6.1 The Probability of an Event

**Definition 6.1.** An outcome in a finite sample space may be assigned a probability mass, a real number in $[0,1]$ such that the sum of the probabilities of all outcomes in $S$ is 1. The probability of an event $A$ is the sum of the probability masses of all outcomes in $A$, subject to

$$0 \leq P(A) \leq 1 \quad P(\emptyset) = 0 \quad P(S) = 1.$$ 

If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots.$$ 

If an experiment can result in $N$ equally likely outcomes, the probability of an event $A$ which contains $n$ outcomes is

$$P(A) = \frac{n}{N}.$$ 

We have the following collection of results.

**Theorem 6.2.** For any two events $A$ and $B$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

This can be proved in a number of ways, for example by use of a Venn Diagram.

**Corollary 6.3.** If $A$ and $B$ are mutually exclusive, $P(A \cup B) = P(A) + P(B)$.

**Corollary 6.4.** If $A_1, A_2, \ldots, A_n$ are mutually exclusive, $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$.

**Theorem 6.5.** $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

This can be proved via Venn diagrams, although a generalized Inclusion-Exclusion rule can be proved by induction. A set-based proof follows from some set operations. See if you can justify each of the steps; you need to know how distribution of intersection and union works: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

$$P(A \cup B \cup C) = P(A \cup (B \cup C))$$

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

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

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

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

**Theorem 6.6.** Complement Rule: $P(A) + P(A') = 1$, i.e. $P(A) = 1 - P(A')$.

This follows from the fact that $A \cup A' = S$ and $P(S) = 1$. 

6-1
6.2 Conditional Probability and Independence

Definition 6.7. The probability that event $A$ occurs when it is known that event $B$ has occurred is a conditional probability and it is written $P(A|B)$, read “The probability of $A$ given $B$”.

Theorem 6.8. The conditional probability $P(A|B)$, assuming $P(B) > 0$, may be calculated as

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

This can be proved in the following way. Because we know that $B$ has occurred, we restrict our sample space to event $B$. Then event $A$ occurring given $B$ means that both $A$ and $B$ occur.

Definition 6.9. Two events $A$ and $B$ are said to be independent if the probability of one event occurring is not affected by whether or not the other event occurs, or in notation

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B).$$

Otherwise, the events are dependent.

**WARNING:** Be careful not to confuse “mutually exclusive” and “independent!” Not only are they not synonymous, two events can never be both independent and mutually exclusive, unless one of the events has probability zero.

This leads to a corollary.

Corollary 6.10. Two events $A$ and $B$ are independent if and only if

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

You can derive this by using the formulaic definition of conditional probability and replacing $P(A|B)$ with $P(A)$.

And the formula for conditional probability may be re-written giving us another corollary.

Corollary 6.11. The Product Rule: For any events $A$ and $B$, $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

This can be extended using induction

Theorem 6.12. For events $A_1, A_2, \ldots, A_k$,

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

If the events are independent then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2) \cdots P(A_k)$$

(but be careful, the converse is not true).

Definition 6.13. The events $\mathcal{A} = \{A_1, A_2, \ldots, A_k\}$ are mutually independent if and only if for any subset $\{A_{i_1}, \ldots, A_{i_r}\}$,

$$P(A_{i_1} \cap \cdots \cap A_{i_r}) = P(A_{i_1}) \cdots P(A_{i_r}).$$
6.3 Bayes Rule

**Definition 6.14.** A partition of a set $S$ is a collection of subsets $E_1, E_2, \ldots, E_k$ such that any two distinct $E_i$ and $E_j$ are mutually exclusive (disjoint) and $E_1 \cup E_2 \cup \cdots \cup E_k = S$.

**Theorem 6.15. The Law of Total Probability:** Let $B_1, \ldots, B_k$ partition $S$. Then for any event $A$,

$$ P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(B_i) P(A|B_i). $$

This can easily be seen to be true since $A \cap B_i$ and $A \cap B_j$ are disjoint for $i \neq j$:

$$(A \cap B_i) \cap (A \cap B_j) = A \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset,$$

and the rest follows from straightforward application of additivity theorems.

**Theorem 6.16. Bayes Rule:** Let $B_1, \ldots, B_k$ partition $S$. Then for any event $A$,

$$ P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r) P(A|B_r)}{\sum_{i=1}^{k} P(B_i) P(A|B_i)}. $$

This is proved by applying the law of total probability to the formula for conditional probability. It can be intuitively justified with a tree diagram, although this leads to some counter-intuitive examples.