STAT1005 Week 3 Cheat Sheet

Conditional Probability

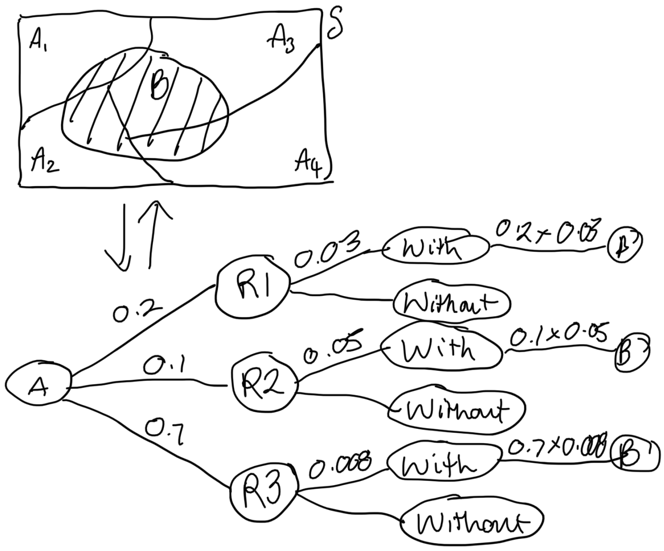
- P(D|T) - is read as the probability of D given T
- This new probability will be (D ∩ T)/T. Essentially T is the new sample space.
- So $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
- P (A|B) is not always ≠ P(B|A)
- If P(B|A) = P(B), the two are said to be independent, because event A occurring did not change the probability of A.

Multiplication Rule

- $P(A \cap B) = P(A|B) \times P(B)$
- $P(A_1 \cap A_2 \cap A_3) = P(A_3|A_1 \cap A_2) \times P(A_1 \cap A_2)$
  OR $P(A_3|A_1 \cap A_2) \times P(A_2|A_1) \times P(A_1)$

The Law of Total Probability

- Let A_1, ..., A_k be mutually exclusive and exhaustive events. Then for any other event B,
- $P(B) = P(B|A_1) \times P(A_1) + \ldots + P(B|A_k) \times P(A_k)$
  $= \sum_{i=1}^{k} P(B|A_i) \times P(A_i)$



Bayes Theorem

- Let A , ..., A be mutually exclusive and exhaustive events with P(A) > 0 for i = 1, ..., k. Then for any other event B for which P(B) > 0
- $P(A_j|B) = \dfrac{P(A_j \cap B)}{P(B)} = \dfrac{P(B|A_j) \times (A_j)}{\sum_{i=1}^{k} P(B|A_i) \times P(A_i)}$ for j = 1, .., k

Variability

- Variability is inevitable, so we have to attach a margin of error to any results we get
- Under random sampling, the variation in the statistic will follow a predictable pattern, and this pattern allows us to calculate margins of error.

Density Curves

- In many instances, observed regularities can be described by a smooth curve, which is a mathematical model for the distribution of the observations
- These density curves will allow us to provide a measure of uncertainty
- A density curve is a curve that:
  - Is always on or above the horizontal axis: $f(X) \geq 0$ for every x defined on the interval [a, b]
  - Has an area of exactly 1 underneath it: $\int_a^b f(x)dx = 1$
  - Caution
    - No set of real data is exactly described by a density curve
    - They are idealised descriptions, accurate enough for practical use

Describing density curves

- The mean and median are the same for a symmetric density curve
- The median cuts the area under the curve in half
- The mean of a skewed curve is pulled away from the median toward the long tail

Normal Distribution

- All normal curves are symmetric, single peaked and bell-shaped
- A specific normal curve is described by its mean μ and standard deviation σ (or variance σ^2)
- The expression for the normal density curve is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} \times e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

- In a normal distribution with mean μ and standard deviation σ
  - Approx 68% of the observations fall within σ of μ
  - Approx 95% of the observations fall within 2σ of μ
  - Approx 99.7% of the observations fall within 3σ of μ
- Normal distributions are good descriptions for some distributions
- They're a good approximation of many kinds of chance outcomes
- In text books you usually see $N(\mu, \sigma^2)$, and in out class the second number will always be variance
- We often use x̄ as an estimate of μ and the sample standard deviation s as an estimate of σ

$$s = \left(\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2\right)^{1/2}$$

Standard Normal Distribution

- The standard normal distribution is the normal distribution with mean 0 and standard deviation 1
- If a variable has a normal distribution $N(\mu, \sigma^2)$, then the standardised variable

$$Z = \frac{X - \mu}{\sigma}$$

has the standard normal distribution, N(0, 1)

- If you have a normal distribution $N(\mu, \sigma^2)$, and take a sample of size n, then to find P(X < or > a) will use the original values, but the P(X < or > a) will use $N(\mu, \frac{\sigma^2}{n})$

Continuous Probability Models

- Where we cannot assign probabilities to each individual value because there is an infinite interval of possible values
- A continuous probability model assigns probabilities as areas under a density curve, The area under the curve and above any range of values is the probability of an outcome in a range
- Normal distributions are continuous probability models

Random Variables

- Random variables that have a finite list of possible outcomes are called discrete
- Random variables that can take on any value in an interval, with probabilities given as areas under a density curve, are called continuous