Classical and Bayesian Statistics

Problems 10

Problem 10.1

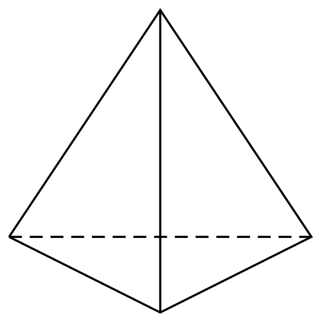
(∗∗) The result of a public-opinion poll for a presidential election in three provinces (A, B and C) are as follow: In province A the percentage of voters supporting the Liberal candidate is 50 %. In province B the percentage of voters supporting the Liberal candidate is 60 %. In province C the percentage of voters not supporting the Liberal candidate is 65 %.

The population of the three provinces is distributed as follows: A has 40 % of the total population in A, B and C, 25 % of the total population live in B and the remaining 35 % live in C.

Let us randomly choose a supporter of the Liberal candidate in province A or B or C. Determine the probability that such a voter was chosen from province B.

Problem 10.2

On a tetrahedral cube, each side is an equilateral triangle.



When you throw the cube, it lands face down and the other three faces are visible as a three-sided pyramid. The faces are labeled 1–4 points, and the value of the bottom face is labeled x.

(∗∗) a) Consider the following three mathematical descriptions of the probabilities of x:

- Model $A : P(x) = 1/4$
- Model $B : P(x) = x/10$
- Model $C : P(x) = 12/(25x)$

For each model, determine the value of $P(x)$ for each value of x. Describe in words what kind of "unfairness" is expressed by each model.

(∗∗) b) Suppose we have a) presented tetrahedral cube, along with the three candidate models for the probabilities of the cube.

Suppose that initially we are not sure what to make of the cube. On the one hand, the die could be fair, with each side landing with the same probability.

On the other hand, the cube could be skewed, so that the faces with more points are more likely to land on the ground (because the points are created by embedding heavy jewels in the cube, so the sides with more points are more likely to land on the ground).

Or, that more dots on a side make it less likely to land at the bottom (because the dots may be made of springy rubber or protrude from the surface).

Initially, then, our beliefs about the three models as

$$p(A) = p(B) = p(C) = \frac{1}{3}$$

can be described.

Now we roll the dice 100 times and find these results:

$$\text{№ } 1 = 25, \quad \text{№ } 2 = 25, \quad \text{№ } 3 = 25, \quad \text{№ } 4 = 25$$

Do these data change our beliefs about the models? Which model now seems most likely?

Suppose that when we rolled the dice 100 times, we found these results:

$$\text{№ } 1 = 48, \quad \text{№ } 2 = 24, \quad \text{№ } = 16, \quad \text{№ } 4 = 12.$$

Which model now seems most likely?

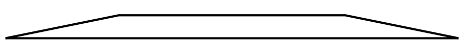
Problem 10.3

(∗∗) In the Malaria example, we examined a positive test twice. Calculate the probability that we test positive first, then negative.

What does it look like if we reverse the order?

Problem 10.4

(∗∗) Consider a coin that has the asymmetric cross section:



We do not know whether H or T is up. What is a reasonable choice for the prior probabilities?

Problem 10.5

School children were asked about their favorite foods. Of the total sample, 20% were first graders, 20% were sixth graders, and 60% were eleventh graders. The following table shows for each grade the percentage of respondents who selected each of the three foods as their favorite:

	Ice cream	Fruit	French fries
1st graders	0.3	0.6	0.1
6th graders	0.6	0.3	0.1
11th graders	0.3	0.1	0.6

From this information, construct a table of joint probabilities of grade and favorite food. Also tell whether or not the grade and favorite food are independent and how you determined the answer.

Note: You will get $P(\text{class})$ and $P(\text{food} \mid \text{class})$. You need to determine $P(\text{food} \cap \text{class})$.

Problem 10.6

This task refers to lecture notes where one coin was tossed once and H occurred. We determined the posterior distribution.

$(***)$ a) Now you toss the coin again and H occurs. What is the new posterior distribution?

$(***)$ b) Instead of an H a T occurs. What is the new posterior distribution? Compare it to the original prior distribution. What can you tell?

Classical and Bayesian Statistic

Sample solution for Problems 10

Solution 10.1

Designations:

- L: Voter supports liberal candidate
- $\overline{L}$: Voter does not support Liberal candidate
- A: Voter is from province A
- B: Voter is from province B
- C: Voter comes from province C

Known (from task):

$$P(L \mid A) = 0.5, \quad P(L \mid B) = 0.6, \quad P(\overline{L} \mid C) = 0.65$$

and

$$P(A) = 0.4, \quad P(B) = 0.25, \quad P(C) = 0.35$$

We are looking for $P(B \mid L)$. According to Bayes' theorem and the marginal probability:

$$\begin{aligned}
P(B \mid L) &= \frac{P(L \mid B)P(B)}{P(L)} \\
&= \frac{P(L \mid B)}{P(L \mid A)P(A) + P(L \mid B)P(B) + P(L \mid C)P(C)} \\
&= \frac{0.6 \cdot 0.25}{0.5 \cdot 0.4 + 0.6 \cdot 0.25 + (1 - 0.65) \cdot 0.35} \\
&= 0.3175
\end{aligned}$$

About 32 % of the voters who support the Liberal candidate are from province B.

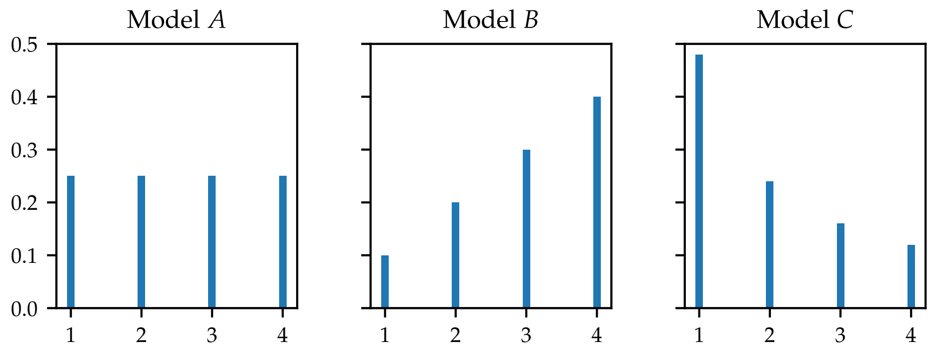
Solution 10.2

a) We can list the probabilities as a table:

	x	1	2	3	4
Model A :	$P(X = x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Model B :	$P(X = x)$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{4}{10}$
Model C :	$P(X = x)$	$\frac{12}{25}$	$\frac{12}{50}$	$\frac{12}{75}$	$\frac{12}{100}$
		$\frac{48}{100}$	$\frac{24}{100}$	$\frac{16}{100}$	$\frac{12}{100}$

Note that all the probabilities in the rows add up to 1.

We still sketch these probabilities:



For model A, all sides have equal probability of being at the bottom.

For model B, side 1 has the smallest probability and this increases linearly with number.

For model C, side 1 has the largest probability and it decreases non-linearly (hyperbolically) with the number. So here the probability $P(1)$ is weighted much more than the other probabilities.

b) Before we roll the dice with the tetrahedron we have no idea about the throw probabilities of the individual sides. In a) we considered three possible models that we can consider as prior probabilities for the sides.

However, we do not know which model best "fits" and before we make any experiments, we make the assumption that all models are equally likely, i.e.

$$p(A) = p(B) = p(C) = \frac{1}{3}$$

which we can consider as the prior probability for the models.

Note that we have two prior probabilities:

- One for the *roll* probability.
- One for the *model* probability.

Now we make a trial of 100 throws and get

$$\text{no. } 1 = 25, \quad \text{no. } 2 = 25, \quad \text{no. } 3 = 25, \quad \text{no. } 4 = 25$$

that is, all sides exit with equal frequency on *this* trial. So we will change our assumption about the probability of the models and assign a larger probability to the model A. However, this probability is not 1, since on another trial the litter numbers will be different.

In the same way, we change the assumption

$$\text{no. } 1 = 48, \quad \text{no. } 2 = 24, \quad \text{no. } = 16, \quad \text{No. } 4 = 12.$$

and assign a greater probability to model C and, most importantly, less to model B, since this model assigns the smallest probability to the number 1.

Solution 10.3

As in theory, we first use the formula for a positive test

$$P(M \mid +) = \frac{P(+ \mid M) \cdot P(M)}{P(+)} = \frac{P(+ \mid M) \cdot P(M)}{P(+ \mid M) \cdot P(M) + P(+ \mid \overline{M}) \cdot P(\overline{M})}$$

and then for the negative test

$$P(M \mid -) = \frac{P(- \mid M) \cdot P(M)}{P(-)} = \frac{P(- \mid M) \cdot P(M)}{P(- \mid M) \cdot P(M) + P(- \mid \overline{M}) \cdot P(\overline{M})}$$

Where $P(M)$ equals $P(M \mid +)$ from the first trial.

Since only the prior probability changes, we do this with Python:

```
# Sensitivity P(+|M)
sens = 0.917

# Specificity P(-|M across)
spec = 0.935

# Formula for positive test
post_pos <- function(prior){
    post = (sens*prior) / (sens*prior + (1-spec)*(1-prior))
    return(post)
}

# Formula for negative test
post_neg <- function(prior){
    post = (1-sens)*prior / ((1-sens)*prior + spec*(1-prior))
    return(post)
}
```

If the test is first positive then negative, we get:

```
post_neg(post_pos(0.03))

[1] 0.03728794
```

In the reverse case:

```
post_pos(post_neg(0.03))

[1] 0.03728794
```

So the order does not matter here.

What happens if there are two negative tests?

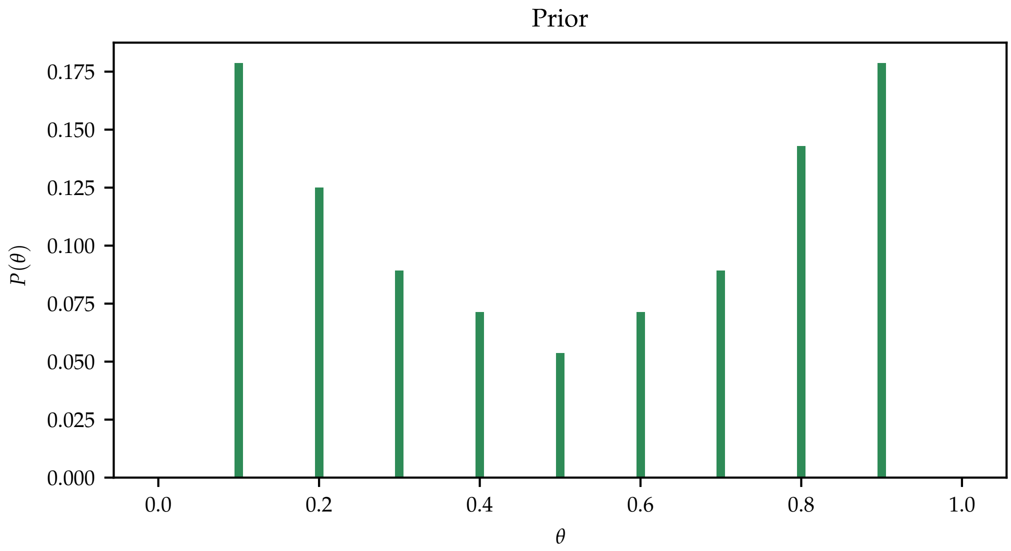
```
post_neg(post_neg(0.03))

[1] 0.0002436557
```

Then we are already pretty certain that we don't have malaria.

Solution 10.4

We need to consider the situation where T or H get a relatively large probability, since we don't know which side H and T are on.

Thus, the prior probability might look like the following:



Solution 10.5

First, a note: the table gives *not* $P(\text{food} \cap \text{class})$. The reason is that the sum of the probabilities in the rows adds up to 1, and in the columns it generally does not.

Thus, the first value in the upper left is the probability that ice cream is the favorite food for a first grader:

$$P(G \mid 1) = 0.3$$

According to the definition of the certain probability

$$P(G \cap 1) = P(G \mid 1) \cdot P(1) = 0.3 \cdot 0.2 = 0.06$$

This is the value in the upper left corner of the following table. The other values are calculated analogously. You have to multiply the 1st line with 0.2, the 2nd with 0.2 and the third with 0.6

	Ice Cream	Fruit	French Fries	
1st graders	0.06	0.12	0.02	0.2
6th graders	0.12	0.06	0.02	0.2
11th graders	0.18	0.06	0.36	0.6
	0.36	0.24	0.4	1

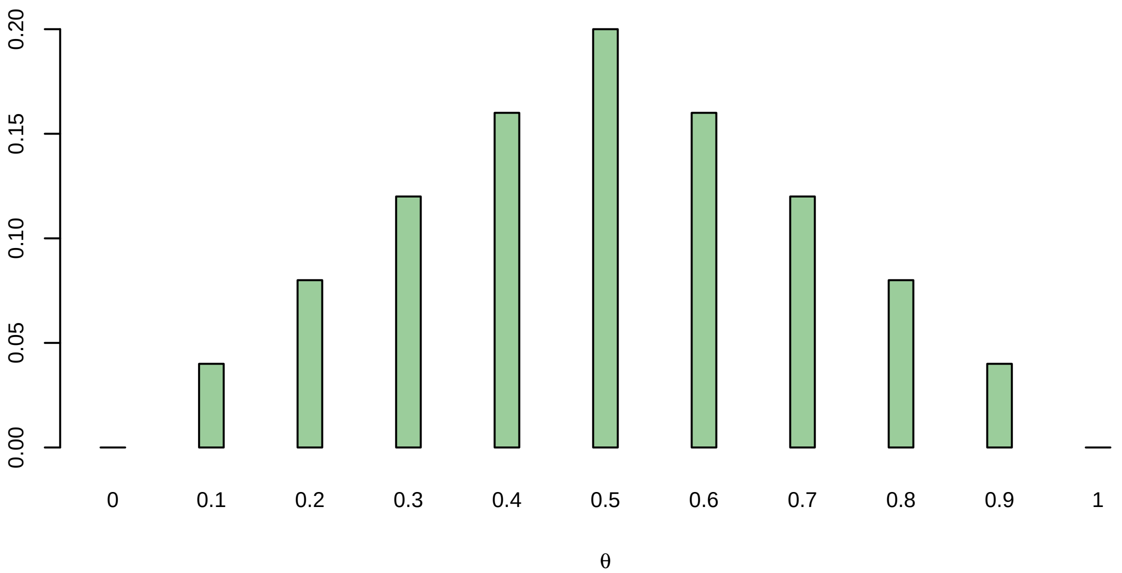
Solution 10.6

We used the prior distribution

```
y <- c(0, 1, 2, 3, 4, 5, 4, 3, 2, 1, 0)
prior <- y / sum(y)
prior

[1] 0.00 0.04 0.08 0.12 0.16 0.20 0.16 0.12 0.08 0.04 0.00
```

Plot:

```
par(mar = c(4,2,1,0))
library(latex2exp)
steps = seq(0, 1, 0.1)
barplot(prior, space = 3, names = steps, col = "darkseagreen3",
        xlab = TeX(r"($\theta$)"))
```

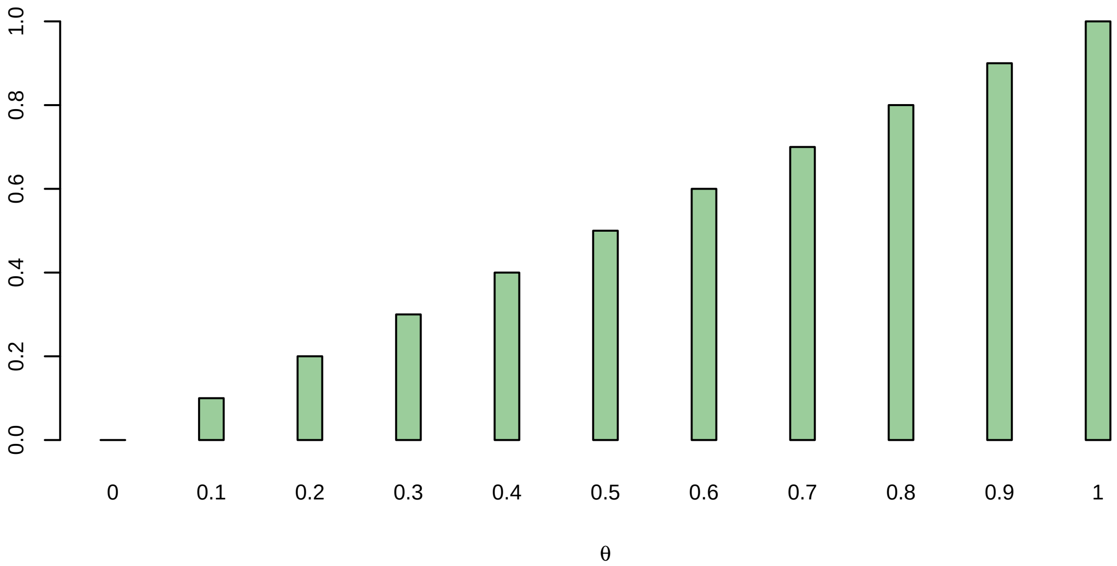


and the likelihood function

```
like <- seq(0, 1, 0.1)
like

[1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

Plot:

```
barplot(like, space=3, names=steps, col="darkseagreen3", xlab=TeX(r"($\theta$)"))
```

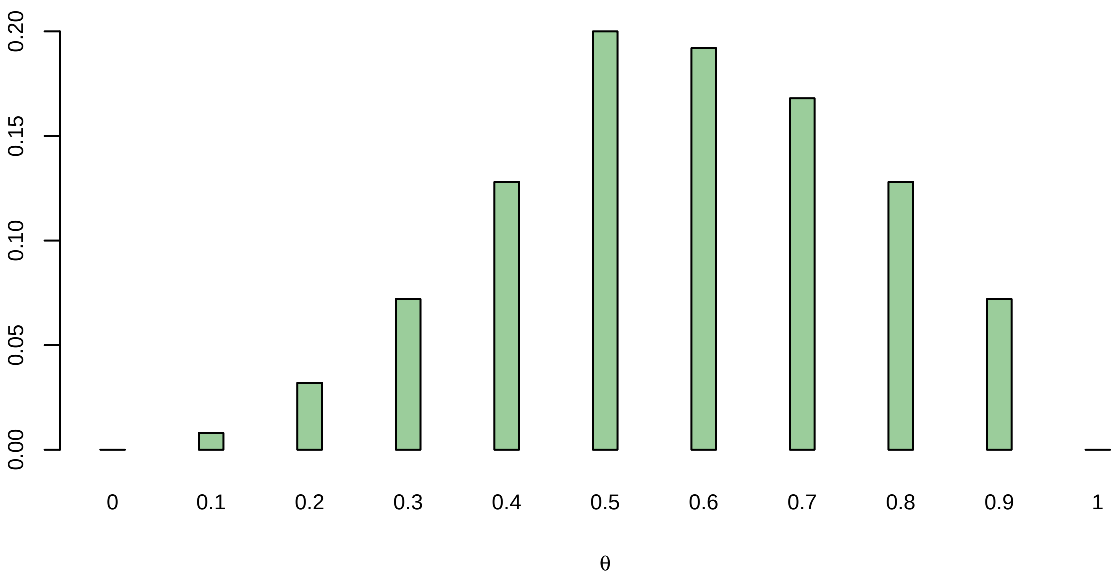


and obtained the posterior distribution

```
margin = sum(prior*like)
post = prior * like / margin
post

[1] 0.000 0.008 0.032 0.072 0.128 0.200 0.192 0.168 0.128 0.072 0.000
```

Plot:

```
barplot(post, space=3, names=steps, col="darkseagreen3", xlab=TeX(r"($\theta$)"))
```

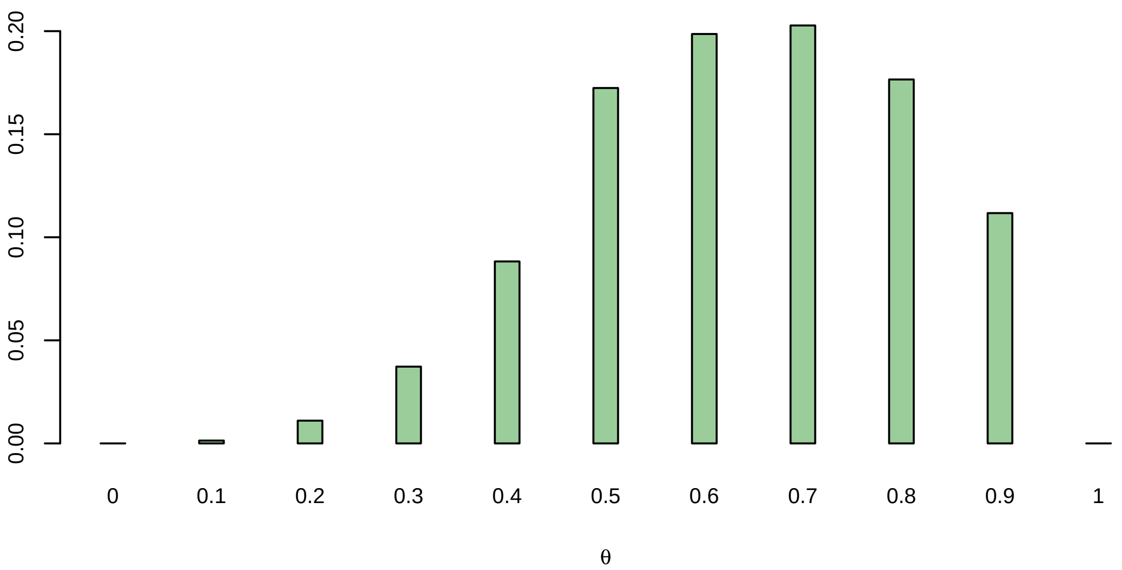


a) The likelihood function stays the same but we use the old posterior as new prior distribution.

```
prior_H = post
margin = sum(prior_H*like)
post_H = prior_H * like / margin
round(post_H, 3)

[1] 0.000 0.001 0.011 0.037 0.088 0.172 0.199 0.203 0.177 0.112 0.000
```

Plot:

```
barplot(post_H, space=3, names=steps, col="darkseagreen3", xlab=TeX(r"($\theta$)"))
```



As expected the distribution moves further to the right because with an additional H, we can assume that the probability θ is greater than 0.5. If $\theta = 0.5$ we would expect that the H and T are evenly distributed.

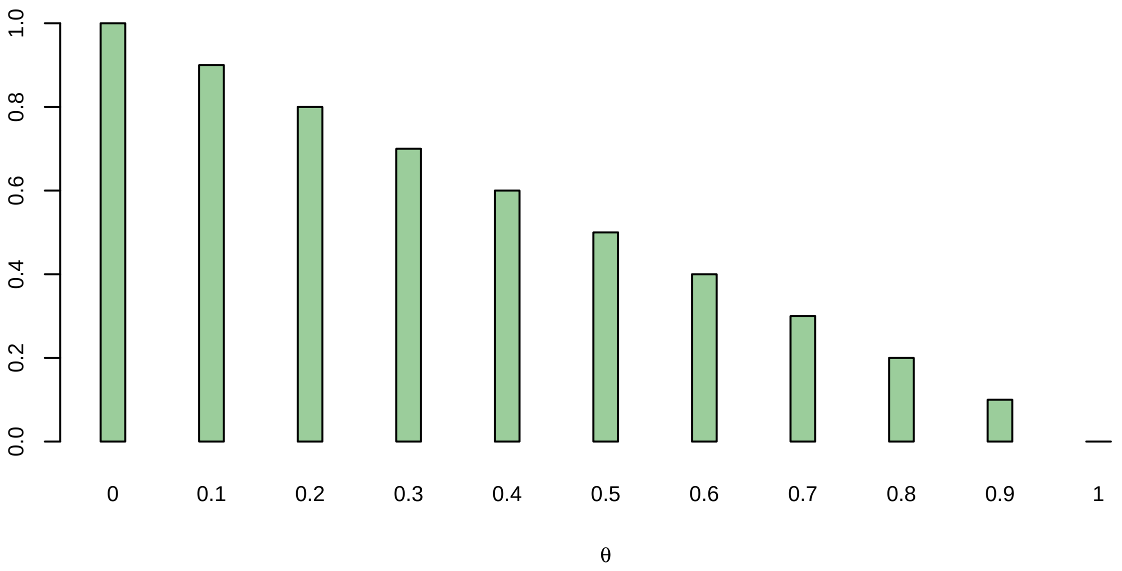
However, it is entirely possible that we toss two H in a row even if $\theta = 0.5$.

b) If we toss a T instead, we have to change the likelihood function to `1-like`.

```
like <- seq(0, 1, 0.1)
like
```

Plot:

```
like_T = 1- like
barplot(like_T, space=3, names=steps, col="darkseagreen3", xlab=TeX(r"($\theta$)"))
```

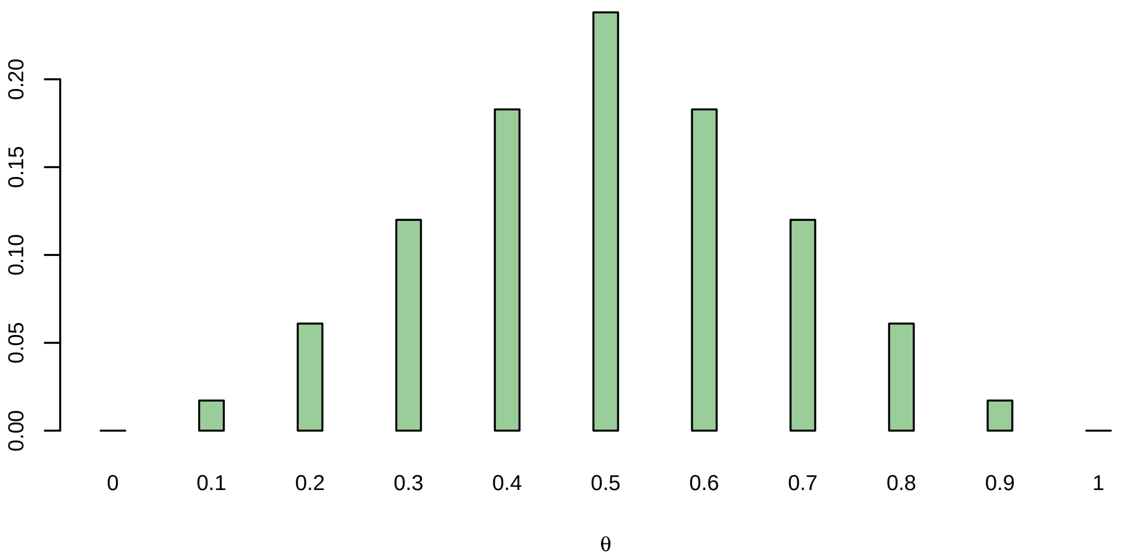


For the new prior distribution we use the old posterior distribution

```
prior_T = post
margin = sum(prior_T*like_T)
post_T = prior_T * like_T / margin
round(post_T, 3)

[1] 0.000 0.017 0.061 0.120 0.183 0.238 0.183 0.120 0.061 0.017 0.000
```

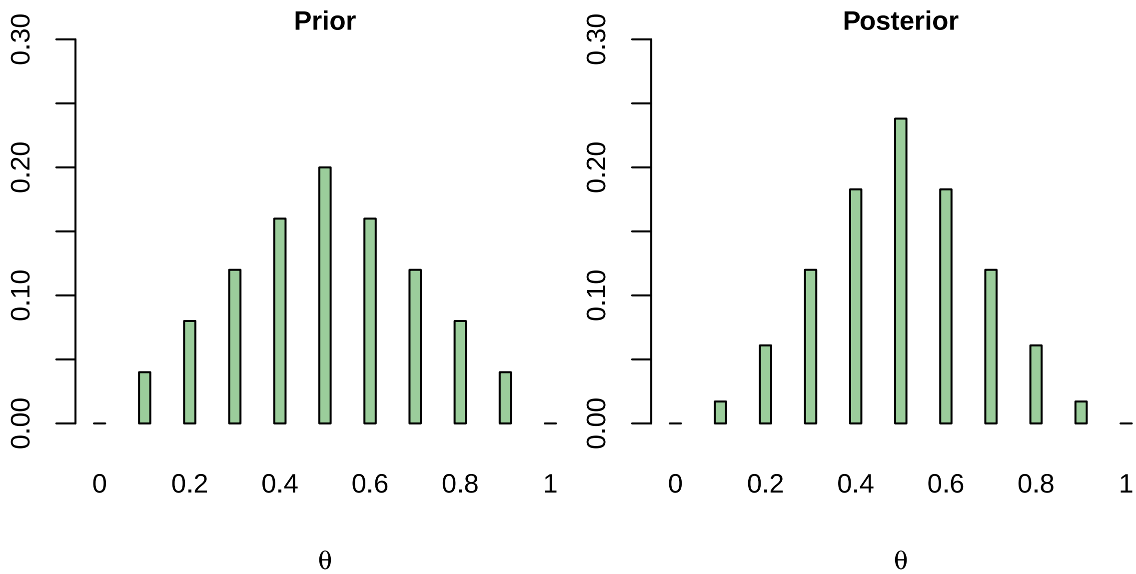
Plot:

```
barplot(post_T, space=3, names=steps, col="darkseagreen3", xlab=TeX(r"($\theta$)"))
```



It *appears* that we are back to square one, namely the original prior distribution. However, this is *not* the case. If we plot the original prior distribution and the posterior distribution we recognize differences.

```r
par(mfrow=c(1,2))
barplot(prior, space=3, names=steps, col="darkseagreen3", xlab=TeX(r"($\theta$)"),ylim=c(0,0.3), main="Prior")
barplot(post_T, space=3, names=steps, col="darkseagreen3", xlab=TeX(r"($\theta$)"),ylim=c(0,0.3),main="Posterior")
```



Both distributions are symmetrical, but the probabilities have reallocated to middle. Why is this the case? If we toss a H and then a T, this is more evidence that the coin is fair, namely $\theta = 0.5$. Hence the θ's in the middle get more weight compared to the prior distribution.