Notes

Week 2

Module 1 Week 2A

Probability and Random Variables

- *Probability Theory* - important in linking *Samples* to underlying *Population* of Interest => important for *Statistical Inference*. ##### Probability
- Values taken on by the variables are not known until they are observed.
- Data is composed of observations of underlying *Random Variables*.
- *Probability Theory* provides the mathematics of this randomness.
- **Probability** - fraction of occurrences an event occurs over many repeated trials.
  - E.g. The probability of a coin flip coming up heads is the fraction of heads that occurs after many, many flips.
  - Note the long-run/repeated trial nature of the concept.
  - This is a *Frequentist* interpretation of *Probability*, adopted in this class.
    - Limitation: need for an objective/repeatable experiment.
    - Alternative: *Bayesian Probability* which views Probabilities as a plausible expectation that an event occurs, which can updated as more information becomes available (critiques cite its subjective nature).
  - Terminology (using the coin flip example):
    - **Experiment** - flipping of the coin.
      - is repeated and conducted under controlled conditions.
    - **Outcome** - result of a single experiment; not known until the coin is flipped; there's a chance involved.
    - **Sample Space** (S) ($\neq$ *Sample* from a *Population*) - all possible outcomes; e.g. $S = \{H, T\}$.
    - **Event** - any combination of outcomes.
      - Denoted with upper-case letters. E.g. A - event of Heads, then $P(A)$ - *Probability* of Heads.
      - $0 \leq \mathbb{P}(A) \leq 1$
      - **Law of Large Numbers** - as the number of trials increases, the *empirical* fraction of occurrences gets closer and closer to the theoretical probability of occurrence; e.g. for an unbiased coin, there more flips you observe, the fraction of tails or heads will get closer to 0.5.
- **Probability Rules**:
  - "and": an outcome is in event A and B if it is in A and B at the same time.
    - $A = \{1, 2, 3, 4, 5, 6\}$.
    - $B = \{4, 5, 6, 7, 8\}$.
    - A and $B = A \cap B = \{4, 5, 6\}$.
  - "or": an outcome is in event A or B if it is in A or B.
    - A or B $= A \cup B = \{1, 2, 3, 4, 5, 6, 7, 8\}$
  - The complement of event A is denoted A'
    - $\mathbb{P}(A') = 1\breve{}\mathbb{P}(A)$
  - **Conditional probability**: $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$
    - A *Conditional Probability* reduces the sample size
      - The $\mathbb{P}(A)$ in the sample space of B, rather than S
    - Two events are **Mutually Exclusive** if $\mathbb{P}(A \cap B) = 0$
    - Two events are **Independent** if $\mathbb{P}(A|B) = \mathbb{P}(A)$
      - If two events are *Independent*, then $\mathbb{P}(A \cap B) = \mathbb{P}(A) * \mathbb{P}(B)$ (Follows from *Conditional Probability*)
    - **Multiplication Rule**: $\mathbb{P}(A \cap B) = \mathbb{P}(B) * \mathbb{P}(A|B)$ (Follows from *Conditional Probability*)
    - **Addition Rule**: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Random Variables

- A **Random Variable** provides a numerical description of the outcomes of an experiment.
- A **Discrete Random Variable** has outcomes that are countable.
  - E.g. Number of children, number of computer crashes
- Uppercase letters denote *Random Variables* (X and Y)
- Lowercase letters denote the outcome (x and y)
  - E.g. we roll a six-sided die and it comes up 4
  - X = rolling a six-sided die
  - $x = 4$
- A **Probability Distribution Function** describes how the probabilities are assigned over all outcomes of a discrete random variable.
  - We use $f(x)$ to denote a *Probability Distribution Function*
    - E.g. if the probability of rolling a 4 with a six-sided die is 1/6
  - $f(4) = \frac{1}{6}$
  - **Uniform Probability Distribution Function -** all outcomes are equally likely.
    - $f(x) = \frac{1}{n}$, if n is the number of possible outcomes.
  - Rules:
    - $0 \leq f(x) \leq 1$
    - $\sum f(x) = 1$
- The center and spread are important characteristics of a probability distribution
  - Center: **Mean** = **Expected Value** $= \mathbb{E}[X] = \sum f(x) * x = \mu$
  - **Law of Large Numbers**: as the number of observations increases the difference between the empirical mean ($\bar{x}$) and the theoretical mean (μ) gets infinitely smaller.
  - **Spread**: **Variance** (i.e. *Second Moment*) $= \mathbb{E}[(X-\mu)^2] = \sum[f(x) * (x-\mu)^2] = \sigma^2$
    - Because square units are hard to interpret, *Standard Deviation* is often used as a measure of *Spread* instead.
- **Binomial distribution** (a *Discrete Distribution*)
  - n independent trials
  - Each trial has the outcome of either "success" or "failure"
  - The probability of "success" is p and the probability of "failure" is q
  - The number of "successes" is denoted k
  - Each trial is repeated under identical conditions
  - The outcomes of a binomial experiment *fit* a binomial distribution
  - $X \sim B(n, p)$: "*Random Variable* X is distributed according to a *Binomial Distribution* with parameters n and p"
  - $f(x) = \binom{n}{k} p^k q^{(n-k)}$, where $q = 1 - p$ and $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
  - $Mean = np$
  - $Variance = npq$
  - A *Binomial* experiment with a *single* trial is known as a **Bernoulli Trial**
- **Poisson Distribution** (a *Discrete Distribution*)s
  - Often used to model the number of occurrences of an event over a specified time interval or space.
  - Known mean rate of occurrence is the same for any two intervals of equal length
  - Occurrence of event in one interval is independent of occurrence in any other
  - x: number of occurrences in an interval
  - λ: average mean rate of occurrence
  - $X \sim Pois(\lambda)$
  - $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$
  - $Mean = \lambda$
  - $Variance = \lambda$