

# AdvStDaAn, Worksheet, Week 9

Michael Lappert

28 April, 2022

## Contents

Task 1 . . . . .	1
Task 2 . . . . .	4

## Task 1

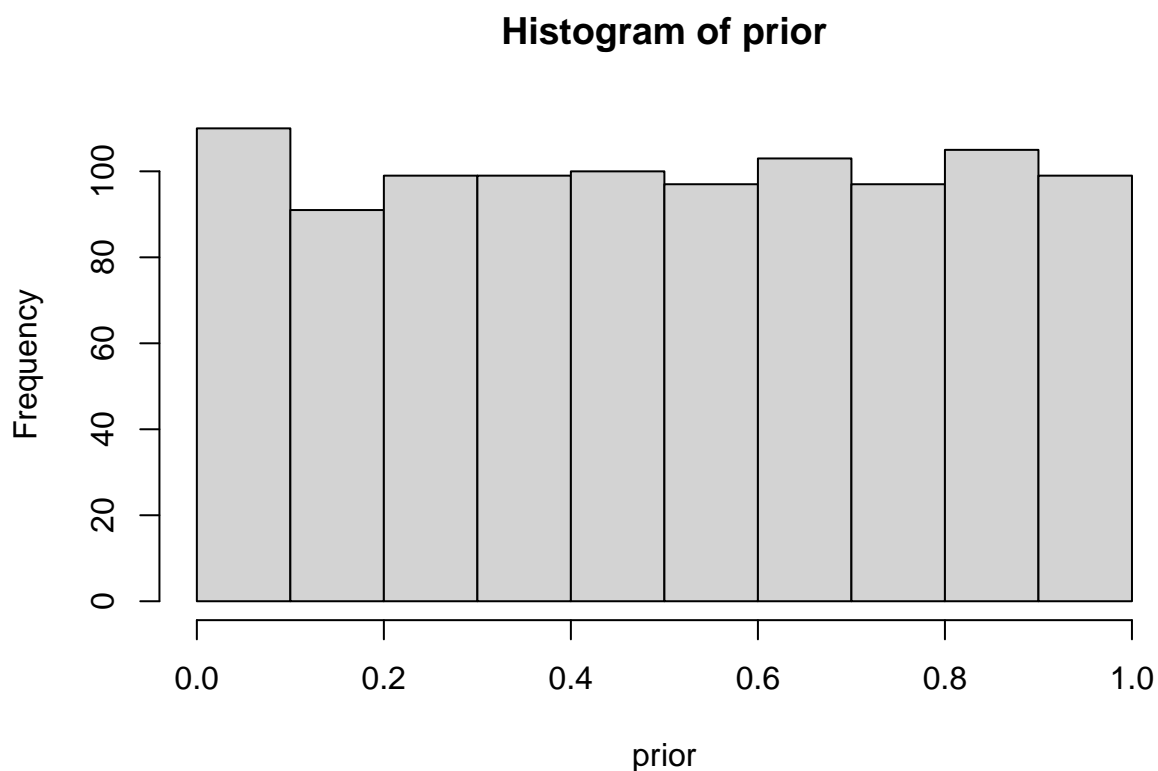
Problem statement: Swedish Fish Incorporated is delivering fish by mail order. They are now trying to enter the Swiss market. The marketing department has done a pilot study with method A: Sending a mail with a colorful brochure that invites people to sign up for a one year salmon subscription. The marketing department sent out 16 mails. Six out of 16 recipients signed up.

Build a Bayesian model that answers the question: What would the rate of sign-up be if method A was used on a larger number of people? Consider you send a mail to 20 people and 8 recipients signed up. Assume that you know nothing about the sign-up rate apriori, i.e. choose a unifom prior.

---

Simulate  $n$  random draws from the prior:

```
n = 1000
prior = runif(n) # Here we sample n draws from the prior
par(mfrow=c(1,1))
hist(prior) # It's always good to eyeball the prior to make sure it looks ok
```



Define the generative model:

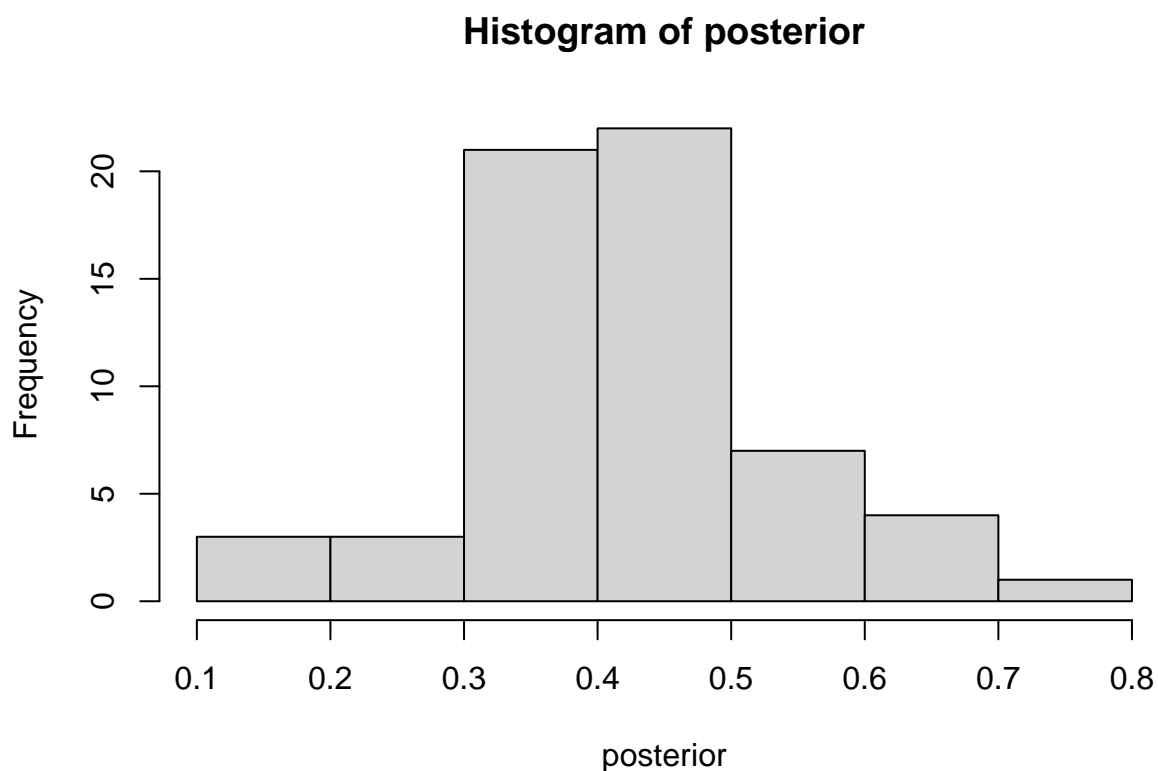
```
generativemodel = function(theta) {  
  rbinom(1, 20, theta)  
}
```

Simulate and store data using parameters from the prior and the generative model:

```
simdata = rep(NA, n)  
for(i in 1:n) {  
  simdata[i] = generativemodel( prior[i] )  
}
```

Filter out all draws that do not match the data:

```
posterior = prior[simdata == 8]  
hist(posterior)
```



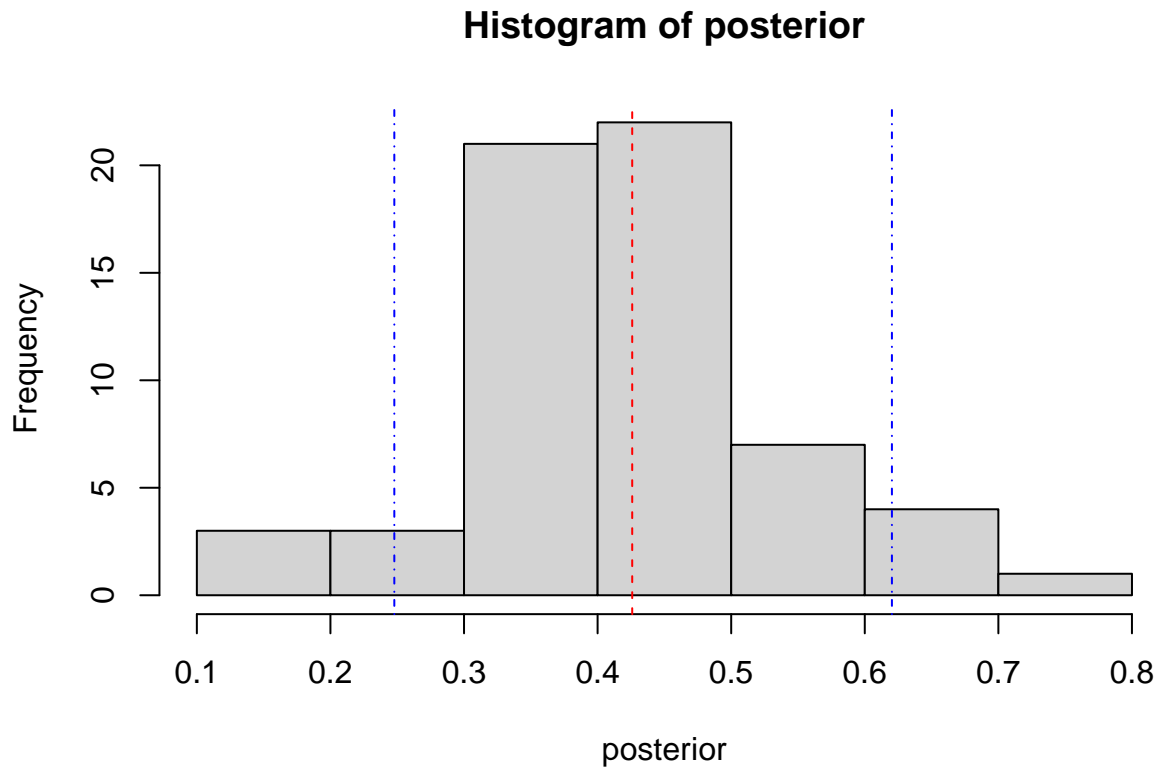
```
# Are there enough draws left after the filtering?  
length(posterior)
```

```
## [1] 61
```

There are no rules about the minimum draws left after filtering, but you probably want to aim for >1000 draws.

Now, summarize the posterior (posterior mean, 90% credible interval):

```
hist(posterior)  
abline(v = mean(posterior), col = "red", lty = 2)  
abline(v = quantile(posterior,c(.05,.95)), col = 'blue', lty = 4)
```



## Task 2

Problem statement: What's the probability that method A is better than telemarketing?

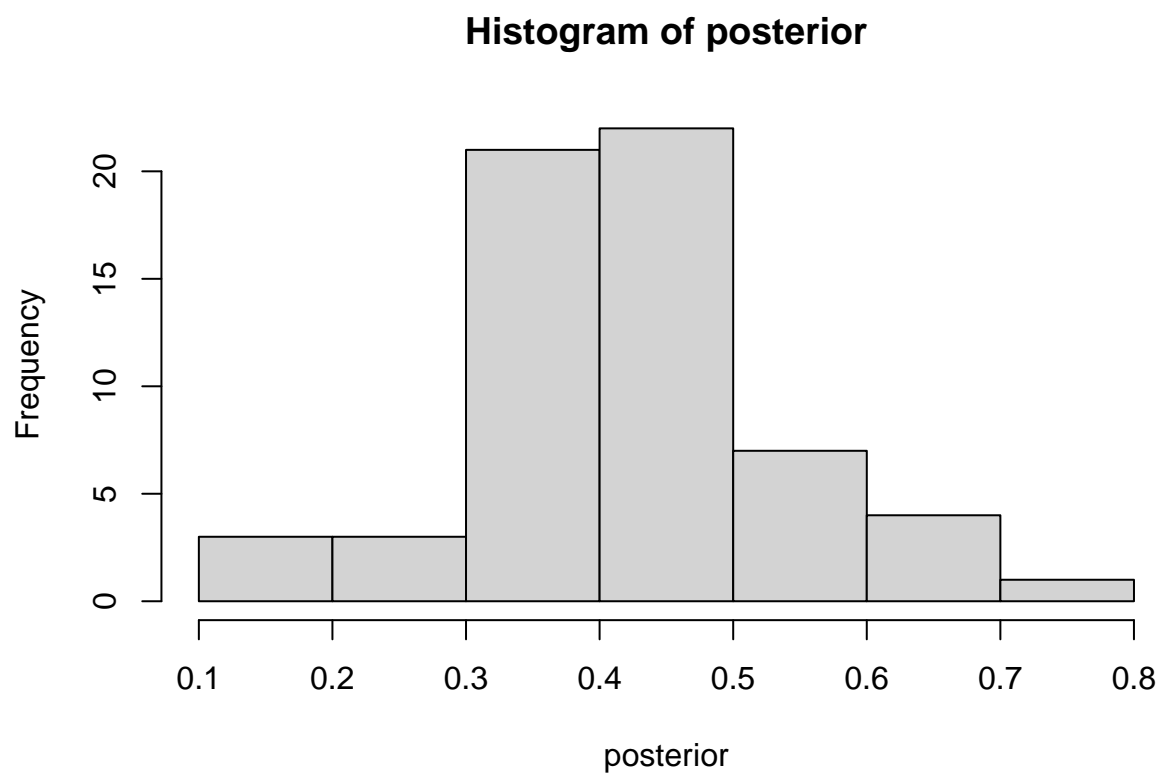
Marketing just told us that the rate of sign-up would be 20% if salmon subscribers were snared by a telemarketing campaign instead (to us it's very unclear where marketing got this very precise number from). So given the model and the data that we developed in the last question, what's the probability that method A has a higher rate of sign-up than telemarketing?

Hint 1: If you have a vector of samples representing a probability distribution, which you should have from the last question, calculating the amount of probability above a certain value is done by simply counting the number of samples above that value and dividing by the total number of samples.

Hint 2: The answer to this question is a one-liner.

---

```
hist(posterior)
```



```
length(posterior)
```

```
## [1] 61
```

```
length(posterior[posterior > 0.2])
```

```
## [1] 58
```

```
sum(posterior[posterior > 0.2]) / length(posterior)
```

```
## [1] 0.4168099
```

So the probability that method A is better than telemarketing is 41.7%.