**Lecture 2 Notes**

Discrete Probability Distributions

* Binomial
  + Binary outcome
  + Probability – assuming theoretical distributions that we study then apply the properties of to a sample
  + Theoretical parameters
    - Eg pregnancy
  + Eg website observation – clicked or not?
    - Binary variable
    - Observation = 1 person
    - Observation = 10 people 🡪 binomial
* Multinomial distributions
* Assume that a probability p is given
  + How many successes do we observe over N number of trials? 🡪 binomial distribution
  + Combinatorics 🡪 important to know the sample space
    - Sample space: all possible outcomes we can observe
    - Eg: yes or no
  + Binomial
    - # of trials
    - Need to know p and number of trials N that we observe
    - X = number of successes

Probability

* P (x >= 2)
* Sample space for binomial distribution, rather than Bernoulli
* All possible combinations are considered (eg YYY, YYN….)

Complement

* Set of outcomes that are not in A
* Probability calculus
* Dividing sample space into 2 events
* Requires some assumptions

Union and Intersection

* Disjoint events
* Disjoint vs independent 🡪 independent is broader scale; if some events are independent that doesn’t mean that they’re disjoint

Application

* Amazon:
  + Uses Jaccard simil arity for their recommender system
  + C1 & C2 = A , C
  + Find most similar customers to you then recommend them

Permutation

* Can write them out 🡪 ordered subsets
* We use permutations in recommender systems

Combination

* Unordered subsets
* Binomial coefficients

Conditional Probability

* Common problem in manufacturing engineering
* 2 lines producing the same product, with one faulty and the other not
* Marginal probability
* Multidimensional probabilities

Multiplication Rule

1. P (A & B)
2. P(R)
3. O(A|Q)

\*using tree diagrams

P(A|B) = P(A&B)/P(B)

P(A&B) = P(A|B)P(B)

🡺 need to take Bayesian Statistics

Independence

* The P(A|B) = P(A)
* Blood types

Binomial Probability Distribution

* Assumption that we don’t have independent variables to estimate
* Think about all the possible outcomes in a sample space
* Binomial coefficient corrects for that

Example

* Add up the probabilities
* Flipping a coin 10x
* Binomial compared to the normal is sufficiently large amount of time

The Poisson Distribution

* Good to analyze independent events

Hypothesis

* Non-directional hypothesis
* Directional hypothesis

Decision Making

* Null hypothesis
* True state
* Type 1 mistake 🡪 alpha is the area under the PDF
* Power