Introduction to Probability

BUS 735: Business Decision Methods and Research

Goals of this section

- Learn basics of probability.

- Learn about how a boring and complicated formula can really mean a lot.

1 Elementary Probability

1.1 Probability of Events

Basic Probability

- **Probability:** numeric value between 0 and 1 (or 0% and 100%) representing the chance, likelihood, or possibility some event will occur.

- **Event:** some possible (or even impossible) outcome occurring.

  - Denote events with capital English letters.

- Example:

  - A: A new restaurant will earn positive profits during its first year.
  - $P(A) = 0.2$ means there is a 20% chance that a new restaurant will earn profits in its first year.

- **A-priori probability:** objective probability of an event that can be stated *before* the event occurs.

- **Relative frequency probability:** long-run expected probabilities based on past data.

- **Subjective probability:** estimated probability based on personal belief, experience, or knowledge of a situation.

Contingency Table

Type of TV	Customer Owns a DVR		
	Yes	No	Total
HDTV	38	42	80
Regular TV	70	150	220
Total	108	192	300

- Define Event A: Customer owns an HDTV.

- What is $P(A)$?

Joint Events

- **Joint Event:** is an event that is composed of two or more events.

- Define event C as any event in either A *or* B.

  - Notation for event C: $C = A \cup B$
  - Notation for probability of event C: $P(C) = P(A \cup B)$.

- Define event C as any event in A *and* B.

  - Notation for event C: $C = A \cap B$
  - Notation for probability of event C: $P(C) = P(A \cap B)$.

Complements of Events

- The **complement** of an event, A, is the outcome of anything *besides* A occurring.

- Notation: $A' = A^c =$ complement of event A.

- $P(A') = 1 - P(A)$.

- Example: what is the complement of Event A: a newborn baby is a female.

Mutually Exclusive Events

- Two events, A and B, are **mutually exclusive** if it is impossible for both A and B to occur at the same time.

- Are the following mutually exclusive?

  - Event A: A person is currently 8 years old. Event B: A person voted for John McCain in the last presidential election.
  - Event A: A person plays football in high school. Event B: A person plays basketball in high school.
  - Event A, Event A'.
  - Event A: We will go out to eat tonight. Event B: We are going out to eat tomorrow night.

Contingency Table

	Customer Owns a DVR		
Type of TV	Yes	No	Total
HDTV	38	42	80
Regular TV	70	150	220
Total	108	192	300

- Define Event A: Customer owns an HDTV.

- Define Event B: Customer owns a DVR.

- Define Event $C = A \cap B$.

- What is $P(C)$?

- Define Event $D = A \cup B$.

- What is $P(D)$?

Venn Diagram
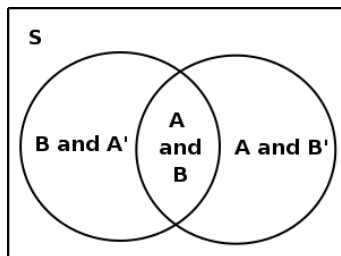
- **Venn diagram:** visualization of all possible events. The areas in the diagram represent the probabilities of those events.

- S: Event that encompasses all possible outcomes.

- Entire area of the left hand circle is $P(B)$.

- Entire area of the right hand circle is event $P(A)$.

- The area that is in both of the circles is $P(A \cap B)$.

Venn Diagram

- From the Venn Diagram we can see that,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Use this equation to find the probability of owning an HDTV (Event A) or owning a DVR (Event B).

| | **Customer Owns a DVR** | | |
Type of TV	Yes	No	Total
HDTV	38	42	80
Regular TV	70	150	220
Total	108	192	300

1.2 Conditional Probability

Conditional probability

- **Conditional probability**: the probability of an event, A, with the additional information that some other event B has already occurred.

- Example:

  - What is the probability a person is female?
    Define event A: a person is a female.
    $P(A) = 0.5$

  - What is the probability a person is female, given he/she is a UW-L college student?
    Define event B: a person is a UW-L college student.
    $P(A|B) = 0.6$

Independence

- Two events A and B are **independent** if knowledge that A happened does not affect P(B), or if knowledge that B happened does not effect P(A).

- In the example above, is being female and being a UW-L college student independent?

- Is owning an HDTV and DVR independent?

- Is the event that someone smokes and the event someone has lung cancer independent?

- Suppose a coin is flipped twice. Is the event the first flip is heads and the event the second flip is heads independent?

- If A and B are independent, then
  $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Bayes Theorem

- This is the coolest thing you'll ever learn a math or stats class:

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

- Why is the cool? Because this proves that:

$$P(A|B) \neq P(B|A)$$

Blood test accuracy

- Suppose a fatal disease breaks out, and a blood test is used to detect the disease.

- The blood test claims to accurately identify the disease 99% of the time.

- Let A be the event you have a disease.

- Let B be the event the blood test came out positive.

$$P(B|A) = 0.99$$

- Suppose you take the blood test and it is positive. What is the probability you have the disease?
$$P(A|B) = \ ?$$

Blood test accuracy

- Suppose 0.2% of people have the disease, and 0.198% have the disease and tested positive.
$$P(A) = 0.002$$
$$P(A \cap B) = 0.00198$$
$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.00198}{0.002} = 0.99$$

Blood test accuracy

- Suppose 5% of people test positive for the disease.

- What is the probability you have the disease given you tested positive?

$$P(A|B) = \frac{P(A \cap B)}{P(A)} = \frac{0.00198}{0.05} = 0.0396$$

- Even though you tested positive, *you still most likely do not have the disease.*

- And the test had the claim of being 99% accurate.

- So what is the real accuracy of things like blood tests, pregnancy tests, and lie detector tests?

2 Probability Distributions

2.1 Random Variables

Definitions

- **Random variable**: a variable that has a single numerical value determined by chance.

- Data is a bunch of realizations of a random variable.

- **Discrete random variable**: an RV that can take on "countable" values.

- **Continuous random variable**: a RV that can take on infinitely many values on a continuous scale.

Probability distribution

- A **probability distribution** is a graph, table, or formula that gives the probability for each value of a random variable.

- The sum of the probabilities for all possible values an RV can take must equal 1 (or 100%).

$$\sum P(x_i) = 1$$

- Each individual probability must be between zero and one.

$$0 \leq P(x_i) \leq 1$$

Mean and variance of a probability distribution

- The mean or **expected value** of a probability distribution is:

$$\mu = \sum x_i P(x_i)$$

- The variance of a probability distribution is given by:

$$\sigma^2 = \sum \left[(x_i - \mu)^2 P(x_i) \right]$$

$$\sigma^2 = \sum \left(x_i^2 P(x_i) \right) - \mu^2$$

2.2   Binomial distribution

Binomial distributions

- A **Bernoulli trial** results in a random variable that can only result in success ($x = 1$) or failure ($x = 0$).

    - Example: an outcome of heads for a single coin flip is a Bernoulli trial.

- A **binomial distribution** is the probability distribution for the number of successes in a fixed number of trials.

- Requirements for a binomial distribution:

    - The experiment must have a fixed number of Bernoulli trials.
    - The trials must be independent.
    - The probability of success must be the same for each trial.

- Example: what are the possible outcomes for the number of successes in 3 coin flips?

Binomial probability distribution

- The binomial probability distribution is given by,

$$P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

    - n: number of trials.
    - $P(x)$: the probability of x number of successes.
    - p is the probability of success for a single Bernoulli trial.

- Calculate the probability distribution of the 5 coin flip experiment.

- Verify $\sum P(x_i) = 1$.

- Calculate the expected value for the number of heads.

- Calculate the variance and standard deviation.

Binomial Distribution

- Recall the mean of a probability distribution:

$$\mu = \sum x_i P(x_i)$$

- For the binomial distribution, this gets more simple:

$$\mu = np$$

- Recall the variance of a probability distribution:

$$\sigma^2 = \sum \left[(x_i - \mu)^2 P(x_i) \right]$$

- It gets simpler:

$$\sigma^2 = np(1 - p)$$

Normal Approximation to a Binomial

- Motivation: Questions like,

- Suppose a study concluded that Burger King fills out 10% of its orders inaccurately. If a particular franchise makes 30,000 orders a month, what is the probability it will make more than 3,100 errors in orders?

- Compute P(x=3100), P(x=3101), P(x=3102), P(x=3103), ... , P(x=30,000).

- That's a ridiculous amount of computations.

- Can you suppose that the number of errors is *normally distributed* with mean equal to $\mu = np$, and variance $\sigma^2 = np(1 - p)$?

Normal Approximation to a Binomial

- Can you suppose that the number of errors is *normally distributed* with mean equal to $\mu = np$, and variance $\sigma^2 = np(1 - p)$?

- Well... the number of errors is truly a *binomial distribution*, not a normal distribution.

- And... the number of errors is a *discrete* random variable, but the normal distribution is for continuous random variables.

- But... the normal distribution will be a good approximation if,

$$np > 5 \text{ and } n(1 - p) > 5$$

3 Combining Random Variables

3.1 Covariance

Covariance

- To measure the variance of combinations of RVs, need to know the covariance.

- **Covariance**: related to Pearson correlation coefficient, it's a measure of how two RVs move together.

- Interpretations:

  - When covariance is negative, variables move in opposite directions.
  - When covariance is positive, variables move in same direction.

3.2 Combining Random Variables

Combining Random Variables

- Suppose two RVs are measured in the same units, can they be combined (added, subtracted, etc)?

- Who would want to do such a thing?

- Addition rule:

    - If $Z = X + Y$, then $E(Z) = E(X) + E(Y)$.
    - More generally, if $Z = aX + bY$, then $E(Z) = aE(X) + bE(Y)$.

Variance of Combinations

- Suppose $Z = X + Y$:

$$VAR(X + Y) \equiv \sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}.$$

- Suppose $Z = aX + bY$:

$$VAR(aX + bY) \equiv \sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_{xy}.$$

3.3 Examples

Portfolio Risk

- Don't put all your eggs in one basket.

- Would it make more sense to put your money in two investments that are negatively correlated (meaning they have a negative covariance)?

- Example, suppose the variance for the quarterly return is equal to 15% for Investment X and 10% for Investment Y, and the covariance is equal to -8%. Suppose you invested have your money in each investment.

- Would it make more sense to put your money in two investments that are positively correlated (meaning they have a positive covariance)?

- Example, suppose the variance for the quarterly return is equal to 18% for Investment X and 12% for Investment Y, and the covariance is equal to 6%. Suppose you invested have your money in each investment.

Another Example

- Suppose you run a website business that produces and sells custom made furniture.

  - When someone submits an order online you must first build the furniture. The average time it takes to build the furniture is 9 business days and the standard deviation is 2.5 business days.
  - After you build the furniture, you must ship it. The average shipping time is 4 business days and the standard deviation is 3 business days.
  - What is the average elapsed time from the time the order is placed until the delivery?
  - What is the variance and standard deviation of the elapsed time from the time the order is placed until the delivery?