Review of Key Concepts

Demetris Athienitis

UNIVERSITY of FLORIDA

Review of key statistical concepts

Number of concepts which will come up quite frequently and it is important to understand them before we begin:

- Random variables
  - Probability mass function
  - Expected value
  - Variance
  - Rules for means and variances
- Conditional probability

Random variables

Definition (Random Variable)

A random variable is a function that assigns a numerical value to each outcome of an experiment. It is a measurable function from a probability space into a measurable space known as the state space.

Example

- Flipping a coin 10 times and counting how many heads turn up
- Rolling two dice and observing the sum

- The range is the set of possible values it can take
  - If we flip a coin 10 times then the range is $\{0, 1, 2, \ldots, 10\}$
  - Sum of two dice $\{2, 3, \ldots, 12\}$
- Seek to understand the probability that the random variable takes

Probability mass function

- Let Y be a categorical random variable of interest
- The probability mass function (PMF) is defined as $P(Y = y)$ for any $y \in \text{Range}(Y)$
- Describes the probability behavior of the random variable
- We will generally denote random variables with capital letters and potential values the random variable can take with lower-case letters

Expected value

- The expected value is denoted by $E(Y) = \sum_{\forall y} yP(Y = y)$
- $E(Y)$ is the mean of the random variable
- If we could re-run the experiment that generates Y an infinitely large number of times, $E(Y)$ would be the average value of the random variable.
- Represents the average value one would expect to see
    - Note that $E(Y)$ does not need to be in the range of Y

Variance

- The variance of a random variable Y is defined as

$$V(Y) = E\left[(Y - \mu)^2\right]$$

  where $\mu = E(Y)$
- Describes the average squared deviation from the mean
- Provides a measure of how uncertain we are about the value of Y
- $SD(Y) = \sqrt{V(Y)}$

Rules for means and variances

Let X and Y be random variables throughout and let a, b be known constants

- $E(aX + b) = aE(X) + b$
- $E(aX + bY) = aE(X) + bE(Y)$
- $V(aX + b) = a^2 V(X)$
- $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2ab\mathrm{Cov}(X, Y)$

Conditional probability

Let X and Y be random variables

- $P(Y = y|X = x)$ is the probability that the random variable Y takes value y conditional on the fact that $X = x$
- Two key properties of conditional probability:

$$P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)}$$

$$P(Y = y|X = x) = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}$$

Statistics and estimators

- A statistic is simply a function of the observed data
- Statistics, before a value is observed is are random variables
- The standard deviation of a statistic is referred to as its *standard error*
- Estimate unknown parameters θ and denote estimates as $\widehat{\theta}$
- $\widehat{\theta}$ will be a function of the observed data

We learned

- Key concepts about (discrete) random variables
- Conditional probabilities