1.2 Interpretations

1. Relative Frequency

One way to interpret probability is the notion of **frequency**, that is, how often a particular event occurs. If a trial is repeated, we can talk about the **relative frequency** of an event. This is an empirical number which is given by a ratio. If event A occurs m times in n total trials, then the relative frequency of event A is $\frac{m}{n}$. On the other hand, there is a theoretical number associated with A, namely the probability that A occurs in a given trial. If we let $P_n(A)$ be the empirical relative frequency of A in n trials and let $P(A)$ be the probability that A occurs, then a theorem called **the law of large numbers** states that as the number of trials increases, $P_n(A)$ most likely is a good approximation for $P(A)$. That is to say that the relative frequency likely approaches the theoretical probability of A occurring as the number of trials is large.

1.3 Distributions

2. Set Theory

In this section we will give some basic set theory notation an definitions:

First, we define a given set using the notation $A = \{x \mid x \text{ satisfies some criteria }\}$. Here A is the set of x which satisfy some membership criteria.

(1) If A is a set then $x \in A$ means that x is a **member** or **element** of A.
(2) If A and B are sets, then $B \subseteq A$ means B is contained in A or is a **subset** of A, with the possibility that $A = B$.
(3) If A and B are sets, $A \cap B$ is the **intersection** of A and B. It is the collection of elements which are common to A and B. As a set $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$. **NOTE:** The book uses AB for $A \cap B$, this is non-standard notation and won't be used in these notes.
(4) If A and B are sets, $A \cup B$ is the **union** of A and B. It is the set of elements which are either in A or B. As a set $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$.
(5) If $A \subset B$ then A^c is the **complement** to A (in B). It is the set of elements of B which are not members of A. As a set $A^c = \{x \in B \mid x \notin A\}$. **Note:** A^c depends on what set it is a subset of.
(6) $\emptyset$ is the **empty set**. It is the set which contains no elements.
(7) We say that A and B are **disjoint** if $A \cap B = \emptyset$. This means that A and B have no common elements.
(8) We say that a collection of sets $\{A_1, A_2, \dots,\}$ are **pairwise-disjoint** if $A_i \cap A_j = \emptyset$ if $i \neq j$.
(9) $\{A_1, A_2, \dots, A_n\}$ is a called a **partition** of B if they are pairwise-disjoint and $A_1 \cup A_2 \cup \cdots \cup A_n = B$.

3. Distributions and Their Properties

A **probability distribution** or **distribution** on a set Ω is a function P on the subsets of Ω which satisfies:

(1) $P(B) \geq 0$ for any $B \subseteq \Omega$
(2) $P(B_1 \cup B_2 \cup \cdots \cup B_n) = P(B_1) + P(B_2) + \cdots + P(B_n)$ if $B_1, B_2, \dots, B_n$ are pairwise-disjoint subsets of Ω.

(3) $P(\Omega) = 1$

These are known as the Kolmogorov axioms. Note that it is possible to define (2.) using an infinite sequence of pairwise-disjoint sets B_i when Ω is infinite. Ω is called the **outcome space** or **outcome set** and the subsets of Ω are called **events**.

Theorem 3.1. *If P is a probability distribution on Ω then the following are true:*

(1) $P(A^c) = 1 - P(A)$ *(Complement Rule)*

(2) $P(B \cap A^c) = P(B) - P(A)$ *if* $A \subseteq B$ *(Difference Rule)*

(3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *(Inclusion/Exclusion Rule)*

Proof: For (1.), note that $\Omega = A \cup A^c$ and $A \cap A^c = \emptyset$ so, using parts (2.) and (3.) of the definition,

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

and the result follows. For (2.), note that $B = (B \cap A^c) \cup A$ and that $(B \cap A^c) \cap A = \emptyset$. So we have:

$$P(B) = P[(B \cap A^c) \cup A] = P(B \cap A^c) + P(A)$$

The result follows by subtracting $P(A)$ from both sides. For (3.), note that $A \cup B = (A \cap B^c) \cup (B \cap A^c) \cup (A \cap B)$ and that these sets are pairwise-disjoint. So

$$P(A \cup B) = P(A \cap B^c) + P(B \cap A^c) + P(A \cap B)$$

By a similar argument

$$P(A) = P(A \cap B^c) + P(A \cap B)$$

$$P(B) = P(B \cap A^c) + P(B \cap A)$$

Combining these three equations gives us our result.

4. Examples

Example 1: Suppose that we roll two 6-sided dice and sum their values. This situation previously was viewed as a case of equally likely outcomes, with the outcome space being the set of ordered pairs of the numbers 1 to 6. Another way to view this is to let $\Omega = \{2, 3, \ldots, 12\}$ (these are the possible sums) and to define $P(\{2\}) = P(\{12\}) = \frac{1}{36}$, $P(\{3\}) = P(\{11\}) = \frac{2}{36}$, $P(\{4\}) = P(\{10\}) = \frac{3}{36}$, $P(\{5\}) = P(\{9\}) = \frac{4}{36}$, $P(\{6\}) = P(\{8\}) = \frac{5}{36}$ and $P(\{7\}) = \frac{6}{36}$. The probabilities of for larger subsets of Ω are defined by addition as in axiom (2.). This is a probability distribution space which is not an equally likely outcome space.

Example 2: Suppose that in a probability distribution space, $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{6}$ and $P(A \cap B) = \frac{1}{12}$, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{4} + \frac{1}{6} - \frac{1}{12} = \frac{1}{3}$$

Example 3: Suppose a box contains 8 cubes, three of which are red, five of which are black. Furthermore two of the red cubes and two of the black cubes are labeled with a 1, the remaining cubes are labeled with a two and a cube is chosen at random. We can construct the outcome space as:

$$\Omega = \{(\text{black}, 1), (\text{black}, 2), (\text{red}, 1), (\text{red}, 2)\}$$

Since there are two black cubes labeled 1, $P(\{(\text{black},1)\}) = \frac{2}{8}$. Similarly, $P(\{(\text{black},2)\}) = \frac{3}{8}$, $P(\{(\text{red},1)\}) = \frac{2}{8}$, $P(\{(\text{red},2)\}) = \frac{1}{8}$. What is the probability of choosing a cube which is either black or is labeled 1? The event of choosing a black cube is given by $A = \{(\text{black},1),(\text{black},2)\}$, the event of choosing a cube labeled 1 is given by $B = \{(\text{black},1),(\text{red},1)\}$ and we want to find $P(A \cup B)$. Note that $A \cap B = \{(\text{black},1)\}$, so

$$P(A \cap B) = \frac{2}{8}$$

Note that

$$P(A) = P(\{(\text{black},1)\}) + P(\{(\text{black},2)\}) = \frac{5}{8}$$

Since the latter events are disjoint. Similarly:

$$P(B) = P(\{(\text{black},1)\}) + P(\{(\text{red},1)\}) = \frac{4}{8}$$

Combining these gives:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{5}{8} + \frac{4}{8} - \frac{2}{8} = \frac{7}{8}$$

5. Special Distributions

In this section we will give a few example of distributions. We will develop more distributions as we go further in the course.

5.1. **Bernoulli** (p) **Distribution.** This is a distribution on the set $\Omega = \{0,1\}$. It is given by $P(\{1\}) = p$ and $P(\{0\}) = 1 - p$. Note that when $p = \frac{1}{2}$, this is just the equally likely outcome case of flipping a coin.

5.2. **Uniform Distribution on a Finite Set.** This is just the equally likely outcome on a set of n points. If $\Omega = \{1,2,\ldots,n\}$ and $A_k = \{k\}$ then A_k is the event that a k occurs in one trial. We then have:

$$P(A_k) = \frac{1}{n}$$

5.3. **Uniform** (a,b) **Distribution.** This is the first infinite case of a distribution. Here Ω is the interval (a,b). Let $a < x < y < b$ and let $A = (x,y)$. Then A corresponds to the event that a randomly chosen value lies in the interval (x,y). Then the probability of A occurring is given by:

$$P(A) = \frac{y-x}{b-a}$$

5.4. **Uniform Area Distribution.** Here Ω is some region in the x,y-plane. Let A be some subregion of Ω. Then A corresponds to the event that a randomly chosen point in Ω actually lies in A. Then, analogously to the uniform (a,b) distribution, we have:

$$P(A) = \frac{\text{Area of } A}{\text{Area of } \Omega}$$

5.5. **Example.** Suppose that Ω is the unit circle. If a point is chosen randomly in this region, what is the probability that it lies in the triangle bounded by the lines $x = 0$, $y = 0$ and $x + y = 1$? This is a case of uniform area distribution. Let A be the region corresponding to this triangle, so A is our event. The area of Ω is π and the area of A is $\frac{1}{2}$, so

$$P(A) = \frac{\text{Area of } A}{\text{Area of } \Omega} = \frac{1}{2\pi}$$

6. Empirical Distributions

An empirical distribution is a way of associating what is essentially a uniform (a, b) distribution to a finite set of data points. Suppose that our data points are $\{x_1, x_2, \ldots, x_n\}$. The we set $\Omega = (-\infty, \infty)$. Let $A_{(a,b)}$ be the event that a randomly chosen data point x_i lies in (a, b), that is $a < x_i < b$. The probability in this case is given by:

$$P(A_{(a,b)}) = \frac{\sharp\{k \mid a < x_k < b\}}{n}$$