To get challenging sampling distributions:

1. Randomization (permutation) ← discussed last video (and in chapter 13)
2. Simulation (Chapter 19)

To get an estimate of precision

1. Bootstrapping (Chapter 19)

Simulation is an important topic in biology. I appreciate the need of W&S to distill the field to a handful of tractable examples but, in case you are interested, I have found a couple of blog posts (because they are free and available) about Markov Chains (I can't emphasize this enough: REALLY important in Biology, Data Science and, yes, even A.I.)

Although not covered in this chapter, here are some good places to start if you are interested in visualizing/understanding these probability chains.

1. https://en.wikipedia.org/wiki/Markov_chain#Biology
2. http://setosa.io/ev/markov-chains/
3. Markov Chains (explains nuances of improvements of Metropolis-Hasting algorithm): https://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/
4. Here is a bunch of visualizations to different types of simulation: https://chi-feng.github.io/mcmc-demo/

Simulation:

Determine the null distribution (from the parameter expectations of the null hypothesis) by simulation of the sampling process

5 main steps

1. **Create and sample imaginary population**
   - parameters specified by null hypothesis
   - Same protocol that was used to collect real data
2. **Calculate test statistic on simulated sample**
3. **Repeat many times**
4. **Form the null distribution**
   - Gather simulated values for the test statistic
5. **Compare test statistic from the actual data to the null distribution**

Drawback --> **more work!**

Simulation:

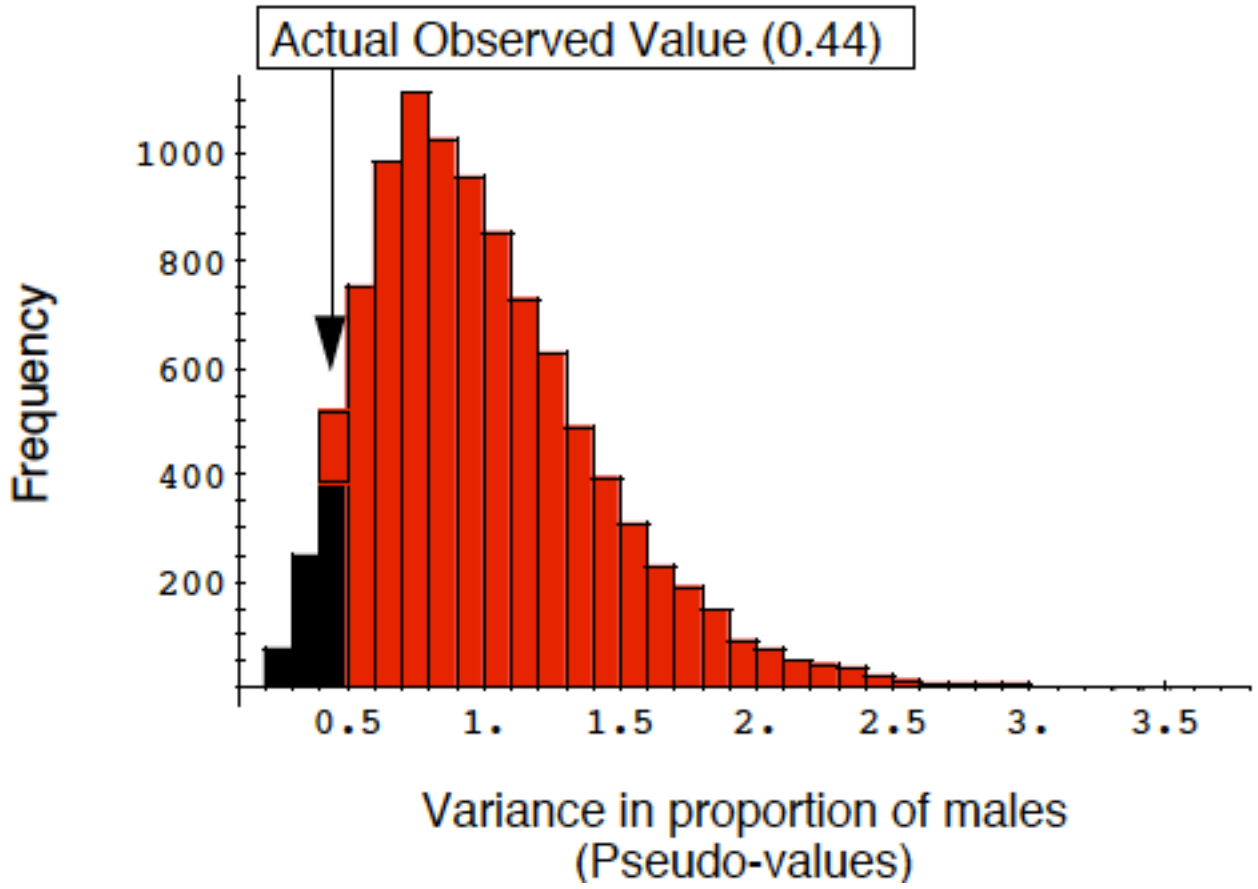
Social spider sex ratio

- Groups are mostly female
- Hypothesis: groups have just enough males to allow reproduction
- Test: whether distribution of number of males is as predicted by chance

→ H_0: variance of males is predicted by overall proportion

- Simulation:
  - Each group has a known number of spiders
  - Overall proportion of males is known, p_m
  - For each group, the computer draws the real number of spiders, and each has p_m probability of being male
  - The variance in proportion of males is calculated
  - This is repeated a large number of times

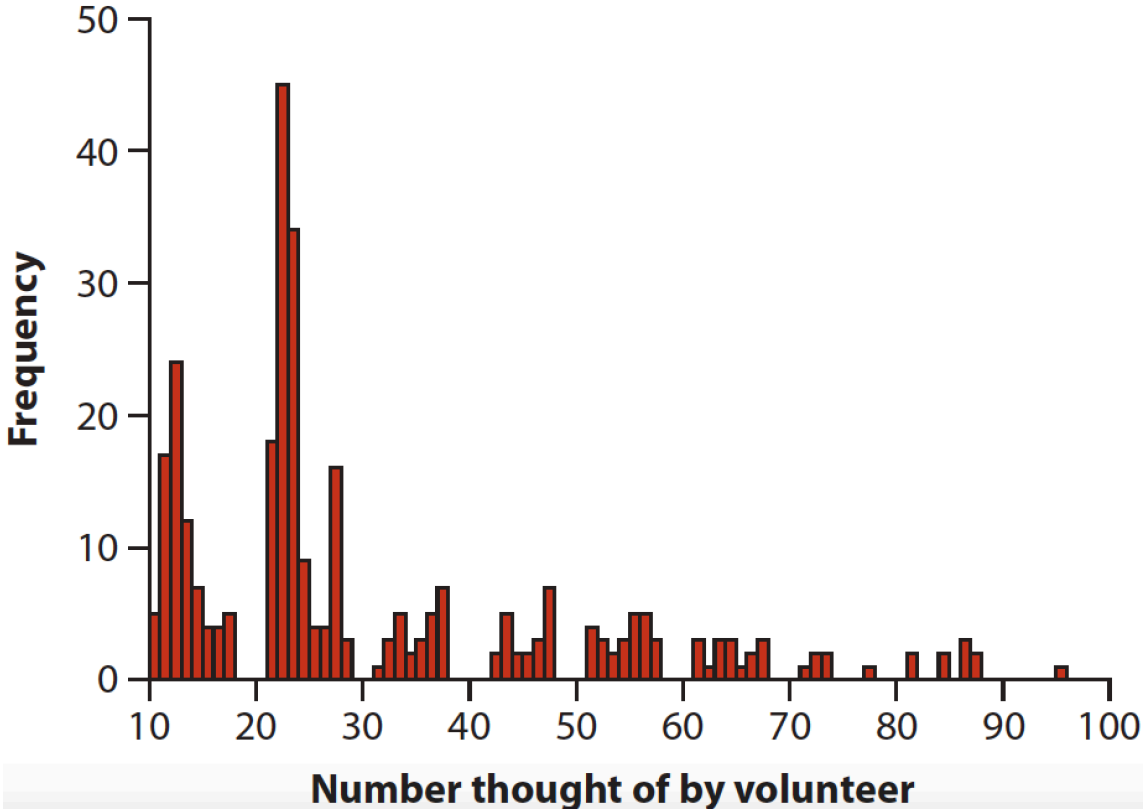
Simulation:
Social spider sex ratio



Simulation:

How did he know?

- People think of two digit numbers with equal probability
- Ho: Two-digit numbers are chosen with equal probability



Simulation:

How did he know?

- Can't use X^2 distribution because expected frequency of each of the 90 categories would be 3.89 (Violation of assumptions: can't have more than 20% of categories have expected values less than 5).

- Can still use X^2 test but not use X^2 distribution

  - Calculate test statistic (X^2 =1111.4)
  - Create null distribution
  - Sample many times from null distribution, calculate X^2 test

Simulation:

How did he know?

- Can still use X^2 test but not use X^2 distribution
  - Calculate test statistic (X^2 =1111.4)
  - Create null distribution
  - Sample many times from null distribution, calculate X^2 test