CS130 - Probability Summary

Concerned with random experiments, which result in elementary outcomes, which make up the sample space.

$$\underbrace{\Omega}_{\text{Sample Space}} = \{ \underbrace{a}, \underbrace{b}, \underbrace{c} \}$$

Elementary outcome

Events are sets of elementary outcomes:

$$\underbrace{E}_{\text{Event}} \subseteq \Omega$$

Permutations of a set are sequences whose length is the size of the set, and whose elements are drawn from the set without repition

Arrangements of a set are sequences of a fixed length, whose elements are drawn from the set without repition.

The number of arrangements of length n of a set of size r is

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

Combinations are subsets of a fixed length. The number of combinations of length n from a set of size r is

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

The probability of an event E occurring is denoted as P(E). It can be conceptualised in a number of ways:

1) The frequency of an event over a number of trials

2) The certainty with which an event will occur, e.g. 1 meaning the event is certain to occur.

We formally model this with "probability spaces":

$$(\underbrace{\Omega}, \underbrace{F}, \underbrace{P})$$

Sample space

Collection of events, 2^{Ω} in our case, but sometimes more complicated

$P : F \rightarrow [0;1]$

Probability measure, a function that maps events to their probabilities.

Probability axioms

1) $P(E) \in [0; 1]$

2) $P(E_1 \cup E_2 \cup \dots \cup E_k) = \sum_{i=1}^{k} P(E_i)$

If all events are disjoint

3) $P(\bar{A}) = 1 - P(A)$

Corrolaries of the axioms

1) $P(A \cup B) = P(A) + P(B)$

If A and B are disjoint, i.e. independent

1.2) $P(A \cup B) = P(A) + P(\bar{A} \cap B)$

Since for (1) to hold A and B are disjoint

1.3) $P(B) = P(\bar{A} \cap B) + P(A \cap B)$

Since
$$
\begin{aligned}
B &= \Omega \cap B \\
&= (A \cup (\Omega \setminus A)) \cap B \\
&= (A \cap B) \cup (\bar{A} \cap B)
\end{aligned}
$$

2) $P(A \cup B) \leq P(A) + P(B)$

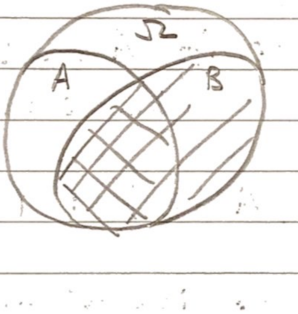
Irrespective of whether A and B are disjoint

Conditional probability concerns finding the probability of an event if another is known to have occurred.

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

<u>Example</u>

Consider a probability space $(\Omega, 2^{\Omega}, P)$ and two events $A, B \in 2^{\Omega}$, $P(B) \neq 0$.



We are essentially "shrinking" the sample space from Ω to B, as we know it occurs, then looking for $P(A)$ within the new sample space.

Law of total probability:

Consider the events $B_1, B_2, \ldots, B_k$ forming a partition of Ω

$$\Rightarrow P(B_i \cap B_j) = 0 \quad \forall i, j, \ i \neq j$$

$$\Rightarrow P\left(\bigcup_{x=1}^{k} B_x\right) = 1$$

And let $P(B_i) > 0 \ \forall i$.

$\forall A \in 2^{\Omega}$,

$$P(A) = \sum_{x=1}^{k} P(A \mid B_x) \cdot P(B_x)$$

Explanation

Each event B_i represents a hypothetical which could precede the event A. So the partition of Ω contains all possible events that could precede A, so the weighted average comes back to $P(A)$.

Bayes theorem

↳ "use when you have a hypothesis, and want to update it given you've observed some new evidence"

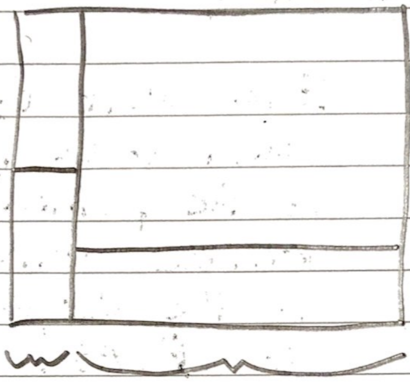
Suppose, as before, we have events B_k which partition Ω, and there is an event $A \in 2^{\Omega}$ which we know occurred

$$P(B_i \mid A) = \frac{P(A \mid B_i) \cdot P(B_i)}{\sum_{j=1}^{k} P(A \mid B_j) \cdot P(B_j)}$$

Normally, we only consider one event, B, not a partition $B_1 \dots B_k$

<u>Explanation</u> Finding: $P(H|E)$ "Posterior"

Difference between $P(E|H)$ and $P(E|\neg H)$ mean how much the evidence changes the likelihood of the event occurring



"Likelihood" $P(E|H)$

$P(E|\neg H)$

Likelihood of hypothesis occurring <u>with</u> evidence,

"Prior" $P(H)$

$P(\neg H)$

Likelihood of hypothesis occurring without evidence

$$P(H|E) = \frac{P(H)\cdot P(E|H)}{P(H)\cdot P(E|H) + P(\neg H)\cdot P(E|\neg H)}$$

$$= \frac{P(H)\cdot P(E|H)}{P(E)}$$

Independent events are when the probability of one event occurring is not affected by whether another event occurs. this can be formalised as:

$$P(A \mid B) = P(A)$$

$$\Rightarrow P(A \cap B) = P(A) \cdot P(B)$$

$P(B) \neq 0$

Bernoulli trials are random experiments with two outcomes (generalised to binomial distribution, $n=2$).

$$\Omega = \{h, t\}$$
$$P(\{h\}) = p, \quad p \in [0;1]$$
$$P(\{t\}) = 1-p$$

We can then write an "indicator function" X, which we say has a Bernoulli distribution with parameter p.

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = h \\ 0 & \text{otherwise.} \end{cases}$$

Averages can be defined in various ways, including mean, median and mode.
We can find the "expectation value" of a random variable as a weighted mean.

$$E(X) = \sum_i \underbrace{i}_{\text{Value}} \cdot \underbrace{P(X=i)}_{\text{Probability that the value occurs}}.$$

The expectation values of the following standard distributions are:

- $X \sim \text{Uniform}(n)$, $E(X) = \frac{n+1}{2}$
  ↳ In this case, equal to median
- $X \sim \text{Binomial}(n, p)$, $E[X] = np$
  ↳ Bernoulli is special case $n=1$

$$E(X) + E(Y) = E(X+Y)$$
$$E(a \cdot X) = a \cdot E(X)$$

$$E(X) \cdot E(Y) \neq E(X \cdot Y)$$

The Variance of a random Variable X measures the "spread" of its events across the range of probabilities

$$\mathrm{Var}[X] = E[X^2] - E(X)^2$$

The Square root of this is called the Standard deviation, and is often used to remove the "squares" introduced.

$E(X^k)$ is called the k^{th} moment of X.

Markov's inequality is a theorem that states that

$$\mathbb{P}[X \geq a] \leq \frac{E[X]}{a}$$

where X is a random variable, and a is a number.