R Tutorial-03

T. Scofield

You may [click here](#) to access the .qmd file.

In this issue, I simulate the probabilities of several events, namely

- being dealt 4-of-a-kind, in a 5-card hand
- getting between 40 and 60 heads when flipping a fair coin 100 times

Simulating the probability of 4-of-a-kind

Our **experiment**: deal 5 cards

Our **random variable**: The largest number of repeat cards

For the purposes of this simulation, the suit of a card is immaterial. So, the vector

```r
deck = rep(c(1:13), 4)
deck
```

```
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13  1  2  3  4  5  6  7  8  9 10 11 12
[26] 13  1  2  3  4  5  6  7  8  9 10 11 12 13  1  2  3  4  5  6  7  8  9 10 11
[51] 12 13
```

serves to simulate a deck of cards. We assume 5 cards are dealt without replacement, so we use `sample()` to simulate this.

```r
hand = sample(deck, size=5)
hand
```

```
[1]  7  9  9 12  3
```

Interpret this as a hand that contains a 7, two 9s, a queen and a 3. To use `tally()` on the hand reveals the same thing: a 3, a 7, two 9s and a queen.

```
tally(~hand)
```

```
hand
 3  7  9 12
 1  1  2  1
```

This hand had a pair of queens. We can use `max()` to see the largest number of repeated cards, the value of our random variable `X`; for the current hand, `X=2`.

```
max(tally(~hand))
```

```
[1] 2
```

We can use `replicate()` to preform any desired number of trials of this experiment.

```
N = 8
values.of.X.in.8.trials = replicate(N, max(tally(~ sample(deck, size=5))) )
values.of.X.in.8.trials
```

```
[1] 2 1 2 2 2 1 1 1
```

In those 8 trials, half of the hands had no pairs, and the other half all had at least one pair, but never 3 or more of the same card.

You may know the *event* "4-of-a-kind" is a rather rarely-occurring one. So, if we are going to approximate the probability of this event, we will need a lot of trials. Each trial results in a value of `X`, so I'll name the result, `manyXs`.

```
N = 200000
manyXs = replicate(N, max(tally(~ sample(deck, size=5))) )
prop(~(manyXs == 4))
```

```
prop_TRUE
 0.000305
```

Wikipedia says (correctly) the probability of 4-of-a-kind is 0.0002401. So, while we are in the right ballpark with this simulation (0.000305), we would need even more trials to tighten up our answer. Generally speaking, one needs extremely large numbers of trials such as this only when the actual probability we we are trying to approximate is near 0 or near 1. For most events, 3000 trials, or even fewer, is sufficient.

Simulating the number of heads in 100 flips

The **experiment**: flip a coin 100 times

The **random variable**: the number of heads

In the vector `coin`, 0 will represent the *tail* face and 1 the *head* face of the coin. We wish to sample from this vector with replacement, so I'll use the `resample()` command.

```
coin = c(0, 1)
singleRun = resample(coin, size=100)
singleRun
```

```
 [1] 0 1 1 1 0 1 0 1 0 0 1 0 0 0 1 1 1 0 1 1 1 0 0 1 0 1 0 0 0 1 0 1 0 0 0 1 1
[38] 1 0 1 0 1 1 1 1 0 1 1 0 0 0 1 1 1 0 0 1 1 1 0 0 1 0 0 0 0 0 1 1 1 0 1 0 1
[75] 0 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0 1 1 1 0 0 1 1 1 1 0
```

We determine the value of our random variable by summing the result of the run:

```
sum(singleRun)
```

```
[1] 48
```

I'll use `replicate()` to achieve a set number of trials, going in for 20000 of them:

```
N = 20000
manyXvals = replicate(N, sum(resample(coin, size=100)))
```

A visual display of the (approximate) **distribution** of X—which values occur, and how often— is provided by a histogram. The **mosaic** package provides `gf_histogram()` (one of many histogram plotters):

```
gf_histogram(~manyXvals, binwidth=1, color="black")
```

3
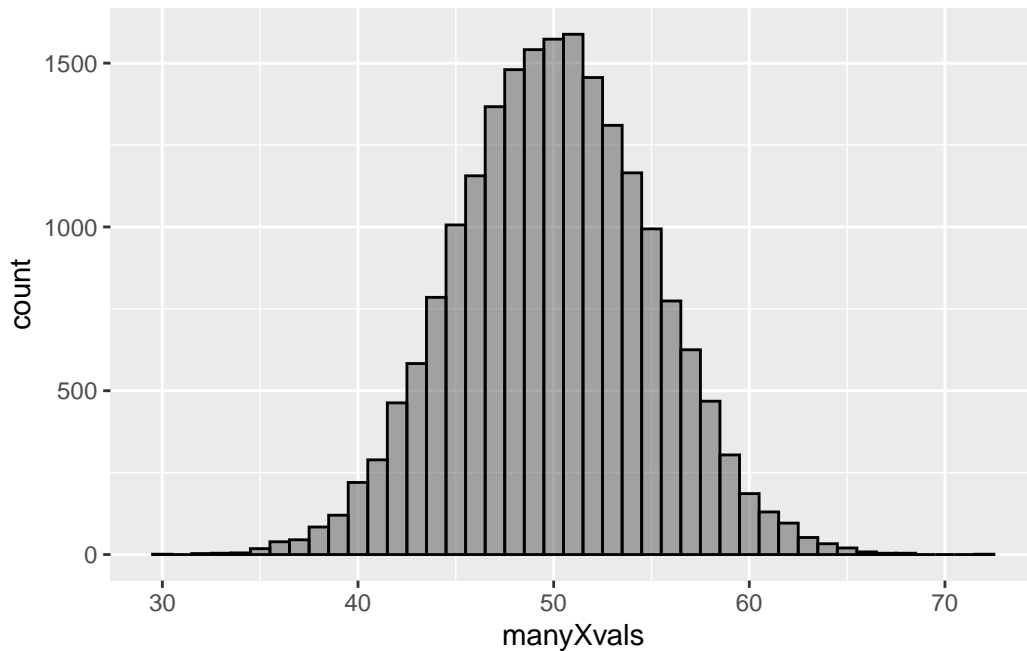
While we could divide the heights of these columns by 20000 to learn the approximate probabilities for various values of X, the goal was to simulate the probability of the *event* "get between 40 and 60 heads."

```
prop(~(manyXvals >= 40 & manyXvals <= 60) )
```

```
prop_TRUE
  0.96665
```

So, one gets between 40 and 60 heads (inclusive) in 100 flips approximately 96 or 97% of the time. Results, of course, vary with each simulation, and our hope is that 20000 trials is enough to make fairly stable this result.