BIM488 Introduction to Pattern Recognition

Review of Probability

Outline

- Sets and Set Operations
- Relative Frequency and Probability

Sets and Set Operations

Probability events are modeled as sets, so it is customary to begin a study of probability by defining sets and some simple operations among sets.

A ***set*** is a collection of objects, with each object in a set often referred to as an ***element*** or ***member*** of the set. Familiar examples include the set of all image processing books in the world, the set of prime numbers, and the set of planets circling the sun. Typically, sets are represented by uppercase letters, such as A, B, and C, and members of sets by lowercase letters, such as a, b, and c.

Sets and Set Operations (con't)

We denote the fact that an ***element a belongs*** to set *A* by

$$\boxed{a \in A}$$

If *a* is not an element of *A*, then we write

$$\boxed{a \notin A}$$

A set can be specified by listing all of its elements, or by listing properties common to all elements. For example, suppose that *I* is the set of all integers. A set *B* consisting the first five nonzero integers is specified using the notation

$$\boxed{B = \{1,2,3,4,5\}}$$

Sets and Set Operations (con't)

The set of all integers less than 10 is specified using the notation

$$\boxed{C = \{c \in I \mid c < 10\}}$$

which we read as "*C* is the set of integers such that each members of the set is less than 10." The "such that" condition is denoted by the symbol " | " . As shown in the previous two equations, the elements of the set are enclosed by curly brackets.

The set with no elements is called the ***empty*** or ***null set***, denoted in this review by the symbol Ø.

Sets and Set Operations (con't)

Two sets *A* and *B* are said to be ***equal*** if and only if they contain the same elements. Set equality is denoted by

$$\boxed{A = B}$$

If the elements of two sets are not the same, we say that the sets are ***not equal***, and denote this by

$$\boxed{A \neq B}$$

If every element of *B* is also an element of *A*, we say that *B* is a ***subset*** of *A*:

$$\boxed{B \subseteq A}$$

Sets and Set Operations (con't)

Finally, we consider the concept of a ***universal set***, which we denote by U and define to be the set containing all elements of interest in a given situation. For example, in an experiment of tossing a coin, there are two possible (realistic) outcomes: heads or tails. If we denote heads by H and tails by T, the universal set in this case is $\{H,T\}$. Similarly, the universal set for the experiment of throwing a single die has six possible outcomes, which normally are denoted by the face value of the die, so in this case $U = \{1,2,3,4,5,6\}$. For obvious reasons, the universal set is frequently called the ***sample space***, which we denote by S. It then follows that, for any set A, we assume that $\emptyset \subseteq A \subseteq S$, and for any element a, $a \in S$ and $a \notin \emptyset$.

Some Basic Set Operations

The operations on sets associated with basic probability theory are straightforward. The ***union*** of two sets *A* and *B*, denoted by

$$\boxed{A \cup B}$$

is the set of elements that are either in *A* or in *B*, or in both. In other words,

$$\boxed{A \cup B = \{z \mid z \in A \text{ or } z \in B\}}$$

Similarly, the ***intersection*** of sets *A* and *B*, denoted by

$$\boxed{A \cap B}$$

is the set of elements common to both *A* and *B*; that is,

$$\boxed{A \cap B = \{z \mid z \in A \text{ and } z \in B\}}$$

Set Operations (con't)

Two sets having no elements in common are said to be ***disjoint*** or ***mutually exclusive***, in which case

$$\boxed{A \cap B = \emptyset}$$

The ***complement*** of set *A* is defined as

$$\boxed{A^c = \{z \mid z \notin A\}}$$

Clearly, $(A^c)^c = A$. Sometimes the complement of *A* is denoted as $\overline{A}$.

The ***difference*** of two sets *A* and *B*, denoted $A - B$, is the set of elements that belong to *A*, but not to *B*. In other words,

$$\boxed{A - B = \{z \mid z \in A,\ z \notin B\}}$$

Set Operations (con't)

It is easily verified that $(A - B) = A \cap B^c$.

The union operation is applicable to multiple sets. For example the union of sets $A_1, A_2, \ldots, A_n$ is the set of points that belong to at least one of these sets. Similar comments apply to the intersection of multiple sets.

The following table summarizes several important relationships between sets. Proofs for these relationships are found in most books dealing with elementary set theory.

Set Operations (con't)

Some Important Set Relationships
$S^c = \emptyset;\ \ \emptyset^c = S;$
$A \cup A^c = S;\ \ A \cap A^c = \emptyset$
$A \cup \emptyset = A;\ \ A \cap \emptyset = \emptyset;\ \ S \cup \emptyset = S;\ \ S \cap \emptyset = \emptyset$
$A \cup A = A;\ \ A \cap A = A;\ \ A \cup S = S;\ \ A \cap S = A$
$A \cup B = B \cup A;\ \ A \cap B = B \cap A$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$
$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$

Set Operations (con't)

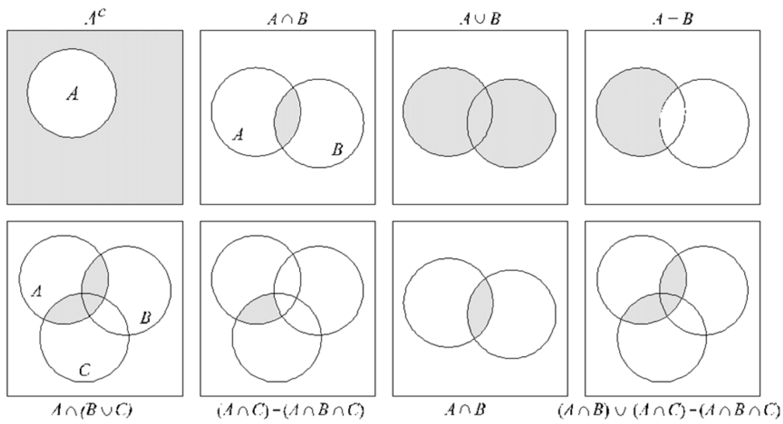
It often is quite useful to represent sets and sets operations in a so-called ***Venn diagram***, in which *S* is represented as a rectangle, sets are represented as areas (typically circles), and points are associated with elements. The following example shows various uses of Venn diagrams.

Example: The following figure shows various examples of Venn diagrams. The shaded areas are the result (sets of points) of the operations indicated in the figure. The diagrams in the top row are self explanatory. The diagrams in the bottom row are used to prove the validity of the expression

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

which is used in the proof of some probability relationships.

Set Operations (con't)



Relative Frequency & Probability

A ***random experiment*** is an experiment in which it is not possible to predict the outcome. Perhaps the best known random experiment is the tossing of a coin. Assuming that the coin is not biased, we are used to the concept that, on average, half the tosses will produce heads (H) and the others will produce tails (T). This is intuitive and we do not question it. In fact, few of us have taken the time to verify that this is true. If we did, we would make use of the concept of relative frequency. Let n denote the total number of tosses, n_H the number of heads that turn up, and n_T the number of tails. Clearly,

$$\boxed{n_H + n_T = n.}$$

Relative Frequency & Probability (con't)

Dividing both sides by n gives

$$\boxed{\frac{n_H}{n} + \frac{n_T}{n} = 1.}$$

The term n_H/n is called the ***relative frequency*** of the event we have denoted by H, and similarly for n_T/n. If we performed the tossing experiment a large number of times, we would find that each of these relative frequencies tends toward a stable, limiting value. We call this value the ***probability of the event***, and denoted it by P(event).

Relative Frequency & Probability (con't)

In the current discussion the probabilities of interest are $P(H)$ and $P(T)$. We know in this case that $P(H) = P(T) = 1/2$. Note that the event of an experiment need not signify a single outcome. For example, in the tossing experiment we could let D denote the event "heads or tails," (note that the event is now a set) and the event E, "neither heads nor tails." Then, $P(D) = 1$ and $P(E) = 0$.

The first important property of P is that, for an event A,

$$\boxed{0 \le P(A) \le 1.}$$

That is, the probability of an event is a positive number bounded by 0 and 1. For the certain event, S,

$$\boxed{P(S) = 1.}$$

Relative Frequency & Probability (con't)

Here the certain event means that the outcome is from the universal or sample set, S. Similarly, we have that for the impossible event, S^c

$$\boxed{P(S^c) = 0.}$$

This is the probability of an event being outside the sample set. In the example given at the end of the previous paragraph, $S = D$ and $S^c = E$.

Relative Frequency & Probability (con't)

The ***event*** that either events A ***or*** B ***or both*** have occurred is simply the union of A and B (recall that events can be sets). Earlier, we denoted the union of two sets by $A \cup B$. One often finds the equivalent notation $A+B$ used interchangeably in discussions on probability. Similarly, the event that ***both*** A ***and*** B occurred is given by the intersection of A and B, which we denoted earlier by $A \cap B$. The equivalent notation AB is used much more frequently to denote the occurrence of both events in an experiment.

Relative Frequency & Probability (con't)

Suppose that we conduct our experiment n times. Let n_1 be the number of times that only event A occurs; n_2 the number of times that B occurs; n_3 the number of times that AB occurs; and n_4 the number of times that neither A nor B occur. Clearly, $n_1+n_2+n_3+n_4=n$. Using these numbers we obtain the following relative frequencies:

$$\frac{n_A}{n} = \frac{n_1 + n_3}{n}$$

$$\frac{n_B}{n} = \frac{n_2 + n_3}{n}$$

$$\frac{n_{AB}}{n} = \frac{n_3}{n}$$

Relative Frequency & Probability (con't)

and

$$\frac{n_{A\cup B}}{n} = \frac{n_1 + n_2 + n3}{n} = \frac{(n_1 + n_3) + (n_2 + n_3) - n_3}{n} = \frac{n_A}{n} + \frac{n_B}{n} - \frac{n_{AB}}{n}.$$

Using the previous definition of probability based on relative frequencies we have the important result

$$\boxed{P(A \cup B) = P(A) + P(B) - P(AB).}$$

If A and B are ***mutually exclusive*** it follows that the set AB is empty and, consequently, $P(AB) = 0$.

Relative Frequency & Probability (con't)

The relative frequency of event *A* occurring, ***given that*** event *B* has occurred, is given by

$$\frac{n_{A/B}}{n} = \frac{\frac{n_{AB}}{n}}{\frac{n_B}{n}} = \frac{n_3}{n_2 + n_3}.$$

This ***conditional probability*** is denoted by $P(A/B)$, where we note the use of the symbol " / " to denote conditional occurrence. It is common terminology to refer to $P(A/B)$ as the ***probability of*** *A* ***given*** *B*.

Relative Frequency & Probability (con't)

Similarly, the relative frequency of *B* occurring, given that *A* has occurred is

$$\frac{n_{B/A}}{n} = \frac{\frac{n_{AB}}{n}}{\frac{n_A}{n}} = \frac{n_3}{n_1 + n_3}.$$

We call this relative frequency ***the probability of*** *B* ***given*** *A*, and denote it by $P(B/A)$.

Relative Frequency & Probability (con't)

A little manipulation of the preceding results yields the following important relationships

$$\boxed{P(A/B) = \frac{P(B/A)P(A)}{P(B)}}$$

and

$$\boxed{P(AB) = P(A)P(B/A) = P(B)P(A/B).}$$

The second expression may be written as

$$\boxed{P(B/A) = \frac{P(A/B)P(B)}{P(A)}}$$

which is known as ***Bayes' theorem***, so named after the 18th century mathematician Thomas Bayes.

Relative Frequency & Probability (con't)

If A and B are ***statistically independent***, then $P(B/A) = P(B)$ and it follows that

$$\boxed{\begin{aligned} P(A/B) &= P(A) \\ P(B/A) &= P(B) \end{aligned}}$$

and

$$\boxed{P(AB) = P(A)P(B).}$$

It was stated earlier that if sets (events) A and B are ***mutually exclusive***, then $A \cap B = \varnothing$ from which it follows that $P(AB) = P(A \cap B) = 0$. As was just shown, the two sets are statistically independent if $P(AB)=P(A)P(B)$, which we assume to be nonzero in general. ***Thus, we conclude that for two events to be statistically independent, they cannot be mutually exclusive***.

Relative Frequency & Probability (con't)

In general, for *N* events to be statistically independent, it must be true that, for all combinations $1 \le i \le j \le k \le \ldots \le N$

$$
\boxed{
\begin{aligned}
P(A_iA_j) &= P(A_i)P(A_j) \\
P(A_iA_jA_k) &= P(A_i)P(A_j)P(A_k) \\
&\vdots \\
P(A_1A_2\cdots A_N) &= P(A_1)P(A_2)\cdots P(A_N).
\end{aligned}
}
$$

Relative Frequency & Probability (con't)

Example: (a) An experiment consists of throwing a single die twice. The probability of any of the six faces, 1 through 6, coming up in either experiment is 1/6. Suppose that we want to find the probability that a 2 comes up, followed by a 4. These two events are statistically independent (the second event does not depend on the outcome of the first). Thus, letting *A* represent a 2 and *B* a 4,

$$P(AB) = P(A)P(B) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

We would have arrived at the same result by defining "2 followed by 4" to be a single event, say *C*. The sample set of all possible outcomes of two throws of a die is 36. Then, $P(C)$=1/36.

Relative Frequency & Probability (con't)

Example (Con't): (b) Consider now an experiment in which we draw one card from a standard card deck of 52 cards. Let *A* denote the event that a king is drawn, *B* denote the event that a queen or jack is drawn, and *C* the event that a diamond-face card is drawn. A brief review of the previous discussion on relative frequencies would show that

$$P(A) = \frac{4}{52},$$

$$P(B) = \frac{8}{52},$$

and

$$P(C) = \frac{13}{52}.$$

Relative Frequency & Probability (con't)

Example (Con't): Furthermore,

$$P(AC) = P(A \cap C) = P(A)P(C) = \frac{1}{52}$$

and

$$P(BC) = P(B \cap C) = P(B)P(C) = \frac{2}{52}.$$

Events *A* and *B* are mutually exclusive (we are drawing only one card, so it would be impossible to draw a king and a queen or jack simultaneously). Thus, it follows from the preceding discussion that $P(AB) = P(A \cap B) = 0$ [and also that $P(AB) \neq P(A)P(B)$].

Relative Frequency & Probability (con't)

Example (Con't): (c) As a final experiment, consider the deck of 52 cards again, and let A_1, A_2, A_3, and A_4 represent the events of drawing an ace in each of four successive draws. If we replace the card drawn before drawing the next card, then the events are statistically independent and it follows that

$$\begin{aligned} P(A_1A_2A_3A_4) &= P(A_1)P(A_2)P(A_3)P(A_4) \\ &= \left[\frac{4}{52}\right]^4 \approx 3.5 \times 10^{-5}. \end{aligned}$$

Relative Frequency & Probability (con't)

Example (Con't): Suppose now that we do not replace the cards that are drawn. The events then are no longer statistically independent. With reference to the results in the previous example, we write

$$\begin{aligned} P(A_1A_2A_3A_4) &= P(A_1)P(A_2A_3A_4/A_1) \\ &= P(A_1)P(A_2/A_1)P(A_3A_4/A_1A_2) \\ &= P(A_1)P(A_2/A_1)P(A_3/A_1A_2)P(A_4/A_1A_2A_3) \\ &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49} \approx 3.7 \times 10^{-6}. \end{aligned}$$

Thus we see that not replacing the drawn card reduced our chances of drawing fours successive aces by a factor of close to 10. This significant difference is perhaps larger than might be expected from intuition.

Summary

- Sets and Set Operations
- Relative Frequency and Probability

References

- R. C. Gonzalez & R. E. Woods, *Digital Image Processing (3rd Edition),* Prentice Hall, 2008.