Probability

DATA606, Lab 2, Zachary Herold

##Hot Hands

Our investigation will focus on the performance of one player: Kobe Bryant of the Los Angeles Lakers. His performance against the Orlando Magic in the 2009 NBA finals earned him the title *Most Valuable Player* and many spectators commented on how he appeared to show a hot hand. Let's load some data from those games and look at the first several rows.

```
load("more/kobe.RData")
head(kobe)
```

```
##    vs game quarter time
## 1 ORL    1       1 9:47
## 2 ORL    1       1 9:07
## 3 ORL    1       1 8:11
## 4 ORL    1       1 7:41
## 5 ORL    1       1 7:03
## 6 ORL    1       1 6:01
##                                         description basket
## 1                Kobe Bryant makes 4-foot two point shot      H
## 2                             Kobe Bryant misses jumper      M
## 3                      Kobe Bryant misses 7-foot jumper      M
## 4 Kobe Bryant makes 16-foot jumper (Derek Fisher assists)      H
## 5                       Kobe Bryant makes driving layup      H
## 6                             Kobe Bryant misses jumper      M
```

In this data frame, every row records a shot taken by Kobe Bryant. If he hit the shot (made a basket), a hit, `H`, is recorded in the column named `basket`, otherwise a miss, `M`, is recorded.

Just looking at the string of hits and misses, it can be difficult to gauge whether or not it seems like Kobe was shooting with a hot hand. One way we can approach this is by considering the belief that hot hand shooters tend to go on shooting streaks. For this lab, we define the length of a shooting streak to be the *number of consecutive baskets made until a miss occurs*.

For example, in Game 1 Kobe had the following sequence of hits and misses from his nine shot attempts in the first quarter:

$$\text{H M} \mid \text{M} \mid \text{H H M} \mid \text{M} \mid \text{M} \mid \text{M}$$

To verify this use the following command:

```
kobe$basket[1:9]
```

```
## [1] "H" "M" "M" "H" "H" "M" "M" "M" "M"
```

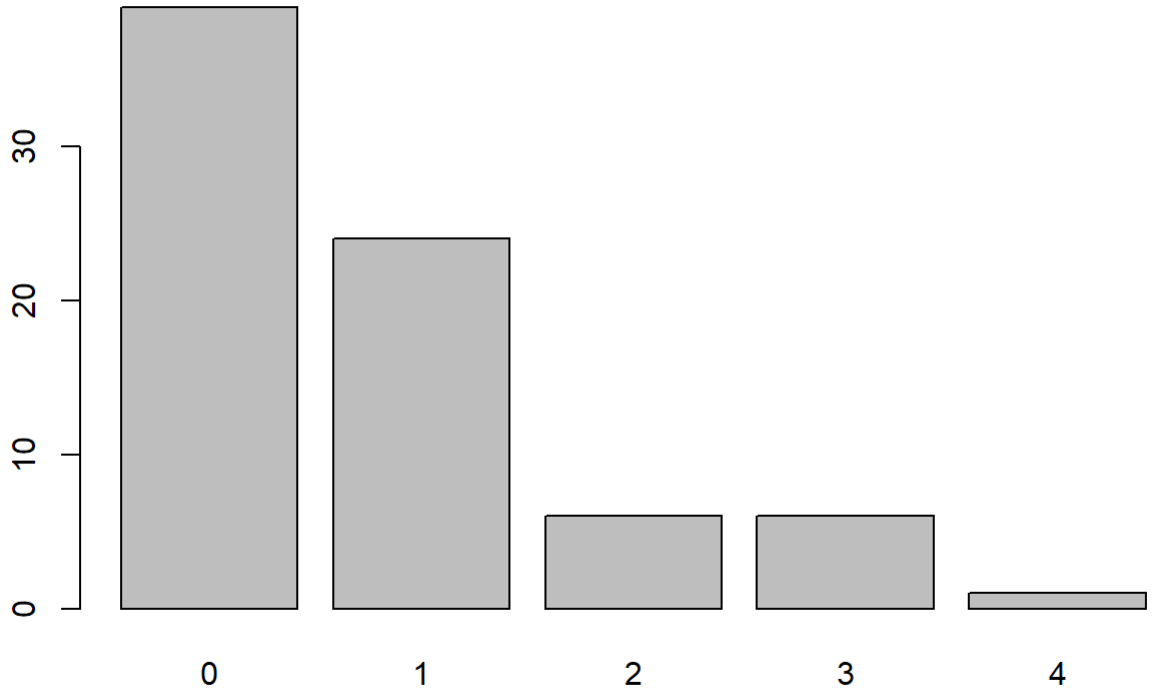
Within the nine shot attempts, there are six streaks, which are separated by a "|" above. Their lengths are one, zero, two, zero, zero, zero (in order of occurrence).

Exercise 1 What does a streak length of 1 mean, i.e. how many hits and misses are in a streak of 1? What about a streak length of 0?

Streak length of 1 means Kobe misses his shot after making one only {(M)HM sequence.} Streak length of 0 means Kobe misses his shot after missing the previous shot (M)M sequence.}

The custom function `calc_streak`, which was loaded in with the data, may be used to calculate the lengths of all shooting streaks and then look at the distribution.

```
kobe_streak <- calc_streak(kobe$basket)
barplot(table(kobe_streak))
```



Exercise 2 Describe the distribution of Kobe's streak lengths from the 2009 NBA finals. What was his typical streak length? How long was his longest streak of baskets?

Follows a lognormal distribution, with mean greater than 0 and less than 1, and with no results less than 0. The longest streak of baskets was 4 in a row before missing.

```
print(paste0("Kobe's typical ('mean') streak is: ", round(mean(kobe_streak),2)))
```

```
## [1] "Kobe's typical ('mean') streak is: 0.76"
```

```
print(paste0("Kobe's max streak is: ", max(kobe_streak)))
```

```
## [1] "Kobe's max streak is: 4"
```

Simulations in R

While we don't have any data from a shooter we know to have independent shots, that sort of data is very easy to simulate in R. In a simulation, you set the ground rules of a random process and then the computer uses random numbers to generate an outcome that adheres to those rules. As a simple example, you can simulate flipping a fair coin with the following.

```
outcomes <- c("heads", "tails")
sample(outcomes, size = 1, replace = TRUE)
```

```
## [1] "heads"
```

To view the results of this simulation, type the name of the object and then use `table` to count up the number of heads and tails.

```
sim_fair_coin <- sample(outcomes, size = 100, replace = TRUE)
sim_fair_coin
```

```
##   [1] "heads" "tails" "tails" "tails" "tails" "heads" "heads" "heads"
##   [9] "heads" "heads" "tails" "tails" "heads" "heads" "heads" "heads"
##  [17] "heads" "tails" "heads" "tails" "heads" "tails" "tails" "tails"
##  [25] "tails" "tails" "heads" "tails" "tails" "tails" "heads" "tails"
##  [33] "tails" "tails" "tails" "tails" "tails" "heads" "tails" "heads"
##  [41] "tails" "tails" "tails" "heads" "tails" "heads" "tails" "heads"
##  [49] "tails" "heads" "tails" "heads" "tails" "heads" "heads" "tails"
##  [57] "tails" "tails" "tails" "heads" "heads" "tails" "tails" "tails"
##  [65] "tails" "tails" "tails" "heads" "heads" "tails" "tails" "heads"
##  [73] "heads" "tails" "heads" "heads" "tails" "tails" "tails" "heads"
##  [81] "heads" "tails" "heads" "heads" "heads" "tails" "tails" "heads"
##  [89] "heads" "heads" "heads" "tails" "tails" "tails" "tails" "tails"
##  [97] "heads" "heads" "heads" "heads"
```

```
table(sim_fair_coin)
```

```
## sim_fair_coin
## heads tails
##    45    55
```

Exercise 3 In your simulation of flipping the unfair coin 100 times, how many flips came up heads?

```
sim_unfair_coin <- sample(outcomes, size = 100, replace = TRUE, prob = c(0.2, 0.8))
sim_unfair_coin
```

```
##   [1] "tails" "heads" "heads" "tails" "tails" "tails" "tails" "tails"
##   [9] "tails" "tails" "tails" "tails" "heads" "tails" "tails" "tails"
##  [17] "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails"
##  [25] "heads" "heads" "tails" "tails" "tails" "tails" "tails" "tails"
##  [33] "tails" "tails" "tails" "heads" "tails" "heads" "tails" "tails"
##  [41] "tails" "tails" "tails" "tails" "heads" "heads" "heads" "tails"
##  [49] "tails" "heads" "tails" "tails" "heads" "tails" "tails" "tails"
##  [57] "tails" "tails" "heads" "tails" "tails" "tails" "tails" "tails"
##  [65] "heads" "tails" "tails" "tails" "tails" "tails" "tails" "heads"
##  [73] "tails" "tails" "tails" "tails" "tails" "tails" "tails" "tails"
##  [81] "tails" "tails" "tails" "tails" "tails" "tails" "heads" "tails"
##  [89] "tails" "tails" "tails" "heads" "tails" "tails" "tails" "tails"
##  [97] "heads" "tails" "tails" "heads"
```

```
table(sim_unfair_coin)
```

```
## sim_unfair_coin
## heads tails
##    19    81
```

Simulating the Independent Shooter

Exercise 4 What change needs to be made to the `sample` function so that it reflects a shooting percentage of 45%? Make this adjustment, then run a simulation to sample 133 shots. Assign the output of this simulation to a new object called `sim_basket`.

```
outcomes <- c("H", "M")
sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45,0.55))
sim_basket
```

```
##   [1] "M" "H" "H" "M" "M" "M" "M" "H" "H" "M" "M" "H" "H" "M" "H" "M" "H"
##  [18] "M" "H" "H" "M" "H" "M" "M" "M" "M" "M" "M" "H" "M" "M" "H" "M" "M"
##  [35] "M" "H" "H" "H" "M" "M" "M" "H" "M" "H" "H" "H" "H" "M" "H" "H" "M"
##  [52] "M" "M" "M" "H" "M" "M" "H" "H" "H" "M" "M" "M" "H" "M" "H" "M" "H"
##  [69] "M" "H" "H" "H" "M" "M" "M" "M" "M" "H" "M" "M" "M" "M" "M" "M" "M"
##  [86] "H" "H" "M" "H" "H" "M" "M" "H" "H" "M" "M" "H" "M" "M" "H" "H" "M"
## [103] "H" "H" "M" "M" "M" "H" "M" "H" "H" "H" "H" "M" "M" "M" "M" "H" "M"
## [120] "H" "M" "H" "H" "M" "M" "M" "H" "M" "M" "H" "H" "H" "M"
```

```
table(sim_basket)
```

```
## sim_basket
##  H  M
## 58 75
```

```
table(kobe$basket)
```

```
##
##  H  M
## 58 75
```

```
table(sim_basket)
```

```
## sim_basket
##  H  M
## 58 75
```

On your own

Comparing Kobe Bryant to the Independent Shooter

Using `calc_streak`, compute the streak lengths of `sim_basket`.

1. Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player's longest streak of baskets in 133 shots?

```
calc_streak(sim_basket)
```

```
##  [1] 0 2 0 0 0 2 0 2 1 1 2 1 0 0 0 0 0 1 0 1 0 0 3 0 0 1 4 2 0 0 0 1 0 3 0
## [36] 0 1 1 1 3 0 0 0 0 1 0 0 0 0 0 0 2 2 0 2 0 1 0 2 2 0 0 1 4 0 0 0 1 1 2
## [71] 0 0 1 0 3 0
```

1. If you were to run the simulation of the independent shooter a second time, how would you expect its streak distribution to compare to the distribution from the question above? Exactly the same? Somewhat similar? Totally different? Explain your reasoning.

```
longest.stk <- NULL
avg.stk <- NULL

for(i in 1:1){
  outcomes <- c("H", "M")
  sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45,0.55))
  r <- calc_streak(sim_basket)
  longest.stk[i] <- max(r)
  avg.stk[i] <- mean(r)
  }


mean(longest.stk)
```

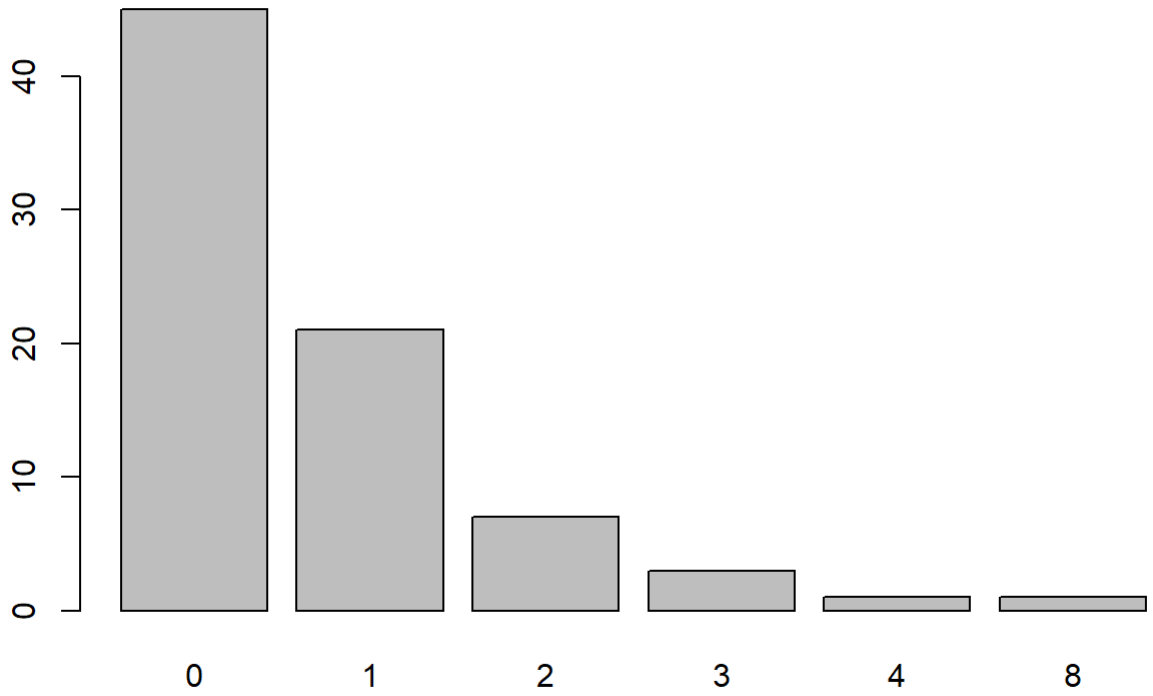
```
## [1] 8
```

```
mean(avg.stk)
```

```
## [1] 0.7179487
```

The following barplot of one simulation reveals the occasional streaks of 5 or 6, more than any streak in Kobe's final performance.

```
barplot(table(calc_streak(sim_basket)))
```



When simulated 10,000 times, the typical streak is shown to be 0.8 makes for the non-hot-hand shooter, with the longest streak of 5.6 on average.

```
for(i in 1:10000){
  outcomes <- c("H", "M")
  sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45,0.55))
  r <- calc_streak(sim_basket)
  longest.stk[i] <- max(r)
  avg.stk[i] <- mean(r) }

mean(longest.stk)
```

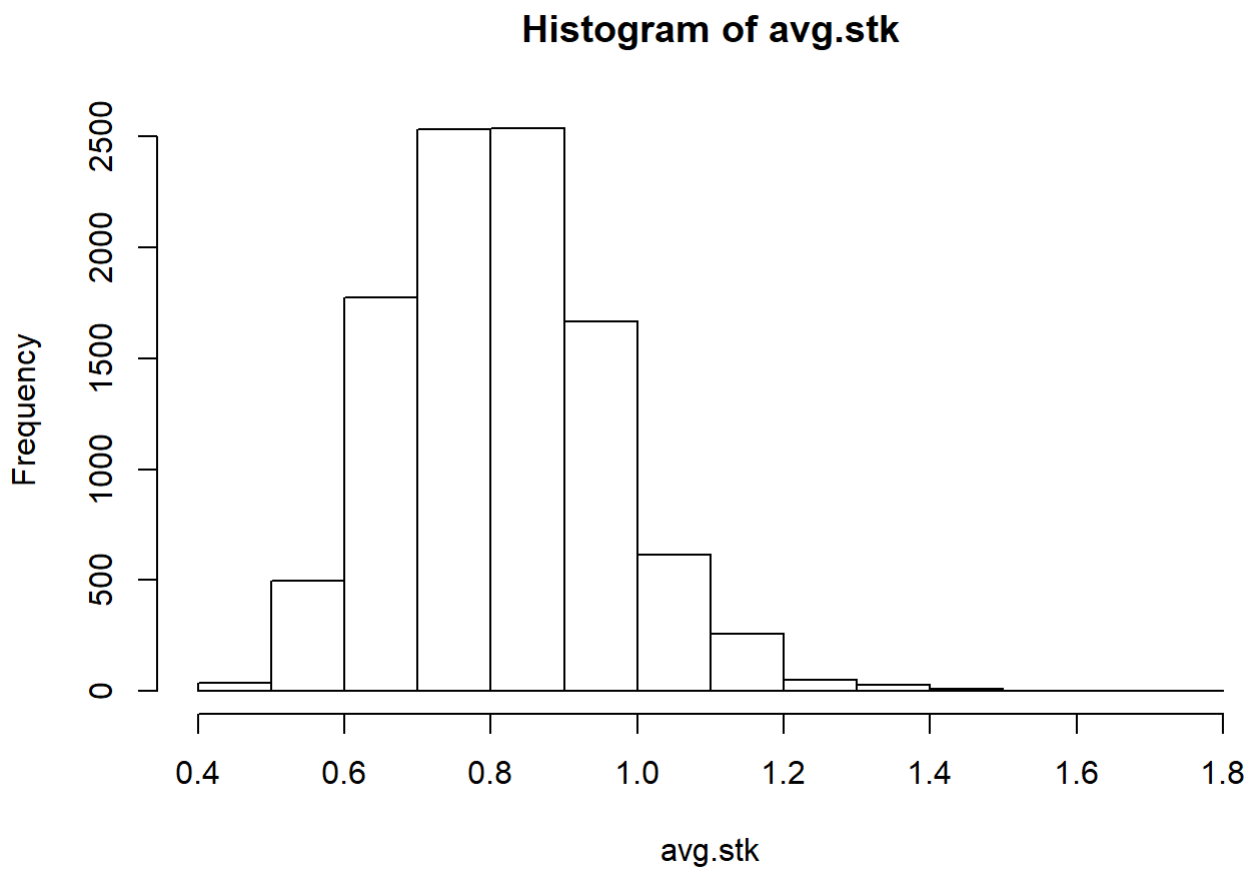
```
## [1] 5.5934
```

```
mean(avg.stk)
```

```
## [1] 0.8178049
```

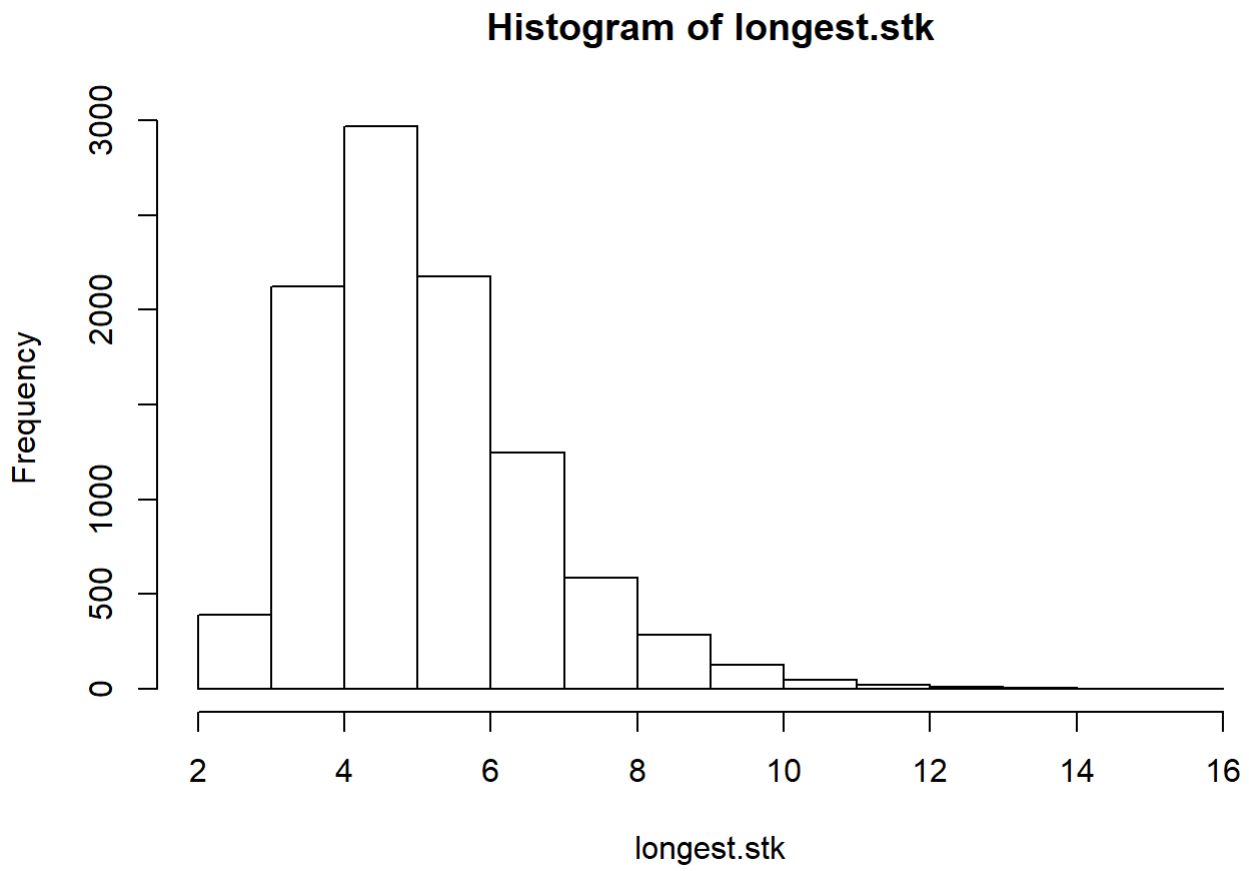
The typical streak (after 10,000 simulations) visualized in a histogram.

```
hist(avg.stk)
```



The longest streak (after 10,000 simulations) visualized in a histogram.

```
hist(longest.stk)
```



1. How does Kobe Bryant's distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter? Using this comparison, do you have evidence that the hot hand model fits Kobe's shooting patterns? Explain.

Kobe's longest streak was 4, with an average of 0.76.

```
q <- calc_streak(kobe$basket)
kobe.longest.stk = max(q)
kobe.avg.stk = mean(q)

kobe.longest.stk
```

```
## [1] 4
```

```
kobe.avg.stk
```

```
## [1] 0.7631579
```

There were only 391 cases out of 10,000 in which there was a shorter longest streak than Kobe's.

```
sum(longest.stk >= 1 & longest.stk <= 3)
```

```
## [1] 391
```

Compared to Kobe's high of 4 in a row, a simulated independent shooter appears to produce longer streaks after 133 shots.

```
head(longest.stk,30)
```

```
##  [1]  5  4  4  5  5  6  7  5  4  4  4  5  5  4  8  5  5  5  4  5  4  8  7
## [24]  4  4 11  4  5  7  9
```

Kobe's performance is a below-average one, even for an independent shooter.

```
print(paste0("Percentage of independent simulations that had a longer streak: ", (sum(longest.stk
  > kobe.longest.stk) / 10000)))
```

```
## [1] "Percentage of independent simulations that had a longer streak: 0.7486"
```

```
print(paste0("Percentage of independent simulations that had a higher mean: ", (sum(avg.stk > kob
  e.avg.stk) / 10000)))
```

```
## [1] "Percentage of independent simulations that had a higher mean: 0.5857"
```

If Kobe evidenced a hot-hand, his streaks on average would be longer, with more lengthy strings of makes. The fact they were not suggests a hot hand (60% make-rate after make) does not apply to his Finals case.