Algebraic & Geometric Methods in Statistics

Sampling distributions - an example

Sonja Petrović
Created for Math/Stat 561

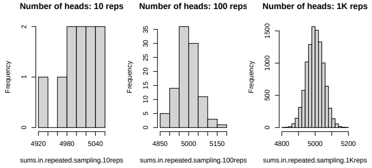
Jan 18, 2023.

PART 1: 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
5008.200
4995.350
5000.736

Vars of Y
2993.139
2993.139
2495.420

Question:

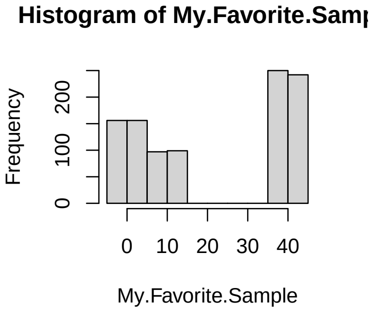
is it possible that something similar to this always happens?

- 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.

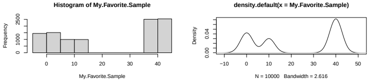
Example




```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-2.7620  0.8092 10.5707 21.6176 39.9569 43.4166
```

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?

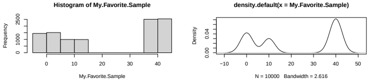


AhaMoment!

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. We will see mixture distributions again in the course, soon. *(See also link in "License" page at the end of these slides, which includes source information.)*

PART 2: conditional independence models

Material is from chapter 4 of the textbook.

Independence

Two independent discrete random variables

Let $X \subset [n] := \{1, \ldots, n\}$ and $Y \subset [m]$.

$$X \perp\!\!\!\perp Y \iff P(X = i, Y = j) = P(X = i)P(Y = J).$$

- In words, the joint probability factorizes as the product of the marginal probabilities. (Conditioning does not have an effect: recall definition of independent events from Lecture 2.)

Two independent continuous random variables

Let $X \subset \mathcal{X}$ and $Y \subset \mathcal{Y}$.

$$X \perp\!\!\!\perp Y \iff f_{X,Y}(x, y) = f_X(x)f_Y(y).$$

- In words, the joint density factorizes as the product of the marginal densities.

→ Extend definition of independence to *sets* of random variables. Example: 3 discrete random varaibles:

Independence

The distribution P is called *independent* if each probability is the product of the corresponding marginal probabilities:

$$P_{ijk} \;=\; P_{i++} \cdot P_{+j+} \cdot P_{++k}$$

Here, for instance,

$$P_{i++} \;= \mathrm{Prob}(X = i) \;=\; \sum_{j=1}^{b} \sum_{k=1}^{c} P_{ijk}$$

The independence model has the parametric representation

$$\begin{aligned} \Theta \;=\; \Delta_{a-1} \times \Delta_{b-1} \times \Delta_{c-1} \quad &\to \quad \Delta \;=\; \Delta_{abc-1} \\ (\alpha, \beta, \gamma) \quad &\mapsto \quad (P_{ijk}) \;=\; (\alpha_i \beta_j \gamma_k) \end{aligned}$$

Figure 1: Source: Bernd Sturmfel's invitatino to Alg Stats lecture, SAMSI 2008

Conditional independence

Recall the definition of a conditional probability from Lecture 2.

Definition 4.1.2. Let $A, B, C \subseteq [m]$ be pairwise disjoint. The random vector X_A *is conditionally independent of* X_B *given* X_C if and only if

$$f_{A\cup B|C}(x_A, x_B \mid x_C) = f_{A|C}(x_A \mid x_C) \cdot f_{B|C}(x_B \mid x_C)$$

for all x_A, x_B, and x_C. The notation $X_A \perp\!\!\!\perp X_B | X_C$ is used to denote that the random vector X satisfies the conditional independence statement that X_A is conditionally independent of X_B given X_C. This is often further abbreviated to $A \perp\!\!\!\perp B | C$.

Figure 2: Source: Algebraic statistics, Seth Sullivant, AMS-GSM book

Example: 3 r.v. and one CI statement

$X_A = X, X_B = Y, X_C = Z$ each vector is a single random variable.

$$f_{X|Y,Z}(x|y,z) = \frac{f_{X,Y|Z}(x,y|z)}{f_{Y|Z}(y|z)} = f_{X|Z}(x|z).$$

- Given Z, knowing X does not give any information about Y.
  $\rightarrow$ *Check discussion after definition 4.1.2, page 73 of the book!*

Marginal independence

Example 4.1.3 (Marginal independence). A statement of the form

$$X_A \perp\!\!\!\perp X_B := X_A \perp\!\!\!\perp X_B | X_\emptyset$$

is called a *marginal independence statement*, since it involves no conditioning. It corresponds to a density factorization

$$f_{A \cup B}(x_A, x_B) = f_A(x_A) f_B(x_B),$$

which should be recognizable as a familiar definition of independence of random variables, as we saw in Chapter 2.

Figure 3: Source: Algebraic statistics, Seth Sullivant, AMS-GSM book

- *Note:* Marginal independence is the same thing as independence of random variables: factorization of joint as product of marginal densities. We use the term 'marginal' when there are more random variables.

The mathematics behind conditional independnece

- Suppose a random vector X satisfies a set of conditional independence statements. Which other conditional independence relations must the same random vector satisfy?
- There is *no finite set of axioms* from which *all* conditional independence relations can be deduced.
- There are some easy conditional independence implications, which are called the conditional independence axioms or **conditional independence rules**.

Conditional independence axioms

Proposition 4.1.4. *Let $A, B, C, D \subseteq [m]$ be pairwise disjoint subsets. Then*

(i) *(symmetry)* $X_A \perp\!\!\!\perp X_B \,|\, X_C \implies X_B \perp\!\!\!\perp X_A \,|\, X_C$;

(ii) *(decomposition)* $X_A \perp\!\!\!\perp X_{B\cup D} \,|\, X_C \implies X_A \perp\!\!\!\perp X_B \,|\, X_C$;

(iii) *(weak union)* $X_A \perp\!\!\!\perp X_{B\cup D} \,|\, X_C \implies X_A \perp\!\!\!\perp X_B \,|\, X_{C\cup D}$;

(iv) *(contraction)* $X_A \perp\!\!\!\perp X_B \,|\, X_{C\cup D}$ *and* $X_A \perp\!\!\!\perp X_D \,|\, X_C \implies X_A \perp\!\!\!\perp X_{B\cup D} \,|\, X_C$.

Figure 4: Source: Algebraic statistics, Seth Sullivant, AMS-GSM book

Work time!

Complete the worksheet 1 handout in lecture.

Problem: complete the steps of the proof of the 4 CI axioms from this slide.

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. That document also contains important questions you may wish to think about.

The worksheet is from Kaie Kubjas, handed out in our class with her permission.