606-01_Lab2_Probability

Shyam BV

September 9, 2016

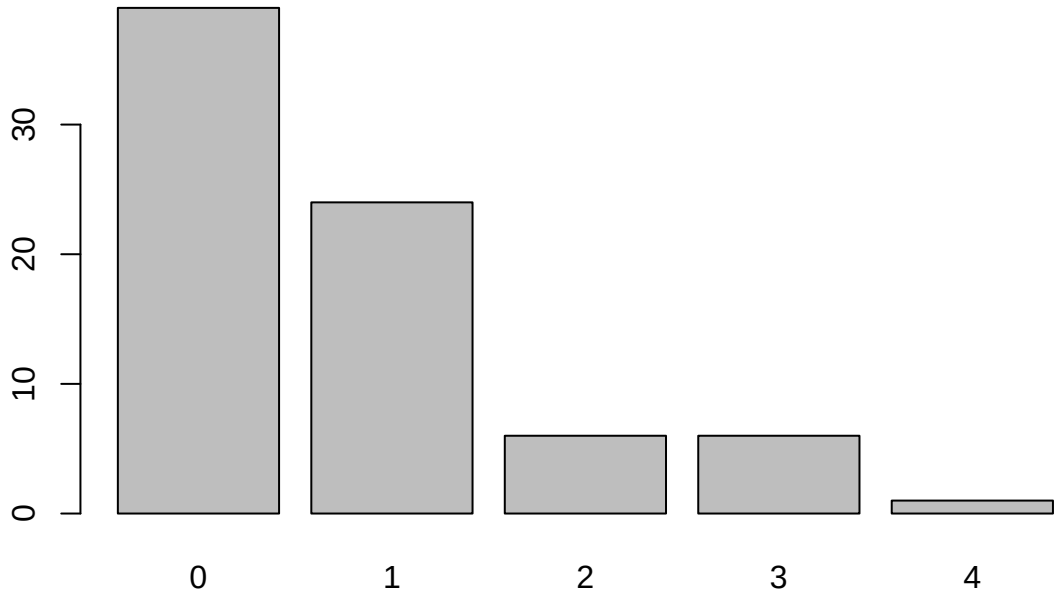
```
library("ggplot2")
library("reshape2")
load("more/kobe.RData")
```

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: Streak length of 1 means, 1 hit which were made. If the number is 2, then there are 2 continuous hits.

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))
```




```
mean(kobe_streak)
```

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

From the bar plot, it shows the distribution of Kobe's streak lengths are right skewed.

His typical streak length is 0.76.

His longest streak of baskets is 4.

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))

sum(sim_unfair_coin =="heads")
```

```
## [1] 20
```

In this simulation, the number of heads is 18. It might vary on each simulation.

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.

```
shooting_outcomes <- c("H","M")

sim_basket <-sample(shooting_outcomes,size = 133,replace = TRUE,prob=c(0.45,0.55))

sum(sim_basket =="H")
```

```
## [1] 65
```

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

On your own:

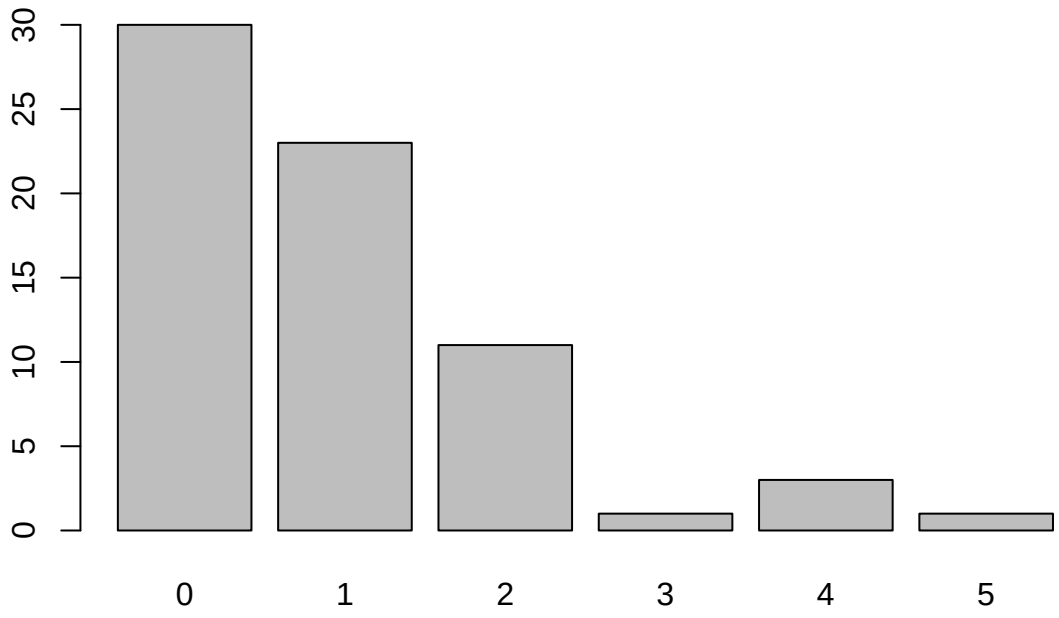
```
sim_streak <- calc_streak(sim_basket)
```

** 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?**

```
table(sim_streak)
```

```
## sim_streak
##  0  1  2  3  4  5
## 30 23 11  1  3  1
```

```
barplot(table(sim_streak))
```




```
mean(sim_streak)
```

```
## [1] 0.942029
```

From the bar plot, it shows the distribution of individual streak lengths are right skewed. There is one outlier with seven streak of baskets.

His typical streak length is 0.

His longest streak of baskets is 5 (This might vary on each run, because it is a simulation).

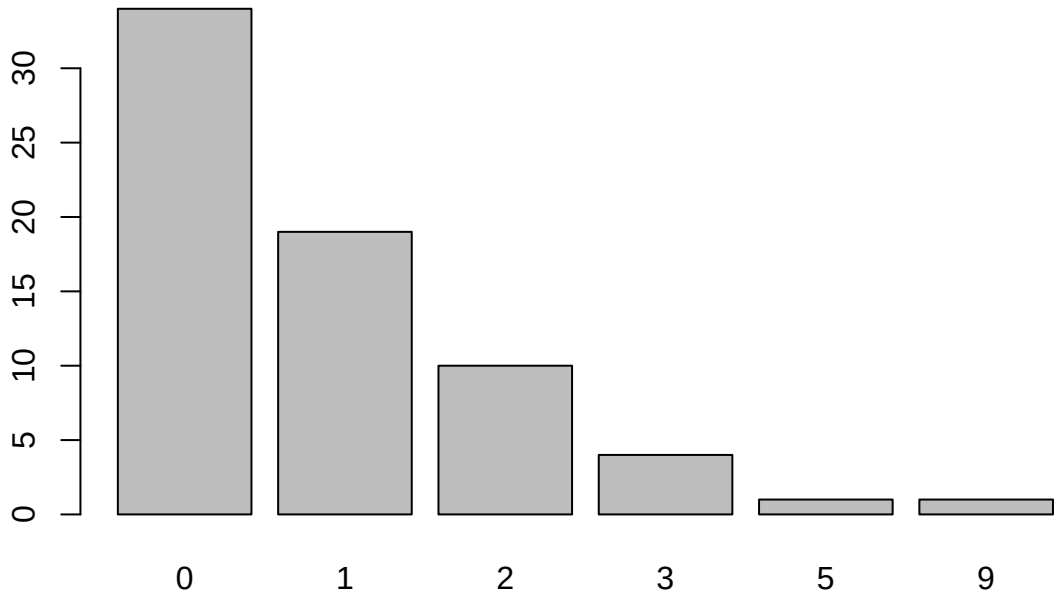
2. If you were to run the simulation of the independent shooter a second time, how would you expect its streak distribution to compare to the distribution from the question above? Exactly the same? Somewhat similar? Totally different? Explain your reasoning.

```
sim_basket <-sample(shooting_outcomes,size = 133,replace = TRUE,prob=c(0.45,0.55))

sum(sim_basket =="H")
```

```
## [1] 65
```

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



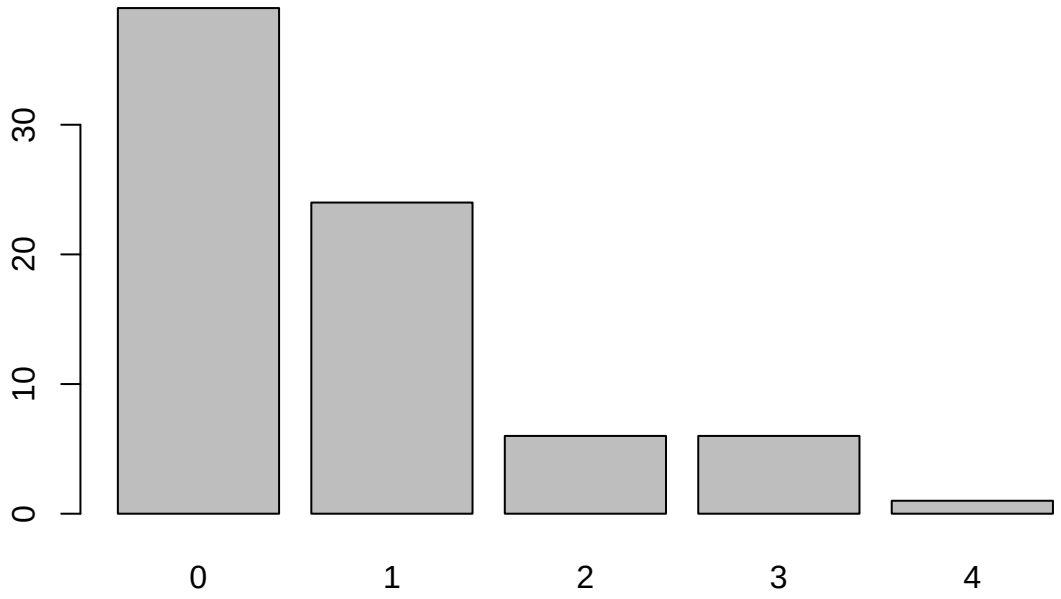

```
mean(sim_streak)
```

```
## [1] 0.942029
```

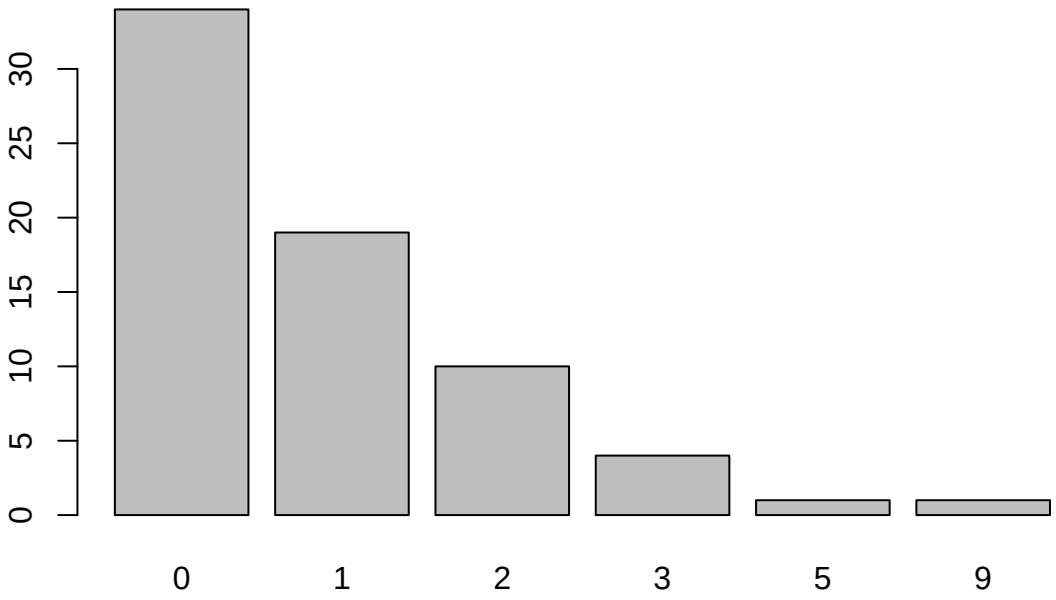
If we run the simulation of the independent shoot for a second time, we will not be getting the exact same results. Because we are only expecting the value to be in that range. But in actual scenario it will not be the exact same number. In each case the actual data might expect from the theoretical data. We can't always expect the same 45% of streak.

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.

```r
barplot(table(kobe_streak))
```




```r
barplot(table(sim_streak))
```



Both distribution are right skewed. The comparison which we have made is not a hot hand model. But it fits the simulation. So from the simulation we can derive that the Kobe's shooting patterns is not an hot hand model.