JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Lecture 1

1. Cover syllabus
    a. mathematical prerequisites
    b. web site
    c. quiz and homework schedule
    d. test schedule
    e. R
2. Abstract the idea of an experiment
3. Develop basic set theory to be used in the development of probability
4. Start discussing probability

Experiments

Consider the outcome of an **experiment** such as:

- a collection of measurements from a sampled population
- measurements from a laboratory experiment
- the result of a clinical trial
- the result from a simulated (computer) experiment
- values from hospital records sampled retrospectively
- ...

Notation

- The **sample space**, Ω, is the collection of possible outcomes of an experiment

  Example: die roll $\Omega = \{1, 2, 3, 4, 5, 6\}$

- An **event**, say E, is a subset of Ω

  Example: die roll is even $E = \{2, 4, 6\}$

- An **elementary** or **simple** event is a particular result of an experiment

  Example: die roll is a four, $\omega = 4$

- $\emptyset$ is called the **null event** or the **empty set**

Interpretation of set operations

Normal set operations have particular interpretations in this setting

1. $\omega \in E$ implies that E occurs when ω occurs
2. $\omega \notin E$ implies that E does not occur when ω occurs
3. $E \subset F$ implies that the occurrence of E implies the occurrence of F
4. $E \cap F$ implies the event that both E and F occur
5. $E \cup F$ implies the event that at least one of E or F occur
6. $E \cap F = \emptyset$ means that E and F are **mutually exclusive**, or cannot both occur
7. E^c or $\bar{E}$ is the event that E does not occur

Fun aside: Russell's paradox

Russell's paradox is one of the most famous results of set theory

- Consider, R, the set containing all sets that do not contain themselves as an element

  Alternatively, consider writing down a catalog of all catalogs who do not have themselves listed as an entry

- Does R contain itself?

  → If yes, then R is not allowed to be in R, by the definition

  → If no, then R has to be in R, by the definition

Set theory facts

- DeMorgan's laws

  $(A \cap B)^c = A^c \cup B^c$

  $(A \cup B)^c = A^c \cap B^c$

  Example: If an alligator or a turtle you are not[$(A \cup B)^c$] then you are not an alligator and you are also not a turtle ($A^c \cap B^c$)

  Example: If your car is not both hybrid and diesel [$(A \cap B)^c$] then your car is either not hybrid or not diesel ($A^c \cup B^c$)

- $(A^c)^c = A$
- $(A \cup B) \cap C = (A \cap C) \cup (A \cap B)$

Probability: some discussion

- Useful strategy used in much of science:
  For a given experiment
  - attribute all that is known or theorized to a mechanistic model (mathematical function)
  - attribute everything else to randomness, *even if the process under study is known not the be "random" in any sense of the word*
  - Use probability to quantify the uncertainty in your conclusions
  - Evaluate the sensitivity of your conclusions to the assumptions of your model

Probability: some discussion

- Probability has been found extraordinarily useful, even if true *randomness* is an elusive, undefined, quantity
- *frequentist* interpretation of probability
  - A probability is the long proportion of times an event will occur in repeated identical repetitions of an experiment
- Other definitions of probability exists
- There is not agreement, at all, in how probabilities should be interpreted
- There is (nearly) complete agreement on the mathematical rules probability must follow

Probability: some discussion

- An alternative interpretation of probability is so-called "Bayesian"
- Named after the 18^{th} century Presbyterian Minister / mathematician Thomas Bayes
- Bayesian interprets probability as a subjective degree of belief
  - For the same event, two separate people could have differing probabilities
  - Bayesian interpretations of probabilities avoid some of the philosophical difficulties of frequency interpretations