PSYC200; Exam 2 Guide

Population Distributions

- In a sample: observed (frequency)
- In a population: theoretical (probability)
- **Continuous Distributions –** describe an infinite number of possible data values
  - Continuous curves
- **Discrete Distributions –** describe a finite number of possible values
  - Histograms

Random Sampling Methods – gives the most unbiased samples; parts of population that are over-represented are by *chance*

- **Simple random sample –** every observation has an equal chance of being sampled; sampled independently (w/ replacement)
- **Systematic random sample –** randomly select a starting point and continue selection by a "rule"
  - ex: every 3rd person in a sample
- **Stratified random sample –** the total population is divided into groups (strata) based on special interests (race, gender)→ samples are taken from each group
  - Used for variables that are difficult to compare
  - Strata are usually deliberately over-represented
- **Cluster sample –** similar to stratified, but division of groups occurs in a convenient way (physical location)

Subjective Probability- personal view/belief of what we think the probability is
Objective Probability- actuality, what the probability actually is

Simple Probability – the likelihood of some event happening compared to all possible events happening

Formula: $P(A) = \frac{\text{number of observations favoring A}}{\text{total number of possible observations}}$ *Probability is a measure

Venn Diagrams

- **Union** ("or") – refers to one or the other event
- **Mutually exclusive** – 2 events have no outcomes in common
  - This intersection in *null*; p(A and B) = 0
- **Intersections** ("and")- when events have outcomes in common
- **Exhaustive –** a set of events that includes *all* possible outcomes

Addition Rule: p(A **OR** B) = p(A) + p(B) – p(A and B)

- If events are mutually exclusive, the addition rule is reduced to: p(A or B)= p(A) + p(B)

Multiplication Rule: p(A **AND** B) = p(A) * p(B | A)

OR → Addition

AND→ Multiplication

- **Joint probabilities** –likelihood of observing each of the 2 events
- **Conditional probabilities** ("given") – the likelihood that an event will occur given that some other event occurs

$$p(B \mid A) = \frac{p(B,A)}{p(A)}$$

$p(A|B) \neq p(B|A)$ UNLESS $p(A) = p(B)$

Independent Events – if occurrence of one event has *no effect* on the probability of occurrence of other event

- Events are independent *if and only if*:
  - p(A and B) = p(A) * p(B)
    - p(B) = p(B|A)

*Equality = independence

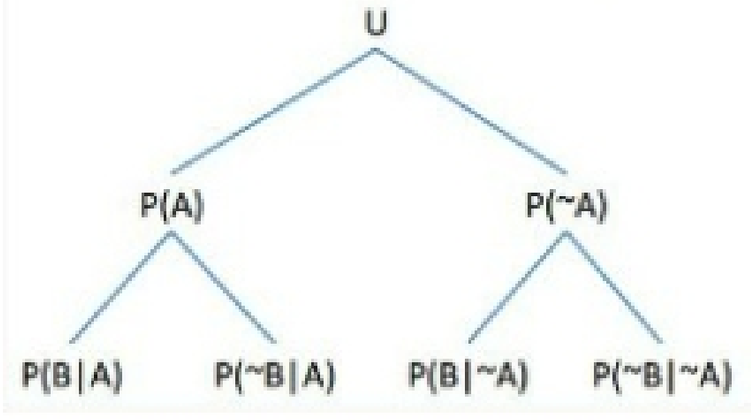
*Inequality = dependence

- If events A and B are independent, then multiplication rule is reduced to:
  - p(A **AND** B) = p(A)*p(B)
- For sampling without replacement= non-independent events
- For sampling with replacement= independent events

Bayes' Rule

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

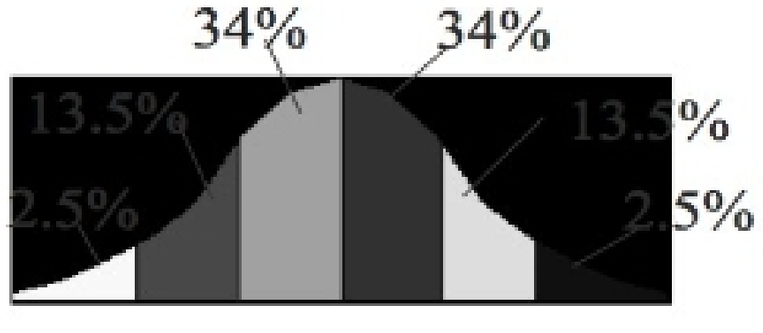
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)\,P(A) + P(B|\sim A)\,P(\sim A)}$$



- Probability of desired event over all possibilities
- Used in diagnosis, medicine, public policy, etc
- Probabilities based on relative frequency data

Probability Distributions

Normal Distributions



- Parameters:
  - μ= population mean
  - σ^2= population variance
- Fully described by its mean and standard deviation
- Shape describes many existing variables (i.e. weight)
- For continuous distribution
- Symmetric about the mean
  - Approx. 68.3% of the observations are within ± 1 s.d. of the mean
  - Approx. 95% of the observations are within ± 2 s.d. of the mean
- Area under the curve = 1

Table A gives the areas of all z scores (z-scores transform raw scores into the # of s.d. away from the mean)

- Column A: z-scores (z = x – mean / standard deviation)
- Column B: area between mean and z-score
- Column C: area above the z-score (z-score to tail)

- p(X or higher) → get (+) z-score→ "C value"

- p(X or higher) → get (-) z-score→ "B value + 0.5"
- p(X or lower) → get (+) z-score→ "B value + 0.5"
- p(X or lower) → get (-) z-score→ "C value"
- p(between X and Y) → get the z scores of X and Y→ add both B values together
- p(not between X and Y) → get the z scores → add both C values together

Binomial Distributions

- When you count how many of a sample of a fixed size have a certain characteristic (i.e. 3 chances to pull out 3 twix); WITH replacement
- For discrete distributions
- N = fixed sample size

  P = success

  Q = failure

$$P(s) = \frac{N!}{s!(N-s)!} p^s q^{N-s}$$

- Requirements:
  - Series of N trials
  - Each trial has only 2 possible outcomes
  - On each trial the two outcomes are mutually exclusive
  - There is independence between the outcomes of each trial
  - The probability of each outcome remains constant from trial to trial
- **Mean** = np

  Variance = npq

  Standard Deviation = √npq

*The binomial *approaches* the normal distribution when p = 0.5, as N (# of trials) gets larger

*Use normal approximation when np ≥ 10 AND nq ≥ 10

Permutations- how many ways can I get X in a certain # of trials?

<u>Discrete Distributions</u> (other than the binomial)

- **Poisson** – # of successes (x) is random, p is fixed, n is fixed (usually a time period), failures aren't measured (i.e. Have 30 seconds to pull out some snickers); WITH replacement
- **Pascal** – # of trials is random (how many trials must I do to get X successes?), p is fixed, # of successes is fixed (i.e. how many trials to get 3 twix from the bag)
  - Mean= x/p
  - Variance= x(1-p)/p^2
  - **Geometric**- only looking for one success (how many trials until I get my 1st success?)
- **Multinomial** – when you have multiple categories/nominal measurement, p is fixed, WITH replacement (i.e. pull out candy, write it down, replace each time)
- **Hypogeometric** – equivalent to the multinomial, except p (probability of success) is not fixed, WITHOUT replacement

<u>Continuous Distributions</u>