EGM101 – Skills Toolbox

Week 7, Part 1: Introduction to Probability

Week 7 Outline

1. Introduction to Probability
2. More Probability
3. Even More Probability!
4. Discrete Probability Distributions
5. Continuous Probability Distributions
6. The Central Limit Theorem

Question

Let’s say that we know that on average, salmon in our favorite stream are around 70 cm long (with a standard deviation of 15 cm)

If we go fishing in our favorite stream, what is the likelihood of catching a fish 100 cm or longer?

Inferential Statistics and Uncertainty

- Recall:
  - Descriptive statistics: organizing and summarizing data
  - Inferential statistics: drawing conclusions from “good” data
- Inherently, there is uncertainty in inferential statistics
- Probability theory helps us:
  - Understand uncertainty
  - Evaluate likelihood
- In other words, probability theory is the foundation for inferential statistics

Some Definitions

- (Statistical) Experiment: the process by which an observation is made
  - Measurements
  - Laboratory experiments
  - Clinical trials
  - Coin flip*
- Outcome: a result of an experiment
- Sample space: the set of all possible outcomes of an experiment
- Event: a *subset* of the sample space
- Probability: the *long-term* relative frequency of an event
  - Remember: relative frequency is always between 0 and 1
  - $P(A) = 0$: impossible for event A to occur
  - $P(A) = 1$: certain that event A will occur

H

T

$$S(1\text{ flip}) = \{H, T\}$$

$$S(2\text{ flips}) = \{HH, TT, HT, TH\}$$

$$P(HH) = \frac{1}{4} = 0.25$$

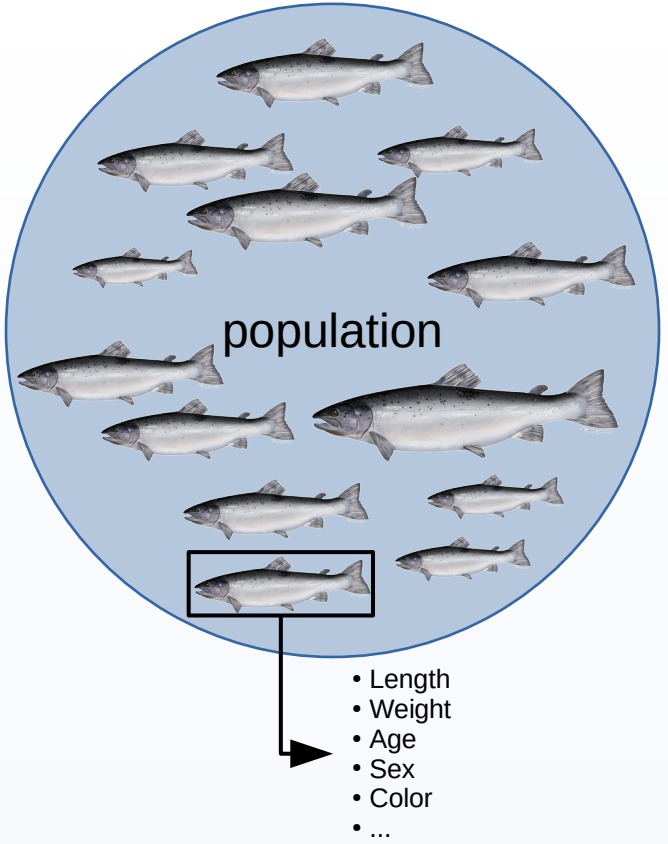
*Like ice cream and shark attacks, I believe that this is a legally-mandated example.

Deterministic, Probabilistic, and Random

- Deterministic: outcome is certain
- Probabilistic/random: outcome is not certain
  - i.e., outcome is not predictable/can't be known beforehand
  - Recall: in a random sample, *all members of population have equal chance of selection*
  - Fair: each outcome has the same probability (equal likelihood)
- Important: referring to individual outcomes, not values

Random Variables

- Recall:
  - Variable: some characteristic that will have different values for each member of the population
- Random variable: the variation in the value of the variable is random (not predictable)
- Examples:
  - Coin flips
  - Dice roll
  - Weight of a *randomly caught* fish
- Note: most of the time, we will assume that we are working with a random variable*



*we do this at our own peril.

The Expected Value

- Expected value of variable X, $E[X]$:
  - A *weighted* average of possible values of X
  - In other words, a measure of central tendency
  - In practice, expected value is just the arithmetic mean
- Example: what is the expected value of rolling a (fair) six-sided die (D6)?

$$E[X] = \sum_{i=1}^{n} x_i P(x_i)$$

$$E[\mathrm{D6}] = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right)$$

$$= \frac{1+2+3+4+5+6}{6}$$

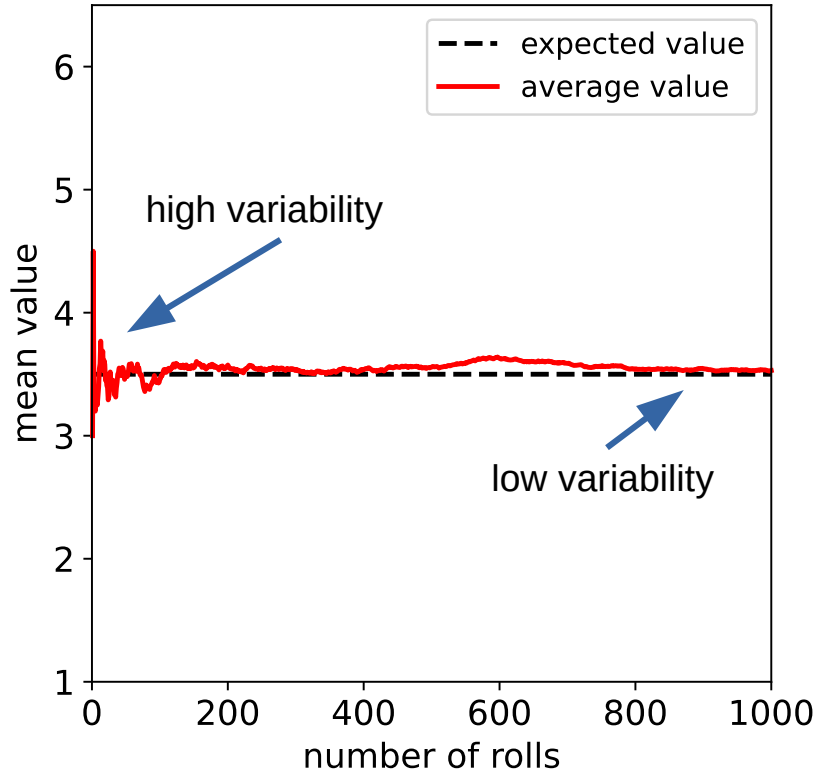
$$= \frac{21}{6} = 3.5$$

The Law of Large Numbers

- What happens to the sample mean as we repeat an experiment (i.e., the number of trials increases)?
- Law of Large Numbers:

  As the number of trials increases, the probability that the sample mean of the random variable equals the expected value within some margin approaches 1.

- Put another way:
  - As number of trials (or samples, measurements…) increases, the average outcome looks more and more like the expected value.
  - Or: the average of many measurements is more accurate than a single measurement.
- Example: D6 rolls



Summary

- Inferential statistics is inherently uncertain
- Probability theory helps us evaluate, understand uncertainty
- Most of the time, we can treat observations/measurements as random variable
  - Though it's always good to check this assumption!
- Law of Large Numbers: average of many observations tends to be more accurate than a single observation

Additional Resources

- Illowsky and Dean, Chapter 3.1
- Caswell, Chapter 12
- Weiss, Chapter 5.1
- Probability Definition and Fundamentals [Jim Frost]