  Posterior Probability
1              0.980387
2              0.019577
3              0.000037
```

47

Posterior Distribution

Value	Prior Probability	Likelihood	Likelihood x Prior	Posterior Probability
4	0.333	1.577910e-20	5.259699e-21	0.980387
6	0.333	3.150830e-22	1.050277e-22	0.019577
7	0.333	5.899288e-25	1.966429e-25	0.000037

This is the way you should write down Posterior Distribution incase of discrete posterior.

Graph of Posterior

```r
barplot(dat$ `Posterior Probability` ,
        # use ` ` if variable name has space in it

        names.arg = dat$Value,
        ylim = c(0,1), xlab = "Parameter Value",
        ylab = "Probability",
        main = "Posterior Probability Mass Function" )
```

Posterior Probability Mass Function
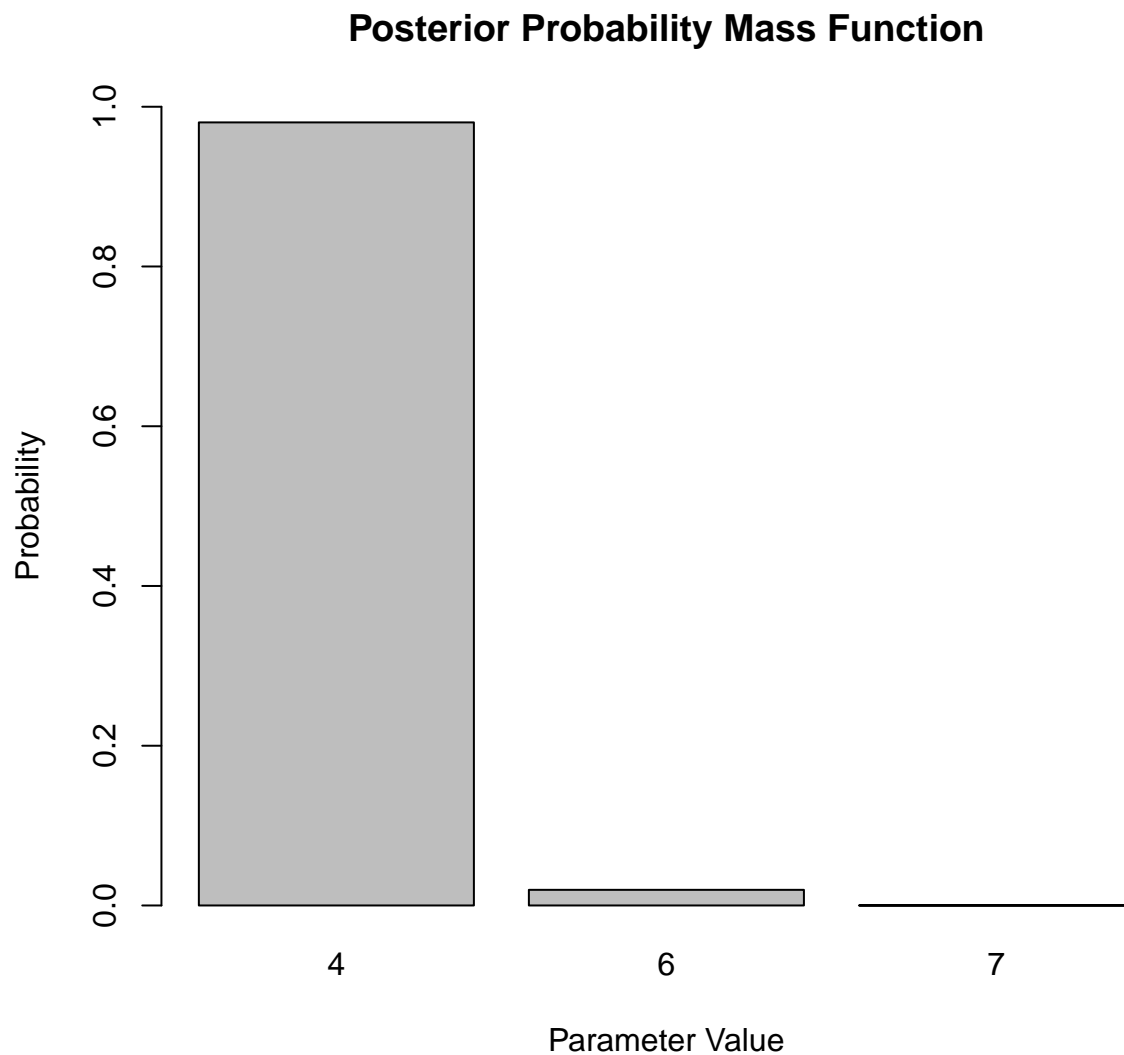
Hope your prior knowledge have been improved by these slides and you have a posterior knowledge now about bayesian estimation. :P

FRK

15 April, 2018