Illapani_Lab2

Srini Illapani

September 12, 2015

Lab2 - Hot Hands

Basketball players who make several baskets in succession are described as having a *hot hand*. Fans and players have long believed in the hot hand phenomenon, which refutes the assumption that each shot is independent of the next. However, a 1985 paper by Gilovich, Vallone, and Tversky collected evidence that contradicted this belief and showed that successive shots are independent events (http://psych.cornell.edu/sites/default/files/Gilo.Vallone.Tversky.pdf). This paper started a great controversy that continues to this day, as you can see by Googling *hot hand basketball*.

We do not expect to resolve this controversy today. However, in this lab we'll apply one approach to answering questions like this. The goals for this lab are to (1) think about the effects of independent and dependent events, (2) learn how to simulate shooting streaks in R, and (3) to compare a simulation to actual data in order to determine if the hot hand phenomenon appears to be real.

Getting Started

Our investigation will focus on the performance of one player: Kobe Bryant of the Los Angeles Lakers. His performance against the Orlando Magic in the 2009 NBA finals earned him the title *Most Valuable Player* and many spectators commented on how he appeared to show a hot hand. Let's load some data from those games and look at the first several rows.

```
load("more/kobe.RData")
head(kobe)
```

In this data frame, every row records a shot taken by Kobe Bryant. If he hit the shot (made a basket), a hit, `H`, is recorded in the column named `basket`, otherwise a miss, `M`, is recorded.

Just looking at the string of hits and misses, it can be difficult to gauge whether or not it seems like Kobe was shooting with a hot hand. One way we can approach this is by considering the belief that hot hand shooters tend to go on shooting streaks. For this lab, we define the length of a shooting streak to be the *number of consecutive baskets made until a miss occurs*.

For example, in Game 1 Kobe had the following sequence of hits and misses from his nine shot attempts in the first quarter:

H M | M | H H M | M | M | M

To verify this use the following command:

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

Within the nine shot attempts, there are six streaks, which are separated by a "|" above. Their lengths are one, zero, two, zero, zero, zero (in order of occurrence).

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?

```
- It means atleast 1 basket was made, there is one hit and one miss in s
streak of 1.
A streak of 0 means, a basket was missed immediately after a previous 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?

```
- The distribution is right skewed. His typical (average) streak length is 1,
and his longest
streak of baskets is 4.
```

Simulations in R

If you wanted to simulate flipping a fair coin 100 times, you could either run the function 100 times or, more simply, adjust the `size` argument, which governs how many samples to draw (the `replace = TRUE` argument indicates we put the slip of paper back in the hat before drawing again). Save the resulting vector of heads and tails in a new object called `sim_fair_coin`.

```
outcomes <- c("heads", "tails")
sample(outcomes, size = 1, replace = TRUE)

## [1] "heads"

sim_fair_coin <- sample(outcomes, size = 100, replace = TRUE)
```

To view the results of this simulation, type the name of the object and then use `table` to count up the number of heads and tails.

```
sim_fair_coin

##   [1] "tails" "heads" "heads" "heads" "heads" "heads" "tails" "heads"
##   [9] "heads" "heads" "heads" "tails" "heads" "tails" "tails" "tails"
##  [17] "heads" "heads" "tails" "heads" "heads" "tails" "tails" "tails"
##  [25] "heads" "tails" "tails" "heads" "tails" "tails" "tails" "heads"
```

```
##  [33] "heads" "heads" "tails" "heads" "tails" "tails" "tails" "tails"
##  [41] "tails" "tails" "heads" "tails" "heads" "tails" "heads" "tails"
##  [49] "tails" "tails" "tails" "heads" "heads" "tails" "tails" "tails"
##  [57] "tails" "tails" "tails" "heads" "tails" "tails" "tails" "tails"
##  [65] "tails" "heads" "tails" "heads" "tails" "heads" "heads" "heads"
##  [73] "heads" "heads" "heads" "heads" "heads" "heads" "heads" "tails"
##  [81] "tails" "heads" "tails" "tails" "heads" "heads" "heads" "heads"
##  [89] "heads" "heads" "heads" "tails" "heads" "tails" "heads" "tails"
##  [97] "heads" "tails" "tails" "heads"
```

```
table(sim_fair_coin)
```

```
## sim_fair_coin
## heads tails
##    50    50
```

Since there are only two elements in `outcomes`, the probability that we "flip" a coin and it lands heads is 0.5. Say we're trying to simulate an unfair coin that we know only lands heads 20% of the time. We can adjust for this by adding an argument called `prob`, which provides a vector of two probability weights.

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

`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?

```
- 27
```

Simulating the Independent Shooter

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`.

```
- Add the 'prob' argument to the sample function with probabilty of Hits at
45%
and Misses as 55%.
```

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

With the results of the simulation saved as `sim_basket`, we have the data necessary to compare Kobe to our independent shooter. We can look at Kobe's data alongside our simulated data.

```r
sim_basket
```

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

Both data sets represent the results of 133 shot attempts, each with the same shooting percentage of 45%. We know that our simulated data is from a shooter that has independent shots. That is, we know the simulated shooter does not have a hot hand.

On your own

Comparing Kobe Bryant to the Independent Shooter

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

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

Q&A

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?

   - The distribution is right skewed. The typical (average) sstreak lenght is 1. The longest streak is 6.

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.

   - The distribution is right skewed, but the streak distribution is not the same. I now see a longest streak lenght if 7. The sampling is random and the streak outcomes will not be the same each time we run the simulation, especially if we are going for 133 shots.

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.

   - Comparing the two bar plots, the distribution pattern is similar but the streak lengths are not the same. There is more evidence of hot hand in the simulated shooter compared to Kobe, as I saw longer streak lenghts in the simulated model, and we know the simulated shots are independent. So Kobe does not have a hot hand, meaning his shots are not dependent (independent) of the previous shot.