Data 606 Lab2

Yuen Chun Wong

September 13, 2017

Exercise

```
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
```

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?

   Ans: streak length of 1 means: how many times get one hit before miss. streak length of 0 means: how many times get zero hit before miss. or it means just miss.

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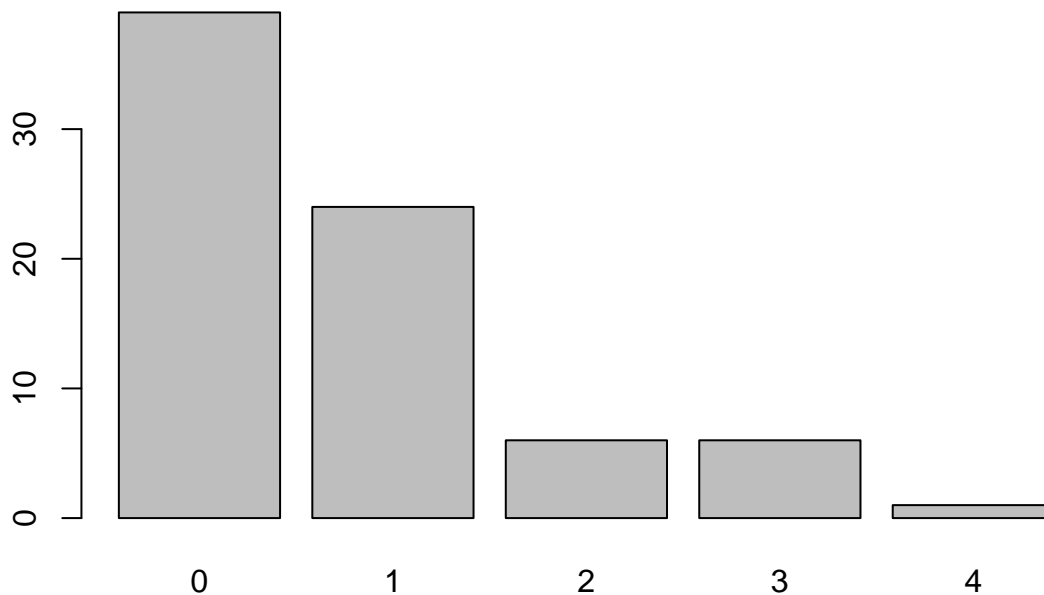
2. Describe the distribution of Kobe's streak lengths from the 2009 NBA finals. What was his typical streak length? How long was his longest streak of baskets?

Ans: The distribution is right skewed. His typical streak is zero and his longest streak is 4

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

```
## sim_unfair_coin
## heads tails
##    20    80
```

`prob=c(0.2, 0.8)` indicates that for the two elements in the `outcomes` vector, we want to select the first one, `heads`, with probability 0.2 and the second one, `tails` with probability 0.8. Another way of thinking about this is to think of the outcome space as a bag of 10 chips, where 2 chips are labeled "head" and 8 chips "tail". Therefore at each draw, the probability of drawing a chip that says "head"" is 20%, and "tail" is 80%.

3. In your simulation of flipping the unfair coin 100 times, how many flips came up heads? Heads: 13.

To make a valid comparison between Kobe and our simulated independent shooter, we need to align both their shooting percentage and the number of attempted shots.

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))
table(sim_basket)
```

```
## sim_basket
##  H  M
## 65 68
```

```
kobe$basket
```

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

```
sim_basket
```

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

Comparing Kobe Bryant to the Independent Shooter

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

- 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?

  Ans: The distribution is right skewed. The typical streak length is zero and the longest streak is 4.

```
sim_streak <- calc_streak(sim_basket)
barplot(table(sim_streak))
```
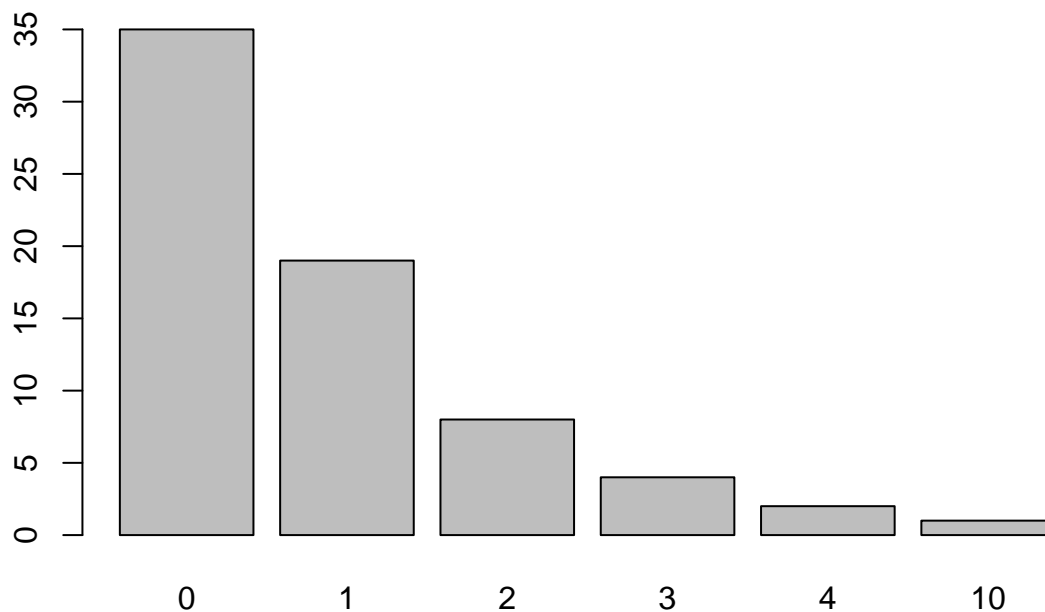
- If you were to run the simulation of the independent shooter a second time, how would you expect its streak distribution to compare to the distribution from the question above? Exactly the same? Somewhat similar? Totally different? Explain your reasoning.

  I would expect the distribution is similar. for independent shooter, the probablityies of: streak 0: 0.55 streak 1: $0.45 \, 0.55$ streak 2: $0.45 \, 0.45 \, 0.55$ streak 3 : $0.45\hat{} 3$ 0.55 streak 4: $0.45\hat{} 4$ * 0.55 This probability shows that it is getting harder and harder to have higher streak number. Therefore the distribution has to be right skewed.

- 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. Even the distribution of kobe and independent shooter are similar. I believe He has hot hand because the streak 2 and streak 3 is more higher and more even than indenpendent shooter. It means he can hit once he make his first hit.