1.4 Joint Probabilities

The joint probability of two or more events is the probability of the intersection of those events. For example consider the events $A_1 = \{2, 4, 6\}$, $A_2 = \{4, 5, 6\}$ in the fair die probability space. Thus, A_1 represents obtaining an even number and A_2 obtaining a number larger than 3.

$$P(A_1) = P(\{2\} \quad \{4\} \quad \{6\}) = 3/6 \tag{1.18}$$

$$P(A_2) = P(\{4\} \quad \{5\} \quad \{6\}) = 3/6 \tag{1.19}$$

$$P(A_1 \cap A_2) = P(\{4\} \quad \{6\}) = 2/6 \tag{1.20}$$

Thus the joint probability of A_1 and A_2 is 1/3.

1.5 Conditional Probabilities

The conditional probability of event A_1 given event A_2 is defined as follows

$$P(A_1 \mid A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} \tag{1.21}$$

Mathematically this formula amounts to making A_2 the new reference set, i.e., the set A_2 is now given probability 1 since

$$P(A_2 \mid A_2) = \frac{P(A_2 \cap A_2}{P(A_2)} = 1 \tag{1.22}$$

Intuitively, Conditional probability represents a revision of the original probability measure P. This revision takes into consideration the fact that we know the event A_2 has happened with probability 1. In the fair die example,

$$P(A_1 \mid A_2) = \frac{1/3}{3/6} = \frac{2}{3} \tag{1.23}$$

in other words, if we know that the toss produced a number larger than 3, the probability that the number is even is 2/3.

1.6 Independence of 2 Events

The notion of independence is crucial. Intuitively two events A_1 and A_2 are independent if knowing that A_2 has happened does not change the probability of A_1. In other words

$$P(A_1 \mid A_2) = P(A_1) \tag{1.24}$$

More generally we say that the events A and A_2 are independent if and only if

$$P(A_1 \cap A2) = P(A1)\, P(A_2) \quad (1.25)$$

In the fair die example, $P(A_1 \mid A_2) = 1/3$ and $P(A_1) = 1/2$, thus the two events are not independent.

1.7 Independence of n Events

We say that the events $A_1, \ldots, A_n$ are independent if and only if the following conditions are met:

1- All pairs of events with different indexes are independent, i.e.,

$$P(A_i \cap A_j) = P(A_i)\mathrm{P}(A_j) \quad (1.26)$$

for all $i, j \in \{1, 2, \ldots, n\}$ such that $i = j$.

2- For all triplets of events with different indexes

$$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k) \quad (1.27)$$

for all $i, j, k \in \{1, \ldots, n\}$ such that $i = j = k$.

3- Same idea for combinations of 3 sets, 4 sets, . . .

4- For the n-tuple of events with different indexes

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2) \cdots P(A_n) \quad (1.28)$$

You may want to verify that $2^n - n - 1$ conditions are needed to check whether n events are independent. For example, $2^3 - 3 - 1 = 4$ conditions are needed to verify whether 3 events are independent.

Example 1: Consider the fair-die probability space and let $A_1 = A_2 = \{1, 2, 3\}$, and $A_3 = \{3, 4, 5, 6\}$. Note

$$P(A_1 \cap A_2 \cap A_3) = P(\{3\}) = P(A_1)P(A_2)P(A_3) = 1/6 \quad (1.29)$$

However

$$P(A_1 \cap A_2) = 3/6 = P(A_1)P(A_2) = 9/36 \quad (1.30)$$

Thus A_1, A_2, A_3 are not independent.

Example 2: Consider a probability space that models the behavior a weighted die with 8 sides: $\Omega = (1, 2, 3, 4, 5, 6, 7, 8)$, $F = \ (\Omega)$ and the die is weighted so that

$$P(\{2\}) = P(\{3\}) = P(\{5\}) = P(\{8\}) = 1/4 \quad (1.31)$$

$$P(\{1\}) = P(\{4\}) = P(\{6\}) = P(\{7\}) = 0 \quad (1.32)$$

Let the events A_1, A_2, A_3 be as follows

$$A_1 = \{1, 2, 3, 4\} \tag{1.33}$$

$$A_2 = \{1, 2, 5, 6\} \tag{1.34}$$

$$A_3 = \{1, 3, 5, 7\} \tag{1.35}$$

Thus $P(A_1) = P(A_2) = P(A_3) = 2/4$. Note

$$P(A1 \cap A2) = P(A1)P(A2) = 1/4 \tag{1.36}$$

$$P(A1 \cap A3) = P(A1)P(A3) = 1/4 \tag{1.37}$$

$$P(A2 \cap A3) = P(A2)P(A3) = 1/4 \tag{1.38}$$

Thus A_1 and A_2 are independent, A_1 and A_3 are independent and A_2 and A_3 are independent. However

$$P(A_1 \cap A_2 \cap A_3) = P(\{1\}) = 0 = P(A_1)\, P(A_2)\, P(A_3) = 1/8 \tag{1.39}$$

Thus A_1, A_2 ,A_3 are not independent even though A_1 and A_2 are independent, A_1 and A_3 are independent and A_2 and A_3 are independent.

1.8 The Chain Rule of Probability

Let $\{A_1, A_2, \ldots, A_n\}$ be a collection of events. The chain rule of probability tells us a useful way to compute the joint probability of the entire collection

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

$$P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \% A_2) \cdots P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \tag{1.40}$$

Proof: Simply expand the conditional probabilities and note how the denominator of the term $P(A_k \mid A_1 \cap \cdots \cap A_{k-1})$ cancels the numerator of the previous conditional probability, i.e.,

$$P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2)\cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1}) = \tag{1.41}$$

$$P(A_1)\frac{P(A_2 \cap A_1)}{P(A_1)}\frac{P(A_3 \cap A_2 \cap A_1)}{P(A_1 \cap A_2)}\cdots\frac{P(A_1 \cap \cdots \cap A_n)}{P(A_1 \cap \cdots \cap A_{n-1})} \tag{1.42}$$

$$= P(A_1 \cap \cdots \cap A_n) \tag{1.43}$$

Example: A car company has 3 factories. 10% of the cars are produced in factory 1, 50% in factory 2 and the rest in factory 3. One out of 20 cars produced by the first factory are defective. 99% of the defective cars produced by the first factory are returned back to the manufacturer. What is the probability that a car produced by this company is manufactured in the first factory, is defective and is not returned back to the manufacturer.

Let A_1 represent the set of cars produced by factory 1, A_2 the set of defective cars and A3 the set of cars not returned. We know

$$P(A_1) = 0.1 \qquad (1.44)$$

$$P(A_2 \mid A_1) = 1/20 \qquad (1.45)$$

$$P(A_3 \mid A_1 \cap A_2) = 1 - 99/100 \qquad (1.46)$$

Thus, using the chain rule of probability

$$P(A \cap A_2 \cap A_3) = P(A_1)\, P(A_2 \mid A1) P(A_3 \mid A_1 \cap A_2) = \qquad (1.47)$$

$$(0.1) \quad (0.05)\,(0.01) = 0.00005 \qquad (1.48)$$

1.9 The Law of Total Probability

Let $\{H_1, H_2, \ldots\}$ be a countable collection of sets which is a partition of Ω. In other words

$$H_i \cap H_j = \emptyset, \text{ for } i = j, \qquad (1.49)$$

$$H_1 \quad H_2 \quad \cdots = \Omega. \qquad (1.50)$$

In some cases it is convenient to compute the probability of an event D using the following formula,

$$P(D) = P(H_1 \cap D) + P(H_2 \cap D) + \cdots \qquad (1.51)$$

This formula is commonly known as the Law of Total Probability (LTP)

Proof: First convince yourself that $\{H_1 \cap D, H_2 \cap D, \ldots\}$ is a partition of D, i.e.,

$$(H_i \cap D) \cap (H_j \cap D) = \emptyset, \text{ for } i = j, \qquad (1.52)$$

$$(H1 \cap D) \quad (H2 \cap D) \quad \cdots = D. \qquad (1.53)$$

Thus

$$P(D) = P((H_1 \cap D) \quad (H_2 \cap D) \quad \cdot \cdot \cdot) = \qquad (1.54)$$

$$P(H_1 \cap D) + P(H_2 \cap D) + \cdot \cdot \cdot \qquad (1.55)$$

We can do the last step because the partition is countable.

Example: A disease called pluremia affects 1 percent of the population. There is a test to detect pluremia but it is not perfect. For people with pluremia, the test is positive 90% of the time. For people without pluremia the test is positive 20% of the time. Suppose a randomly selected person takes the test and it is positive. What are the chances that a randomly selected person tests positive?:

Let D represent a positive test result, H_1 not having pluremia, H_2 having pluremia. We know $P(H_1) = 0.99$, $P(H_2) = 0.01$. The test specifications tell us: $P(D \mid H_1) = 0.2$ and $P(D \mid H_2) = 0.9$. Applying the LTP

$$P(D) = P(D \cap H_1) + P(D \cap H_2) \qquad (1.56)$$

$$= P(H_1)P(D \mid H_1) + P(H_2)P(D \mid H_2) \qquad (1.57)$$

$$= (0.99)(0.2) + (0.01)(0.9) = 0.207 \qquad (1.58)$$

1.10 Bayes' Theorem

This theorem, which is attributed to Bayes (1744-1809), tells us how to revise probability of events in light of new data. It is important to point out that this theorem is consistent with probability theory and it is accepted by frequentists and Bayesian probabilists. There is disagreement however regarding whether the theorem should be applied to subjective notions of probabilities (the Bayesian approach) or whether it should only be applied to frequentist notions (the frequentist approach).

Let $D \in F$ be an event with non-zero probability, which we will name . Let $\{H_1, H_2, \ldots\}$ be a countable collection of disjoint events, i.e,

$$H_1 \quad H_2 \quad \cdot \cdot \cdot = \Omega \qquad (1.59)$$

$$H_i \cap H_j = \varnothing \quad if\ I\ =\ j \qquad (1.60)$$

We will refer to $H_1, H_2, \ldots$ as "hypotheses", and D as "data". Bayes' theorem says that

$$P(H_i \mid D) = \frac{P(D \mid H_i)P(H_i)}{P(D \mid H_1)P(H_1) + P(D \mid H_2)P(H_2) + \cdots} \qquad (1.61)$$

where

- $P(H_i)$ is known as the prior probability of the hypothesis H_i. It evaluates the chances of a hypothesis prior to the collection of data.
- $P(H_i \mid D)$ is known as the posterior probability of the hypothesis H_i given the data.
- $P(D \mid H_1)$, $P(D \mid H_2)$, . . . are known as the likelihoods.

Proof: Using the definition of conditional probability

$$P(H_i \mid D) = \frac{P(H_i \cap D)}{P(D)} \qquad (1.62)$$

Moreover, by the law of total probability

$$P(D) = P(D \cap H_1) + P(D \cap H_2) + \cdots = \qquad (1.63)$$

$$P(D \mid H_1)P(H_1) + P(D \mid H_2)P(H_2) + \cdots \qquad (1.64)$$

Example: A disease called pluremia affects 1 percent of the population. There is a test to detect pluremia but it is not perfect. For people with pluremia, the test is positive 90% of the time. For people without pluremia the test is positive 20% of the time. Suppose a randomly selected person takes the test and it is positive. What are the chances that this person has pluremia?:

Let D represent a positive test result, H_1 not having pluremia, H_2 having pluremia. Prior to the the probabilities of H_1 and H_2 are as follows: $P(H_2) = 0.01$, $P(H_1) = 0.99$. The test specifications give us the following likelihoods:

$P(D \mid H_2) = 0.9$ and $P(D \mid H_1) = 0.2$. Applying Bayes' theorem

$$P(H_2 \mid D) = \frac{(0.9)(0.01)}{(0.9)(0.01) + (0.2)(0.99)} = 0.043 \qquad (1.65)$$

Knowing that the test is positive increases the chances of having pluremia from 1 in a hundred to 4.3 in a hundred.