Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professors David Karger and Tom Leighton

Handout 49
November 14, 1995

Notes 19

1 Probability and Set Theory

Before discussing some more complicated probability problems, let's develop some useful connections between probabiliy and set theory. Recall from the last lecture that

Definition 1.1 *A* sample space *for a problem (or experiment) is a set consisting of all possible outcomes.*

Definition 1.2 *A* probability space *consists of a sample space S and a probability function $Pr : S \to \mathbb{R}$, such that*

1. *For all $s \in S$, $0 \leq Pr(S) \leq 1$.*
2. $\sum_{s \in S} Pr(s) = 1$.

Definition 1.3 *An* event *is a subset of the sample space.*

Definition 1.4 *The probability of an event S' is the sum of the probabilities of all outcomes in the event.*

Using these definitions, we can establish some basic facts.

Theorem 1.5 *Let A and B be events such that $A \subseteq B$. Then $Pr(A) \leq Pr(B)$.*

Proof.

$$Pr(A) = \sum_{s \in A} Pr(s) \leq \sum_{s \in B} Pr(s) = Pr(B).$$

∎

For an event A, let $\bar{A}$ denote the *complement* of A.

Theorem 1.6 *For any event A,*

$$Pr(\bar{A}) = 1 - Pr(A).$$

Proof.

$$\begin{aligned} Pr(\bar{A}) + Pr(A) &= \sum_{s \in A} Pr(s) + \sum_{s \notin A} Pr(s) \\ &= \sum_{s \in S} Pr(s) \\ &= 1 \end{aligned}$$

where the first equality follows from the definition of complementation and the second follows from the definition of a probability space. ∎

Also this has a simple proof, often it's much easier to calculate a probability by first calculating the probability of the complement and then subtracting from 1.

We can use the inclusion-exclusion principle to get probabilities of unions.

Theorem 1.7 *For any events A and B, $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.*

Proof.

$$\begin{aligned} Pr(A \cup B) &= \sum_{s \in A \cup B} Pr(s) \\ &= \sum_{s \in A} Pr(s) + \sum_{s \in B} Pr(s) - \sum_{s \in A \cap B} Pr(s) \\ &= Pr(A) + Pr(B) - Pr(A \cap B). \end{aligned}$$

∎

Sometimes it's hard to calculate the exact probabilities of unions and intersections. But the advantage of a result like Theorem 1.7 is that it allows us to give bounds on these probabilties.

Corollary 1.8 (Boole's Inequality) *For any events A and B,*

$$Pr(A \cup B) \leq Pr(A) + Pr(B).$$

Note that equality holds if and only if $Pr(A \cap B) = 0$. By induction, Boole's Inequality can be easily extended to n events.

Theorem 1.9 *For all $n \geq 1$ and any events $A_1, \ldots, A_n$, we have*

$$Pr(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} Pr(A_i).$$

Proof. We prove this by induction on n. For $n = 1$, we just have $Pr(A_1) \leq Pr(A_1)$. Now, assuming the theorem is true for $n-1$, we apply Boole's Inequality followed by the induction hypothesis:

$$\begin{aligned} Pr(\cup_{i=1}^{n} A_i) \leq & \leq Pr(\cup_{i=1}^{n-1} A_i) + Pr(A_n) \\ & \leq \sum_{i=1}^{n-1} Pr(A_i) + Pr(A_n) \\ & = \sum_{i=1}^{n} Pr(A_i). \end{aligned}$$

■

As an example, let's suppose you're wiring up a circuit. Overall, you are supposed to make n connections, and for each connection, the probability that it is made incorrectly is p. What is the probability that any of the connections is made incorrectly?

Well, if we let A_i denote the event that connection i is made incorrectly, we are just asking for $Pr(\cup_{i=1}^{n} A_i)$. Without knowing the relationship among these events, we can't give an exact answer — it could be that messing up one connection will cause you to mess up others as well.

However, we can still gives bounds on this probability. Using Theorem 1.9, we know

$$Pr(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} Pr(A_i) = np.$$

Also, since $A_1 \subseteq \cup_{i=1}^{n} A_i$, and $Pr(A_1) = p$, we can use Theorem 1.5 to conclude

$$Pr(\cup_{i=1}^{n} A_i) \geq Pr(A_1) = p.$$

So for example, if $n = 10$ and $p = .01$, we get the following bounds:

$$.01 \leq Pr(\cup_{i=1}^{10} A_i) \leq .10.$$

Now the event that the chip gets wired correctly is the just the complement of the event that a connection is made incorrectly; so by Theorem 1.6, the probability that the chip gets wired correctly in this example is between .9 and .99.

2 Conditional Probability

In a lot of problems (including the one we looked at in the last lecture), we want to consider the probability that an event A happens, *given* that we know a certain other event B has happened. We write

$$Pr(A \mid B)$$

to denote the probability of event A, given B.

There is a very simple formula for this, which can be viewed as follows. Once we know that event B has happened, we are in the subset of the sample space equal to B. We now want to know what fraction of this set consists of outcomes in A. Thus,

$$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)}.$$

We will take this formula as the definition of the *conditional probability of A given B.*

Let's do an example. Suppose a baseball team is playing a best-2-out-of-3 series. The probability it wins the first game is $\frac{1}{2}$. The probability of winning a game following a win in the previous game is $\frac{2}{3}$; the probability of winning following a loss is only $\frac{1}{3}$.

The question is: what is the probability the team wins the series if it wins the first game?

There are actually two types of conditioning going on here:

1. The probability of winning a game depends on the probability of having won the previous game.
2. We're asked for the probability of winning the series, *given* that the team wins the first game.

We solve this problem in the following step-by-step way.

1. Draw the sample space. (See the figure below.)
2. Assign probabilities to the outcomes. These are all calculated using the same principle; let's show we calculate the probability of the outcome (tree branch) WLL.

$$\begin{aligned} Pr(WLL) &= Pr(\text{Win 1st}) \cdot Pr(\text{Lose 2nd} \mid \text{Win 1st}) \cdot Pr(\text{Lose 3rd} \mid \text{Win 1st, Lose 2nd}) \\ &= \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} \\ &= \frac{1}{9} \end{aligned}$$

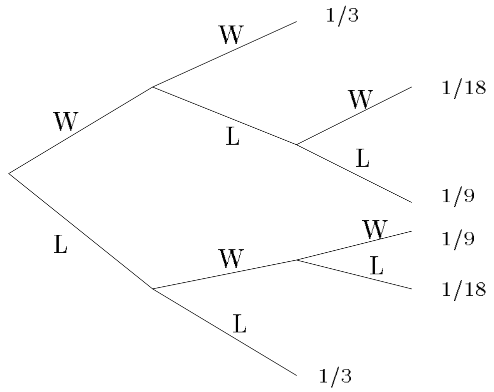

Figure 1: Why the Red Sox didn't make it to the World Series

3. Now we identify the events. A is the event that the team wins the series, and B is the event that the team wins the first game.

4. Finally, the answer we want is $Pr(A \mid B)$. We compute this using the formula, and adding up the probabilities from the figure:

$$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{\frac{1}{3} + \frac{1}{18}}{\frac{1}{3} + \frac{1}{9} + \frac{1}{18}} = \frac{\frac{7}{18}}{\frac{9}{18}} = \frac{7}{9}.$$

3 Another Example

Conditional probability can lead to some counter-intuitive answers; but things really do work out if you just follow the definition. Here's a problem that always confuses people. It's based on calculating the probability of an event A, given an event B that comes *after* it in time. In such cases, the conditional probability is often called the *a posteriori probability* — though it's really the same thing mathematically.

Suppose I have two coins — one is a fair coin, which comes up heads or tails with probability $\frac{1}{2}$ each, and the other is an unfair coin, which always comes up heads. Now I choose one of the two coins with probability $\frac{1}{2}$ each, and I flip this chosen coin once. If it comes up heads, what is the probability that I picked the fair coin?

The trickiness here is that our intuitive notion of conditional probability does not always match the definition as $Pr(A \cap B)/Pr(B)$ — in this case, the notion of "causality" seems to go backwards. But in such cases, we always have to follow the definition.

As in the previous example, we solve this problem in the following steps.

1. Draw the sample space. (See the figure below.)

2. Assign probabilities to the events, again as in the previous example.

3. Define the relevant events. Here, A is the event that we choose the fair coin, and B is the event that the result is heads.

4. $$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2}} = \frac{1}{3}.$$

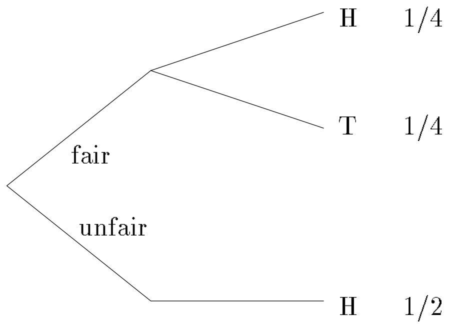

Figure 2: Flipping Coins

The key point here is that we knew the probability of picking the fair coin was exactly $\frac{1}{2}$. Without this, we could not have computed $Pr(A \mid B)$.