MAT 105 - Summary of Lecture 7
Tuesday, January 26, 2015

The theory of probability as a mathematical discipline can and should be developed from axioms in exactly the same way as geometry and algebra. (Andrey Kolmogorov)

Probability – definition

There are three views of probability.

- **axiomatic approach:** each event E is associated to a probability $P(E)$ which measures the chance of this event occurring. In the case of a finite state space, and equally likely outcomes, this probability is given by
$$P(E) = \frac{\text{\# outcomes in } E}{\text{total \# of outcomes}}.$$
- **frequency interpretation:** if we repeat an experiment a large number of times, then the fraction of times the event E occurs will be close to $P(E)$. This is supported by the Law of Large Numbers, as we will see in a later lecture.
- **measure of belief:** we associate a likelihood of occurrence to an event using our intuition and understanding of the world. This is not a good idea, since our intuition is sometimes wrong.

For example, if we want to find the probability of a die landing on 6, we would reason as follows:

- **axiomatic approach:** since each outcome is equally likely, we assign probability $1/6$ to landing on 6.
- **frequency interpretation:** we roll the die many times and observe that the die lands on 6 about $1/6$ of the time, hence the probability of a 6 is $1/6$.
- **measure of belief:** each side looks the same, so intuitively, a die lands on 6 with probability $1/6$.

Roughly, a probability is a measure of the occurrence of an event. Formally, we will opt for the axiomatic approach and define probability as follows:

Definition: A probability P is a function from the set of outcomes into the closed interval $[0, 1]$, $P : S \to [0, 1]$ satisfying the axioms:

(i) $P(S) = 1$ (since the chance *something* happens is 1)

(ii) $P(\emptyset) = 0$ (since the chance *nothing* happens is 0)

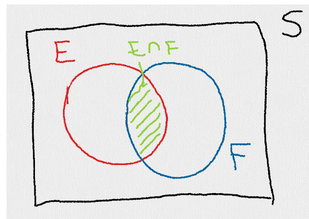
(iii) $0 \leq P(E) \leq 1$ for any event E.

(iv) If E^c denotes the complement of E, then $P(E^c) = 1 - P(E)$.

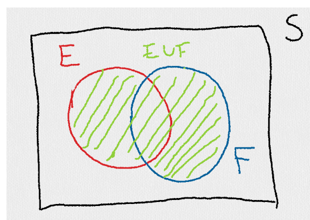
(v) If E and F are *disjoint* (meaning $E \cap F = \emptyset$), then $P(E \cup F) = P(E) + P(F)$.

Notes:

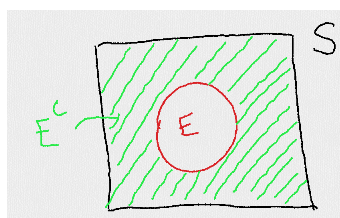
- $\emptyset$ refers to the emptyset (a set with no elements)
- $E \cap F$ denotes the *intersection* of E and F (the overlap of E and F) and includes elements that are in E AND F



- $E \cup F$ denotes the *union* of E and F and includes elements that are in E OR F (all outcomes that are in E, or F, or both)



- The complement E^c contains all outcomes that are in S but NOT in E.



It might be easy to look at a diagram and interpret probability as the "weight" of an event, with the understanding that the "weight" of the entire sample space S is 1. Using this approach, it is not difficult to see that when $E \cap F \neq \emptyset$ (E and F are not disjoint),

$$P(E \cup F) = P(E) + P(F) - P(E \cap F). \tag{1}$$

This can be proved by noting that when we add the weights of E and F, we add the weight of the overlap $E \cap F$ twice, so we need to subtract one copy of it.

It might be useful to remind ourselves of the following rules for sets:

(i.) $A \cap A = A$ and $A \cup A = A$.

(ii.) $A \cap \emptyset = \emptyset$ and $A \cup \emptyset = A$.

(iii.) If all elements of A are in B, we say A is a subset of B and write $A \subseteq B$. In this case,

$$A \cap B = A \quad \text{and} \quad A \cup B = B.$$

Examples:

1. This example reviews working with sets. Let $A = \{1, 2, 7\}$ and $B = \{1, 2, 15, 29\}$. Then
$$A \cap B = \{1, 2\} \qquad A \cup B = \{1, 2, 7, 15, 29\}.$$
Furthermore, if $C = \{6\}$, then $A \cap C = \emptyset$.

2. Roll two dice (red and blue). Let A = red die lands on 6, B = blue die lands on 6, E = at least one die lands on 6. Then there are 36 possible outcomes for this experiment and
$$P(A) = \frac{6}{36} = \frac{1}{6}, \quad P(B) = \frac{6}{36} = \frac{1}{6},$$
$P(A \cap B) = \frac{1}{36}$ (since both dice land on 6) and $P(A \cup B) = \frac{11}{36}$ by looking at all outcomes when the red die, the blue die or both land on 6. Note that $E = A \cup B$, so then we found $P(E) = P(A \cup B) = \frac{11}{36}$ and using the 4th axiom, we get
$$P(E^c) = 1 - P(E) = 1 - \frac{11}{36} = \frac{25}{36}$$
which makes sense because E^c is the event where no die lands on 6, which occurs in 25 outcomes. Let us also verify equation (1):
$$P(A \cup B) = \frac{11}{36}, \quad P(A) + P(B) - P(A \cap B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36} \;\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

3. Flip 3 coins. Let A = first two coins land on Heads, B = last two coins land on Heads. Then the sample space is
$$S = \{HHH, THH, HTH, HHT, HTT, THT, TTH, TTT\}.$$
We find the following events
$$A = \{HHH, HHT\}, \quad B = \{HHH, THH\}, \quad A \cup B = \{HHH, HHT, THH\}, \quad A \cap B = \{HHH\}$$
Therefore, $P(A) = 1/4$, $P(B) = 1/4$, $P(A \cup B) = 3/8$ and $P(A \cap B) = 1/8$. Note that (1) holds true, since
$$P(A) + P(B) - P(A \cap B) = \frac{1}{4} + \frac{1}{4} - \frac{1}{8} = \frac{3}{8} = P(A \cup B).$$

4. Suppose we pick one card from a deck of 52.

   (a) Let A = it is a face card, B =it is a spade. Then $P(A) = 12/52$, $P(B) = 13/52$ and $P(A \cap B) = P(\text{face on spade card}) = 3/52$. Thus, $P(A \cup B) = (12 + 13 - 3)/52 = 22/52$.

   Note that this makes sense, since counting all cards that are spades or face cards (or both), we can first take the 13 spades, and to those add the 9 remaining face cards not yet picked, to get 22 cards out of 52.

   (b) Let A = it is a heart, B =it is a spade. Then $P(A) = 13/52$, $P(B) = 13/52$ and $P(A \cap B) = P(\text{heart and spade card}) = 0$. Thus, $P(A \cup B) = (13 + 13)/52 = 1/2$, which seems reasonable since the hearts and spades make up half the deck of cards.

   Remark: It is very important to remember to subtract the overlap, otherwise the probabilities will be overestimated! When the events are disjoint (or mutually exclusive), the probability of the overlap will be zero, as indicated in the 5th axiom for probability measures.

Random Variables

Sometimes, we might want to associate a real value to an outcome. For example, we place a wager on certain outcomes, or we want to count the number of successes in an experiment with many trials. In such an instance, we define a **random variable** $X : S \to \mathbb{R}$ to be a function that takes on values from the set of outcomes and maps into the reals. It is a *variable* because it takes on different values, and it is *random* because it depends on outcomes of random experiments.

When we list all possible values for X along with the corresponding probabilities, we describe the distribution of X. More precisely, for a (discrete) random variable X, the **probability distribution of** X is a listing of probabilities for all possible values of X.

Example:

(a) You flip a coin. If the coin lands on Heads, you win \$10 and if the coin lands on Tails, you lose \$1. Let X = the net amount of money won in this game. Then we define

$$X(H) = 10 \qquad\qquad X(T) = -1$$

We write $X \in \{-1, 10\}$ to mean that X can take on the values 10 or -1. We want to find probabilities for X taking on each value:

$$P(X = 10) = P(\text{Heads}) = 1/2, \qquad P(X = -1) = P(\text{Tails}) = 1/2.$$

These probabilities describe the distribution of X. A good question to ask is "how much money would I make on this game, on average"? We can say that, since half the time we make \$10, and half the time we lose \$1, we expect to make \$4.5 on average. We will formalize this concept next lecture.

(b) You roll 1D6. If the die lands on 6, you win \$10 and if it lands on a value other than 6, you lose \$1. Let X = the net amount of money won in this game. Then we define

$$X(\text{land on } 6) = 10 \qquad\qquad X(\text{not land on } 6) = -1$$

Again, we write $X \in \{-1, 10\}$ to mean that X can take on the values 10 or -1. We want to find probabilities for X taking on each value:

$$P(X = 10) = P(\text{lands on } 6) = 1/6, \qquad P(X = -1) = P(\text{not land on } 6) = 5/6.$$

Is it advantageous to play this game? Is the game "fair"? How would you measure the fairness of the game?