Bootstrap – Background and Intuition

SDS 291

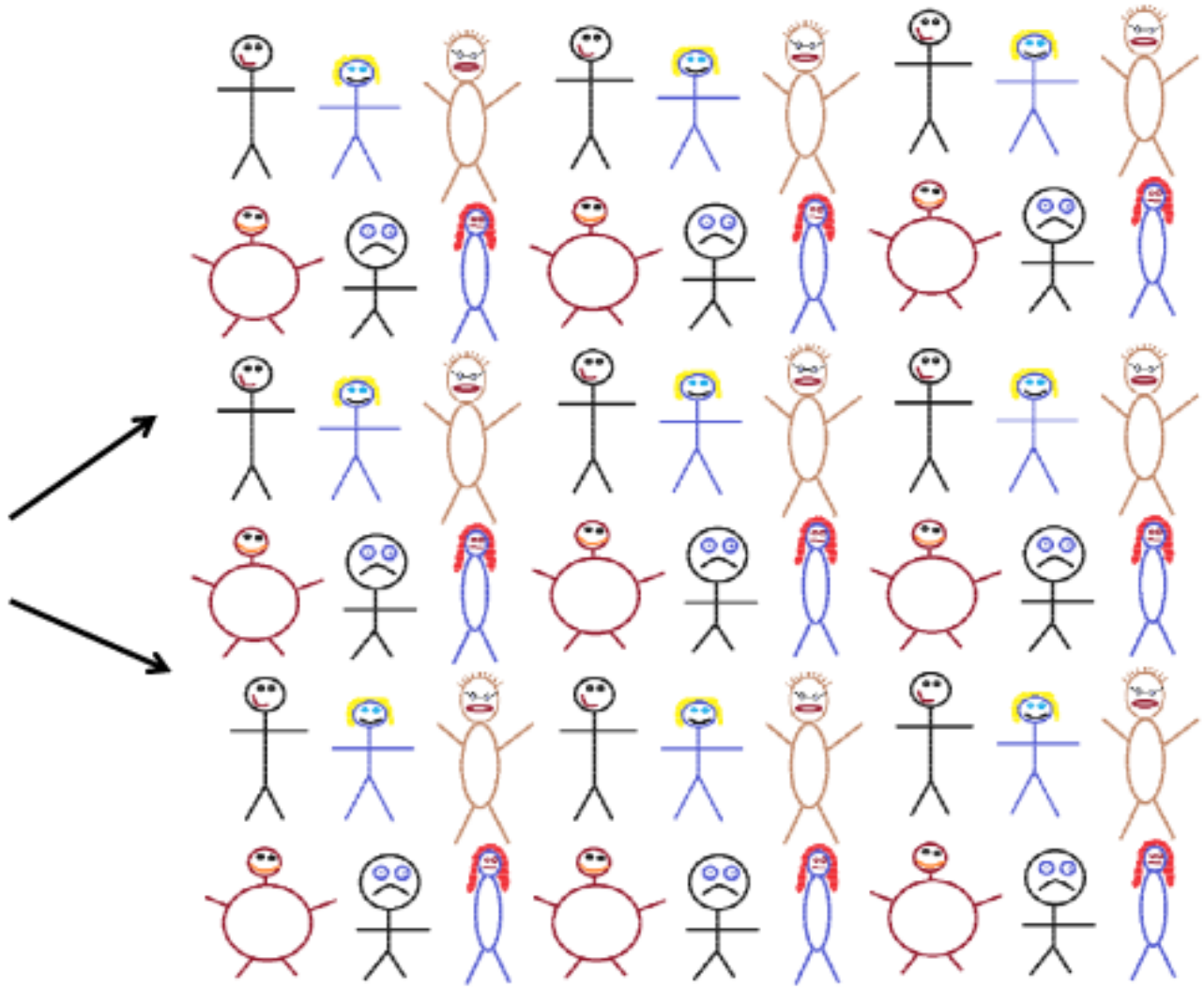
March 4, 2020

What's the problem?

- We want to make inferences to a larger population
- But!
  - We only have one sample
  - Some tests/statistics don't have standard deviations/errors to use
    - Or the assumptions of regression are violated
  - So we don't have a way to conduct formal tests, **construct CIs**, etc.

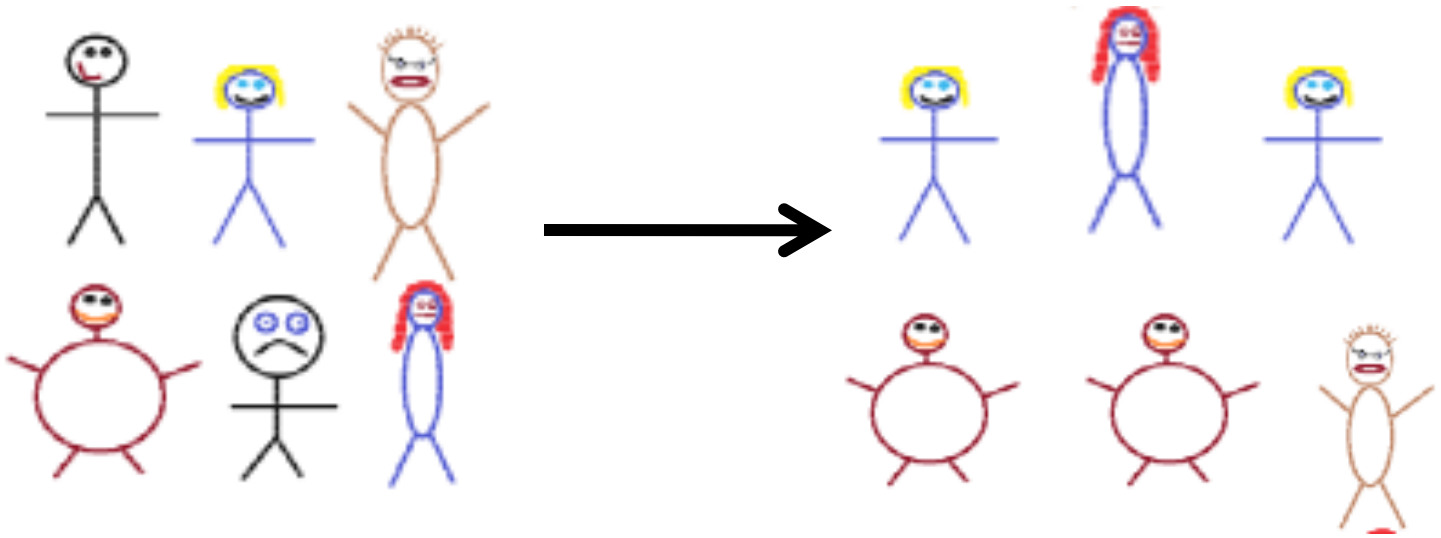
Suppose we have a random sample of 6 people:

Original Sample



A simulated “population” to sample from

Bootstrap Sample: Sample ***with replacement*** from the original sample, using the same sample size.



Original Sample

Bootstrap Sample

Let's make our own Bootstrap Sampling Distribution!

Our original sample has 2 people who have the flu and four people who don't.

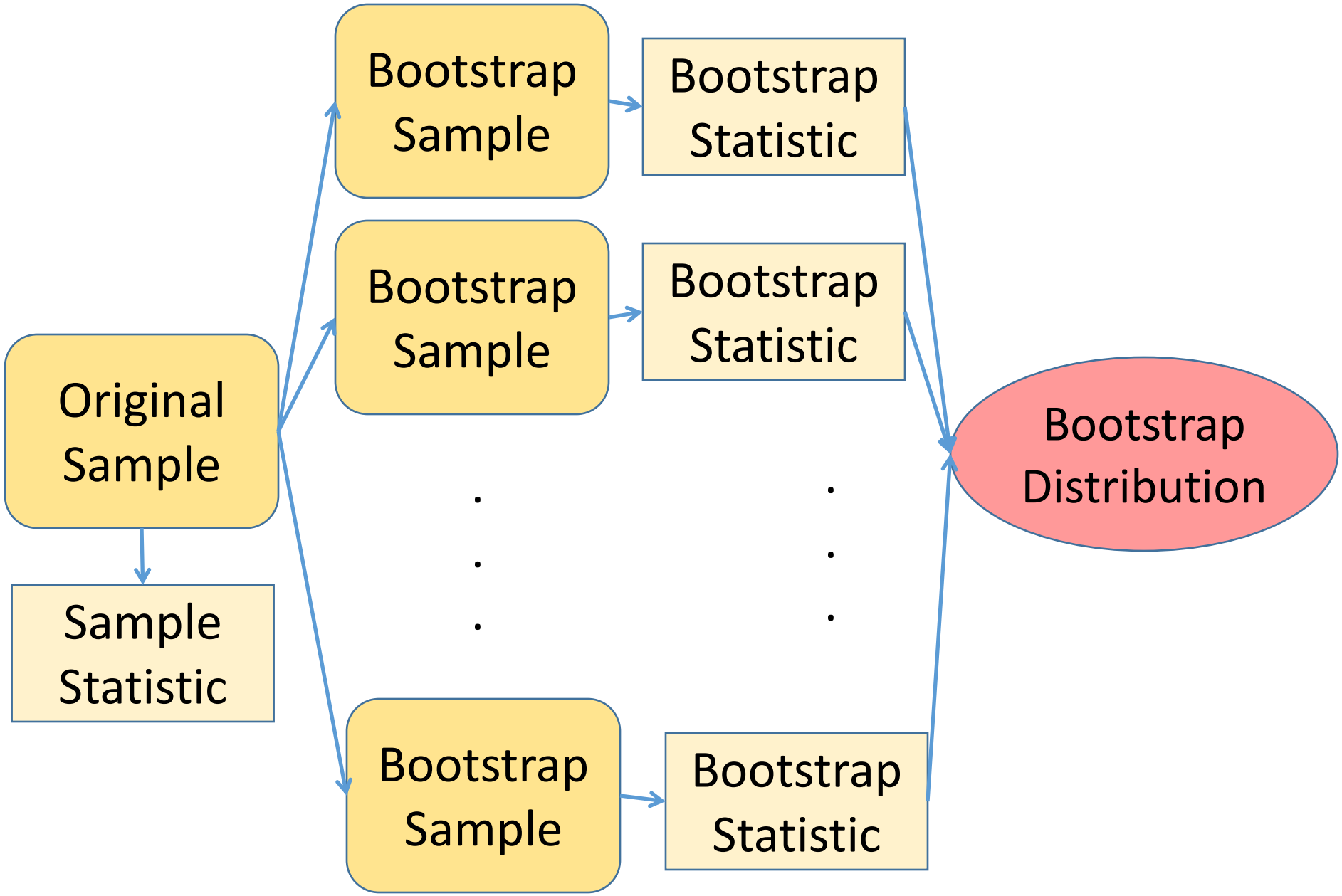
1. With a partner: Divide your paper into 6 even pieces. Write "Flu" on two of them, and "no flu" on 4 of them.
2. Take a sample of size 6 with replacement from your cards, and record what proportion have the flu in this new bootstrap sample
3. Take a total of 5 bootstrap samples and calculate 5 proportions of flu prevalence in the sample.
4. Make a dot for each of your 5 proportions on the board

Bootstrap

A ***bootstrap sample*** is a random sample taken with replacement from the original sample, of the same size as the original sample

A ***bootstrap statistic*** is the statistic computed on a bootstrap sample

A ***bootstrap distribution*** is the distribution of many bootstrap statistics


Original Sample
Sample Statistic
Bootstrap Sample
Bootstrap Statistic
Bootstrap Sample
Bootstrap Statistic
Bootstrap Sample
Bootstrap Statistic
Bootstrap Distribution