MTHM508 Coursework

2022-11-09

```
library(tidyr)
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.8      v dplyr   1.0.10
## v readr   2.1.3      v stringr 1.4.1
## v purrr   0.3.4      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Problem

Lucy has had 65 “yes” responses to baby Percy’s Birthday party invitation. She is planning the catering and there is a sickness bug going around that means it is likely a number of people will cancel.

Suppose that she judges that the probability of any given guest dropping out due to illness is 1 - θ (so probability of attending is θ).

We are making the tacit assumptions that,

i. Out of the 65 people who have responded “yes”, those who will drop out will do so only because of the sickness, i.e. people will not be absent from the party for any other reason.
ii. Those who have been infected will not show up, i.e. people will not come to the party while having contracted the sickness.

1. Describe the exchangeability judgement that Lucy is making.

Definition 2.9. A sequence $(Y_1, Y_2, ...)$ of random quantities is said to be exchangeable if the joint probability distribution of each sub-collection of n quantities $(Y_{i_1}, ..., Y_{i_n})$ is the same.

i. **Comment**. To say that a set of random quantities is exchangeable is a subjective judgement.

ii. **Comment**. Exchangeability amounts to indifference over re-labelling the random quantities or permuting them.

To simplify, for 2 r.v.s, x_1 and x_2, to be exchangeable, their joint probability distribution $P(x_1, x_2) = P(x_2, x_1)$. Therefore, *independent* and *identically distributed* random variables are exchangeable, by definition.

The event of someone attending the event can be modeled as a random variable, $X \sim Bern(\theta)$ with $0 \leq \theta \leq 1$. Thus, $P(X = x_i) = \theta^{x_i}(1-\theta)^{(1-x_i)}$. Here, Lucy views all x_i to be i.i.d., thereby allowing her to make the exchangeability judgement. In simple terms, she makes a symmetry judgement that the probability of **any given guest attending** is θ, i.e., she views the chance of one person attending the event as equal to that of any other person and thus, the joint probability does not change with any shuffling of the i’s in the sequence of x_is.

2. Suppose that you make the same exchangeability judgement as Lucy yourself.

Derive and carefully justify your Beta prior distribution for θ, $\pi(\theta)$.

b. Use a "pseudo data" argument to select the Beta parameters (expressing current knowledge as if updating a uniform prior with "pseudo observations" as discussed in lectures)

Say we first have a uniform prior θ, such that $\pi(\theta) \propto 1$. By Bayes we have,

$$\pi(\theta|\boldsymbol{y}) \propto \pi(\theta)p(\boldsymbol{y}|\theta)$$

$$\pi(\theta|\boldsymbol{y}) \propto 1 \cdot \prod_{i=1}^{n} p(y_i|\theta)$$

$$\pi(\theta|\boldsymbol{y}) \propto \prod_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i}$$

$$\pi(\theta|\boldsymbol{y}) \propto \theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i}$$

Writing $s = \sum_{i=1}^{n} y_i = n\overline{y}$,

$$\pi(\theta|\boldsymbol{y}) \propto \theta^{(s+1)-1}(1-\theta)^{(n-s+1)-1}$$

$\pi(\theta|\boldsymbol{y})$ is proportional to a Beta distribution

Then, our posterior $\pi(\theta|\boldsymbol{y}) \sim Beta(s+1, n-s+1)$ with $s = n\overline{y}$.

Let's fabricate some data. $\boldsymbol{y} = (1,0,1,1,1,0,1,1,1,1,1,1,1,1,0,1,1,1,0,0)$

From here, we say that $n = 20$ and $s = 15$, such that $\overline{y} = 0.75$, which gives us $a = s+1$ and $b = n-s+1$.

Now we have $\pi(\theta|y) \sim Beta(16, 6)$ with $s = n\overline{y}$

c. Use the prior predictive distribution to justify your choice.

$p(y) = \int_{-\infty}^{\infty} p(y|\theta)\pi(\theta)d\theta$

Plugging in the prior predictive distribution,

$$p(y) = \int_{-\infty}^{\infty} \theta^{y}(1-\theta)^{(1-y)}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_{-\infty}^{\infty}\theta^{a+y-1}(1-\theta)^{b-y}d\theta$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_{-\infty}^{\infty}\theta^{(a+y)-1}(1-\theta)^{(b-y+1)-1}d\theta$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+y)\Gamma(b-y+1)}{\Gamma(a+b+1)}\int_{-\infty}^{\infty}\frac{\Gamma(a+b+1)}{\Gamma(a+y)\Gamma(b-y+1)}\theta^{(a+y)-1}(1-\theta)^{(b-y+1)-1}d\theta = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\frac{\Gamma(a+y)\Gamma(b-y+1)}{\Gamma(a+b+1)}$$

$$= \frac{1}{a+b}\frac{\Gamma(a+y)\Gamma(b-y+1)}{\Gamma(a)\Gamma(b)}$$

In the second to last step, we integrate the pdf of a *Beta* distribution to be 1. Then, plugging in values of a and b,

$= \frac{1}{22}\frac{\Gamma(16+y)\Gamma(7-y)}{\Gamma(16)\Gamma(6)}$

Therefore, $p(y=0) = \frac{1}{22}\frac{\Gamma(16)\Gamma(7)}{\Gamma(16)\Gamma(6)}$ $p(y=0) = \frac{6}{22}$

And, $p(y=1) = \frac{1}{22}\frac{\Gamma(17)\Gamma(6)}{\Gamma(16)\Gamma(6)}$ $p(y=0) = \frac{16}{22}$, which is consistent with the expectation of a $Beta(16,6)$ distribution, i.e. $\frac{a}{a+b}$

3. Lucy attends 5 children's birthday parties in the run up to Percy's and she judges that each one has the same probability of an attendee dropping out due to sickness. The number of expected attendees at each was 10, 50, 35, 25 and 40. There were 2, 8, 12, 6 and 8 absences respectively.

What is your posterior distribution for θ (*which specific Beta distribution is it*)?

By Bayes,

$$\pi(\theta|\boldsymbol{y}) \propto p(\boldsymbol{y}|\theta)\pi(\theta)$$

$$\pi(\theta|\boldsymbol{y}) \propto \pi(\theta)\prod_{i=1}^{n} p(y_i|\theta)$$

$$\pi(\theta|\boldsymbol{y}) \propto \theta^{a-1}(1-\theta)^{b-1}\prod_{i=1}^{n}\theta^{y_i}(1-\theta)^{1-y_i}$$

$$\pi(\theta|\boldsymbol{y}) \propto \theta^{a-1}(1-\theta)^{b-1}\theta^{\sum_{i=1}^{n} y_i}(1-\theta)^{n-\sum_{i=1}^{n} y_i}$$

Let us substitute $s = \sum_{i=1}^{n} y_i = n\overline{y}$

$$\pi(\theta|\boldsymbol{y}) \propto \theta^{a-1}(1-\theta)^{b-1}\theta^{s}(1-\theta)^{n-s}$$

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

Hence, $\pi(\theta|\boldsymbol{y})$ follows a Beta distribution

$\theta|\boldsymbol{y} \sim Beta(a+s, n-s+b)$

Plugging in our determined values for a and b, we get $\theta|\boldsymbol{y} \sim Beta(16+s, n-s+6)$

We know from our data that

```
y_trials = c(10, 50, 35, 25, 40)
y_failures = c(2, 8, 12, 6, 8)
y_successes = y_trials - y_failures
print(sum(y_trials)) # Value for n
```

```
## [1] 160
```

```
print(sum(y_successes)) # Value for s
```

```
## [1] 124
```

```
print(sum(y_successes)/sum(y_trials)) # Value for y_bar
```

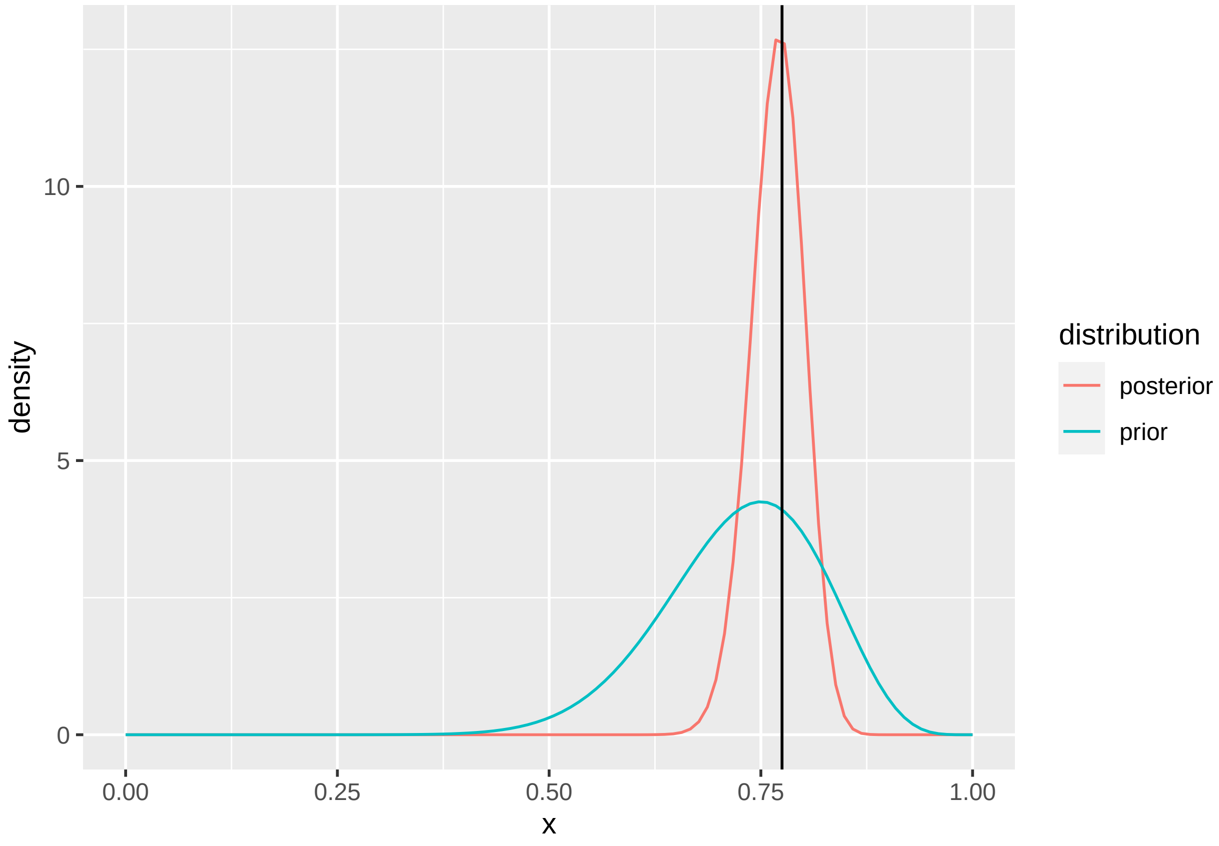
```
## [1] 0.775
```

We know the values of n as 160 and s as 124

Plugging in values for s and n, we get $\theta|\boldsymbol{y} \sim Beta(140, 42)$

4. Plot your prior and posterior on the same figure and give critical comment on the influence of the data on your prior beliefs.

```r
a = 16
b = 6

plot_beta_prior = function(successes , trials , a , b , n){

  binary_data <- tibble(x=seq(from = 0 , to = 1 , len = n)) %>%
    mutate(prior = dbeta(x , a , b)) %>%
    mutate(posterior = dbeta(x , successes + a , trials - successes + b))%>%
    pivot_longer(cols = c(2,3) , names_to = 'distribution' , values_to = 'density')

  ggplot(binary_data) +
    geom_line(aes(x = x , y = density , col = distribution)) +
    geom_vline(xintercept = successes/trials)
}

# Prior has an effect
plot_beta_prior(sum(y_successes) , sum(y_trials) , a , b , 100)
```



We see that the data has significantly changed our prior beliefs.

Critical Comments:

Although the prior and posterior means are not that far apart, the prior is relatively spread out over the range $(0, 1)$. On the other hand, the posterior has condensed the probabilities around $\frac{140}{182} = 0.769$, which happens to be very close to the average number of success rate observed by Lucy $\frac{124}{160} = 0.775$. Incorporating data from the observations has thus allowed us to have a much closer approximation of what the 'true' θ parameter might be.

5. Using only uniform random numbers, estimate the posterior predictive probability that more than 50 people attend Percy's birthday party using Monte Carlo, ensuring that the error on your estimate is bounded above by 0.01.

You must:

- Write down the probability you require as the appropriate integral.
- State clearly how your uniform samples are converted to samples from the right distributions.
- Only use sampling algorithms shown during the course and when more than one sampler is available, you must choose the most efficient.
- Report the Monte Carlo Error of your estimate.
- Include the code from your sampler in your report.
- Whilst you may not use any inbuilt R samplers, you may use their inbuilt density/distribution functions (so rbinom cannot be used but dbinom and pbinom can be).

The posterior predictive distribution is as follows: $p(\tilde{y}|\boldsymbol{y}) = \int_{-\infty}^{\infty} p(\tilde{y}|\theta)\pi(\theta|\tilde{\boldsymbol{y}})d\theta$

With T as an arbitrary threshold,

$p(\tilde{y} > T|\boldsymbol{y}) = \int_{T}^{\infty} p(\tilde{y}|\boldsymbol{y})d\tilde{y}$

$= \int_{T}^{\infty}\int_{-\infty}^{\infty} p(\tilde{y}|\theta)\pi(\theta|\tilde{\boldsymbol{y}})d\theta d\tilde{y}$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\tilde{y} > T)p(\tilde{y}|\theta)\pi(\theta|\tilde{\boldsymbol{y}})d\theta d\tilde{y}$

$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\tilde{y} > T)p(\tilde{y},\theta|\boldsymbol{y})d\theta d\tilde{y}$

where I(A) is the indicator random variable for event A occuring, i.e.

$$I(A) = \begin{cases} 1, & if\ A\ happens \\ 0, & otherwise \end{cases}$$

We take

$$g(\theta|\tilde{y}) = I(\tilde{y} > T),\ as\ our\ functionandf(\theta, y) = p(\tilde{y},\theta|\boldsymbol{y}),\ as\ our\ sampling\ distribution$$

To apply the Monte Carlo algorithm here, we need to

1. sample $(y_1, \theta_1), ..., (y_n, \theta_n)$ from the conditional joint distribution $p(\tilde{y},\theta|\boldsymbol{y})$
2. Pass the values through the function $g(.)$
3. Compute average of samples making it above T

Sequentially sample θ_* from the posterior $\pi(\theta|\boldsymbol{y})$, i.e. $Beta(140, 42)$ and then use said θ to sample y_i from $Binom(n, \theta_*)$. Plugging $n = 65$, we have $Binom(65, \theta_*)$

Compute MC estimate as $\hat{g} = \frac{1}{N}\sum_{i=1}^{N} I(y_i > T)$. Plugging $T = 50$, we have $\hat{g} = \frac{1}{N}\sum_{i=1}^{N} I(y_i > T)$.

Note: $N = 65$ here.

We will use **Box Muller** for sampling Normal distribution and **Rejection Sampling with Normal Envelope** for sampling Beta distribution.

Rejection Sampling can be done with a bounding box method as well, but using a Normal Envelope would lead to higher acceptance rates, and thus lower compute.

Before coding up the samplers, let's take a quick look at how **uniform samples are converted to samples from these distributions.**

For $\mathcal{N}(0, 1)$ distribution,

1. Generate θ from $Unif(0, 2\pi)$, (basically $Unif(0, 1)$ scaled up by 2π)
2. Generate W as $-2log(1 - U)$ where $U \sim Unif(0, 1)$
3. Generate X_1 as $\sqrt{W}cos\theta$ and X_2 as $\sqrt{W}sin\theta$

4. Generate normal draw Z as $\frac{1}{2\pi}e^{-\frac{1}{2}(X_1^2+X_2^2)}$

For $Beta(a, b)$ distribution,

1. $A = a - 1$
2. $B = b - 1$
3. $\hat{x} = \frac{A}{A+B}$ and $H = -(A^{\frac{1}{3}} + B^{\frac{1}{3}})^3$
4. Draw X_u from $\mathcal{N}(\hat{x}, -\frac{1}{H})$
5. Sample U from $Unif(0, 1)$
6. Define $g(x) = x^A(1 - x)^B$
7. Define $g_u(x) = g(\hat{x}) \cdot e^{\frac{H}{2}(x-\hat{x})^2}$
8. Accept X_u as sample if following criteria is met:

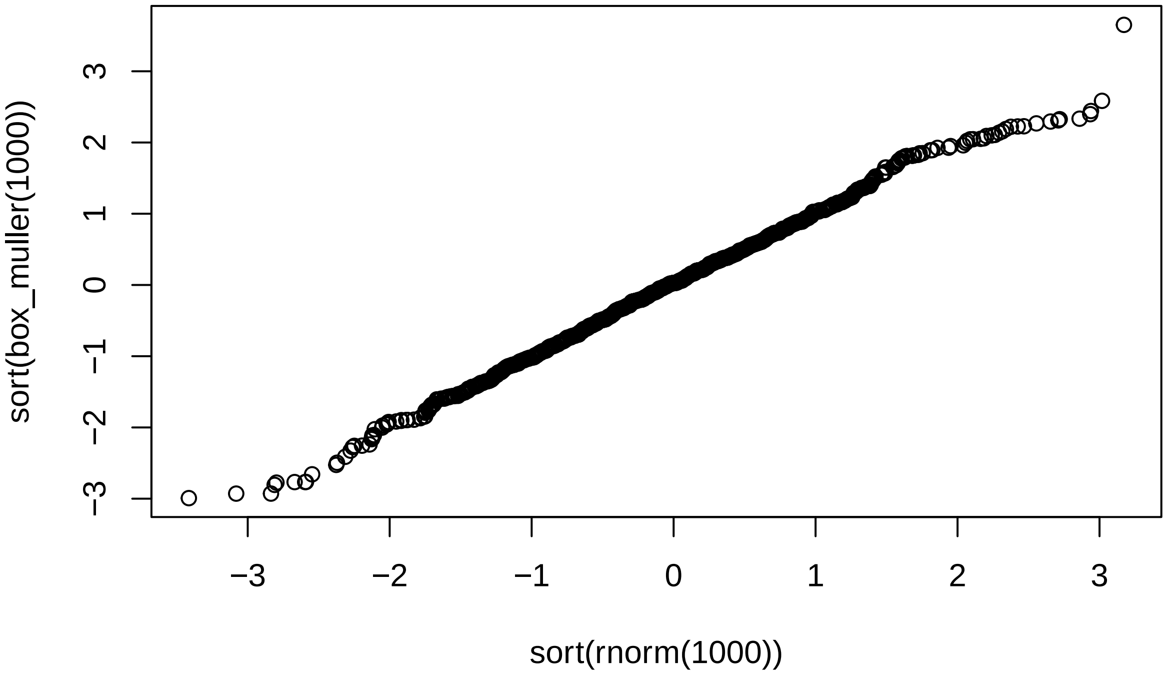
$$U \cdot g_u(X_u) \leq g(X_u)$$

We want to generate a $Beta(140, 42)$ by using a standard uniform distribution $U(0, 1)$

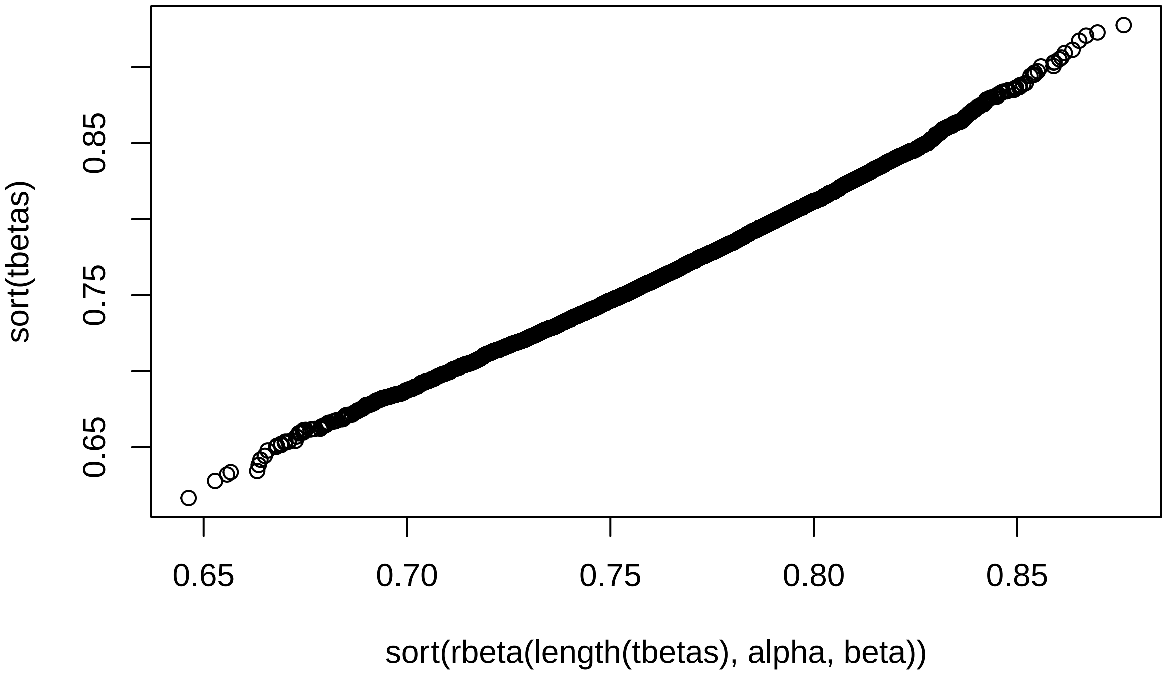
Creating Beta and Binomial Samples without using native R samplers

In order to create a $Beta$ distribution, we need to be able to sample from a Normal distribution, which in turn, we will build via the Box-Muller method.

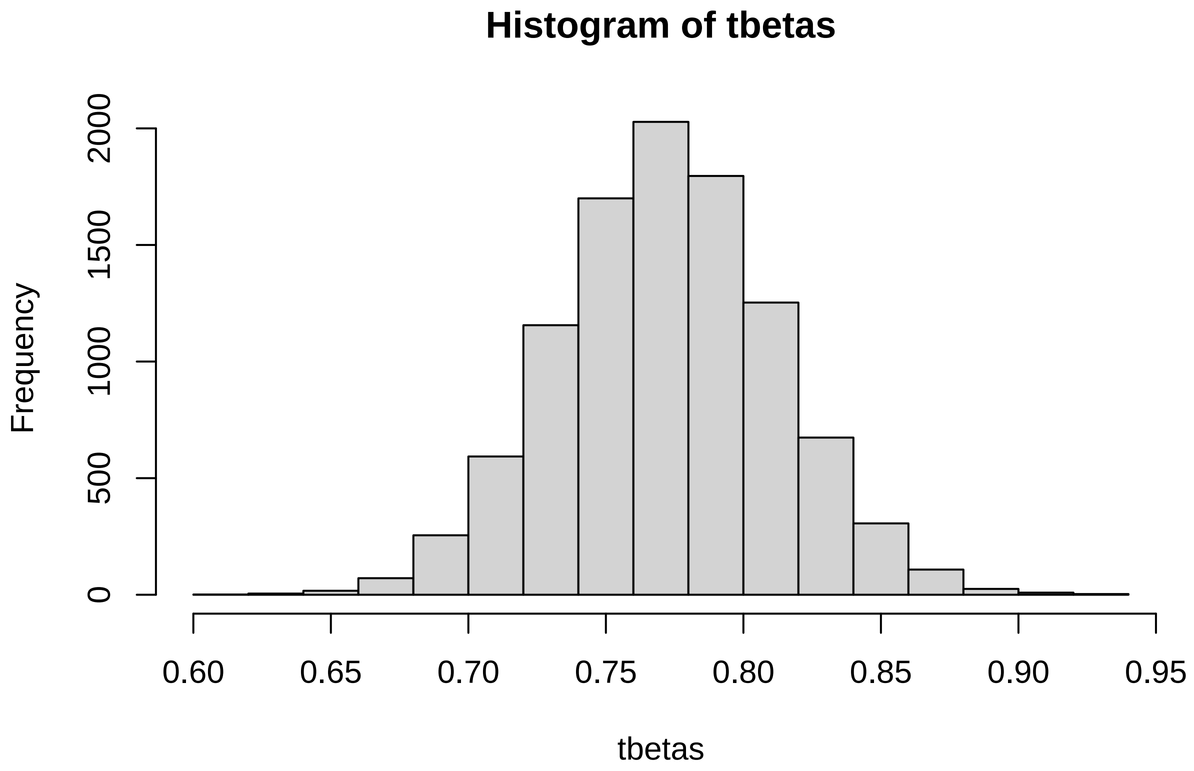
```
N = 1e4
alpha = 140
beta = 42

# Box Muller for sampling normals
box_muller = function(N){
  M = ceiling(N/2)
  thetas = runif(M)*2*pi
  Us = runif(M)
  Ws = -2*log(1-Us)
  X1 = sqrt(Ws)*cos(thetas)
  X2 = sqrt(Ws)*sin(thetas)
  c(X1 , X2)[1:N]
}
# Check how well it approximates a normal dist
plot(sort(rnorm(1000)) , sort(box_muller(1000)))
```



```r
# Normal Envelope for Beta
betaSams = function(n , a, b){
  A = a - 1
  B = b - 1
  xhat = A/(A+B)
  H = -1*(A^(1/3) + B^(1/3))^3
  Zs = box_muller(n) # substitute Normal sampler via Box Muller
  Xus = xhat + Zs*(sqrt(-1/H))
  hxhat = A*log(xhat)+ B*log(1-xhat)
  c = gamma(a)*gamma(b)/gamma(a+b)
  k = exp(hxhat)*sqrt(-2*pi/H)
  # Any negative numbers are already not worth considering
  Xus = Xus[Xus >= 0]
  # Initiate just as many random unifs (0,1)
  Us = runif(length(Xus))
  # Acceptance-Rejection step
  Xus[Us*exp(hxhat)*exp(H/2)*(Xus - xhat)^2 <= (Xus^A)*((1-Xus)^B)]
}

# Get the accepted samples
x = betaSams(N , alpha , beta)
tbetas = x[!is.na(x)]

# Check how well it approximates a Beta dist
plot(sort(rbeta(length(tbetas) , alpha , beta)) , sort(tbetas))
```


sort(tbetas)
0.65
0.75
0.85
0.65
0.70
0.75
0.80
0.85
sort(rbeta(length(tbetas), alpha, beta))

```
hist(tbetas)
```



Now that we have a sample of Beta-distributed random variables, let us sample a binomial conditional on this Beta. The easiest way to do this would be by generating n i.i.d. Bernoulli random quantities.

```
# Variables
trials = 65

# Bernoulli from Uniform
binarify = function(num , theta){
  # theta is the probability of success
  # though not enforced by the function, we will use this in a context where
  # theta is a probability and num is a random draw from a U(0,1) distribution
  if (num >= theta){
    return(0)
  }
  else{
    return(1)
  }
}

# Binomial Sampler - Count successes
binomial = function(trials , theta){
  # trials is number of bernoulli experiments
  # theta is the probability of a success
  # Generate 'trials' number of samples from U(0,1)
  Us = runif(trials)
  # create an empty vec - this will store the results of our 'trials' number
  # of Bernoulli experiments
  bin_vec = c()
  for (i in seq_along(Us)){
    # perform bernoulli experiment
    val = sapply(Us[i] , binarify , theta)
    # store results
    bin_vec = c(bin_vec , val)
  }
  # return the number of successes
  return(sum(bin_vec))
}

binomial_samples_gen = function(N , trials , thetas){
  # Run the binomial sampler N times. See distribution
  ySamples = c()
  for (i in 1:N){
    ySamples = c(ySamples , binomial(trials , sample(thetas , 1)))
  }
  return(ySamples)
}

# Generate
ySamples = binomial_samples_gen(N , trials , tbetas)

# Monte Carlo estimate
phat = sum(ySamples>50)/N
# Monte Carlo error
MCerror = sqrt(phat*(1-phat)/N)

print(paste('The MC estimate is', round(phat, 6), 'and the MC error is' ,round(MCerror , 6)))
```

```
## [1] "The MC estimate is 0.4839 and the MC error is 0.004997"
```

```
print(paste('(',round(phat-1.96*MCerror , 6), ',',round(phat+1.96*MCerror , 6),') is a 95% C.I'))
```

```
## [1] "( 0.474105 , 0.493695 ) is a 95% C.I"
```