Algebraic & Geometric Methods in Statistics

Sampling distributions - an example

Sonja Petrović

Created for Math/Stat 561

Jan 18, 2023.

Warning

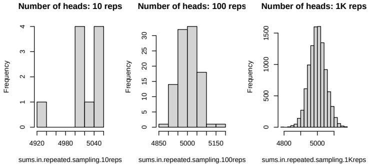
This is only a part of the lecture.

Why do we care about distributions?

Who cares about model fitting and testing whether we have the correct model in the first place?
Why do I have to understand a model?

Simulation of a coin toss

- Let x be the random variable recording the outcome of a coin toss:
  - $x_i = 0$ if we see Tail on the i-th trial (toss),
  - $x_i = 1$ if we see heads on the i-th trial.
- Fix $n = 10000$.
- Y = the number of heads.
  - Is the number of heads supposed to be $n/2$? How far off is it? Does it vary? What does this mean?



- Sampling distribution of Y appears to have a mean around *the expected number of heads when a fair coin is tossed*, which is about $n/2$.
- The more times we repeat the experiment of n coin tosses, the closer Y gets to its expected value – this can be measured by looking at both the mean and the variance of Y.

Means of Y
5020.600
5004.340
5000.699

Vars of Y
2535.015
2535.015
2461.334

Question:

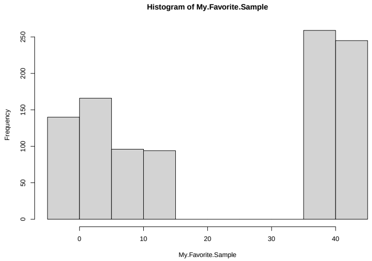
is it possible that something similar to this always happens?

- As we will see, the sampling distribution of Y is approximately *normal* with mean equal to the expected value of X.
- In other words, the example above illustrates a known result–**the Central Limit Theorem**, one of the cornerstone results used in inference.
- You should already be familiar with it from your probability class.

Importance of sampling distributions

- Sampling distributions tell a story about the model behind the data (i.e., the probability distribution or population from which the data was sampled);
- they give a glimpse into how it was generated.

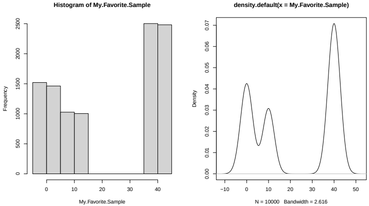
Example




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.5055  0.9193 37.4835 22.0968 39.9636 43.0376
```

Hmm. . .

- Is it strange to see "two bumps" in the histogram instead of one, as usual?
- Maybe the sample size is too small, we need to simulate more data?

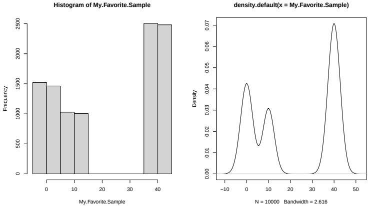


AhaMoment!

What do you see?

Hmm...

- Is it strange to see “two bumps” in the histogram instead of one, as usual?
- Maybe the sample size is too small, we need to simulate more data?



AhaMoment!

What do you see?

- This data is *not* being drawn from anything like a normal distribution.
- Consequently, knowing simply the mean and the variance ... is not enough to understand the data, that is, the data-generating mechanism behind it.

… Wait, what was that?!

This was an example of a **mixture** of normal distributions.

Check out the handout on Campuswire to see more, and ponder some important questions.

License

This document is created for Math/Stat 561, Spring 2023, at Illinois Tech.

Examples are drawn from other sources; for details see this file with full references.