dsmath-practice

1 Review DataCamp lesson 1 (binomial distribution)

1. What is the main theme of this lesson?

   - Probability simulation vs. exact computation of probabilities
   - Generating data from a given distribution `X = Binomial(size,p)`

2. What is "inference" vs. "probability"? Why is this important?

   - Inference means using observed data to build an underlying model (an abstraction from reality) - e.g. the *sample average*.
   - Probability means predicting data (events)from a model - e.g. *outcomes* of rolling a die with a probability of 1/6.
   - High tension between "data-centric" vs. "model-centric" people right now (resource struggle): e.g. "one-shot" vs. deep learning.
   - I've lately moved from the algorithmic/modeling into the data-first camp, against "black-box" algorithms.

3. What is a coin flip before you look at the outcome?

   Before looking at the outcome, the coin flip represents a random variable - i.e. the outcome of a single draw is not fixed (or is it?)

4. What is a coin flip once you've looked at the outcome?

   An event that is no longer random, or an observation that can be recorded.

5. What are the possible outcomes of `rbinom(1,1,0.5)`?

```
## outcomes: 0 (tails) or 1 (heads)
rbinom(n=1,  # number of coin flips
       size=1, # number of coins
       prob=0.5)  # probability of "heads"
```

```
[1] 1
```

6. What are the possible outcomes of `rbinom(1,10,0.5)`?

```
## outcomes = no. of heads when flipping 10 coins
rbinom(1,10,0.5)
```

```
[1] 7
```

7. What are the possible outcomes of `rbinom(10,10,0.5)`?

```
## no. of heads when flipping 10 coins 10 times in a row
r <- rbinom(10,10,0.5)
r
mean(r)
```

```
 [1] 8 5 2 6 4 2 7 5 4 4
[1] 4.7
```

8. What data structure is `rbinom`? What about `rbinom(1,1,0.5)`?

```
str(rbinom)
is.vector(rbinom(1,1,0.5))
```

```
function (n, size, prob)
[1] TRUE
```

9. What is an 'unfair' coin? How is this simulated in R?

```
## unfair coin = biased with prob \ne 0.5
rbinom(10,1,0.05)
```

```
[1] 0 0 0 0 0 0 0 0 0 0
```

10. What does "coin flip with probability 25%" mean? How to show this?

```
## prob = 0.25: only 25% of flips are expected to be 'heads'
flips <- rbinom(100000,1,0.25)
mean(flips)
```

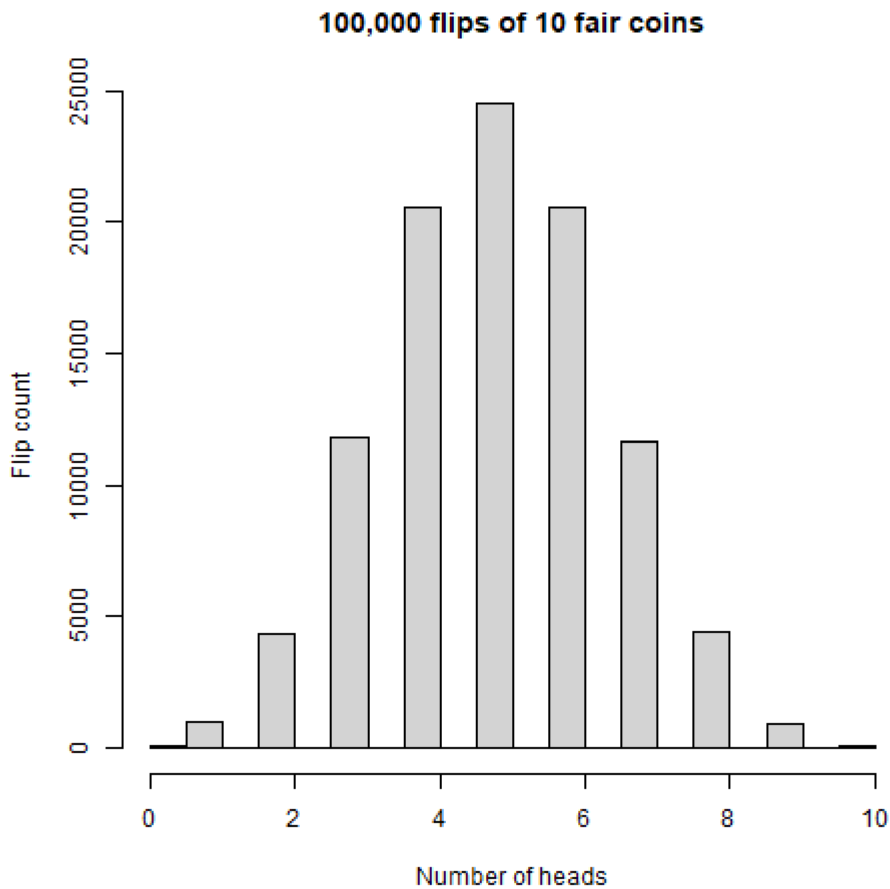
```
[1] 0.2514
```

11. Which plot type is used to visualize many coin flips?

    - The histogram - it plots counts/frequency of events against the event categories, e.g. for 10 coins the events `N \in {0,...,10}`.
    - The result is a distribution of probabilities `Pr(X=N)` - the random values are drawn from a binomial distribution

12. Plot 100,000 simultaneous flips of 10 fair coins!

```
flips <- rbinom(100000,10,.5)
hist(flips,
     main="100,000 flips of 10 fair coins",
     xlab="Number of heads",
     ylab="Flip count")
```



13. How can you inspect the frequencies without plotting them?

```
table(rbinom(100000,10,0.5))  ## see also prop.table
```

```

 0     1     2     3     4     5     6     7     8     9    10 
96  1001  4327 11614 20684 24671 20464 11770  4319   966    88 
```

14. What is the simulated density of the binomial distribution at `x=5`?

```
## density = 'degree of compactness' of information
flips <- rbinom(100000,10,0.5)
mean(flips==5)
```

```
[1] 0.2466
```

15. What is the exact binomial probability density at `x=5`?

```
## probability of getting 5 heads when flipping 10 fair coins
dbinom(x = 5, size = 10, prob = 0.5) # Pr(X=5)
mean(rbinom(100000,10,.5)==5) # simulation
```

```
[1] 0.2461
[1] 0.2473
```

16. What is cumulative density? How do you simulate/compute it?

```
## Pr(X \le 4) - simulation
flips <- rbinom(100000, 10, .5)
mean(flips <= 4)
## Pr(X \le 4) - computation
pbinom(4, 10, .5)
```

```
[1] 0.3757
[1] 0.377
```

17. What is the relationship between `Pr(X \ge 5)` and `Pr(X < 4)`?

```
##  Pr(X \ge 5) = 1 - Pr(X < 4)
1 - pbinom(4,10,.5) # Pr(X \ge 5)
pbinom(4,10,.5)     # Pr(X < 4)
1 - pbinom(4,10,.5) + pbinom(4,10,.5)
```

```
[1] 0.623
[1] 0.377
[1] 1
```

18. What's the relationship between sample average and expected value?
    - The expected value E[X] is the center of the distribution
    - The sample average is the arithmetic mean of all observations
    - As the sample size goes up, the mean approaches the expected value (= Law of large numbers).
19. What's the formula for the expected value, and how does it relate to the definition of the `binom` family of functions?

    E[X] = sample `size` $\times$ `prob` ability
20. What's the definition of the variance of a distribution?
    - The variance is the average squared distance of each value from the mean of the sample
    - `var(X) = ~size` $\times$ `prob` $(1 -$ `prob`$)$

Created: 2022-11-08 Tue 06:51