AdvStDaAn, Worksheet, Week 11

Michael Lappert

05 Juni, 2022

Contents

Task 1

You can choose from two different types of light-emitting diodes.

You order 100 diodes of type 1 and 100 diodes of type 2. 8 diodes of type 1 and 12 diodes of type 2 were insufficient for your purposes. However diodes of type 2 are 10% cheaper.

Is it beneficial to buy diodes of type 2? Reject rates usually vary around 10% $\pm$ 5%. Propose an appropriate prior and perform a Bayesian analysis.

First one can calculate a and b for the beta prior.

$E(X) = \frac{a}{a+b} = 0.1$ b = 9a

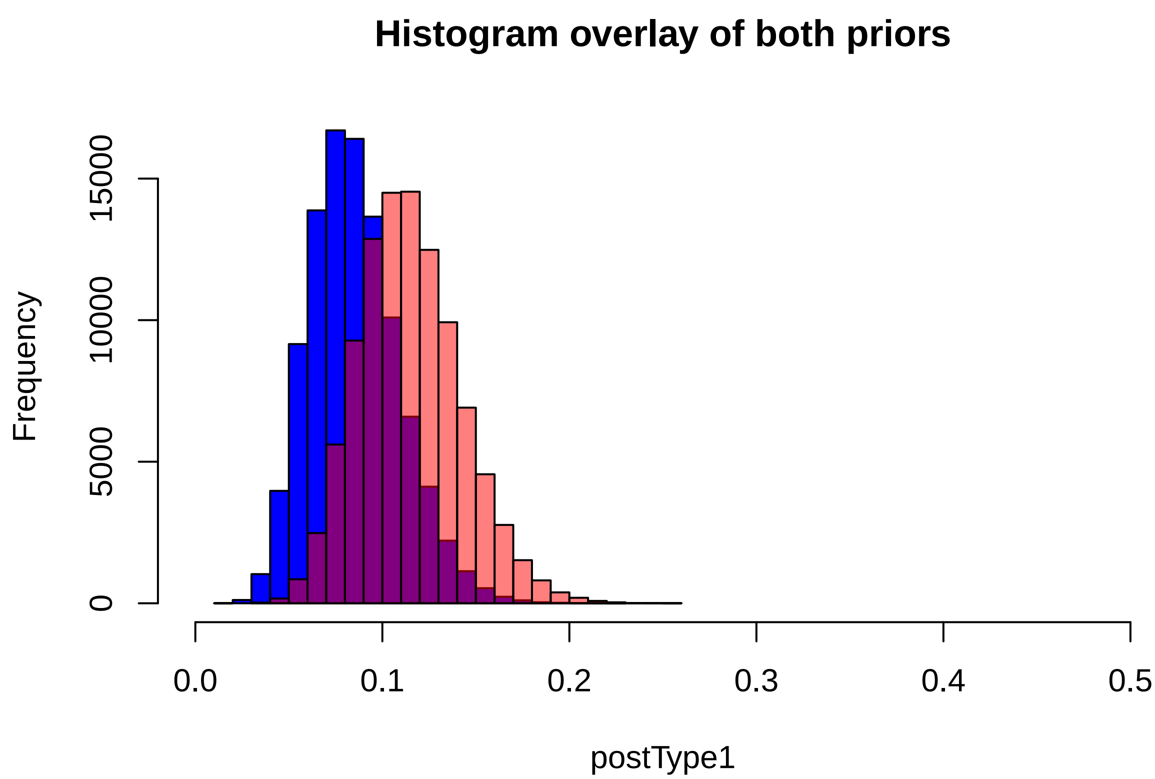
$Var(X) = \frac{ab}{(a+b+1)(a+b)^2} = 0.05^2$

set b = 9a

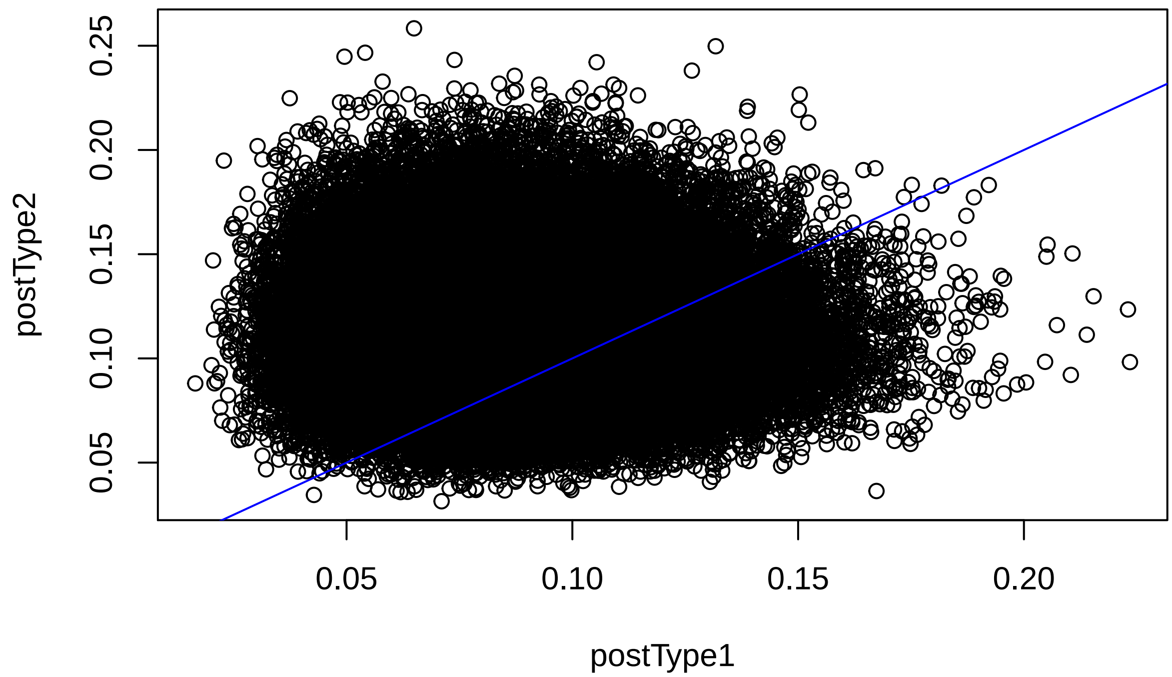
$\frac{a*9a}{(a+9a+1)(a+9a)^2} = 0.05^2$

a = 3.5
b = 31.5

```r
# Simulate reject rates for diodes of type 1 and type 2
nSamples = 100000
postType1 = rbeta(nSamples, 3.5 +  8, 31.5 + 92)
postType2 = rbeta(nSamples, 3.5 + 12, 31.5 + 88)

hist(postType1, xlim = c(0, .5), col = 'blue', main = 'Histogram overlay of both priors')
hist(postType2, xlim = c(0, .5), col = alpha('red', 0.5), add = TRUE)
```




```r
plot(postType1, postType2)
abline(0,1,col="blue")
```




```
sum(postType1< postType2)/nSamples
```

```
## [1] 0.7938
```

```
# Calculate expected reject rates
expType1 = (3.5 +   8)/ (3.5 +   8 +31.5 + 92) # -> E(X)
expType2 = (3.5 +  12)/ (3.5 +  12 +31.5 + 88) # -> E(X)

# To compare, include that diodes of type 2 are 10% cheaper
(1-expType1)
```

```
## [1] 0.9148148
```

```
(1-expType2)/.9
```

```
## [1] 0.9835391
```

=> Type 2 is better than Type 1.

Task 2

Given are 6 trials and their numbers of trys before a success occurs: 5, 8, 3, 9, 2 and 6.

- In what range is the rate of success θ?
- What's the probability, that the rate of success is bigger than 25%?

```r
n <- 6
obs <- c(5, 8, 3, 9, 2, 6)
obs_sum <- sum(obs)
obs_sum
```

```
## [1] 33
```

```r
# calculating the range of the rate of success
qbeta(c(0.05, 0.95), 1 + n, 1 + obs_sum)
```

```
## [1] 0.0851325 0.2747445
```

```r
# calculating the probability that the rate of success is bigger than 25%
1 - pbeta(0.25, 1+n, 1+obs_sum)
```

```
## [1] 0.09622459
```

A flat prior seems not very plausible, as data indicates that the success rate should be small. Using a flat prior probably gives too much weight to high success rates and therefore overestimates the real success rate.

Task 3

You have two machines producing a good. Your quality control tracks the number of proper goods until the machine produces a poor good.

- The counts of machine 1 are: 10, 15, 18, 20, 5, 12 and 3.
- The counts of machine 2 are: 23, 16, 19, 28 and 37.

Assume a Beta(1, 20) prior on the failure rate and calculate an equal-tailed 95% credible interval for each machine. What's the expected failure rate for each machine? Calculate the maximum a posteriori probability (MAP) estimate for each failure rate.

```r
# number of observations
nM1 <- 7
nM2 <- 5

# sum of failures
(sumObsM1 <- sum(c(10, 15, 18, 20, 5, 12, 3)))
```

```
## [1] 83
```

```r
(sumObsM2 <- sum(c(23, 16, 19, 28, 37)))
```

```
## [1] 123
```

```r
# parameters of both distribution
aM1 <- 1+nM1
bM1 <- 20+sumObsM1
aM2 <- 1+nM2
bM2 <- 20+sumObsM2

# equal-tailed 95% credible interval
qbeta(c(0.025, 0.975), 1+nM1, 20+sumObsM1)
```

```
## [1] 0.03192018 0.12672593
```

```r
qbeta(c(0.025, 0.975), 1+nM2, 20+sumObsM2)
```

```
## [1] 0.01502071 0.07707699
```

```r
# expected failure rate (E(X))
expFailM1 <- (1+nM1) / ((1+nM1) + 20+sumObsM1)
expFailM1
```

```
## [1] 0.07207207
```

```r
expFailM2 <- (1+nM2) / ((1+nM2) + 20+sumObsM2)
expFailM2
```

```
## [1] 0.04026846
```

```r
# maximum a posteriori probability (MAP) -> mode(X)
mapM1 <- ((1+nM1) - 1) / ((1+nM1) + (20+sumObsM1) - 2)
mapM1
```

```
## [1] 0.06422018
```

```r
mapM2 <- ((1+nM2) - 1) / ((1+nM2) + (20+sumObsM2) - 2)
mapM2
```

```
## [1] 0.03401361
```

Task 4

- What's the probability of observing two failures before a success occurs if θ is the probability of success?
- Calculate the maximum likelihood estimate for the parameter θ of the geometric distribution, when you observed two failures before a success occurs.

- What's the posterior distribution of θ conditioned on two failures before a success occurs, when you assume a flat prior?
- What's the posterior mean of θ?
- What's the maximum a posteriori probability estimate of θ?

Probability of observing two failures before a success occurs: $\theta(1-\theta)^2$

mle estimate for θ:

```r
1/3
```

```
## [1] 0.3333333
```

Posterior distribution of θ conditioned on two failures before a success occurs with a flat prior? $\theta|D \sim$ beta(1 + 1, 1 + 2)

Posterior mean of θ? $E(X) = \frac{a}{a+b} = \frac{2}{2+3} = 0.4$

Maximum a posteriori probability(MAP -> mode(X) estimate of θ? $mode(X) = \frac{a-1}{a+b-2} = \frac{2-1}{2+3-2} = 0.3333333$

Question Task 4

|>Q1: How do we come up with the equations in the mle calculation? Where comes the argmax argument from? ***

Task 5

Your business lends special machines on a daily basis. You have 5 machines to satisfy the demand. A staff member recorded the requests for these special machines on 10 successive days:

5, 7, 3, 9, 2, 6, 6, 5, 8, 4.

An additional machine would be profitable, if you could lend the machine for at least 183 days per year (365 day).

What's the posterior distribution on requests per day λ? Calculate a 90% credible interval. Assume a Gamma(5, 1) prior on λ.

Would it be beneficial to invest in an additional machine?

Posterior distribution on requests per day? $\theta|D \sim$ gamma(5 +)

```r
obs <- c(5, 7, 3, 9, 2, 6, 6, 5, 8, 4)
(numObs <- length(obs))
```

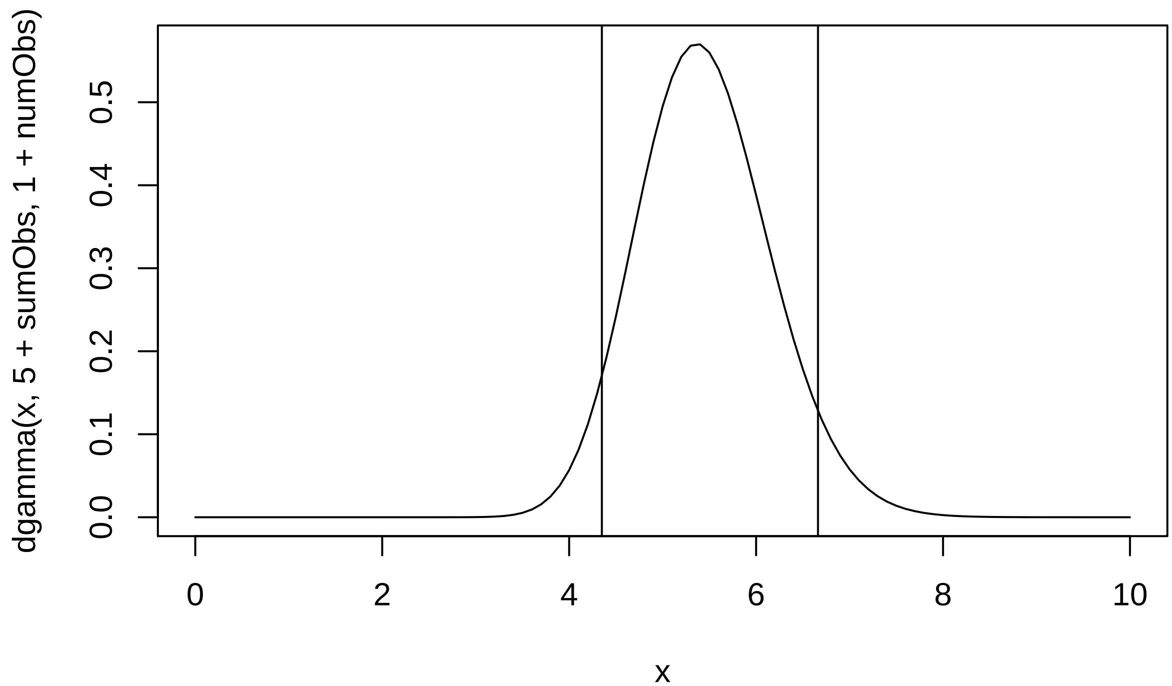
```
## [1] 10
```

```r
(sumObs <- sum(obs))
```

```
## [1] 55
```

$\theta|D \sim \text{Gamma}(5 + \text{sumObs}, 1 + \text{numObs})$

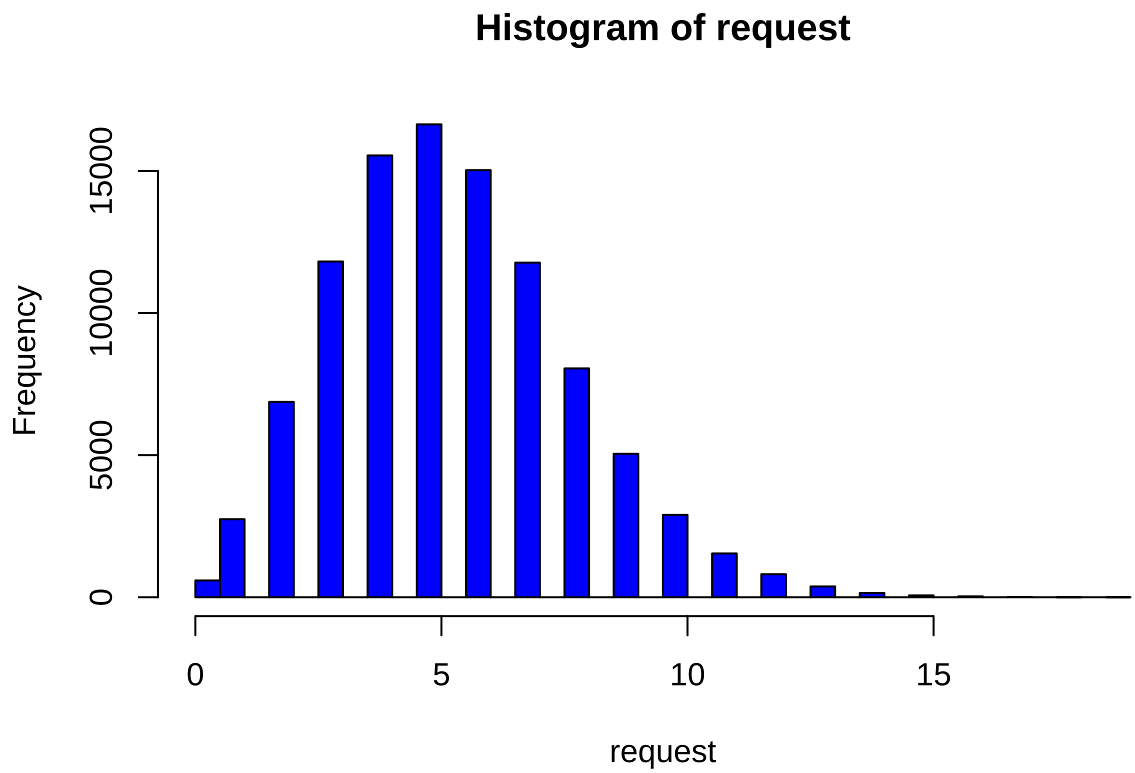
```r
curve(dgamma(x, 5+sumObs, 1 + numObs), 0, 10)

# 90% credible interval
abline(v=qgamma(c(0.05,0.95),5+55,1+10))
```




```r
qgamma(c(0.05,0.95),5+55,1+10)
```

```
## [1] 4.350211 6.662153
```

```r
# beneficial?
n=100000
# simulate the requests per day
request = rpois(n,rgamma(n,5+55,1+10))
hist(request,breaks=30,col="blue")
```




```
# are there more than 5 requests per day in more than 50% of all days?
sum(request>5)/n
```

```
## [1] 0.4579
```

This means an additional machine would not be beneficial.

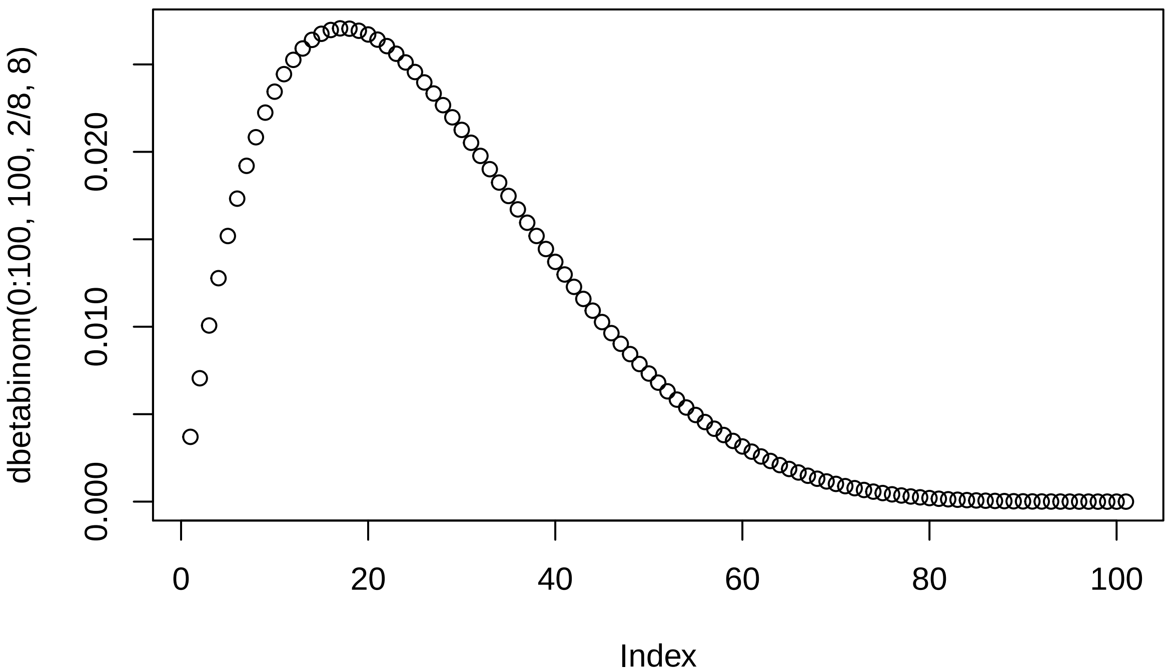
Task 6

A random sample of size 10 from a container with 100 electronic elements is evaluated: 3 of 10 elements had failures.

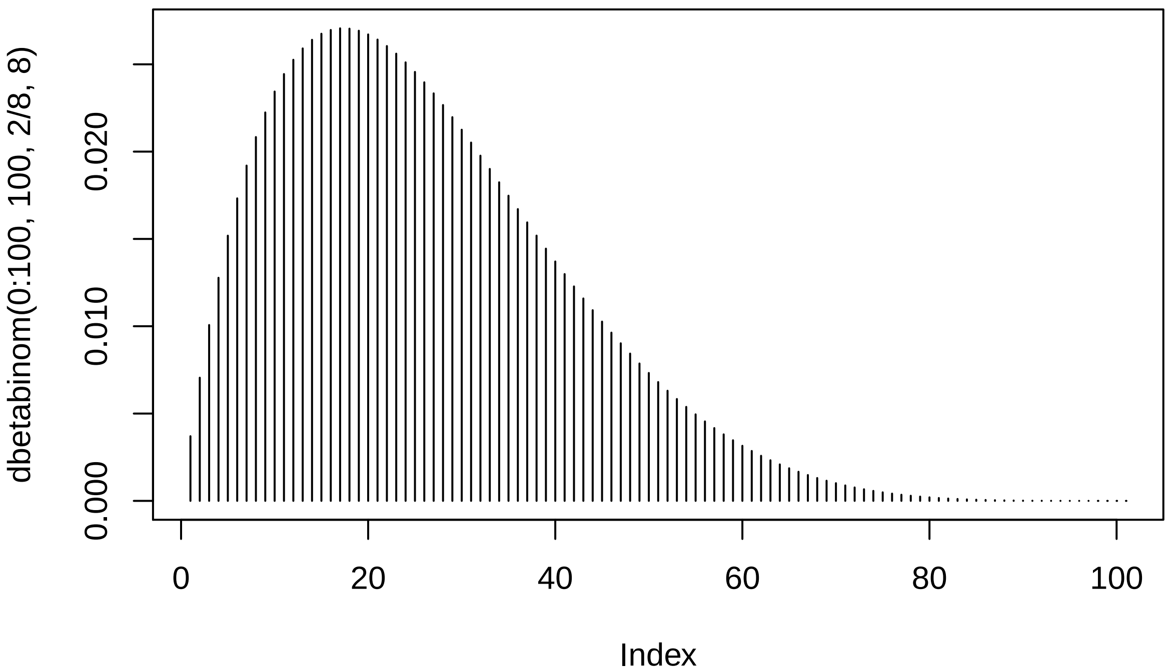
Determine the uncertainty for the number of elements with failures in the container, when the prior distribution on the number of elements with failures is BetaBinom(100, 2, 6). Calculate the equal-tailed 90% credible interval.

With a BetaBinom(100, 2, 6) prior the posterior becomes:
M-3 ~ BetaBin(100(N) - 10(n), 2(a) + 3(m), 6(b) + 10(n) - 3(m))

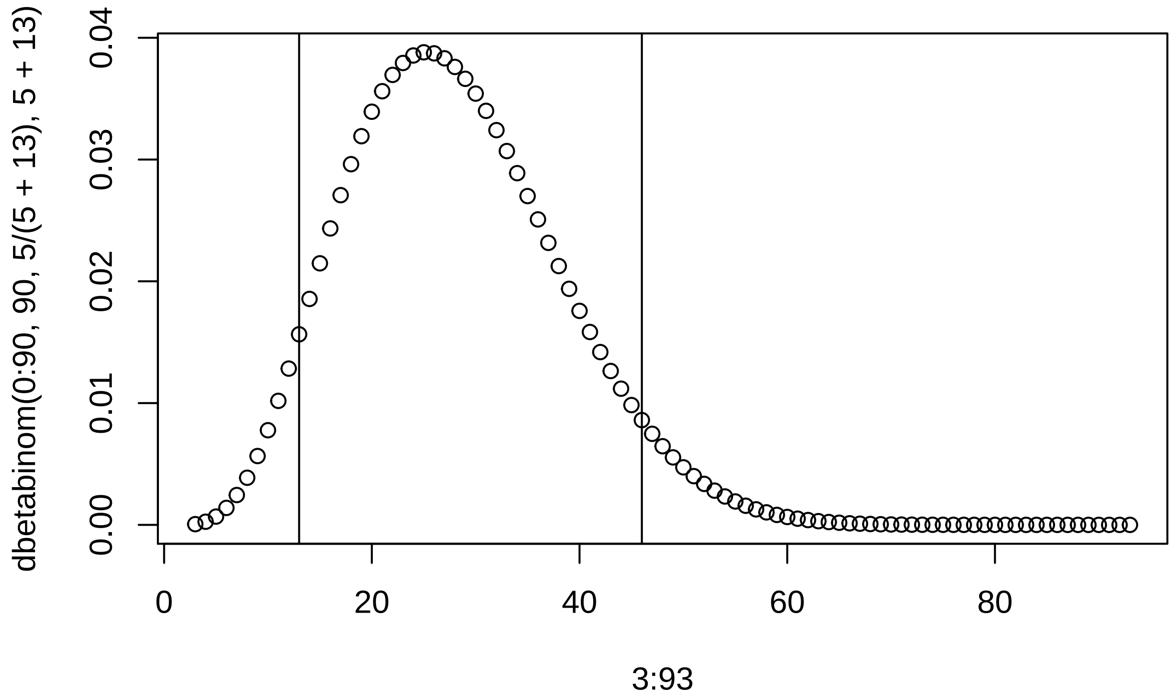
```
# prior: a=2, b=6
plot(dbetabinom(0:100,100,2/8,8))
```




```
plot(dbetabinom(0:100,100,2/8,8), type = 'h')
```




```
# posterior: a=2+3, b=6+7 -> explanation about implementation in the question header
plot(3:93, dbetabinom(0:90, 90, 5/(5+13), 5+13))

abline(v=3+qbetabinom(c(0.05,0.95),90,5/18,18))
```




```
3+qbetabinom(c(0.05,0.95),90,5/18,18)
```

```
## [1] 13 46
```

Question Task 6

|>Q1: Why do we have to divide the vector of probabilities? Like 5/(5+13)? |»A1: According to the documentation of betabinom (type ?dbetabinom in the console) the inputs for it are dbetabinom(y, size, m, s). Therby it is stated that
m = a/(a+b) and s = (a+b) ***