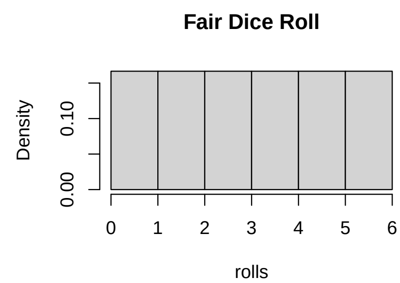
Snake Eyes and the Bell Shape

Soumya Ray

2019-02-24

Consider rolling a die (singular of dice). It produces a uniform distribution of values between 1 and 6, with each value having a 1/6 likelihood.

```
rolls <- sample(1:6, 6, replace = FALSE)
hist(rolls, prob = TRUE, col = "lightgray", breaks = 0:6,
    main = "Fair Dice Roll")
```



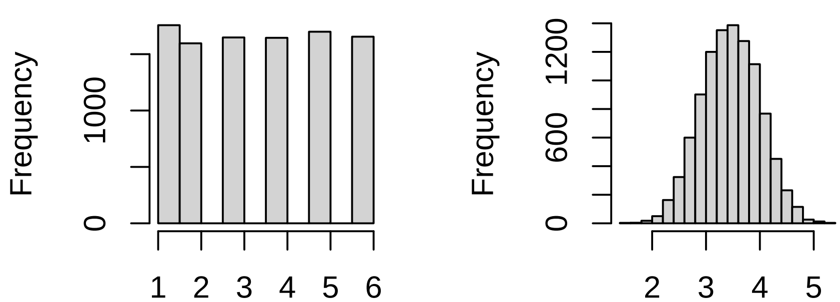
Now consider rolling two dice together and taking the mean of their values. Their mean would be between 1 and 6. But will the mean be uniformly distributed?

Let's simulate the rolling of multiple dice together and sum their results. We will do it with one dice and then with ten dice. For each set, let's repeat it 1000 times. Let's see how the total of each set of dice turns out over a 1000 rounds:

```
num_dice <- c(1, 10)
die_rolls <- function() {
    round(runif(10000, min = 0.5, max = 6.5))
}
dice_means <- function(num_dice) {
    rowMeans(replicate(num_dice, die_rolls()))
}

dice_games <- lapply(num_dice, dice_means)

par(mfrow = c(1, 2))
invisible(sapply(dice_games, hist, col = "lightgray",
    main = NULL, xlab = NULL))
```




```
par(mfrow = c(1, 1))
```

For one die, we see the results split evenly between all six possibilities. However, when we sum the rolls of two and then ten dice, we see the totals begin to take a bell-shape. Why does this bell-shape show up in aggregated distributions so often?

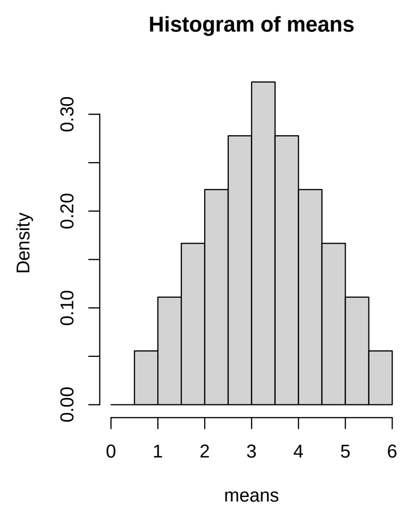
To investigate dice roll means in detail, let's list all possible means that *two* dice can possibly produce together:

```
mean_rolls <- matrix(NA, nrow = 6, ncol = 6)
for (i in 1:6) {
    for (j in 1:6) {
        mean_rolls[i, j] = mean(c(i, j))
    }
}
mean_rolls
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]  1.0  1.5  2.0  2.5  3.0  3.5
## [2,]  1.5  2.0  2.5  3.0  3.5  4.0
## [3,]  2.0  2.5  3.0  3.5  4.0  4.5
## [4,]  2.5  3.0  3.5  4.0  4.5  5.0
## [5,]  3.0  3.5  4.0  4.5  5.0  5.5
## [6,]  3.5  4.0  4.5  5.0  5.5  6.0
```

We can see that some means are easier to arrive at than others. Let's use a histogram to visualize the distribution of *possible means*:

```
bins <- seq(from = 0, to = 6, by = 0.5)
means <- mean_rolls[mean_rolls > 0]
hist(means, breaks = bins, prob = TRUE, col = "lightgray")
```



We can see that means towards the middle of the matrix, such as `3.0`, `3.5`, and `4.0` can be produced by many combinations of dice roll. For example, a mean of `3.0` can be gotten from 5 combinations of rolls: `[5,1]`, `[4,2]`, `[3,3]`, `[2,4]`, `[1,5]`. In contrast, means far from the center, can only be produced by fewer combinations. At the extreme corners, a mean of `6.0` can only produced by rolling `[6,6]`, while a mean of 1.0 requires snake eyes (`[1,1]`).

In a sense, this illustrates the central limit theorem. Values that are towards the center of possible outcomes are more likely to occur because there are more *random combinations* with which they can be produced. For many phenomena, this gives the impression that their means are converging in a near normal distribution towards central values.