

Compte Rendu

Lab4 : probability & distributions

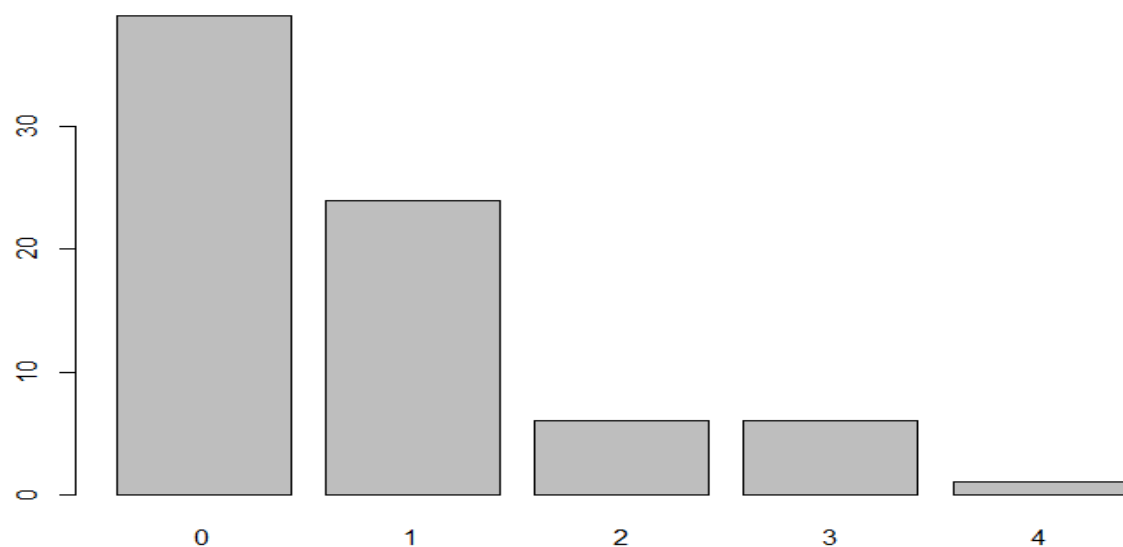
Part 1: Probability

a) The Data

Q1: 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?

A streak length of 2 means there was two total hits in the series (two in a row)

```
>
> load("kobe.RData")
> head(kobe)
  vs game quarter time      description basket
1 ORL    1         1 9:47 Kobe Bryant makes 4-foot two point shot      H
2 ORL    1         1 9:07      Kobe Bryant misses jumper          M
3 ORL    1         1 8:11      Kobe Bryant misses 7-foot jumper     M
4 ORL    1         1 7:41 Kobe Bryant makes 16-foot jumper (Derek Fisher assists) H
5 ORL    1         1 7:03      Kobe Bryant makes driving layup      H
6 ORL    1         1 6:01      Kobe Bryant misses jumper          M
> view(kobe)
>
> kobe_streak=calc_streak(kobe$basket)
> barplot(table(kobe_streak))
.
```



Q2. 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?

From the chart above, his typical streak length was 0 to 1 as most of the data falls into those buckets. His longest streak was 4 baskets in a row.

b) Simulations in R : Example

Q3. In your simulation of flipping the unfair coin 100 times, how many flips came up heads?

Heads (18) and tails (82)

c) Exercise :

Q1. Simulate a single shot from an independent shooter with a shooting percentage of 50%

To simulate a single shot from an independent shooter with a shooting percentage of 50% we type,

```
> outcomes <- c("H", "M")
> sim_basket1 <- sample(outcomes, size = 1, replace = TRUE)
>
> sim_basket1
[1] "H"
>
```

Q2. 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. What change needs to be made to the sample function so that it reflects a shooting percentage of 45%?

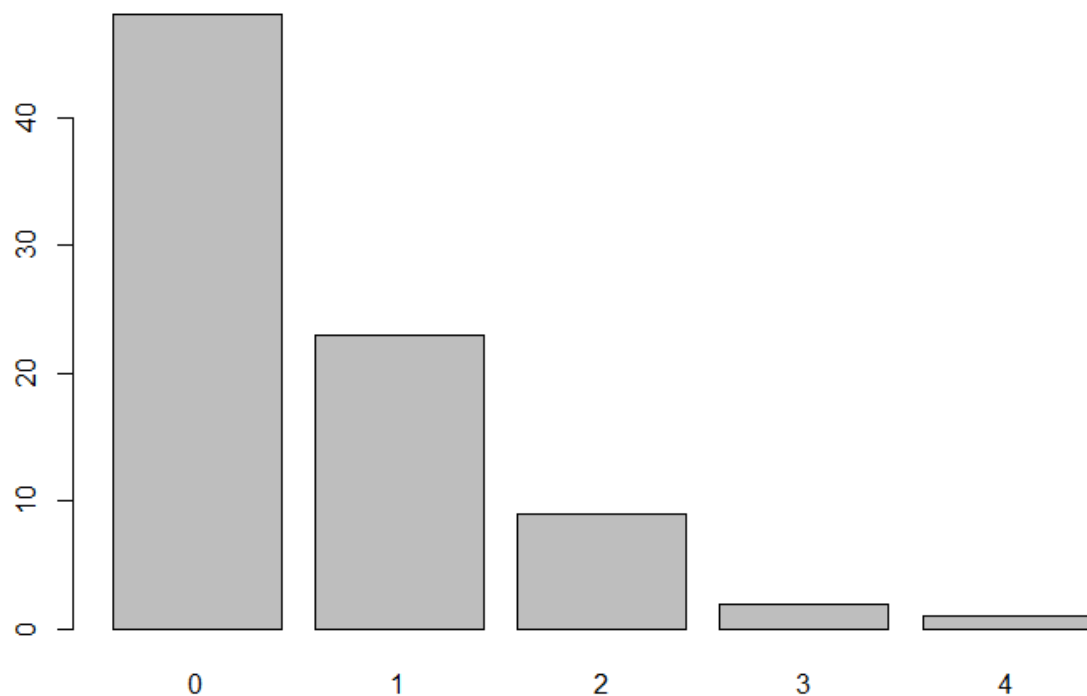
Q3. Run a simulation to sample 133 shots. Assign the output of this simulation to a new object called sim_basket.

```
>
> outcomes <- c("H", "M")
> sim_basket1 <- sample(outcomes, size = 133, replace = TRUE, prob = c(0.45, 0.55))
> sim_basket1
 [1] "M" "H" "M" "H" "M" "H" "H" "H" "M" "M" "M" "M" "H" "M" "H" "M" "M" "H" "H" "M" "M" "M" "M" "M" "M" "H"
[33] "H" "H" "M" "M" "H" "M" "H" "M" "M" "M" "H" "H" "M" "M" "M" "H" "M" "M" "H" "H" "M" "H" "H" "M" "H" "M"
[65] "M" "H" "M" "H" "M" "H" "H" "M" "M" "M" "M" "H" "M" "M" "M" "H" "M" "M" "H" "H" "M" "M" "M" "M" "M" "H"
[97] "M" "M" "H" "H" "H" "H" "M" "H" "M" "H" "M" "M" "M" "H" "M" "H" "M" "H" "M" "H" "H" "M" "M" "M" "H" "M"
[129] "M" "H" "H" "H" "M"
>
> kobe$basket
 [1] "H" "M" "M" "H" "H" "M" "M" "M" "M" "H" "H" "H" "M" "H" "H" "M" "M" "H" "M" "H" "H" "H" "M" "M" "M" "M"
[33] "M" "H" "M" "H" "M" "M" "H" "H" "H" "H" "M" "H" "M" "M" "H" "M" "M" "H" "H" "M" "H" "M" "H" "H" "M" "H"
[65] "M" "M" "M" "H" "M" "M" "M" "H" "M" "H" "M" "M" "H" "M" "M" "H" "H" "H" "M" "M" "H" "M" "M" "H" "M" "M"
[97] "H" "H" "M" "H" "M" "H" "M" "M" "M" "H" "M" "H" "H" "H" "H" "M" "H" "H" "M" "H" "M" "M" "M" "H" "M" "H"
[129] "M" "M" "M" "M" "H"
>
```

Both data sets represent the results of 133 shot attempts, each with the same shooting percentage of 45%.

Q4. Using `calc_streak`, compute the streak lengths of `sim_basket`.

```
[129] "M" "M" "M" "M" "H"  
> sim_streak <- calc_streak(sim_basket1)  
> barplot(table(sim_streak))  
> |
```

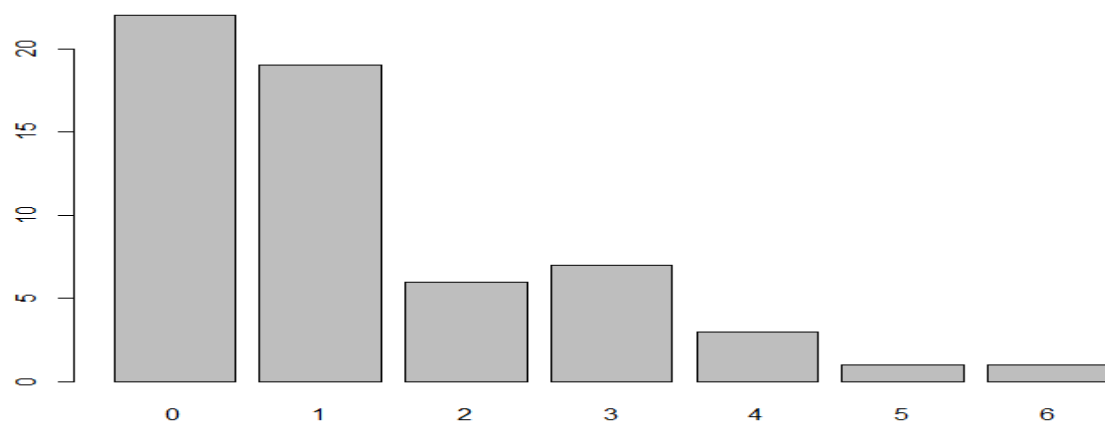


Q5. Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player's longest streak of baskets in 133 shots?

```
>  
>  
> cat("Longest streak is:", max(sim_streak))  
Longest streak is: 4  
> |
```

Q6. 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("M", "H")  
> sim_basket2 <- sample(outcomes, size = 133, replace = TRUE, prob = (c(0.45, 0.55)))  
>  
> sim_streak2 <- calc_streak(sim_basket2)  
> barplot(table(sim_streak2))  
> |
```



```
> barplot(table(sim_streak2))  
> cat("Longest Streak is:", max(sim_streak2))  
Longest Streak is: 6  
> |
```

After running the simulation again, it seems my hypothesis was correct - most streaks were around 0-1, with a few high streaks