LECTURE NOTE 1 *
SET AND PROBABILITY THEORY

MIT 14.30 SPRING 2006

HERMAN BENNETT

1 Set Theory

1.1 Definitions and Theorems

1. Experiment: any action or process whose outcome is subject to uncertainty.

2. Sample Space: collection of all possible outcomes (or elements) of the experiment (set S). [Finite vs. Infinite; Discrete vs. Continuous]

*Caution: These notes are not necessarily self-explanatory notes. They are to be used as a complement to (and not as a substitute for) the lectures.

3. Event: collection of elements (subset A, B, etc.) contained in the sample space (S).

4. $s \in S$: The outcome s belongs to the sample space S. The contrary is defined by the symbol $\notin$.

5. $\emptyset = \{\}$: Denotes the empty set (the set of no elements). It also defines the set of elements of an impossible event; *e.g.*: the event 'generating a negative number' when rolling a die.

6. Union: The union of event A and event B, denoted $A \cup B$, is the collection (or set) of elements that belong to either A or B or both. $[A \cup B = \{x : x \in A \text{ or } x \in B\}]$

   Properties: $A \cup A = A$; $A \cup S = S$; $A \cup \emptyset = A$.

7. Intersection: The intersection of event A and event B, denoted $A \cap B$, is the collection (or set) of elements that belong to A and B. $[A \cap B = \{x : x \in A \text{ and } x \in B\}]$

   Properties: $A \cap A = A$; $A \cap S = A$; $A \cap \emptyset = \emptyset$.

8. Complement: The complement of event A, denoted A^c (or equivalently A'), is the set of all elements that are not in A. $[A^c = \{x : x \notin A\}]$

   Properties: $(A^c)^c = A$; $\emptyset^c = S$; $S^c = \emptyset$; $A^c \cup A = S$; $A^c \cap A = \emptyset$.

9. $A \subset B$: The event A is contained in event B, $A \subset B$, if every element of A also belongs to B.

   Properties: – If $A \subset B$ and $B \subset A \Rightarrow A = B$;
   – If $A \subset B$ and $B \subset C \Rightarrow A \subset C$;
   – $\emptyset \subset A$, for any event A.

10. Disjoint: Event A and event B are disjoint, or mutually exclusive, if A and B have no outcome in common. [$A \cap B = \emptyset \Leftrightarrow A$ and B are disjoint events]

11. Exhaustive: Event A and event B are exhaustive if their union is S. [$A \cup B = S \Leftrightarrow$ A and B are exhaustive events]

12. Finally, some additional results (HOMEWORK: think of them in terms of Venn diagrams):

    – Commutativity: $A \cup B = B \cup A$; $A \cap B = B \cap A$.
    – Associativity: $A \cup (B \cup C) = (A \cup B) \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C$.
    – Distributive Laws: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
    – DeMorgan's Laws: $(A \cup B)^c = A^c \cap B^c$; $(A \cap B)^c = A^c \cup B^c$.

2 Probability Theory

2.1 Definition of Probability

How likely is it for event A to occur? This concept is represented by the *probability that event A will take place*, which is denoted by $P(A)$ and can take any value from 0 to 1.

The mathematical definition of probability function is based on 3 axioms, which are based on our intuitive notion of probability. $[P() : \{\text{set of all possible events}\} \rightarrow [0,1]]$

– Axiom 1: For any event A, P(A)$\geq$ 0 (nonnegative).
– Axiom 2: P(S)=1.
– Axiom 3: For any sequence of disjoint sets $A_1, A_2, ..., A_n$, $P(A_1 \cup A_2 \cup ... A_n) = \sum_{i=1}^{n} P(A_i)$ where n is the total number of disjoint sets in the sequence.

Properties (for events A and B):
– $P(A) = 1 - P(A^c)$; $P(A \cup B) = P(A) + P(B) - P(A \cap B)$; $P(\emptyset) = 0$;
– If A and B are disjoint $\Rightarrow P(A \cap B) = 0$;
– If $A \subset B \Rightarrow P(A) \leq P(B)$.

Example 2.1. Show that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

2.2 Counting Techniques to Compute $P(A)$ when each Possible Outcome is Equally Likely – (same Probability)

$$P(A) = \frac{N(A)}{N} \tag{1}$$

Where N is the number of outcomes contained in S and $N(A)$ is the number of outcomes contained in event A.

When the sample space is small and the outcomes are equally likely (same probability), just count. For example, rolling a die: $N = 6$, $P(3) = \frac{1}{6}$. If the case you are dealing with is not so simple, you can use the following techniques to count.

1. General Product Rule: If a process has multiple stages (call the number of stages k) and if stage i can be completed in n_i ways, regardless of which outcomes occur in earlier stages, then the process itself can be completed in $\mathbf{n_1 n_2 ... n_k}$ ways. Note that the choices are not necessarily the same in each stage (although they could be).

   Example 2.2. Assume a box that contains 7 balls of different colors. How many ways are there to take 3 balls from the box, if each ball taken is immediately returned to the box?

2. Permutations: Suppose the outcome is constructed by selecting k objects from a total of n and *without replacement.* The total number of permutations (which means that *order matters*) is $\mathbf{n(n-1)...(n-k+1)}$. General formula: $\mathbf{P_{k,n} = \frac{n!}{(n-k)!}}$ [Following Example 2.2: $P_{3,7} = \frac{7!}{(7-3)!} = 7 \cdot 6 \cdot 5$]

Example 2.3. How many ways are there to rank 4 different dogs? How many ways are there to rank 4 different dogs out of a total of 10 dogs?

3. Combinations: Now assume the outcome is constructed in the same way as before: selecting k objects from a total of n and *without replacement.* The total number of combinations (which means that *order does not matter*) is: $\mathbf{C_{k,n}} = \binom{\mathbf{n}}{\mathbf{k}} = \frac{\mathbf{n!}}{\mathbf{k!(n-k)!}}$. The symbol $\binom{n}{k}$ is read "n choose k"; is the number of ways a group of k objects can be selected from a collection of n objects. [Following Example 2.2: $C_{3,7} = \frac{7!}{3! \cdot (7-3)!}$]

Example 2.4. How many possible combinations of 3 books are there in a set of 5 books? How many possible combinations of 5 books are there in a set of 5 books? (Note the difference from permutation.)

Wrap-up: When simple counting is not practical, we use techniques 1-3 to compute N, the number of outcomes contained in the sample space, and to compute N(A), the number of outcomes contained in event A. With this information we can compute $P(A)$.

Example 2.5. A deck of 52 cards has 4 aces. Assume you give 13 cards each to 4 players. What is the probability that each player gets exactly 1 ace?

Example 2.6. A fair coin is tossed 7 times. What is the probability of obtaining 3 heads? What is the probability of obtaining at most 3 heads?

2.3 Conditional Probability

We use probabilities because we are uncertain about the exact outcome of an experiment. However, this does not mean that we are completely ignorant about the process. The belief about the likelihood of an event, $P(A)$, is based on the information at hand when the assignment of probability is made. New information can be available, which could make us modify our belief (probability). Conditional Probability, $P(A|B)$, is the name given to the new belief after receiving the new information, in this case that event B occurred. [1]

$$\text{Definition: } P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \text{for } P(B) > 0 \tag{2}$$

Note that:

- $P(A|B)P(B) = P(B|A)P(A)$.
- If events $A_1, A_2, ...A_k$ are disjoint and exhaustive, then:

$$P(A_1|B) + P(A_2|B)...+ = 1 \text{ and}$$

$$\sum_{i=1}^{k} P(B|A_i)P(A_i) = P(B) \text{ (Law of Total Probability).}$$

Bayes Theorem. Let the events $A_1, A_2, ...A_k$ be disjoint and exhaustive events in the sample space S, such that $P(A_i) > 0$, and let B be an event such that $P(B) > 0$. Then,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)} \quad \left(= \frac{P(A_i \cap B)}{P(B)} \quad \frac{\text{cond. prob.}}{\text{law of total prob.}} \right) \tag{3}$$

This way of updating the probability of event A is usually call *Bayesian updating*.

[1] $P(A|B)$ and $P(A)$ are also called *posterior* and *prior*, respectively.

Example 2.7. There is a new music device in the market that plays a new digital format called MP∞. Since it's new, it's not 100% reliable. You know that 20% of the new devices don't work at all, 30% last for only 1 year, and the rest last for 5 years. If you buy one and it works fine, what is the probability that it will last for 5 years?

2.4 Independence

Two events A and B are said to be independent if $P(A|B) = P(A)$; otherwise they are dependent.

• For example, tossing a fair coin twice. The probability of getting H or T on the second toss does not depend on whether you got H or T in the first. Another way to see this: the result of the first toss does not provide any additional information about the result of the second one: $P(A|B) = P(A)$.

• If A and B are independent, then $P(A \cap B) = P(A)P(B)$ (by definition of conditional probability).

• If A and B are independent, then A and B^c are also independent $[P(A \cap B^c) = P(A)P(B^c)]$.

• General definition of independence between 2 or more events: Events $A_1, A_2, ..., A_n$ are mutually independent if, for all possible subcollections of $k \leq n$ events: $P(A_i \cap A_j \cap ... A_k) = P(A_i)P(A_j)...P(A_k)$.

Example 2.8. Events associated with the experiment of rolling a die: $A = \{2, 4, 6\}$ $B = \{1, 2, 3, 4\}$ $C = \{1, 2, 4\}$. Are events A and B independent? What about A and C?