

Let us first define our three sequences in R and then create a big matrix with all possible sequences:

```
A <- c("H", "T", "H", "T", "H", "T", "H", "T", "H", "T", "T", "H", "T",
"H", "H", "T", "T", "T", "H")
B <- c(rep("H", 19))
C <- c(rep("H", 10), rep("T", 9))
```

```
M <- as.matrix(expand.grid(rep(list(c("H", "T")), 19)))
head(M, 10)
##          Var1 Var2 Var3 Var4 Var5 Var6 Var7 Var8 Var9 Var10 Var11 Var12
Var13
## [1,] "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [2,] "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [3,] "H"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [4,] "T"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [5,] "H"  "H"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [6,] "T"  "H"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [7,] "H"  "T"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [8,] "T"  "T"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [9,] "H"  "H"  "H"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
## [10,] "T"  "H"  "H"  "T"  "H"  "H"  "H"  "H"  "H"  "H"  "H"  "H"
"H"
##          Var14 Var15 Var16 Var17 Var18 Var19
## [1,] "H"      "H"      "H"      "H"      "H"      "H"
## [2,] "H"      "H"      "H"      "H"      "H"      "H"
## [3,] "H"      "H"      "H"      "H"      "H"      "H"
## [4,] "H"      "H"      "H"      "H"      "H"      "H"
## [5,] "H"      "H"      "H"      "H"      "H"      "H"
## [6,] "H"      "H"      "H"      "H"      "H"      "H"
## [7,] "H"      "H"      "H"      "H"      "H"      "H"
## [8,] "H"      "H"      "H"      "H"      "H"      "H"
## [9,] "H"      "H"      "H"      "H"      "H"      "H"
## [10,] "H"      "H"      "H"      "H"      "H"      "H"
```

```
tail(M, 10)
##          Var1 Var2 Var3 Var4 Var5 Var6 Var7 Var8 Var9 Var10 Var11
Var12 Var13
## [524279,] "H"  "T"  "T"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524280,] "T"  "T"  "T"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
```

```

"T"      "T"
## [524281,] "H"  "H"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524282,] "T"  "H"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524283,] "H"  "T"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524284,] "T"  "T"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524285,] "H"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524286,] "T"  "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524287,] "H"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
## [524288,] "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"  "T"
"T"      "T"
##
      Var14 Var15 Var16 Var17 Var18 Var19
## [524279,] "T"   "T"   "T"   "T"   "T"   "T"
## [524280,] "T"   "T"   "T"   "T"   "T"   "T"
## [524281,] "T"   "T"   "T"   "T"   "T"   "T"
## [524282,] "T"   "T"   "T"   "T"   "T"   "T"
## [524283,] "T"   "T"   "T"   "T"   "T"   "T"
## [524284,] "T"   "T"   "T"   "T"   "T"   "T"
## [524285,] "T"   "T"   "T"   "T"   "T"   "T"
## [524286,] "T"   "T"   "T"   "T"   "T"   "T"
## [524287,] "T"   "T"   "T"   "T"   "T"   "T"
## [524288,] "T"   "T"   "T"   "T"   "T"   "T"

```

We see that there are exactly 524,288 sequences. That is because there are two possibilities for the first coin times two possibilities for the second times... and so on, which give $2^{19} = 524,288$ sequences altogether

Now let us find out how often each sequence appears:

```

sum(colSums(t(M) == A) == 19) # how many possibilities are exactly
equal to A
## [1] 1

sum(colSums(t(M) == B) == 19)
## [1] 1

sum(colSums(t(M) == C) == 19)
## [1] 1

```

Each sequence appears exactly once. The probability for each sequence therefore is...

```

1 / nrow(M)
## [1] 1.907349e-06

```

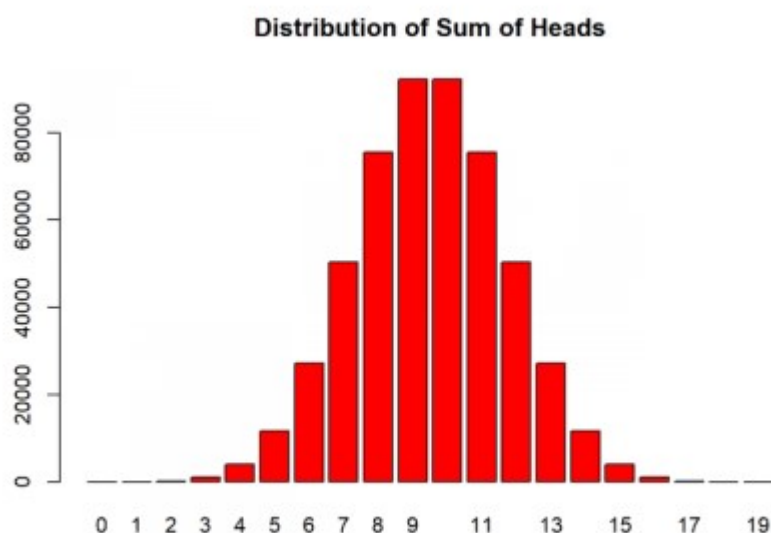
...which is exactly $1/2^{19}$ for all three cases because the coin tosses are independent of each other.

But why do so many people, also some renowned data scientists and even colleagues of mine,

chose A? My guess is that they confuse the actual sequences with some derived feature, like the sum of heads. Counting the number of heads in each sequence shows a completely different picture, which can also be derived analytically by using the binomial distribution:

```
tab <- rbind(counting = table(colSums(t(M) == B)), binomial =
dbinom(0:19, 19, 0.5) * nrow(M))
tab
##           0   1   2   3   4   5   6   7   8   9   10  11
12
## counting 1  19 171 969 3876 11628 27132 50388 75582 92378 92378 75582
50388
## binomial 1  19 171 969 3876 11628 27132 50388 75582 92378 92378 75582
50388
##           13   14   15   16   17  18  19
## counting 27132 11628 3876 969 171 19  1
## binomial 27132 11628 3876 969 171 19  1
```

```
barplot(tab[1, ], col = "red", main = "Distribution of Sum of Heads")
```



So, the probability of seeing *some mix* of “H” and “T” is much higher than seeing only heads or only tails... but “some mix” is something completely different than an *exact sequence* like A (which all have the same probability as we just saw).