Stats 102C - Lecture 5-3: The Metropolis Algorithm

Miles Chen

Week 5 Friday

Last time

On Wednesday we saw that if a Markov chain with transition matrix $\mathbb{P}$ is ergodic, it will eventually reach its stationary distribution $\boldsymbol{\pi}$

Once it converges to the stationary distribution, the values that it generates from that point on will come from the stationary distribution.

Today we introduce the Metropolis Algorithm which is a way to design the transition probabilities of a Markov chain so that the stationary distribution of the Markov chain will be equal to a desired distribution.

Section 1

Metropolis Algorithm - Island Hopping Example

A silly example

The following example comes from the textbook *Doing Bayesian Data Analysis* Chapter 7.

Imagine a nation consisting of islands in a chain. A politician travels from island to island trying to stay in the public eye. At the end of each day, he has to decide if he should:

- stay at the current island
- move to the adjacent island to the west
- move to the adjacent island to the east

His goal is to visit all the islands proportionally to their relative population so he spends the most time at the most populated island and less time at the less populated islands.

Unfortunately, he does not know the total population of the island chain. However, he is able to find out the population of the island he is currently on and the population of the adjacent islands.

The politician's algorithm

- Each morning the politician flips a coin:
  - if it lands heads he proposes the island on the west.
  - if it lands tails he proposes the island on the east.
- He checks the population of the proposed island and the population of the current island.
  - if the proposed island has a larger population than the current island, he definitely goes to the proposed island
  - if the proposed island has a smaller population than the current island, he goes to the island probabilistically. That is his probability of moving that that island is:
    - ★ $p_{move} = \frac{p_{proposed}}{p_{current}}$
    - ★ To do this, he can select a random uniform value. If $u < p_{move}$ he moves to the proposed island. If not, he stays at the current island.

The system works! In the long run, the probability the politician is on any one of the islands exactly matches the relative population of the island!

A slighlty more concrete example

Let's consider the follow example with concrete numbers.

Let's say our island chain has 7 islands. The islands are numbered by θ From $\theta = 1$ in the west to $\theta = 7$ in the east.

The relative populations of each island are proportional their island numbers. So island $\theta = 1$ has a population that is 1/7th the size of island $\theta = 7$'s population.

The target distribution, which want to be our stationary distribution is:

$$\boldsymbol{\pi} = \left\{ \frac{1}{28}, \frac{2}{28}, \frac{3}{28}, \frac{4}{28}, \frac{5}{28}, \frac{6}{28}, \frac{7}{28} \right\}$$

Starting at island 4 going to island 5

Let's say the politician starts at island 4.

What is the probability he moves to island 5?

He flips a coin and will propose island 5 if it lands tails. The probability he proposes island 5 is 0.5.

If he proposes island 5, he will move to island 5 with probability 1 because the population of island 5 is greater than the population of island 4.

Total probability of moving from island 4 to island 5 $= P_{45}$

$$P_{45} = \Pr(\text{proposing island 5}) \Pr(\text{moving to island 5} | \text{island 5 proposed}) = 0.5 \times 1 = 0.5$$

Starting at island 4 going to island 3

Let's say the politician starts at island 4.

What is the probability he moves to island 3?

He flips a coin and will propose island 3 if it lands heads. The probability he proposes island 3 is 0.5.

If he proposes island 3, he will move to island 3 with probability:

$$p_{move} = \frac{p_{proposed}}{p_{current}} = \frac{3/28}{4/28} = \frac{3}{4}$$

Island 3's population relative to island 4's population is 3/4.

Total probability of moving from island 4 to island 3 $= P_{43}$

$$P_{43} = \Pr(\text{proposing island 3}) \Pr(\text{moving to island 3}|\text{island 3 proposed}) = 0.5 \times \frac{3}{4} = 0.375$$

Starting at island 4 staying at island 4

Let's say the politician starts at island 4.

What is the probability he stays at island 4?

The politician stays at island 4 if he proposes an island but does not go to it.

$$
\begin{aligned}
P_{44} &= \Pr(\text{proposing island 3}) \Pr(\text{NOT moving to island 3}|\text{island 3 proposed}) + \\
&\quad \Pr(\text{proposing island 5}) \Pr(\text{NOT moving to island 5}|\text{island 5 proposed}) \\
&= 0.5(1 - 3/4) + 0.5(0) \\
&= 0.125
\end{aligned}
$$

Building our transition probability matrix

We can start to fill in the transition probability matrix with our findings

$$\mathbb{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} & P_{17} \\ P_{21} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} & P_{27} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} & P_{36} & P_{37} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} & P_{46} & P_{47} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} & P_{56} & P_{57} \\ P_{61} & P_{62} & P_{63} & P_{64} & P_{65} & P_{66} & P_{67} \\ P_{71} & P_{72} & P_{73} & P_{74} & P_{75} & P_{76} & P_{77} \end{bmatrix} = \begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0.375 & 0.125 & 0.5 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Starting at island 1

What if he starts at island 1?

- probability he goes to island 2:

$$P_{12} = \Pr(\text{proposing island 2}) \Pr(\text{moving to island 2}|\text{island 2 proposed}) = 0.5 \times 1 = 0.5$$

- probability he goes to island 0:
  - Island 0 does not exist. We can treat it like it has a population of 0

$$P_{10} = \Pr(\text{proposing island 0}) \Pr(\text{moving to island 0}|\text{island 0 proposed}) = 0.5 \times 0 = 0$$

- probability he stays at island 1:

$$\begin{aligned} P_{11} &= \Pr(\text{proposing island 0}) \Pr(\text{NOT moving to island 0}|\text{island 0 proposed}) + \\ &\quad \Pr(\text{proposing island 2}) \Pr(\text{NOT moving to island 2}|\text{island 2 proposed}) \\ &= 0.5(1) + 0.5(0) \\ &= 0.5 \end{aligned}$$

Add to our transition probability matrix

$$\mathbb{P} = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} & P_{17} \\ P_{21} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} & P_{27} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} & P_{36} & P_{37} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} & P_{46} & P_{47} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} & P_{56} & P_{57} \\ P_{61} & P_{62} & P_{63} & P_{64} & P_{65} & P_{66} & P_{67} \\ P_{71} & P_{72} & P_{73} & P_{74} & P_{75} & P_{76} & P_{77} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0.375 & 0.125 & 0.5 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix}$$

Starting at island 2

What if he starts at island 2?

- probability he goes to island 3:

$$P_{23} = \Pr(\text{proposing island 3})\Pr(\text{moving to island 3}|\text{island 3 proposed}) = 0.5 \times 1 = 0.5$$

- probability he goes to island 1:
  - $$p_{move} = \frac{p_{proposed}}{p_{current}} = \frac{1/28}{2/28} = \frac{1}{2}$$

$$P_{21} = \Pr(\text{proposing island 1})\Pr(\text{moving to island 1}|\text{island 1 proposed}) = 0.5 \times 0.5 = 0.25$$

- probability he stays at island 2:

$$\begin{aligned} P_{22} &= \Pr(\text{proposing island 1})\Pr(\text{NOT moving to island 1}|\text{island 1 proposed}) + \\ &\qquad \Pr(\text{proposing island 3})\Pr(\text{NOT moving to island 3}|\text{island 3 proposed}) \\ &= 0.5(0.5) + 0.5(0) \\ &= 0.25 \end{aligned}$$

Starting at island 7

What if he starts at island 7?

- probability he goes to island 8:
  - Island 8 does not exist. We can treat it like it has a population of 0

$$P_{78} = \Pr(\text{proposing island 8})\Pr(\text{moving to island 8}|\text{island 8 proposed}) = 0.5 \times 0 = 0$$

- probability he goes to island 6:
  - $$p_{move} = \frac{p_{proposed}}{p_{current}} = \frac{6/28}{7/28} = \frac{6}{7}$$

$$P_{76} = \Pr(\text{proposing island 6})\Pr(\text{moving to island 6}|\text{island 6 proposed}) = 0.5 \times \frac{6}{7} = \frac{3}{7} \approx 0.42857$$

- probability he stays at island 7:

$$\begin{aligned} P_{77} = {} & \Pr(\text{proposing island 8})\Pr(\text{NOT moving to island 8}|\text{island 8 proposed}) + \\ & \Pr(\text{proposing island 6})\Pr(\text{NOT moving to island 6}|\text{island 6 proposed}) \\ & = 0.5(1) + (0.5)\frac{1}{7} \\ & = \frac{8}{14} \approx 0.5714 \end{aligned}$$

Filling in the rest of the probability matrix

Following the rules, we can fill in the rest of the probability matrix.

$$\mathbb{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{6} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{8} & \frac{1}{8} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{2}{5} & \frac{1}{10} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{5}{12} & \frac{1}{12} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{3}{7} & \frac{4}{7} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.33\bar{3} & 0.166\bar{6} & 0.5 & & 0 & 0 \\ 0 & 0 & 0.375 & 0.125 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.1 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0.4166\bar{6} & 0.0833\bar{3} & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.42857 & .57143 \end{bmatrix}$$

The stationary distribution of this matrix

```r
P = rbind(c( 0.5,  0.5,    0,    0,     0,     0,    0),
          c(0.25, 0.25, 0.5,    0,     0,     0,    0),
          c(   0,  1/3, 1/6, 0.5,     0,     0,    0),
          c(   0,    0, 3/8, 1/8,   0.5,     0,    0),
          c(   0,    0,    0, 0.4,   0.1,   0.5,    0),
          c(   0,    0,    0,    0, 5/12, 1/12, 0.5),
          c(   0,    0,    0,    0,     0,  3/7, 4/7))

r_eigen <- eigen(t(P))$vectors[ ,1] # right eigenvector of P^T
stationary <- t(r_eigen) / sum(r_eigen) # normalize the transpose
stationary # the stationary distribution found by finding the left eigenvector
```

```
##            [,1]       [,2]      [,3]      [,4]      [,5]      [,6] [,7]
## [1,] 0.03571429 0.07142857 0.1071429 0.1428571 0.1785714 0.2142857 0.25
```

```r
stationary %*% P # stationary dist times P remains the same
```

```
##            [,1]       [,2]      [,3]      [,4]      [,5]      [,6] [,7]
## [1,] 0.03571429 0.07142857 0.1071429 0.1428571 0.1785714 0.2142857 0.25
```

The target distribution

Island populations

The target distribution is a distribution proportional to the island populations:

$$\pi = \left\{ \frac{1}{28}, \frac{2}{28}, \frac{3}{28}, \frac{4}{28}, \frac{5}{28}, \frac{6}{28}, \frac{7}{28} \right\}$$

```r
target <- matrix( (1:7)/28, nrow = 1)
all.equal(stationary, target)
```

```
## [1] TRUE
```

Our stationary distribution is equal to the target distribution.

Running the Markov Chain

The politician's algorithm is:

1. Flip a coin to propose the next island.
   - $X_{proposed} = X_t + 1$ with probability 0.5.
   - $X_{proposed} = X_t - 1$ with probability 0.5.
2. Calculate probability p_{move}
   - $p_{move} = \min\left\{1, \dfrac{p_{proposed}}{p_{current}}\right\}$ (If $p_{proposed} > p_{current}$ we move with probability 1.)
3. Decide to move by sampling $U \sim \text{Unif}(0, 1)$.
   - If $U \leq p_{move}$ then move: $X_{t+1} = X_{proposed}$
   - If $U > p_{move}$ then do not move: $X_{t+1} = X_t$

Code for Markov Chain

```
# proposed is a function that returns current value plus or minus one
propose <- function(current) {  current + sample(c(1, -1), size = 1) }

# p is function that returns target probability, if x is 1 through 7, the target prob is x/28
p <- function(x) { ifelse(x %in% 1:7, x/28, 0) }

n <- 10^3
x <- c(4, rep(NA, n-1)) # start at 4
set.seed(1)
for (t in 1:(n-1)) {
  current <- x[t]
  proposed <- propose(current)
  pmove <- p(proposed) / p(current)
  # pmove is min(1, p(propose)/p(current)), but if pmove > 1, random U will be less than it anyway
  u <- runif(1)
  if(u < pmove) {
    x[t + 1] <- proposed
  } else {
    x[t + 1] <- current
  }
}
```

Results

```r
counts <- table(x)
results <- rbind( counts/n , (1:7)/28 ); rownames(results) <- c("empirical", "target")
round(results, 4)
```

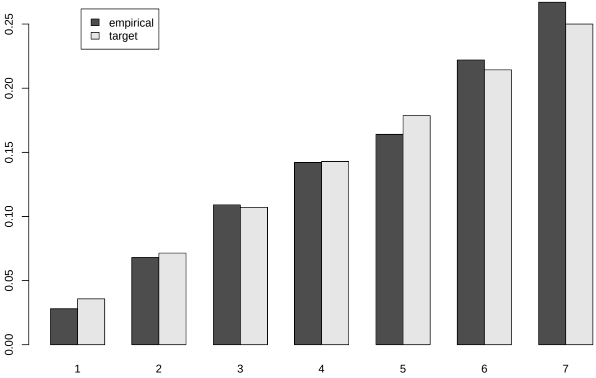
```
##                1      2      3      4      5      6     7
## empirical 0.0280 0.0680 0.1090 0.1420 0.1640 0.2220 0.267
## target    0.0357 0.0714 0.1071 0.1429 0.1786 0.2143 0.250
```

```r
chisq.test(counts, p = (1:7)/28 )
```

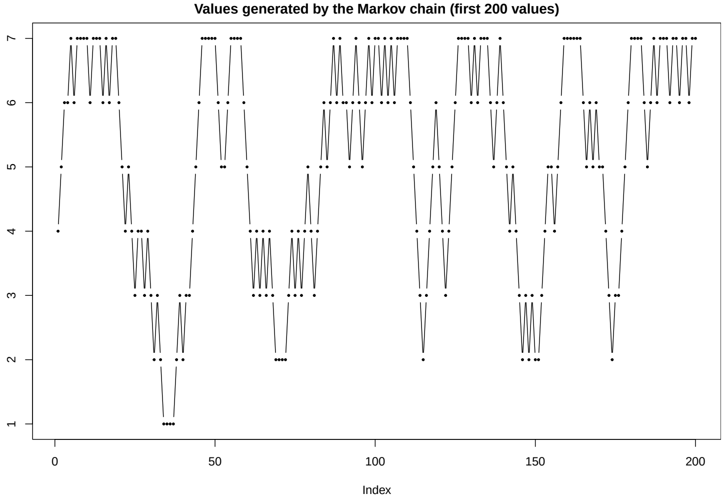
```
##
##  Chi-squared test for given probabilities
##
## data:  counts
## X-squared = 4.4909, df = 6, p-value = 0.6105
```

We see close alignment between the empirical probabilities and target distribution. The chi-squared goodness of fit test produces a very large p-value indicating that we do not have enough evidence to say that the values generated by the Markov chain do not come from the target distribution.

```r
barplot(results, beside = TRUE, legend.text = row.names(results), args.legend = list(x = 5))
```




```r
par(mar = c(4, 2, 2, 0))
plot(x[1:200], type = "b", pch = 19, cex = 0.4,
     main = "Values generated by the Markov chain (first 200 values)")
```



Running the chain longer

```r
n <- 10^5
x <- c(7, rep(NA, n-1)) # start at 7
set.seed(5)
for (t in 1:(n-1)) {
  current <- x[t]
  proposed <- propose(current)
  pmove <- p(proposed) / p(current)
  u <- runif(1)
  if(u < pmove) {
    x[t + 1] <- proposed
  } else {
    x[t + 1] <- current
  }
}
```

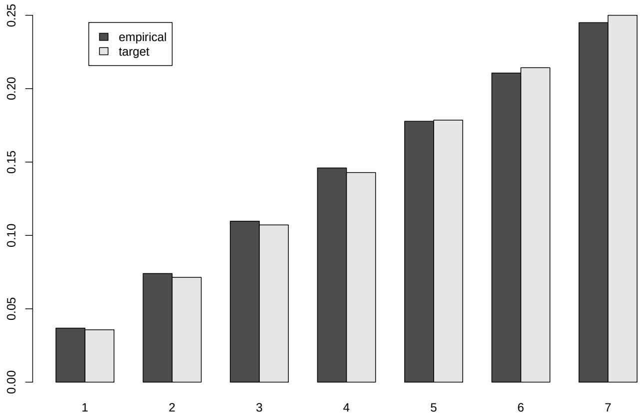
Results

```r
counts <- table(x)
results <- rbind( counts/n , (1:7)/28 )
rownames(results) <- c("empirical", "target")
round(results, 4)
```

```
##                1      2      3      4      5      6     7
## empirical 0.0368 0.0740 0.1097 0.1460 0.1778 0.2107 0.245
## target    0.0357 0.0714 0.1071 0.1429 0.1786 0.2143 0.250
```

Comparing the empirical and theoretic probabilities, we see the values align quite closely.

```r
barplot(results, beside = TRUE, legend.text = row.names(results), args.legend = list(x = 5))
```



Comparing the empirical and theoretic probabilities, we see the values align quite closely.

```r
chisq.test(counts, p = (1:7)/28 )
```

```
##
##  Chi-squared test for given probabilities
##
## data:  counts
## X-squared = 42.452, df = 6, p-value = 1.497e-07
```

The chi-squared goodness-of-fit test for this sample, however, gives a very small p-value. The problem is that the chi-squared test is very sensitive to sample size. With large samples ($n > 2000$), the chi-squared test will detect even the smallest of differences.

In these situations, we might consider using a p-value much smaller than 5%. The discussion is a bit beyond the scope of the course, but is worth reading.

https://stats.stackexchange.com/questions/74780/goodness-of-fit-for-very-large-sample-sizes

In short, don't worry if your p-value ends up very small if you have a very large sample size.

You might be better served simply comparing the empirical and theoretic probabilities using your best judgement.

Distributions one day at a time

In the next few slides, we'll take a look at the distribution over states by multiplying the distribution by the transition probability matrix.

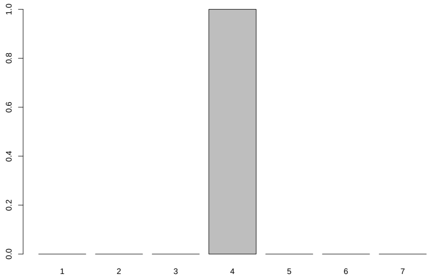
Distribution step by step: Day 1

Let's see the probability distribution of where the politician will be if he starts at island 4.

```
pi_n <- matrix(rep(0, 7*100), nrow = 100)
colnames(pi_n) <- 1:7
pi_n[1, ] <- c(0, 0, 0, 1, 0, 0, 0) # at initial state politician is at island 4
```

On day 1, the politician has probability 1 of being at island 4.

```
barplot(pi_n[1,])
```

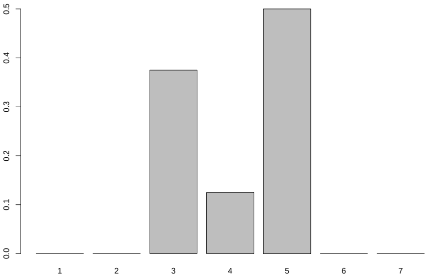


Day 2

```
pi_n[2, ] <- pi_n[1, ] %*% P
pi_n[2, ]
```

```
##     1     2     3     4     5     6     7
## 0.000 0.000 0.375 0.125 0.500 0.000 0.000
```

```
barplot(pi_n[2,])
```

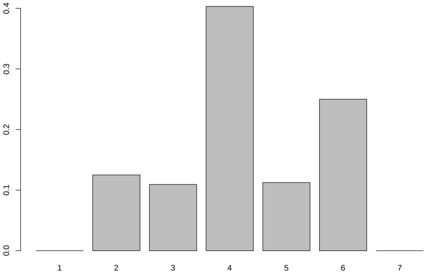


Day 3

```
pi_n[3, ] <- pi_n[2, ] %*% P
pi_n[3, ]
```

```
##        1        2        3        4        5        6        7
## 0.000000 0.125000 0.109375 0.403125 0.112500 0.250000 0.000000
```

```
barplot(pi_n[3,])
```

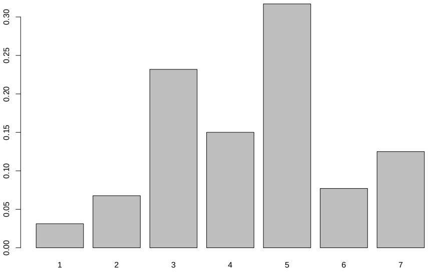


Day 4

```
pi_n[4, ] <- pi_n[3, ] %*% P
pi_n[4, ]
```

```
##          1          2          3          4          5          6          7
## 0.03125000 0.06770833 0.23190104 0.15007813 0.31697917 0.07708333 0.12500000
```

```
barplot(pi_n[4,])
```

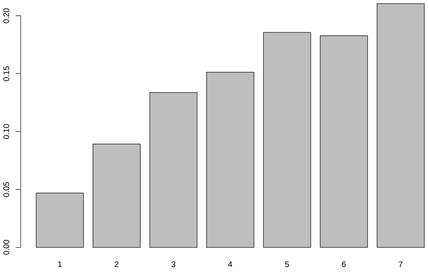


Repeatedly applying the transition matrix

```
for (i in 5:100) {
  pi_n[i, ] <- pi_n[i-1, ] %*% P
}
```

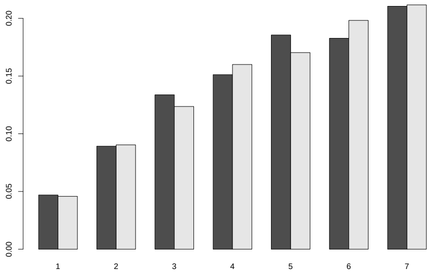
Day 10

```
barplot(pi_n[10,])
```



Day 10 vs Day 11

```
barplot(pi_n[10:11,], beside = TRUE)
```

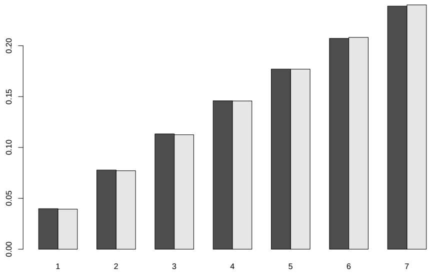



```
all.equal(pi_n[10,], pi_n[11,])
```

```
## [1] "Mean relative difference: 0.05307472"
```

Day 20 vs Day 21

```
barplot(pi_n[20:21,], beside = TRUE)
```

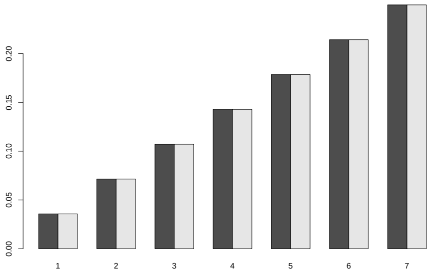



```
all.equal(pi_n[20,], pi_n[21,])
```

```
## [1] "Mean relative difference: 0.004388209"
```

Day 99 vs Day 100

```r
barplot(pi_n[99:100,], beside = TRUE)
```




```r
all.equal(pi_n[99,], pi_n[100,])
```

```
## [1] "Mean relative difference: 3.436071e-07"
```