

This is obtained from the work of Rasmus Bååth: http://sumsar.net/files/posts/2017-bayesian-tutorial-exercises/modeling_exercise2.html

Exercise 1: Bayesian A testing for Swedish Fish Incorporated (B comes later)

Swedish Fish Incorporated is the largest Swedish company delivering fish by mail order. They are now trying to get into the lucrative Danish market by selling one year Salmon subscriptions. The marketing department have done a pilot study and tried the following marketing 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 of type A. Six Danes that received a mail signed up for one year of salmon and marketing now wants to know, how good is method A?

At the bottom of this document you'll find a solution. But try yourself first!

Question I) 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?

Hint 1: The answer is not a single number but a distribution over probable rates of sign-up.

Hint 2: As part of your generative model you'll want to use the binomial distribution, which you can sample from in R using the `rbinom(n, size, prob)`. The binomial distribution simulates the following process `n` times: The number of "successes" when performing `size` trials, where the probability of "success" is `prob`.

Hint 3: A commonly used prior for the unknown probability of success in a binomial distribution is a uniform distribution from 0 to 1. You can draw from this distribution by running `runif(1, min = 0, max = 1)`

Hint 4: Here is a code scaffold that you can build upon.

```
# Number of random draws from the prior
n_draws <- 10000

prior <- ... # Here you sample n_draws draws from the prior
hist(prior) # It's always good to eyeball the prior to make sure it looks ok.

# Here you define the generative model
generative_model <- function(parameters) {
  ...
}

# Here you simulate data using the parameters from the prior and the
# generative model
sim_data <- rep(NA, n_draws)
for(i in 1:n_draws) {
  sim_data[i] <- generative_model(prior[i])
}

# Here you filter off all draws that do not match the data.
posterior <- prior[sim_data == observed_data]

hist(posterior) # Eyeball the posterior
```

```
length(posterior) # See that we got enough draws left after the filtering.
                  # There are no rules here, but you probably want to aim
                  # for >1000 draws.

# Now you can summarize the posterior, where a common summary is to take the mean
# or the median posterior, and perhaps a 95% quantile interval.
median(posterior)
quantile(posterior, c(0.025, 0.975))
```

Question II) What's the probability that method A is better than telemarketing?

So 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.

Question III) If method A was used on 100 people what would be number of sign-ups?

Hint 1: The answer is again not a single number but a distribution over probable number of sign-ups.

Hint 2: As before, the binomial distribution is a good candidate for how many people that sign up out of the 100 possible.

Hint 3: Make sure you don't "throw away" uncertainty, for example by using a summary of the posterior distribution calculated in the first question. Use the full original posterior sample!

Hint 4: The general pattern when calculating "derivatives" of posterior samples is to go through the values one-by-one, and perform a transformation (say, plugging in the value in a binomial distribution), and collect the new values in a vector.

Solutions (but this can be done in many ways)

###Question I

```
n_draw <- 10000

# Defining and drawing from the prior distribution
prior_rate <- runif(n_draw, 0, 1)

# Defining the generative model
gen_model <- function(rate) {
  subscribers <- rbinom(1, size = 16, prob = rate)
  subscribers
}

# Simulating the data
subscribers <- rep(NA, n_draw)
```

```

for(i in 1:n_draw) {
  subscribers[i] <- gen_model(prior_rate[i])
}

# Filtering out those parameter values that didn't result in the
# data that we actually observed
post_rate <- prior_rate[subscribers == 6]

# Checking that there enough samples left
length(post_rate)

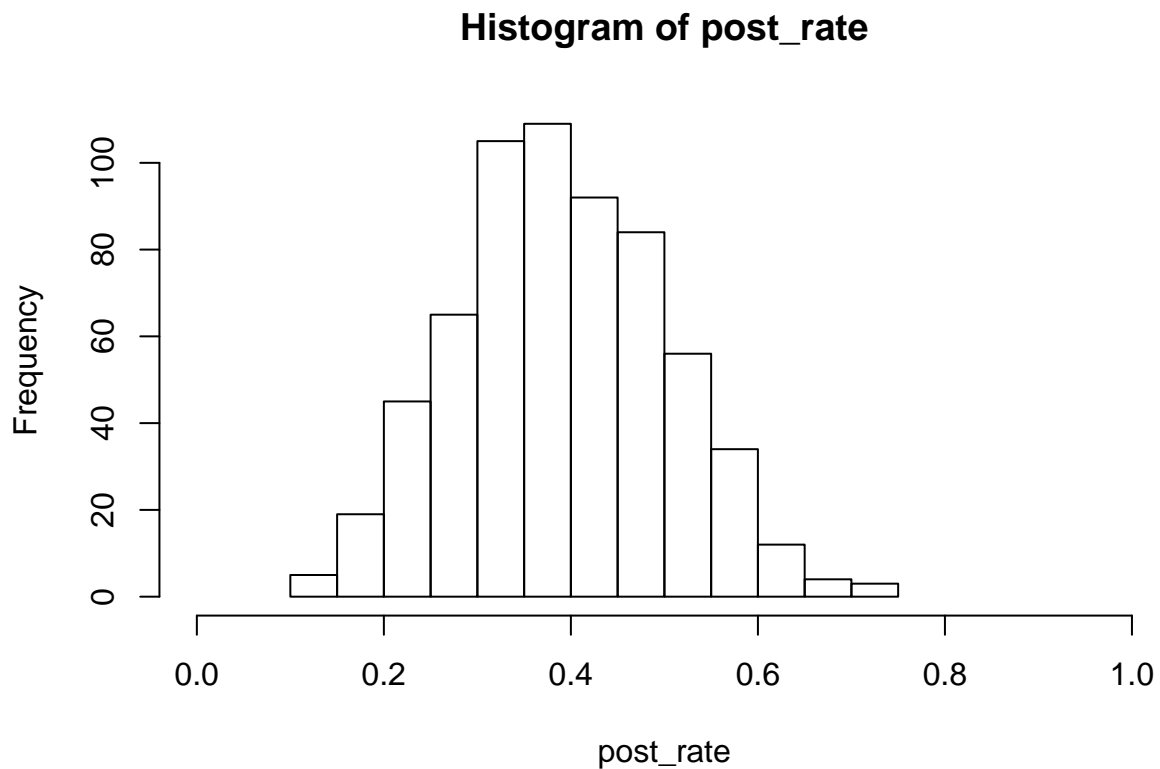
```

```
## [1] 633
```

```

# Plotting and summarising the posterior.
hist(post_rate, xlim = c(0, 1))

```



```
mean(post_rate)
```

```
## [1] 0.3906014
```

```
quantile(post_rate, c(0.025, 0.975))
```

```
##      2.5%      97.5%
## 0.1763407 0.6128725
```

Question II

```
sum(post_rate > 0.2) / length(post_rate)
```

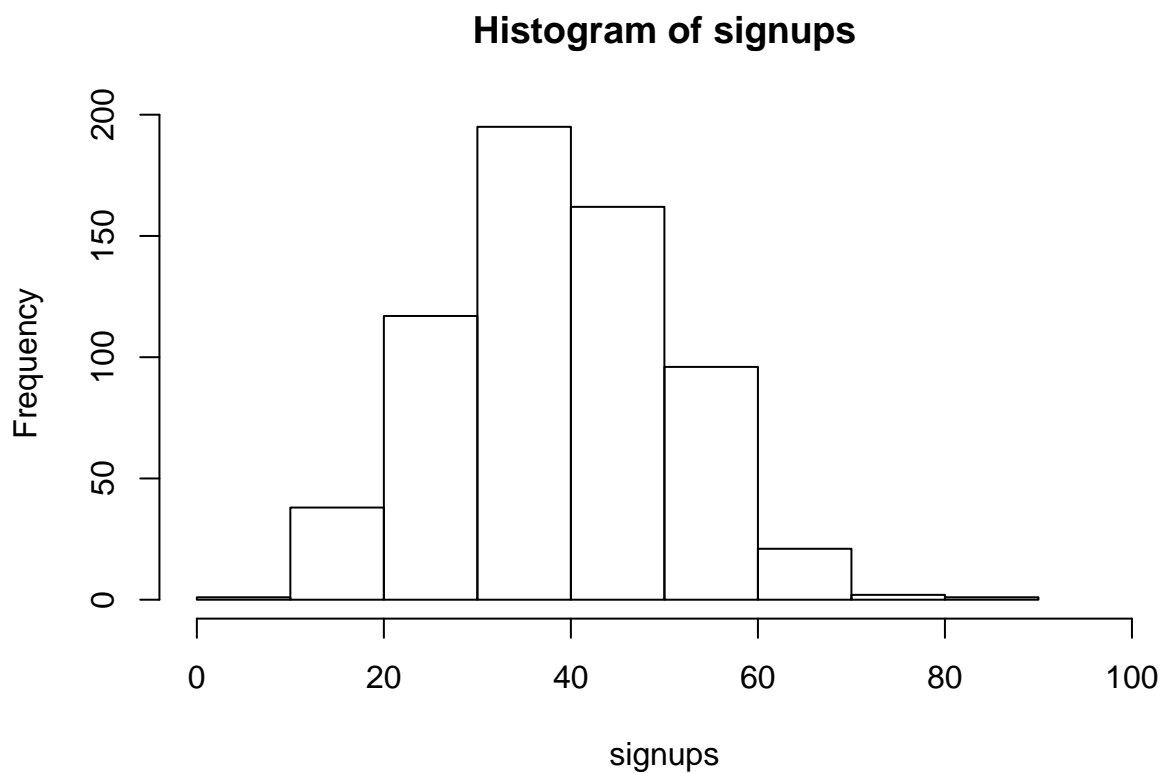
```
## [1] 0.9620853
```

Question III

```
# This can be done with a for loop
signups <- rep(NA, length(post_rate))
for(i in 1:length(post_rate)) {
  signups[i] <- rbinom(n = 1, size = 100, prob = post_rate[i])
}

# But since rbinom is vectorized we can simply write it like this:
signups <- rbinom(n = length(post_rate), size = 100, prob = post_rate)

hist(signups, xlim = c(0, 100))
```



```
quantile(signups, c(0.025, 0.975))
```

```
## 2.5% 97.5%
```

```
## 18 63
```

```
# So a decent guess is that it would be between 20 and 60 sign-ups.
```