

Exercise2 : Probability

Shikhar Dhwaaj

26 September 2017

Reading the data file

```
download.file("http://www.openintro.org/stat/data/kobe.RData", destfile = "kobe.RData")
load("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
```

Q-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?

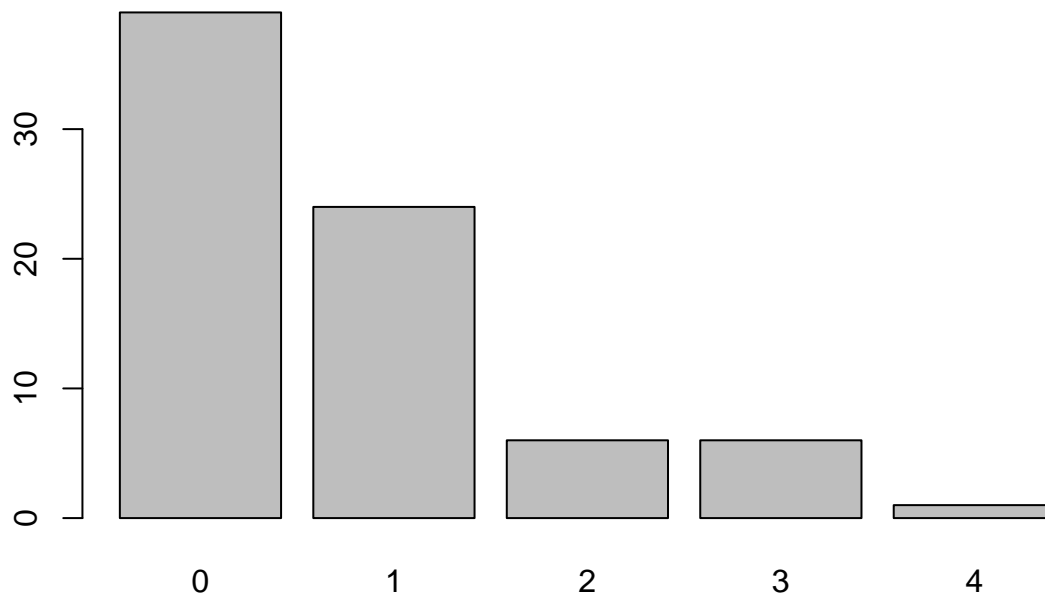
```
kobe$basket[1:9]
```

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

Streak length of 1 means hit followed by a miss Streak length of 0 means a miss followed by a miss

Q-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?

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



Distribution is right skewed Typical length of streak as measured by median is 0 Streak Lengths of 3 and 4 are usually high as compared to rest of the distribution

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

```
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
##      26     74
```

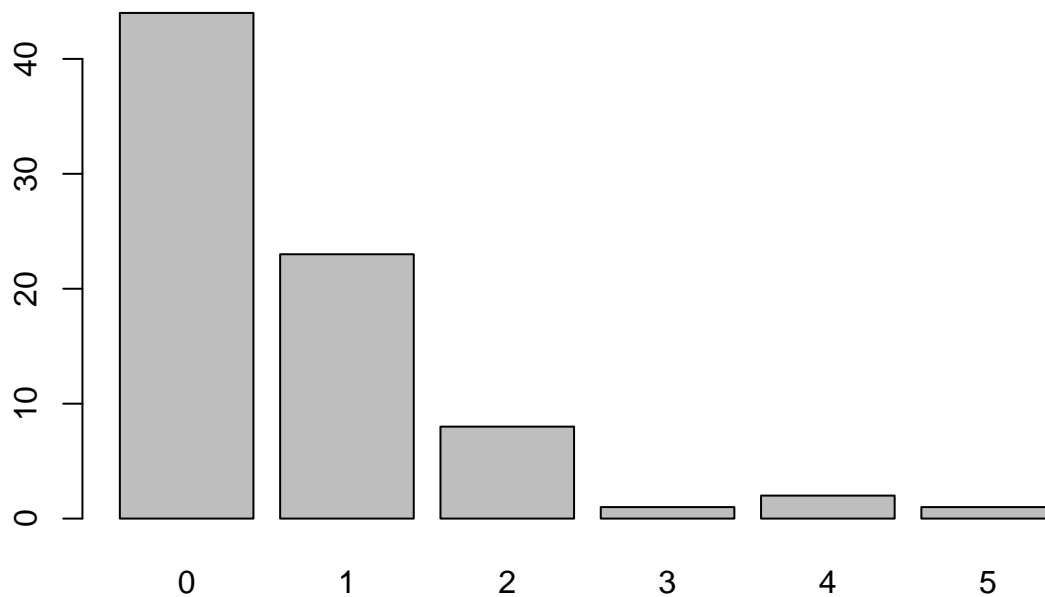
18 heads

Q-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))
```

Q-5) 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?

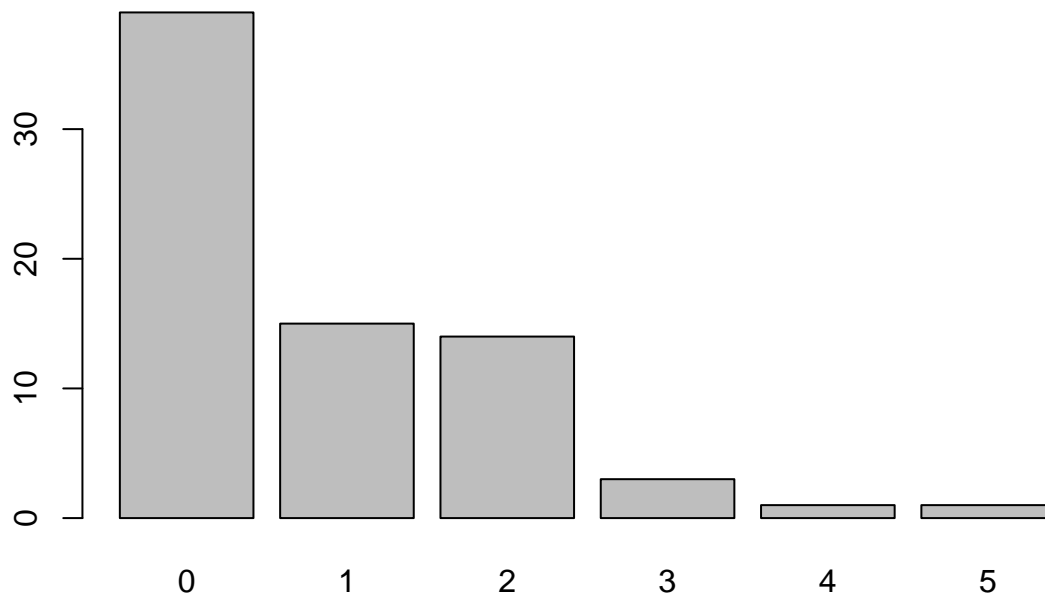
```
kobe_streak <- calc_streak(sim_basket)
barplot(table(kobe_streak))
```



Distribution of streak length for simulated player is similar to kobe's streak length distribution. Longest streak is 6. It represents 6 hits followed by a miss.

Q-6) 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.

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



We would expect somewhat similar distribution

Q-7)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.

Since Kobe's streak length distribution looks very similar to the independent shooter's simulated steak length distribution, we can conclude that Kobe Bryant likely does not have a "hot hand"