Probability and Statistics*

Roshan Rai

2025-01-20

*This note contains parts that I learnt from the Probability and Statistics course of Georgia Tech university in edx.org.

Contents

1. Pre-requisites

1.1 Bootcamp: Set

Set is a collection of objects. Members of set are called elements.

Notation:

For sets, $A, B, C, ...$

For elements, $a, b, c, ...$

For membership, $\in$ e.g. $a \in A$

For non membership, $\notin$.

For universal set, $\mathbb{U}$ i.e. everything.

For null set, ϕ.

Example:

$B = \{x/0 \leq x \leq 1\}$ where / means such that.

$C = \{x/x \in \mathbf{R}, x^2 = -1\} = \phi$

Definition: If every element of A is an element of B then A is subset of B. i.e. $A \subset B$.

Definition: $A = B$ iff (if and only if) $A \subset B$ and $B \subset A$.

Properties:

- $\phi \subset A$; $A \subset U$; $A \subset A$
- $A \subset B$, $B \subset C \implies A \subset C$

Remark: The order in which the elements of set are listed is immaterial. E.g. $\{a, b, c\} = \{b, c, a\}$.

Definition: The complement of A with respect to U is $A^{\mathbf{c}} = \{x \mid x \in U \text{ and } x \notin A\}$.

Definition: The intersection of A and B is $A \cap B = \{x \mid x \in B \text{ and } x \in B\}$.

Definition: The union of A and B is $AUB = \{x \mid x \in A \text{ or } x \in B\}$.

If $A \cap B = \phi$, then A and B are **disjoint** or **mutually exclusive**.

Definition:

- Minus: $A - B = A \cap B^{\mathbf{c}}$
- Symmetric difference or XOR: $A \triangle B = (A - B) \cup (B - A) = (A \cup B) - (A \cap B)$

- The **cardinality** of A, denoted by $|A|$ is the number of elements in A. A is finite if $|A| < \infty$.

  $B = \{1, 2, 3, ...\}$ is **countably infinite** i.e. $|B| = \aleph_0$

  $C = \{x|x \in [0, 1]\}$ is **uncountably infinite** i.e. $|C| = \aleph_1$

Laws of Operation:

- **Complement Law**: $A \cup A^{\mathbf{c}} = U$, $A \cap A^{\mathbf{c}} = \phi$, $(A^{\mathbf{c}})^{\mathbf{c}} = A$
- **Commutative Law**: $A \cup B = B \cup A$, $A \cap B = B \cap A$
- **Associative Law**: $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$
- **Distributive Law**: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- **De-Morgan's Law**: $(A \cup B)^{\mathbf{c}} = A^{\mathbf{c}} \cap B^{\mathbf{c}}$, $(A \cap B)^{\mathbf{c}} = A^{\mathbf{c}} \cup B^{\mathbf{c}}$

1.2 Bootcamp: Derivative

Definition: The function $f(x)$ maps values of X from a certain domain X to a certain range Y which can be denoted $f : x \to Y$.

If $f(x) = x^2$ then the function takes x-values from the real line $\mathbb{R}$ to the non-negative portion of real line $\mathbb{R}^+$.

Definition: We say that $f(x)$ is **continuous** function if for any x_0 & $x \in X$, we have $lim_{x \to 0} f(x) = f(x_0)$ where $f(x)$ is assumed to exist for all $x \in X$.

The function $f(x) = 3x^2$ is continuous for all x. The function $f(x) = \lfloor x \rfloor$ i.e. round down to nearest integer e.g. $\lfloor 3.4 \rfloor = 3$. This function has discontinuity at any integer x.

Definition: The **inverse** of function $f : X \to Y$ is reverse mapping of $g :\to X$ such that $f(x) = y$ iff $g(y) = x$ for all appropriate x and y. The inverse is often written as f^{-1} and is especially useful if $f(x)$ strictly increasing or decreasing function. Note that $f^{-1}(f(x)) = x$.

Defintion: If $f(x)$ is continuous, then it is **differentiable** if,

$$\frac{d}{dx} f(x) = f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists and is well defined for given x. The derivative of $f(x)$ is slope of the function.

$$[x^k]' = kx^{k-1}$$

$$[e^x]' = e^x$$

$$[sin(x)]' = cos(x)$$

$$[cos(x)]' = -sin(x)$$

$$[ln(x)]' = \frac{1}{x}$$

$$[arctan(x)]' = \frac{1}{1+x^2}$$

Theorem: Some properties of derivatives

$$[af(x) + b]' = af'(x)$$

$$[f(x) + g(x)]' = f'(x) + g'(x)$$

$$[f(x) + g(x)]' = f'(x)g(x) + f(x)g'(x)$$

$$[\frac{f(x)}{g(x)}]' = \frac{g(x)f'(x) - f(x)g'(x)}{g^2(x)}$$

$$[f(g(x))]' = f'(g(x))g'(x)$$

Remark: The second derivative $f''(x) = \frac{d}{dx}f'(x)$ and is the "slope of slope". If $f(x)$ is position, then $f'(x)$ can be regarded as "velocity" and $f''(x)$ as "acceleration".

The minimum or maximum of $f(x)$ can only occur when slope of $f(x)$ is 0, i.e. only when $f'(x) = 0$, say at the critical point $x = x_0$. Exception: Check the endpoints of your intervals of interest as well.

If $f''(x) < 0$, you get maximum, if $f''(x) > 0$, you get a minimum. If $f''(x) = 0$, you get a **point of inflection**.

1.3 Bootcamp: Integration

Definition: The function $F(x)$ having derivative $f(x)$ is called the **anti-derivative** or **indefinite integral**. It is denoted by $F(x) = \int f(x)dx$.

Fundamental Theorem of Calculus: If $f(x)$ is continuous, then the area under the curve for $x \in [a, b]$ is denoted and given by the **definite integral**.

$$\int_a^b f(x)dx = F(x)|_a^b = F(b) - F(a)$$

$$\int x^k dx = \frac{x^{k+1}}{k+1} + c \qquad \text{for } \ k \neq 1 \text{ where c is arbitrary constant}$$

$$\int \frac{dx}{x} = ln|x| + c$$

$$\int e^x dx = e^x + c$$

$$\int cos(x)dx = sin(x) + c$$

$$\int \frac{1}{1+x^2} dx = arctan(x) + c$$

Theorem: Some well known properties of definite integrals

$$\int_a^a f(x)dx = 0$$

$$\int_a^b f(x)dx = -\int_b^a f(x)dx$$

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

Theorem: Some other properties of general integrals:

$$\int [f(x) + g(x)]dx = \int f(x)dx + \int g(x)dx$$

$$\int f(x)g'(x)dx = f(x)g(x) - \int g(x)f'(x)dx \quad \text{integration by parts}$$

$$\int f(g(x))g'(x)dx = \int f(u)du \quad \text{Substitution rule with u = g(x)}$$

Defiinition: Derivative of arbitrary order K can be written as $f^k(x)$ or $\frac{d^k}{dx^k}f(x)$. By convention $f^0(x) = f(x)$.

The **Taylor Series Expansion** of $f(x)$ about a point a is given by

$$f(x) = \sum_{k=0}^{\infty} \frac{f^k(a)(x-a)^k}{k!}$$

The **Maclaurin Series** is simply Taylor expanded around $a = 0$.

Some famous **Maclaurin Series**;

$$sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{(2k+1)!}$$

$$cos(x) = \sum_{k=0}^{\infty} \frac{(-1)^x x^{2k}}{(2k)!}$$

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Here are some miscellaneous sums:

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1}{6}$$

$$\sum_{k=0}^{\infty} = \frac{1}{1-p} \quad (\text{ for } \quad -1 < p < 1)$$

Theorem: Occasionally, we run into trouble when taking indeterminate ratios of form $\frac{0}{0}$ or $\frac{\infty}{\infty}$. In such cases, **L' Hospital Rule** is useful. If the limits $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both go to 0 or both go to ∞, then,

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \lim_{x\to a}\frac{f'(x)}{g'(x)}$$

Example:

$$\lim_{x\to 0}\frac{sin(x)}{x} = \lim_{x\to 0}\frac{cos(x)}{1} = 1$$

Double Integration:

Whereas single integrals get us the area under a curve, double integrals represent the volume under a three dimensional function.

The volume under $f(x,y) = 8xy$ over region $0 < x < y < 1$ is given by

$$\int_0^1 \int_0^y f(x,y)dxdy = \int_0^1 \int_0^y 8xydxdy = \int_0^1 4y^3dy = 1$$

We can swap the order of integration to get same answer.

$$\int_0^1 \int_x^1 8xydydx = \int_0^1 4x(1 = x^2)dx = 1$$

2. Introduction to Probability

2.1 Introduction

Mathematical models are either

- Deterministic (no uncertainty/randomness)
- Probabilistic (have some uncertainty)

Q. A couple has two kids and at least one is boy. What is the probability that both are boys?

Possibilities: GG, BG, GB, BB. Eliminate GG since we know that there's at least one boy. Then $P(BB) = \frac{1}{3}$.

Probability is methodology that describes the random variation in systems. **Statistics** uses data (sample) to draw conclusion about population.

Definition: A **sample space** associated with an experiment E is the set of all possible outcome of E. It's usually denoted by S or Ω.

Coin Toss: $S = \{H, T\}$

Toss a coin 2 times: $S : \{HH, HT, TH, TT\}$

Definition: An **event** is a set of possible outcomes. Thus, any subset of S is event.

Toss a dice, $S = \{1, 2, ...\}$

If A is event "odd number occurs", $A = \{1, 3, 54\}$

The **empty set** ϕ is an event of S.

S is an event of S.

If A is an event, then A^c is the **complementary** event.

If A and B are events, then $A \cup B$ and $A \cap B$ are events.

Definition: The **Probability** of a generic event $A \subset S$ is a function that adheres to following axioms:

- $0 \leq P(A) \leq 1$ (probabilities are always between 0 and 1)
- $P(S) = 1$ (probability of some outcome is 1)
- If A and B are disjoint events, i.e. $A \cap B = \phi$ then, $P(A \cap B) = P(A) + P(B)$.
- Suppose A_1, $A_2, ...$ is a sequence of disjoint events, i.e. $A_i \cap A_J = \phi$ for $i \neq j$.

$$
\begin{aligned}
P(S) &= P(U_{i=1}^{\infty} A_i) \\
&= \Sigma_{i=1}^{\infty} P(A_i) \\
&= \Sigma_{i=1}^{\infty} \frac{1}{2^i}
\end{aligned}
$$

Theorem: $P(A^c) = 1 - P(A)$

Proof:

$$
\begin{aligned}
1 &= P(S) \\
&= P(A \cup A^c) \\
&= P(A) + P(A^c) \quad \therefore A \cap A^c = \phi
\end{aligned}
$$

Corollary: $P(\phi) = 0$

Proof: By definition, $\phi = S^c$; so the result follows the theorem and axiom 2. **Remark:** The converse is false: $P(A) = 0$ doesn't imply $A = \phi$.

Theorem: For any two events A and B,

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

Proof: Use Venn-diagram.

Remark: Axiom 3 is special case of this theorem with $A \cap B = \phi$.

Theorem: For any three events A, B and C,

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

Theorem: Here is the **Principle of inclusion-exclusion**:

$$
\begin{aligned}
P(A_1 \cup A_2 \cup .... \cup A_n) = \Sigma_{i=1}^{n} P(A_i) - \Sigma\Sigma_{i<j} P(A_i \cap A_j) + \Sigma\Sigma\Sigma_{i<j<k} P(A_i \cap A_j \cap A_k) \\
+ ..... + (-1)^{n-1} P(A_1 \cap A_2 \cap .... A_n)
\end{aligned}
$$

Remark: You "include" all of the "single" events, "exclude" the double events, include the triple events etc.

Finite Sample Space:

Suppose S is finite $S = S_1, S_2, ... S_n$. Finite sample space often allows us to calculate the probabilities of certain events more efficiently. To illustrate, let $A \subset S$ be any event, then $P(A) =$

$\Sigma_{S_i \in A} P(S_i)$.

You have 2 red cards, a blue and a yellow card. Pick a card at random then,

$S = \{S_1, S_2, S_3\} = \{red, blue, yellow\}$

$P(S_1) = \frac{1}{2}$ $P(S_2) = \frac{1}{4}$ $P(S_3) = \frac{1}{4}$

$P(\text{red or yellow}) = \frac{1}{2} + \frac{1}{4}$

Definition: A **simple sample space (SSS)** is a finite sample space in which outcomes are equally likely.

Remark: In above example, S is not simple sample space since $P(S_1) \neq P(S_2)$.

Example: Toss 2 fair coins,

$S = \{HH, HT, TH, TT\}$ is a SSS (all probabilities are $\frac{1}{4}$).

Theorem: For any event A in SSS,

$P(A) = \frac{|A|}{|S|} = \frac{\text{no. of elements in A}}{\text{no. of elements in S}}$

2.1.1 Counting Techniques

Muffin (blueberry or oatmeal) or a bagel (sesame, plain, salt, garlic) but not both. You have $2 + 4 = 6$ choices in total.

$n_{AB} = 3$ ways to go from city A to B (walk, car, bus) and $n_{BC} = 4$ ways to go from B to C (car, bus, train, plane). Then you can go from A to C (via B) using $n_{AB}.n_{BC} = 3 * 4 = 12$ ways.

Roll two dice. How many outcomes?

$(3, 2) \neq (2, 3)$ so, answer $= 6 * 6 = 36$ ways.

Toss n dice. Outcome $= 6^n$ possibilities.

Toss n coins. Outcome $= 2^n$ possibilities.

2.1.2 Permutation

An arrangement of n symbols in a **definite order** is a **permutation** of n symbols.

Example: How many ways to arrange $1, 2, 3$?

Answer: 6 ways: 123, 132, 213, 312, 321, 231

- **Number of ways to arrange $1, 2, ..., n = n * (n-1) * (n-2) * ..... * 2 * 1 = n!$

Definition: The number of **r-tuples** we can make from n different symbols (each used at most once) is called the **number of permutations of n things taken r at a time**.

$$P_{n,r} = \frac{n!}{(n-r)!}$$

Note: $0! = 1$ & $P_{n,n} = n!$

Proof:

$$\begin{aligned} P_{n,r} &= \text{(choose first)(choose secoond)....(choose} r^{th}) \\ &= n(n-1)(n-2)....(n-r+1) \\ &= \frac{n(n-1)...(n-r+1)(n-r)...2*1}{(n-r)...2*1} \\ &= \frac{n!}{(n-r)!} \end{aligned}$$

Example: How many license plates of 6 digits can be formed from numbers \${1,2,....9}? + with no repetitions: $P_{9,3} = 60480$ + with repetitions: $9 * .... * 9 = 9^6$ ways + containing repetitions: $9^6 - 60480 = 470961$

2.1.3 Combination

How may subsets of $\{1, 2, 3\}$ contain exactly 2 elements? (order isn't important)

Answer: 3 subsets - $\{1, 2\}, \{1, 3\}, \{2, 3\}$

Definition: The number of subsets with r elements of a set with n elements is called **number of combinations of n things taken r at a time**.

Notation: $C_{n,r}$ or $\binom{n}{r}$. These are also called **binomial coefficients**.

$$C_{n,r} = \frac{n!}{r!(n-r)!}$$

The difference between permutation and combination:

- Combination: $(a, b, c) = (b, a, c)$ i.e. order doesn't concern,
- Permutation: $(a, b, c) \neq (b, a, c)$ i.e. concerned with order.

Choosing a permutation is same as first choosing a combination and putting the elements in order.

$$\frac{n!}{(n-r)!} = \binom{n}{r} r!$$
$$\frac{n!}{(n-r)!r!} = \binom{n}{r}$$

Following results should be intuitive:

- $\binom{n}{r} = \binom{n}{n-r}$
- $\binom{n}{0} = \binom{n}{n} = 1$
- $\binom{n}{1} = \binom{n}{n-1} = n$

2.1.4 Binomial Theorem

$$(x+y)^n = \Sigma_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

This is where **Pascal's triangle** comes from.

Corollary: Surprising fact

$$\Sigma_{i=0}^{n} \binom{n}{i} = 2^n$$

Proof: By the binomial theorem:

$$\begin{aligned} 2^n &= (1+1)^n \\ &= \Sigma_{i=0}^{n} \binom{n}{i} 1^i 1^{n-i} \end{aligned}$$

2.1.5 Problems

Q. Select 2 cards from a dect without replacement and care about order? Possibilities $= 52 * 51 = 2652$ ways.

Q. Box of 10 sox - 2 red and 8 black. Pick 2 without replacement.

- Let A be event that both are red.

  $P(A) = \frac{\text{ways to pick 2 reds}}{\text{ways to pick 2 sox}} = \frac{2*1}{10*9} = \frac{1}{45}$
- Let B be event that both are black.

  $P(B) = \frac{8*9}{10*9} = \frac{28}{45}$

- Let C be one of each color. Since, A and B are disjoint,

  $P(C) = 1 - P(C^c) = 1 - P(A \cup B) = 1 - \frac{1}{45} - \frac{28}{45} = \frac{16}{45}$

Q. An NBA team has 12 players. How many ways can the coach choose the starting 5?

$\binom{12}{5} = \frac{12!}{5!7!} = 792$

Q. Smith is one of the players on the team. How many of 792 starting lineup include him?

$\binom{11}{4} = \frac{11!}{4!7!} = 330$

Q. 4 red marbles, 2 whites. Put them in random order.

a. P(2 end marbles are W)

$S = \{$Possible pairs of slots that W's occupy$\}$

$|S| = \binom{6}{2} = \frac{6!}{2!(6-2)!} = 15$

Since, W's must occupy end slots so, $|A| = \binom{2}{2} = 1$

$P(A) = \frac{|A|}{|S|} = \frac{1}{15}$

b. $P($2 end marbles aren't both W$\} = 1 - P(A) = \frac{14}{15}$

c. $P($2 W's are side by side$\}$

WWRRRR or RWWRRR or RRWWRR or RRRWWRR or RRRRWW

$|B| = 5$

$P(B) = \frac{5}{15}$

2.2 Hypergeometric Distribution

Definition: You have a objects of type 1 and b objects of type 2. Select n objects **without replacement** from $a+b$ objects. Then,

$$
\begin{aligned}
P(\text{k type 1's were picked}) &= \frac{\text{(Number of ways to choose k type 1's out of a)(Choose n-k type 2's out of b)}}{\text{(Number of ways to choose n out of a+b)}} \\
&= \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}
\end{aligned}
$$

The number of type 1's chosen is said to have the **hypergeometric distribution**.

Example: 3 sox in box with $a = 2$ red, $b = 1$ blue. Pick $n = 3$ without replacement.

$$
\begin{aligned}
P(\text{Exactly k=2 reds are picked}) &= \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}} \\
&= \frac{\binom{2}{2}\binom{1}{1}}{\binom{3}{3}} \\
&= 1
\end{aligned}
$$

2.3 Binomial Distribution

Definition: You again have a objhects of type 1 and b objects of type 2. Now, select n objects **with replacement** from $a+b$ objects.

$$
\begin{aligned}
P(\text{k type 1's were picked}) = &\text{(Number of ways to choose k 1's and n-k 2's)} \\
&P(\text{Choose k 1's in a row then n-k 2's in a row})
\end{aligned}
$$

$$
P(\text{k type 1's were picked}) = \binom{n}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k}
$$

2.4 Multinomial Coefficients

Example: n_1 blue sox, n_2 reds. The number of assortments is $\binom{n_1+n_2}{n_1}$. Generalization for k types of objects: $n = \Sigma_{i=1}^{k} n_i$ The number of arrangements is

$$
\binom{n}{n_1, n_2, \ldots.n_k} = \frac{n!}{n_1!n_2!\ldots.n_k!}
$$

This is known as **multinomial coefficient**.

Example: How many ways letters in "MISSISSIPPI" be arranged?

$$\frac{\text{Number of permutations of 11 letters}}{\text{(Number of M's)(Number of P's)(Number of I's)(Number of S's)}}$$

$$= \frac{11!}{1!2!4!4!}$$

2.5 Conditional Probability

The probability of A occurs given B occurs is

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

Definition: If $P(B) > 0$, the conditional probability of A given B is

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

Remark: If A and B are disjoint, then $P(A/B) = 0$. If B occurs, there is no chance that A can occur.

What happens if $P(B) = 0$? In that case, no need to consider $P(A/B)$.

Example: Toss 2 dice and take the sum.

A: odd toss $= \{3, 5, 7, 9, 11\}$

B: $\{2, 3\}$

$P(A) = P(3) + ... + P(11) = \frac{2}{36} + \frac{4}{36} + ... + \frac{2}{36} = \frac{1}{2}$

$P(B) = \frac{1}{36} + \frac{2}{36} = \frac{1}{12}$

$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{36}}{\frac{1}{12}} = \frac{2}{3}$

Example: A couple has two kids and at least one is boy. What's the probability that both are boys?

$S = \{GG, GB, BG, BB\}$

C : Both are boys $= \{BB\}$

D : At least 1 boy $= \{GB, BG, BB\}$

$P(C/D) = \frac{P(C \cap D)}{P(D)} = \frac{P(C)}{P(D)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$

Example: A couple has two kids and at least one is born on tuesday. What is the probability that both are boys?

$B_x[G_x] = Boy[Girl]$
born on day x; x = 1,2,....7

$x = 3$ is Tuesday.

$S = \{(G_x, G_y), (G_x, B_y), (B_x.G_y), (B_x, B_y), x, y = 1, 2, ...7\}$

So, $|S| = 4 * 49 = 196$

i.e. 4 combination of B and G and 49 combination of x and y.

C: Both are boys (with at least one born on tuesday)

$= \{(B_x, B_3), x = 1, 2, ..., 7\} \cup \{(B_3, B_y), y = 1, 2, ...7\}$

Note: $|C| = 13\{$to avoid double counting$(B_3, B_3)\}$

D: There is at least one boy born on Tuesday.

$= C \cup \{(G_x, B_3), (B_3, G_y), x, y = 1, 2, ...7\}$

$|D| = 27$

$P(C/D) = \frac{P(C \cap D)}{P(D)} = \frac{P(C)}{P(D)} = \frac{\frac{13}{196}}{\frac{27}{197}} = \frac{13}{27}$

Properties: Analogous to axioms of probability

- $0 \leq P(A/B) \leq 1$
- $P(S/B) = 1$
- $A_1 \cap A_2 = \phi \rightarrow P(A_1 \cap A_2/B) = P(A_1/B) + P(A_2/B)$
- If A_1, $A_2, ...$ are all disjoint then

$$P(U_{i=1}^{\infty} A_i/B = \Sigma_{i=1}^{\infty} P(A_i/B)$$

2.6 Independence

Any unrelated events are independent.

Example:

A: It rains on Mars tomorrow.

B: Coin lands on H.

Definition: A & B are independent iff $P(A \cap B) = P(A).P(B)$

Remark: If $P(A) = 0$, then A is independent of any other event.

Remark: Events don't have to be physically unrelated to be independent.

Theorem: Suppose $P(B) > 0$. Then A and B are independent $\leftrightarrow P(A/B) = P(A)$.

Proof: A & B independent $\leftrightarrow P(A \cap B) = P(A).P(B) \leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A)$

Remark: So, if A and B are independent, the probability of A doesn't depend on whether or not B occurs.

Bayes Theorem: A and B are independent $\leftrightarrow A'$ and B' are also independent.

Proof: Only need to prove in $\rightarrow$ direction (then $\leftarrow$ follows trivially).

$$P(A) = P(A \cap B') + P(A \cap B)$$

So,

$$\begin{aligned}
P(A \cap B') &= P(A) - P(A \cap B) \\
&= P(A) - P(A).P(B) \qquad \{\text{A, B are independent}\} \\
&= P(A)\{1 - P(B)\} \\
&= P(A).P(B')
\end{aligned}$$

Don't confuse independence with disjointness!

Theorem: If $P(A) > 0$ and $P(B) > 0$, A and B can't be independent and disjoint at the same time.

Proof: Suppose A and B are disjoint, $A \cap B = \phi$. Then, $P(A \cap B) = 0 < P(A).P(B)$. Thus, A and B aren't independent. Similarly, independent doesn't imply disjoint.

Remark: In fact, independence and disjointness are almost opposite. If A and B are disjoint and A occurs, then you have information that B cannot occur. So, A and B can't be independent.

Extension to more than two events:

Definition: A, B, C are independent iff

- $P(A \cap B \cap C) = P(A).P(B).P(C)$
- All pairs are independent:

  $P(A \cap B) = P(A).P(B)$

  $P(A \cap C) = P(A).P(C)$

  $P(B \cap C) = P(B).P(C)$

General Definition: $A_1, ..., A_k$ are independent iff $P(A_1 \cap ..... \cap A_k) = P(A_k)$ and all subsets of $\{A_1, ..., A_k\}$ are independent.

Independent Trials: Perform n trials of an experiment such that the outcome of one trial is independent of outcomes of other trials. Eg. Flip 3 coins independently.

Remark: For independent trials, you just multiply the individual probabilities.

Eg. Flip a coin infinitely many times (each flip is independent of others).

$$
\begin{aligned}
P_n &= P(\text{First H on nth trial}) \\
&= P(\underbrace{TT...TH}_{n-1}) \\
&= \underbrace{P(T).P(T)...P(T)}_{n-1}.P(H) \\
&= (\frac{1}{2})^{n-1}.\frac{1}{2} = \frac{1}{2^n} \\
&= \frac{1}{2^n} \quad \{\text{Each has probability } 1/2\} \\
P(H\ eventually) &= \Sigma_{n=1}^{\infty} P_n \\
&= \Sigma_{n=1}^{\infty} 2^{-n} \\
&= 1
\end{aligned}
$$

2.7 Partitions and laws of probability

Partition of Sample Space split the sample space into disjoint, yet all encompassing subsets.

Definition: The events $A_1, A_2, ..., A_n$ form a partition of sample space S if

- $A_1, A_2, ..., A_n$ are disjoint.
- $U_{i=1}^{n} A_i = S$
- $P(A_i) > 0$ for all i.

Remark: When an experiment is performed, exactly one A_i's occur.

Example: A and A' form partition.

Suppose $A_1, A_2, ..., A_n$ form partition of S and B is arbitrary event. Then,

$$B = U_{i=1}^{n}(A_i \cap B)$$

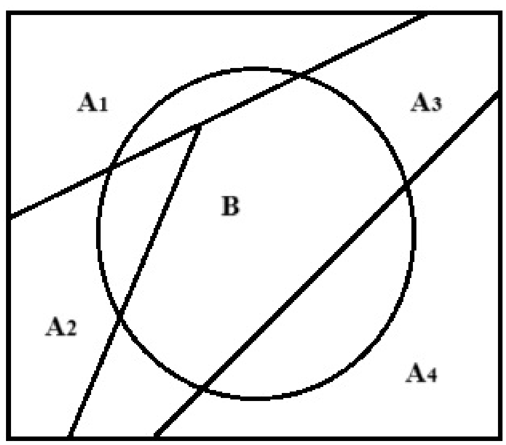

Figure 1: Partitions

$$\begin{aligned} P(B) &= P[U_{i=1}^{n}(A_i \cap B)] \\ &= \Sigma_{i=1}^{n} P(A_i \cap B) \quad (\text{Since} A_1, A_2, ...A_n \text{ are disjoint}) \\ &= \Sigma_{i=1}^{n} P(A_i)P(B/A_i) \quad (\text{Definition of conditional Probability}) \end{aligned}$$

This is **law of probability**.

Example: Suppose we have 10 Georgia Tech students and 20 University of Georgia students taking a test. GT students have 95% chance of passing but UGA have 50%. Determine probability that he/she passes.

$P(passes) = P(GT)P(passes/GT) + P(UGA)P(passes/UGA)$

2.8 Bayes Theorem

Immediate consequence of law of total probability.

Bayes Theorem: If $A_1, A_2, ...A_n$ form partition of S and B is any event then,

$$\begin{aligned} P(A_j/B) &= \frac{P(A_j \cap B)}{P(B)} \\ &= \frac{P(A_j)P(B/A_j)}{\Sigma_{i=1}^{n} P(B/A_i)} \end{aligned}$$

The $P(A_j)$'s are prior probabilities ("before B").

The $P(A_j/B)$'s are posterior probabilities ("after B").

The $P(A_j/B)$'s add up to 1.

2.9 Probability Problems