Grando-2

John Grando

February 6, 2017

Grando-2 Lab

Set the working directory and RData file

```
setwd("~/Documents/Masters/DATA606/Week2/Lab/Lab2")
load("more/kobe.RData")
require(ggplot2)
```

```
## Loading required package: ggplot2
```

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?

Answer:

A streak length of 1 is when kobe made only one basket before missing his next shot (i.e. one hit and one miss). A streak length of zero would consist of two misses in a row.

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?

Answer:

```
kobe_streak <- calc_streak(kobe$basket)
summary(kobe_streak)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.0000  0.0000  0.0000  0.7632  1.0000  4.0000
```

The distribution has a right skew with a median of 0, median of 0.7632, IQR of 1, and max of 4. The typical streak length was 0 (using median) and 0.7632 (using mean). The longest streak was 4.

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

```
outcomes <- c("H", "M")
sim_unfair_coin <- sample(outcomes, size = 100, replace = TRUE,
    prob = c(0.2, 0.8))
number_of_heads <- table(sim_unfair_coin)[["H"]]
number_of_heads
```

```
## [1] 18
```

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.

Answer:

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

Question 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?

Answer:

I will re-run the sample command; however, this time I will use set.seed() to freeze the results so I can comment on them.

```
set.seed(100)
outcomes <- c("H", "M")
sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45,
    0.55))
sim_basket_streak <- calc_streak(sim_basket)
sim_basket_streak
```

```
##  [1] 0 0 1 0 0 1 0 0 2 0 2 0 0 1 1 1 0 2 0 0 1 3 2 0 4 2 0 0 0 0 0 1 0 0 1
## [36] 0 3 0 0 0 0 1 0 7 0 3 2 0 2 0 0 1 0 0 0 1 0 0 0 4 0 2 0 1 2 1 0 0 1 0
## [71] 3 0 2
```

```
barplot(table(sim_basket_streak))
```



```
summary(sim_basket_streak)
```
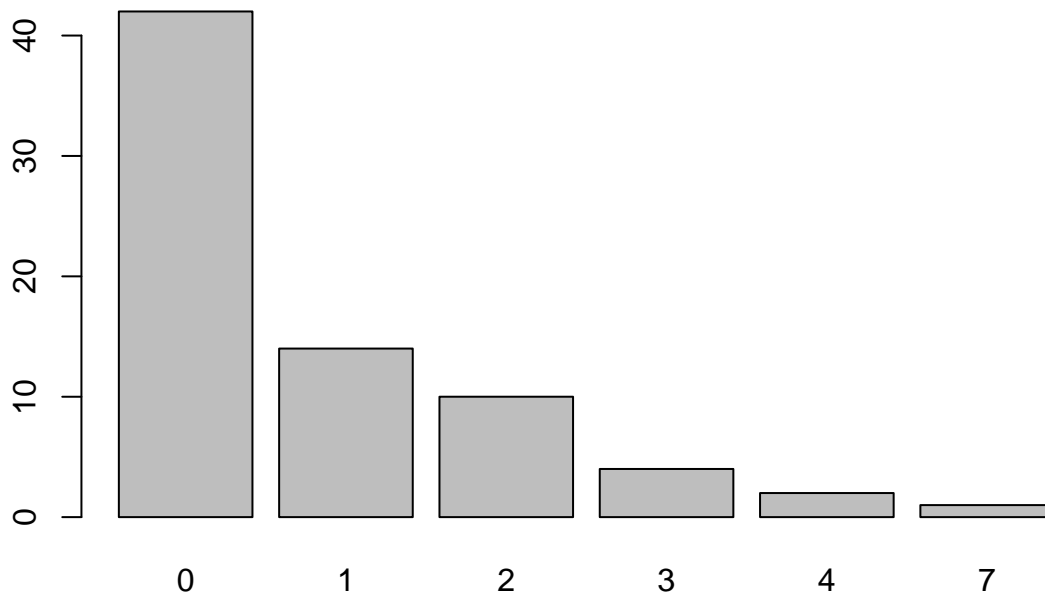
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.8356  1.0000  7.0000
```

The distribution has a right skew with a median of 0, median of 0.8356, IQR of 1, and max of 7. The typical streak length was 0 (using median) and 0.8356 (using mean). The longest streak was 7.

Question 2 - 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.

Answer:

My expectation would be that a second simulation would result in similar, but not exactly the same, distribution. We would expect simulation results to be more similar as the number of samples increases; however, since the probability of making each shot is independent of the others, it is unlikely we would get a result with exactly the same number of makes and misses.

I have run another simulation below which verifies that we would get a similar, but not exactly the same, result:

```
set.seed(200)
outcomes <- c("H", "M")
sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45,
    0.55))
sim_basket_streak <- calc_streak(sim_basket)
sim_basket_streak
```

```
##  [1] 0 6 0 0 0 1 1 0 1 2 0 0 1 0 1 0 0 0 0 0 0 2 1 0 3 2 3 0 2 0 0 0 2 5 0
## [36] 0 1 0 0 0 0 0 1 1 3 0 1 1 1 0 0 0 1 0 0 0 3 1 0 1 0 0 1 0 1 2 0 0 2 3
## [71] 1 0 0 2
```

```
barplot(table(sim_basket_streak))
```



```
summary(sim_basket_streak)
```
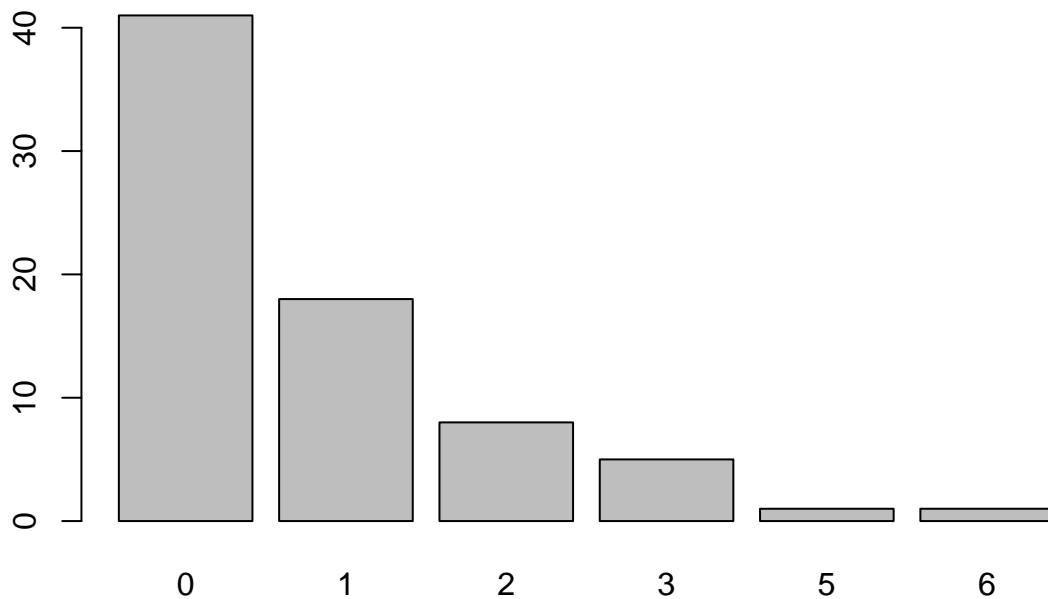
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.8108  1.0000  6.0000
```

Quesiton 3 - 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.

Answer:

Let's make a graph to compare our first simulation with Kobe's baskets

```r
set.seed(100)
outcomes <- c("H", "M")
sim_basket <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45,
    0.55))
sim_basket_streak <- calc_streak(sim_basket)
sim_basket_table_f <- table(sim_basket_streak)/length(sim_basket_streak)
kobe_streak <- calc_streak(kobe$basket)
kobe_basket_table_f <- data.frame(table(kobe_streak)/length(kobe_streak),
    "kobe")
names(kobe_basket_table_f) <- c("Streak", "Frequency", "Data")

table(sim_basket_streak)
```

```
## sim_basket_streak
##  0  1  2  3  4  7
## 42 14 10  4  2  1
```

```r
sim_basket_table_f
```

```
## sim_basket_streak
##          0          1          2          3          4          7
## 0.57534247 0.19178082 0.13698630 0.05479452 0.02739726 0.01369863
```

```r
table(kobe_streak)
```

```
## kobe_streak
##  0  1  2  3  4
## 39 24  6  6  1
```

```r
table(kobe_streak)/length(kobe_streak)
```

```
## kobe_streak
##          0          1          2          3          4
## 0.51315789 0.31578947 0.07894737 0.07894737 0.01315789
```

```r
combined_table <- data.frame(sim_basket_table_f, "Sim")
names(combined_table) <- c("Streak", "Frequency", "Data")
combined_table <- rbind(combined_table, kobe_basket_table_f)

ggplot(combined_table, aes(x = Streak, y = Frequency, fill = Data)) +
    geom_bar(stat = "identity", position = "dodge")
```
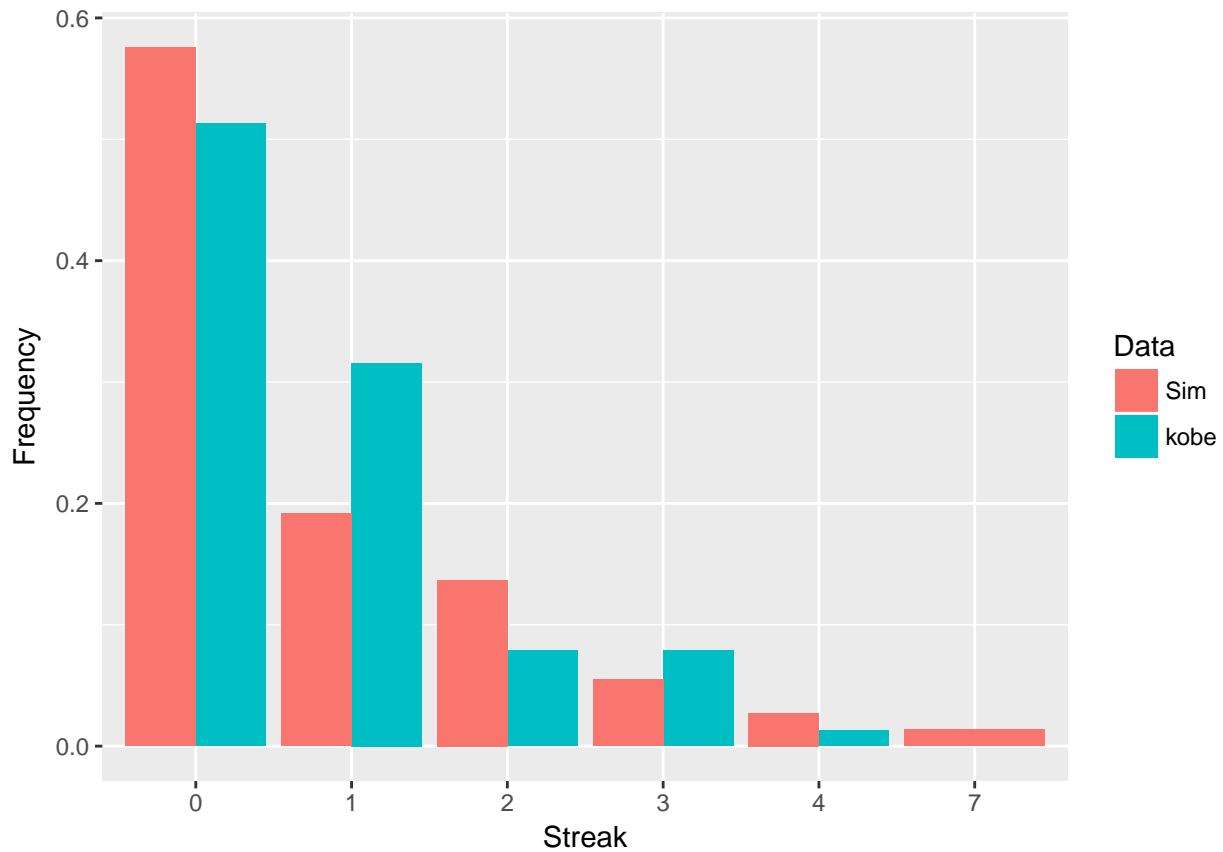
I made a bar plot that summarizes the relative frequency of each type of streak. As you can, a streak of zero happened over 50% of the time for both cases, followed by decreasing frequencies for each streak type. As expected, Kobe's distribtuion appears to be similar to the simulation performed. No, it does not appear the hot hand model, where a shooter's percent chance of making a basket increases on successive shots, fits Kobe's shooting pattern. His shots appear to be independent of each other.