Part 1

Week 1: Counting & Permutation

(Basic) Sum Rule of Counting
Set A of n outcomes, set B of m outcomes
No outcome in A is in B, vice versa - when $A \cap B$ is the empty set
An experiment is performed by drawing one outcome from **either** A or B
There are m + n possible outcomes
Example: servers

Product Rule of Counting
Experiment A of n outcomes, Experiment B of m outcomes
An experiment is performed by drawing one outcome from **both** A and B
For the two experiments together there are mn possible outputs
Example: bits in a byte

- Generalised Product rule of counting = extending the Product Rule of counting to many experiments i.e. for the r experiments together there are $n1 \times n2 \times \cdots \times nr$ possible outputs.
- Generalised Sum Rule of Counting = when $A \cap B$ is not the empty set. For the experiments together there are $|A| + |B| - |A \cap B|$ possible outputs

Set {A, B} has two permutations {A,B}, {B,A} and one combination {A,B}

Number of permutations, with/without repeated objects
Permutation - Counting the number of ways to generate an ordered (order matters and is counted) subset of size k from a set of n distinguishable objects.

In general, number of permutations of n objects is n! – by direct application of product rule.
E.g the number of ways we can arrange a, b, c = 3!
However when repeated objects are concerned e.g m,o,o there are 3!/2! Ways to arrange the letters

$n! / (n1!n2! \cdots nk!)$

Number of combinations (n choose k)
Combination - • Counting the number of ways to generate an unordered(order does not matter and is not counted) subset of size k from a set of n distinguishable objects

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

E.g. Number of distinct pizzas we can create by selecting 4 toppings from 6 available. 6 choose 4 = 15

Effect of simple constraints on counting e.g. two people must sit together or must not
Add up all possible combinations that take into account the constraint or minus the constraint combination from all possible combinations. NB Test Questions 1,Question 4

Week 2: Axioms of Probability, Conditional Probability and Bayes Theorem

Sample Space = The set of all possible outcomes of an experiment
Random Event = subset of sample space

Sets:
Know set operations: union, intersection, complement and combinations of these
Basic Properties of sets: draw venn diagrams to figure out
DeMorgan's laws:
E and F are events,
$(E \cup F)^c = E^c \cap F^c$
$(E \cap F)^c = E^c \cup F^c$

Axioms of Probability:

Axiom 1: $0 \le P(E) \le 1$
Axiom 2: $P(S) = 1$, where S is sample space (set of all possible outcomes)
Axiom 3: If E and F are mutually exclusive ($E \cap F = \varnothing$, they cannot occur at the same time) then $P(E \cup F) = P(E) + P(F)$.

Immediate Consequences/Implications of Axioms incl proof:

$P(E^c) = 1 - P(E)$
since $S = E \cup E^c$ and $E \cap E^c = \varnothing$ then $P(S) = 1 = P(E) + P(E^c)$
$E \subset F$ implies that $P(E) \le P(F)$
since $F = E \cup (E^c \cap F)$ and $E \cap E^c = \varnothing$
then $P(F) = P(E) + P(E^c \cap F))$
$P(E^c \cap F) \ge 0$ so $P(E) = P(F) - P(E^c \cap F)) \le P(F)$
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$
$E \cup F = E \cup (E^c \cap F)$ and $E \cap (E^c \cap F) = \varnothing$ (mutually exclusive)
$F = (E \cap F) \cup (E^c \cap F)$, also mutually exclusive
So $P(E \cup F) = P(E) + P(E^c \cap F)$
and $P(F) = P(E \cap F) + P(E^c \cap F)$ i.e. $P(E^c \cap F) = P(F) - P(E \cap F)$
Equally Likely Outcomes
P(S) = 1, P(Heads) = P(Tails) = 1, 2p = 1, p = 1/2

Sampling with and without replacement

Conditional Probability = the probability that event E occurs given that event F has already occurred. Written as P(E|F)**.** It is a probability where Sample space is restricted to $S \cap F$ and event space is restricted to $E \cap F$

General definition:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

where $P(F) > 0$. Implies

$$P(E \cap F) = P(E|F)P(F)$$

known as the chain rule – its important !
If $P(F) = 0$?

- $P(E|F)$ is undefined
- Can't condition on something that can't happen

NB CHAIN RULE - P(E∩F) = P(E|F)P(F)
Conditional prob is a probability (satisfies axioms, incl proof.)
0 ≤ P(E|F) ≤ 1
E ∩ F ⊂ F so P(E ∩ F) ≤ P(F) and P(E|F) = P(E∩F) P(F) ≤ 1
P(S|F) = 1
P(S|F) = P(S∩F) / P(F) = P(F) /P(F) = 1 (chain rule)
If E1, E2 are mutually exclusive events then P(E1 ∪ E2|F) = P(E1|F) + P(E2|F)
P(E1 ∪ E2|F) = P((E1∪E2)∩F) P(F) = P((E1∩F)∪(E2∩F)) P(F) = P(E1∩F)+P(E2∩F) / P(F)

Marginalisation incl proof:
Suppose we have mutually exclusive events F1, F2, ..., Fn such that F1 ∪ F2 ∪ ... ∪ Fn = S. Then P(E) = P(E ∩ F1) + P(E ∩ F2) + · · · + P(E ∩ Fn)
Prove by chain rule - see slide 15

NB Special Case: use in HIV/Guilty question
P(E) = P(E|F)P(F) + P(E|F c)P(F c) = P(E|F)P(F) + P(E|F c)(1 − P(F))

Marginalisation example:
Roll two coins. What is the probability that the first coin is heads ?
Event E is first coin heads, F1 is second coin heads, F2 is second coin tails
P(E) = P(E ∩ F1) + P(E ∩ F2) = (1/2 × 1/2) + (1/2 × 1/2) = 1/2

Bayes Rule:

P(E|F) = posterior = updated probability of E after observing F
P(F|E) = likelihood = probability of F given E
P(E) = prior = our guess with no extra info
P(F) = evidence = probability with no extra info

Week 3: Independance

Two events E and F are independent if the order in which they occur doesn't matter. Alternatively, if observing one doesn't affect the other.

Definition: Two events E and F are independent if $P(E \cap F) = P(E)P(F)$

NB three events are independent if they are pairwise independent <u>and</u> triply independent
$P(E \cap F \cap G) = P(E)P(F)P(G)$ and
$P(E \cap F) = P(E)P(F)$,
$P(E \cap G) = P(E)P(G)$,
$P(F \cap G) = P(F)P(G)$,

Caution re fragility of independence assumptions
Multiplying can result in very small probabilities e.g. probability of defaulting on a mortgage

Definition of conditional independence
NB: Independent events can become dependent when we condition on additional information. Also dependent events can become independent.

Two events E and F are called <u>conditionally independent</u> given G if:
$P(E \cap F|G) = P(E|G)P(F|G)$

NB: if two events are independent it does not follow that they are conditionally independent and vice versa

Part 2

Week 4: Random Variables, Bernoulli and Binomial RVs

Random Variable = Effectively mapping every event to a real number
- **Discrete Random Variable =** X can only take on discrete values
- **Continuous Random Variable =** X can take on continuous data

Indicator Random Variable = takes value 1 if event E occurs and 0 if event E does not occur. The sample space, I = {1,0}

Random Variable X can be associated with outcomes of an experiment e.g. X = 0 when outcome is (T,T), X = 1 when outcome is (H,T) or (T, H), X = 2) when outcome is (H, H).

Events are linked to values of random variables, so can apply ideas for events directly to RVs (chain rule, Bayes, marginalisation)

Probability Mass Function:
A probability is associated with each value that a discrete random variable can take

Cumulative Distribution Function:
$F(a) = P(X \leq a)$ where a is real-valued.
Use PMF to calculate e.g. $F(2) = 3/10 = P(X = 0) + P(X = 1) + P(X = 2)$ so $P(X = 2) = 2/10$

Bernoulli Random Variable: when an experiment results in success or failure, X ~ Ber(p).

Binomial Random Variable: when the number of successes or failures is counted i.e. a sum of n bernoulli random variables.

- X is a **Binomial** random variable: $X \sim Bin(n, p)$

$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \; i = 0, 1, \cdots, n$$

(recall $\binom{n}{i}$ is the number of outcomes with exactly i successes and $n - i$ failures)

Examples:

- number of heads in n coin flips
- number of 1's in randomly generated bit string of length n
- number of packets erased out of a file of n packets

Suppose X ~ Bin(n1, p) and Y ~ Bin(n2, p) (its important than p is the same for both) Then Z = X + Y ~ Bin(n1 + n2, p)
Z = X1 + X2 + · · · + Xn1 + Y1 + Y2 + · · · + Yn2 . All terms are independent, all are Ber(p).
Use in voters question.

Simple stochastic simulation: generating bernoulli and binomial samples in matlab

Week 5: Mean, Variance, Correlation and Conditional Expectation

Definition of expected value = Sometimes we have a number of measurements that we want to summarise by a single value, also referred to as mean or average

Viewing the probability of an event as the frequency with which it occurs when an experiment is repeated many times, the expected value tells us about the overall outcome we can expect.

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i)$$

I.e. Multiply each value times its respective probability.
Interpretation of expected value in games of chance/reward
NB - if it costs to play the game you must take this into account when calculating reward
Expected value of an indicator RV
Suppose I is the indicator variable for event E (so I = 1 if event E occurs, I = 0 otherwise).
Then E[I] = 1 × P(E) + 0 × (1 − P(E)) = P(E).

Expected value of number of iterations of repeated game (coin tossing etc)
Limitations of expected value in games of chance/reward e.g. gamblers ruin

Linearity of expected value incl proof.

E[aX + b] = aE[X] + b

$$
\begin{aligned}
E[aX+b] &= \sum_{i=1}^{n}(ax_i+b)P(X=x_i) \\
&= \sum_{i=1}^{n} ax_i P(X=x_i) + \sum_{i=1}^{n} bP(X=x_i) \\
&= a\sum_{i=1}^{n} x_i P(X=x_i) + b\sum_{i=1}^{n} P(X=x_i) \\
&= aE[X] + b
\end{aligned}
$$

Use example = changing currency

Use of linearity in expected value of sums of random variables

E[aX + bY] = aE[X] + bE[Y]

for any two random variables X and Y and constants a and b.
Proof:

$$
\begin{aligned}
E[aX+bY] &= \sum_x\sum_y (ax+by)P(X=x \text{ and } Y=y) \\
&= a\sum_x\sum_y xP(X=x \text{ and } Y=y) + b\sum_y\sum_x yP(X=x \text{ and } Y=y) \\
&\overset{(a)}{=} a\sum_x xP(X=x) + b\sum_y yP(Y=y) \\
&= aE[X] + bE[Y]
\end{aligned}
$$

(a) Recall marginalising, $\sum_y P(X=x \text{ and } Y=y) = P(X=x)$

Expected value of product of independent RVs, incl proof.

[XY] = E[X]E[Y]

- Take two **independent** random variables X and Y
- $E[XY] = E[X]E[Y]$
- Proof:

$$\begin{aligned} E[XY] &= \sum_x \sum_y xyP(X = x \text{ and } Y = y) \\ &= \sum_x \sum_y xyP(X = x)P(Y = y) \\ &= \sum_x xP(X = x) \sum_y yP(Y = y) \\ &= E[X]E[Y] \end{aligned}$$

Definition of variance = Variance is a summary value (a statistic) that quantifies “spread”
Let X be a random variable with mean μ.
The variance of X is Var(X) = E[(X − μ)^2]
Variance is mean squared distance of X from the mean μ
Var(X) ≥ 0
Standard deviation is square root of variance p Var(X)

Alternative expression for discrete random variables : Var(X) = E[X^2] − (E[X])^2

Proof:

$$\begin{aligned} Var(X) &= \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i) \\ &= \sum_{i=1}^{n} (x_i^2 - 2x_i\mu + \mu^2) p(x_i) \\ &= \sum_{i=1}^{n} x_i^2 p(x_i) - 2\sum_{i=1}^{n} x_i p(x_i)\mu + \mu^2 \sum_{i=1}^{n} p(x_i) \\ &= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 \\ &= E[X^2] - (E[X])^2 \end{aligned}$$

var(aX+b)=a^{2}var(X), incl proof.
Unlike expectation, variance is not linear. Instead we have the above ^. Observe that b does not affect variance.

Proof:

$$
\begin{aligned}
Var(X) &= \sum_{i=1}^{n}(x_i - \mu)^2 p(x_i) \\
&= \sum_{i=1}^{n}(x_i^2 - 2x_i\mu + \mu^2)p(x_i) \\
&= \sum_{i=1}^{n} x_i^2 p(x_i) - 2\sum_{i=1}^{n} x_i p(x_i)\mu + \mu^2 \sum_{i=1}^{n} p(x_i) \\
&= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 \\
&= E[X^2] - (E[X])^2
\end{aligned}
$$

Variance of sum of independent RVs, incl proof.

Var(X + Y) = Var(X) + Var(Y).

$$
\begin{aligned}
Var(X+Y) &= E[(X+Y)^2] - E[X+Y]^2 \\
&= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2 \\
&= E[X^2] + 2E[XY] + E[Y^2] - E[X]^2 - 2E[X]E[Y] - E[Y]^2 \\
&= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 + 2E[XY] - 2E[X]E[Y] \\
&= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 \\
&= Var(X) + Var(Y)
\end{aligned}
$$

(recall $E[XY] = E[X]E[Y]$ when X and Y are independent)

- Bernoulli random variable, $X \sim Ber(p)$:

$$
\begin{aligned}
E[X] &= p \\
Var(X) &= E[X^2] - (E[X])^2 = p - p^2 = p(1-p)
\end{aligned}
$$

- Binomial random variable, $X \sim Bin(n, p)$. Sum of n $Ber(p)$ independent random variables so:

$$
\begin{aligned}
E[X] &= np \\
Var(X) &= np(1-p)
\end{aligned}
$$

Definition of joint PMF
$P(X = x \text{ and } Y = y)$

PMF of X given Y = y is $P(X = x|Y = y) = P((X=x \text{ and } Y =y) / P(Y =y)$

Definition of covariance
Say X and Y are random variables with expected values μX and μY .
The covariance of X and Y is defined as:
$Cov(X, Y) = E[(X - \mu X)(Y - \mu Y)] = E[XY] - E[X]E[Y]$

Recall when X and Y are independent then $E[XY] = E[X]E[Y]$, so $Cov(X, Y) = 0$. But $Cov(X, Y) = 0$ does not imply that X and Y are independent

NB - $Cov(X, X) = Var(X)$

Definition of correlation

The correlation is another example of a summary statistic. It indicates the strength of a linear relationship between X and Y .

Correlation says nothing about the slope of line (other than its sign).
When relationship between X and Y is not roughly linear, correlation coefficient tells us almost nothing

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Correlation varies between -1 and 1
If $X = Y$ then $corr(X, Y) = 1$. If $X = -Y$ then $corr(X, Y) = -1$.

Recall when X and Y are independent then $E[XY] = E[X]E[Y]$, so $corr(X, Y) = 0$. But $corr(X, Y) = 0$ does not imply that X and Y are independent.

Correlation vs causality
Correlation does not imply causality

Conditional expectation

$$E[X|Y = y] = \sum_x xP(X = x|Y = y)$$

Conditional expectation can be used to make predictions
Linearity of conditional expectation, incl proof.
Marginalisation and conditional expectation, incl proof
Use of conditional expectation in random sums