Smallest font

Calibration slide

Smallest font

Scientific Programming and Computing for Vision and Neuroscientists

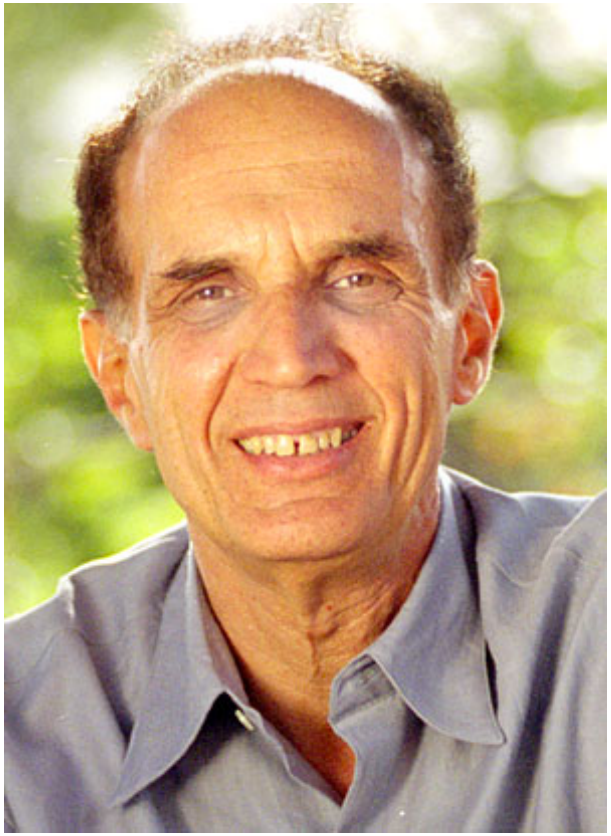
Resampling methods
The Bootstrap
Permutation tests

Resampling methods

- You are interested in whether adding lithium to the water supply reduces the suicide rate.
- So you do a trial. Adding lithium to the water supply of randomly selected households, but not others.
- After 10 years, there are 104 suicides among the 11,037 households that were lithium treated, but 189 suicides among the 11,034 untreated households.
- This is a ratio of 0.55 – in other words, lithium seems to half the suicide rate.
- At this point, the FDA stops the trial because it would be unethical to withhold lithium from the untreated households.
- But suicides are – thankfully – rare events. We have no idea how this ratio distributes – and whether this result is statistically reliable.
- In other words, was the FDA justified in stopping the trial and how would you know?

Enter the bootstrap

Bradley Efron

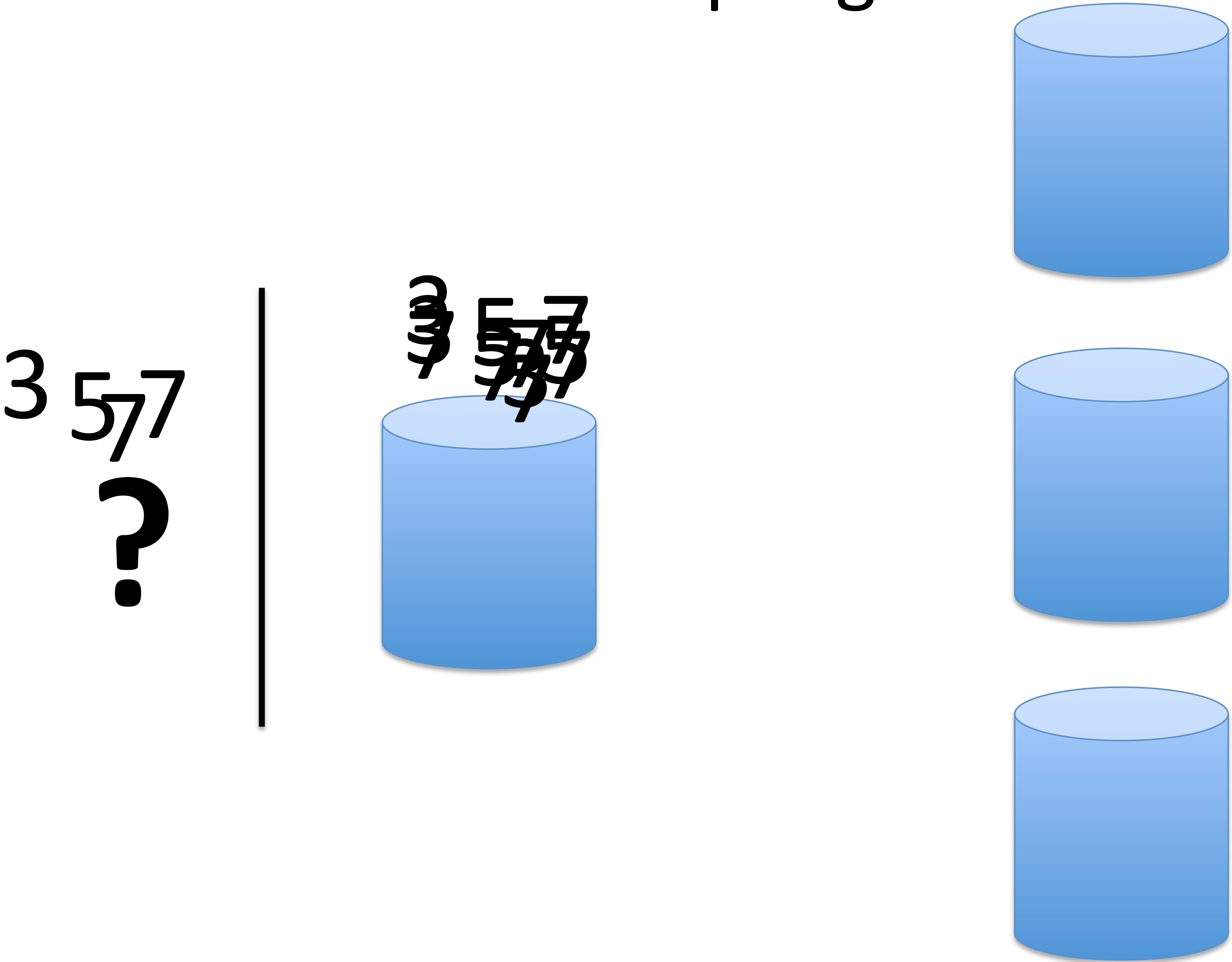


The computer

The procedure is quite straightforward

- To obtain a sense of how stable a sample mean is (how it distributes).
- Simply *resample* (with replacement!) from the existing sample.
- Do this n times (10k → 100k → 1m)
- Calculate the standard deviation of these resampled means.
- Compare with the figure of interest.

Bootstrap logic



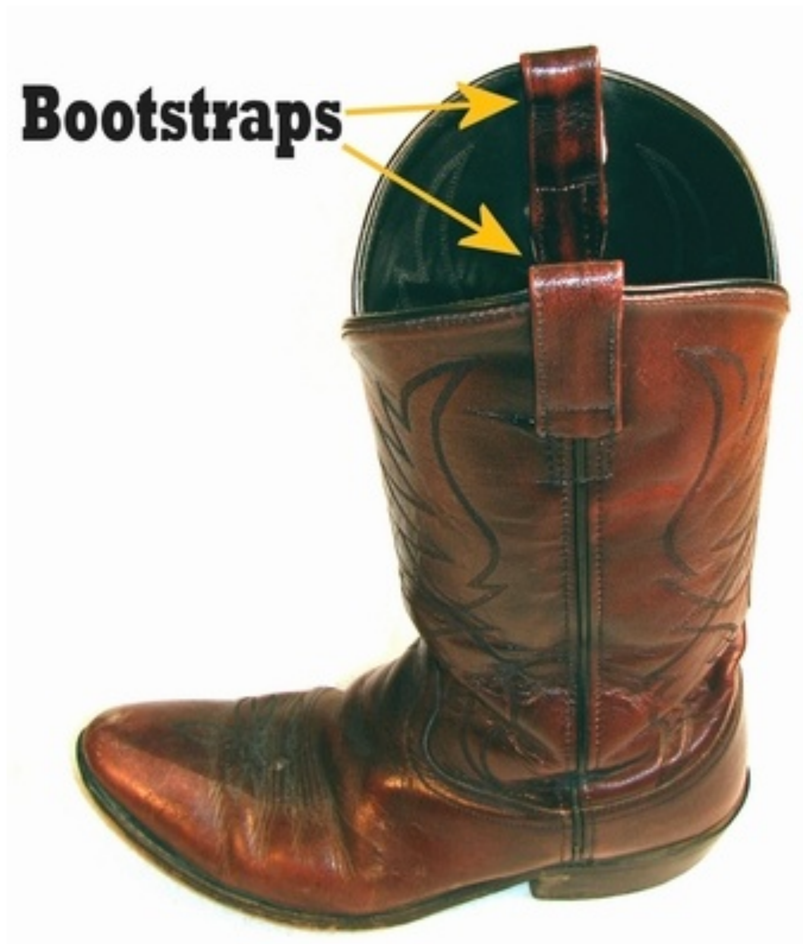
Use cases

- When there is only limited data (controversial)
- When the underlying distribution is not normal and/or not known
- When estimating sample means of rare events

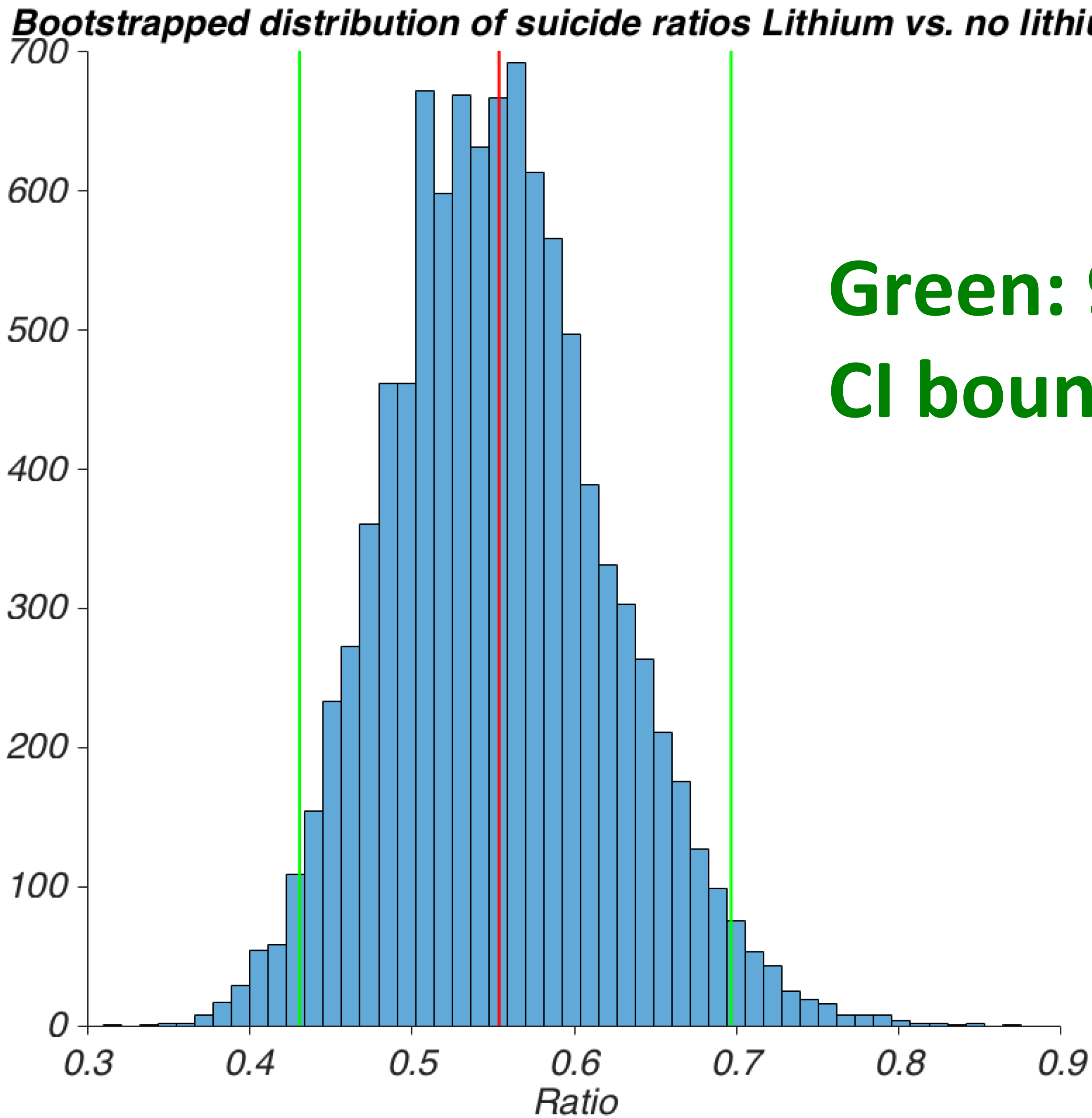
Bootstrap drawback/assumption

- This works if and only if the sample truly is representative of the population.
- In other words, if something didn't happen in the sample, it can't ever happen.
- "The data we have is the only data that can ever be"
- So small samples necessarily over- or underestimate the probability of rare events.

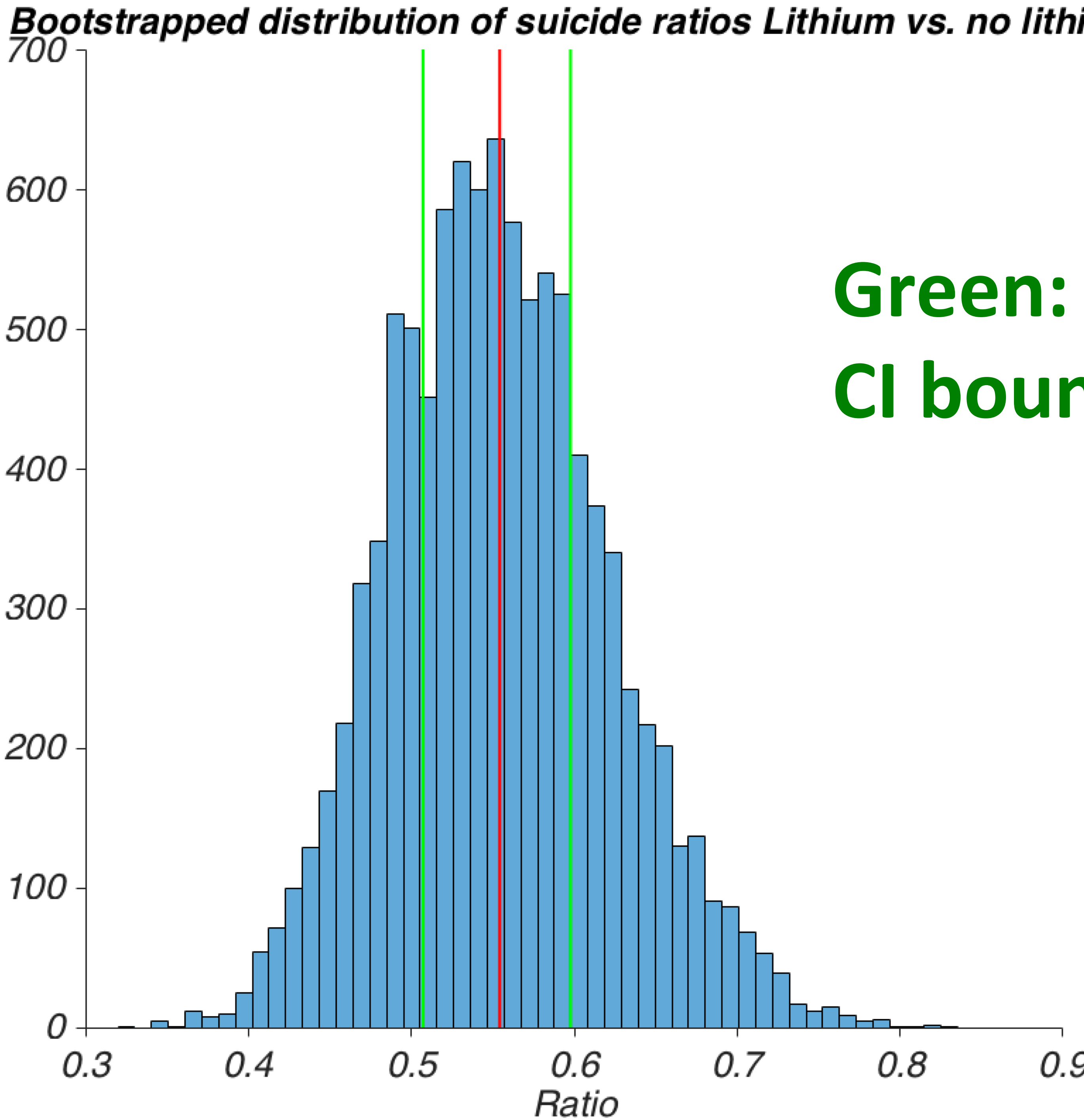
It seems like a miracle



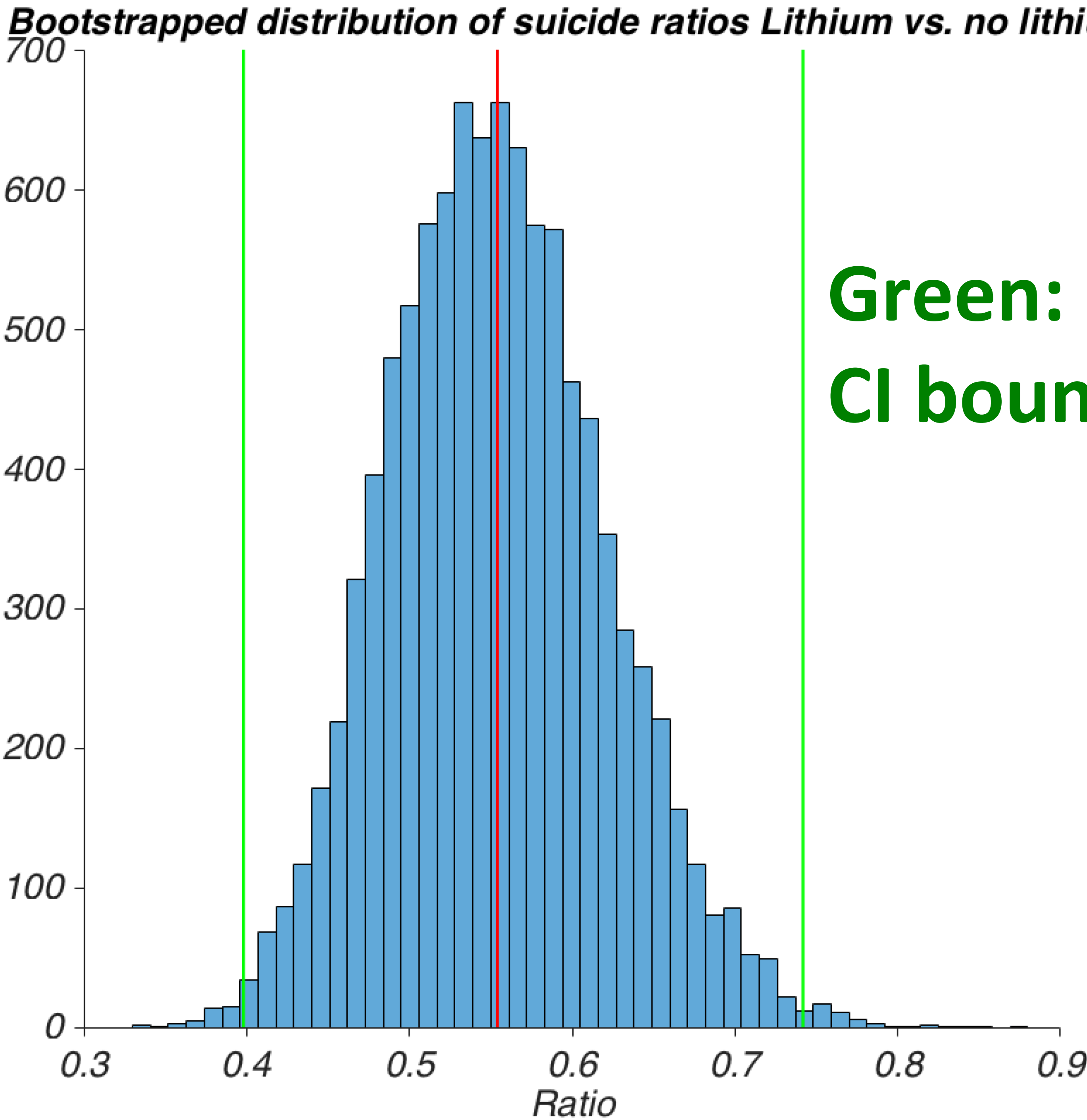
But it works (usually)!



Green: 95% CI bounds



Green: 50% CI bounds



Green: 99% CI bounds

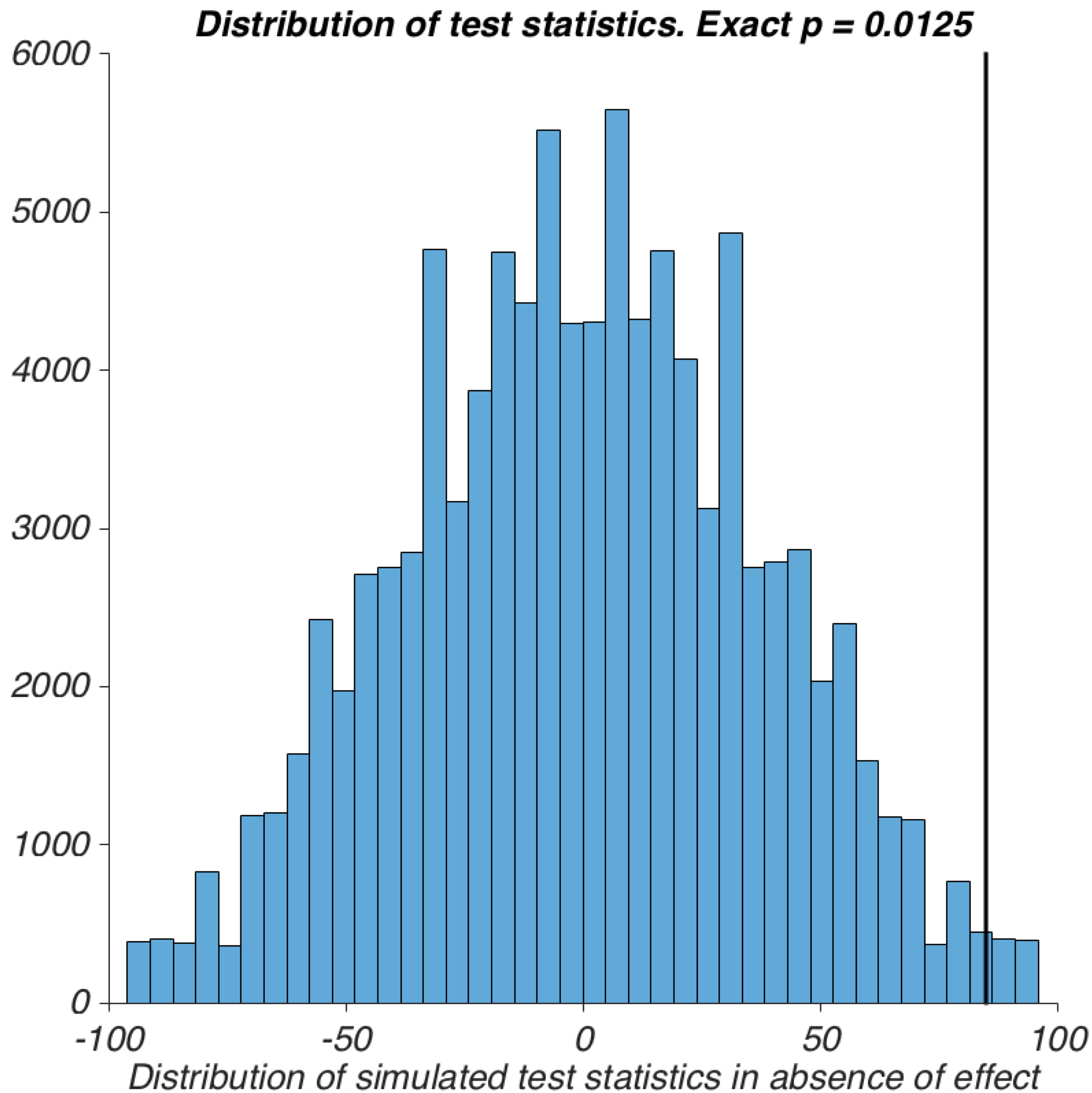
Permutation tests

- Classical tests (e.g. t-test) assume that the data is distributed in a certain way.
- If the distribution of the data is not what is assumed, the reported p-value by the test is not the real p-value (!)
- Permutation tests use the actual data to estimate how likely a given result is.
- Logic: We pretend we lost the labels (which group data came from), then create a null distribution (by random arrangement of groups) and compare with empirical result.

Example

- A rat is stressed out for 2 weeks
- 10 neurons are then taken out
- 5 we treat with Ketamine and 5 we don't treat at all
- We then count the number of dendritic spines of each neuron
- Hypothesis: Ketamine works by growing dendritic spines
- K = [117   123   111   101   121]
- C = [98   104   106   92   88]
- Test statistic: sum(K) – sum(C)
- We can now calculate exact p-value by determining the null distribution through resampling methods.

The result



Distribution of test statistics. Exact p = 0.0125
6000
5000
4000
3000
2000
1000
0
-100
-50
0
50
100
Distribution of simulated test statistics in absence of effect