Tutorial 10

1 Probability and Cardinality

This week (and next week) we'll develop some relations between probability and set theory. We'll see that the methods and rules we developed for counting sets carry over to calculating probabilities.

Recall from lectures that

Definition 1 *A* sample space *for a problem (or experiment) is the set of all possible outcomes.*

Definition 2 *A* probability space *consists of a sample space S and a probability function $Pr : S \to \mathcal{R}$, such that*

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

Definition 3 *An* event *is a subset of the sample space.*

Thus an event is analogous to a set. Using these definitions, we can establish some basic facts.

Theorem 1 *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 2 *For any event A,*

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

Proof.

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

■

As we will soon see, it is sometimes easier to calculate a probability by calculating the probability of the complement.

Next, we derive an inclusion-exclusion principle to calculate the probability of a union of events. This is exactly analogous to the inclusion-exclusion principle for sets.

Theorem 3 *For any events A and B, $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.*

Proof.

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

■

Similarly,

Theorem 4 *For a set of n events, $A_1, A_2, \ldots, A_n$,*

$$Pr(\cup_{i=1}^{n} A_i) = \sum_{k=1}^{n} \sum_{S \subseteq \{1,2,\ldots,n\} : |S| = k} (-1)^{k+1} Pr(\cap_{i \in S} A_i)$$

Sometimes it is hard to calculate the exact probabilities of unions and intersections. But Theorem 3 allows us to get bounds on these probabilties.

Corollary 5 (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 $Pr(A \cap B) = 0$. Boole's Inequality can be extended to n events.

Theorem 6 *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. By induction on n.
I.H. : $P(n) : Pr(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} Pr(A_i)$
Base: $n = 1$, we just have $Pr(A_1) \leq Pr(A_1)$.
Ind. Step: Assume $P(n-1)$, to prove $P(n)$,

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

Here the first inequality follows from Boole's inequality and the second inequality from the induction hypothesis $P(n-1)$. ■

2 Wiring circuits

As an example, let us suppose you are wiring up a circuit. There are a total of n connections to be made. From past experience we assume that any particular connection is made *incorrectly* with probability p, for some $0 \leq p \leq 1$. That is,

$\forall i$, Pr(i^{th} connection is wrong) $= p$.

What is the probability that the circuit is wired correctly, i.e., there are no incorrect connections?

To solve this problem we set up a sample space and define an event for each connection. Let A_i denote the event that connection i is made correctly. Then $\bar{A}_i$ is the event that connection i is made incorrectly. Then Pr(all connections are OK) =

$$Pr(A_1 \cap A_2 \ldots \cap A_n)$$

This can be hard to evaluate exactly. Since we don't know the relationships 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 give bounds on this probability. By Theorem 2

$$Pr(A_1 \cap A_2 \ldots \cap A_n) = 1 - Pr(\overline{A_1 \cap A_2 \ldots A_n})$$

$$= 1 - Pr(\bar{A}_1 \cup \bar{A}_2 \cup \ldots \cup \bar{A}_n)$$

Using Theorem 6, we know

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

Hence, $Pr(A_1 \cap A_2 \ldots \cap A_n) \geq 1 - np$. So for example, if $n = 10$ and $p = .01$, we get the following bounds:

$$Pr(all \quad connections \quad OK) \geq .90$$

What about an upper bound?

$$Pr(all \quad OK) = Pr(A_1 \cap \ldots \cap A_n)$$

Since $A_1 \cap \ldots \cap A_n \subseteq A_1$, and $Pr(A_1) = 1 - p$, we can use Theorem 1 to conclude

$$Pr(all \quad OK) \leq Pr(A_1) = 1 - p.$$

So the probability that all connections are okay is between 90% and 99%.

Could it actually be as high as 99% ? Yes, if the errors are in such a way that whenever you make 1 connection wrong, all the connections are wrong. That is if $\bar{A}_1 = \bar{A}_2 = \ldots = \bar{A}_n$ then $A_1 = A_2 = \ldots = A_n$ and $A_1 \cap A_2 \ldots \cap A_n = A_1$. And $Pr(all \quad OK) = Pr(A_1) = 1 - p = 0.99$

Could it be 90% ? Yes, suppose the errors are such that we never make two wrong connections. So the events $\bar{A}_i$ are all disjoint and the probability of getting it right is

$$Pr(A_1 \cap \ldots \cap A_n) = 1 - Pr(\bar{A}_1 \cup \ldots \cup \bar{A}_n)$$

$$= 1 - \sum_{i=1}^{n} Pr(\bar{A}_i) = 1 - np$$

Here we used the inclusion-exclusion formula with the observation that for 2 or more sets, $Pr(\cap_{i \in S} A_i) = 0$.

Note that the sample space for this problem is

$$S = \{(a_1, a_2, \ldots, a_n) \mid a_i \in \{0, 1\}\}$$

Where each sample point $(a_1, a_2, \ldots, a_n)$ corresponds to a way of wiring the circuit; with $a_i = 0$ if the i^{th} connection is bad and $a_i = 1$ if it is OK. Although we don't have enough information to calculate the probability of individual sample points, we were able to conclude something useful by using simple facts from set theory.

3 Nontransitive dice

As a second application of Boole's inequality, recall the "nontransitive" dice we talked about in lecture. There were 3 dice a, b, c. They had the property that on any roll the probability that a beats b is 5/9 (i.e., a comes up higher than b), the probability that b beats c was 5/9 and that c beats a was 5/9.

How high can we make this probability? E.g. can we construct dice so that the probability of a beating b, b beating c, c beating a are all arbitrarily high? say 3/4 each?

To answer this, define A to be the event that a beats b, B the event that b beats c and C the event that c beats a.

Theorem 7

$$\min\{Pr(A), Pr(B), Pr(C)\} \leq \frac{2}{3}$$

Proof.

$$\begin{aligned} Pr(A \cap B \cap C) &= 1 - Pr(\bar{A} \cup \bar{B} \cup \bar{C}) \\ &\geq 1 - (Pr(\bar{A}) + Pr(\bar{B}) + Pr(\bar{C})) \end{aligned}$$

using Theorem 2 and Boole's inequality. But $Pr(A \cap B \cap C) = 0$ since it is impossible that on any single roll a is higher than b and b is higher than c *and* c is higher than a. Also $Pr(\bar{A}) = 1 - Pr(A)$.

Hence

$$\begin{aligned} 0 &\geq 1 - (3 - (Pr(A) + Pr(B) + Pr(C))) \\ 2 &\geq Pr(A) + Pr(B) + Pr(C) \\ &\geq 3 \min\{Pr(A), Pr(B), Pr(C)\} \\ \rightarrow 2/3 &\geq \min\{Pr(A), Pr(B), Pr(C)\} \end{aligned}$$

■

4 Conditional probability and Bayes rule

Recall the formula for conditional probability:

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

As an example, consider rolling two dice. What is the probability of having rolled at least one 3, given that total of the two dice is 8?

Assume each die gives each outcome (i.e., 1,2,3,4,5, or 6) with probability 1/6, no matter what the outcome of the other die is. Let A be the event that total seen is 8, and let B be the event that at least one 3 has been rolled. The sample space is $S = \{(i,j) \mid i,j \in \{1,2,3,4,5,6\}\}$, so that $|S| = 36$.

$$A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}, \ \ B \cap A = \{(3,5),(5,3)\},$$

Each outcome has probability 1/36, so $Pr(A) = 5/36$ and $Pr(B \cap A) = 2/36$. Therefore $Pr(B|A) = \frac{2/36}{5/36} = 2/5$.

Are $Pr(A|B)$ and $Pr(B|A)$ related? Yes, both relate to $Pr(A \cap B)$:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}, \text{ and } Pr(B|A) = \frac{Pr(B \cap A)}{Pr(A)},$$

so

$$Pr(A|B)Pr(B) = Pr(A \cap B) = Pr(B|A)Pr(A).$$

This yields the famous *Bayes Rule:*

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

Example: At a school 40% of students are male, and 60% of males and 80% of all students play sports. What is the probability that a randomly chosen student is a male, given that the student plays sports?

Let A be the event that a student is male. Let B be the event that a student plays sports. We are given $P(A) = 40\%$, $P(B) = 80\%$ and $Pr(B|A) = 60\%$. We want to compute $Pr(A|B)$. By Bayes Rule

$$Pr(A|B) = \frac{Pr(B|A)}{P(B)}Pr(A) = \frac{0.6 \times 0.4}{0.8} = 30\%.$$

5 Team problems

1. Let $A_1, \ldots, A_n$ be a collection of disjoint sets such that their union is equal to the sample space. Prove by induction that for any event B

   $$Pr(B) = Pr(B \cap A_1) + \ldots + P(B \cap A_n)$$

   And note the corollary:

   $$Pr(B) = Pr(B \cap A) + Pr(B \cap \bar{A})$$

2. Use part 1 to show that if $A_1, \ldots, A_n$ is a disjoint collection that covers the sample space, then for any events B, C such that $P(C) \neq 0$

   $$Pr(B \mid C) = Pr(B \cap A_1 \mid C) + \ldots + Pr(B \cap A_n \mid C)$$

   And note the corollary:

   $$Pr(B \mid C) = Pr(B \cap A \mid C) + Pr(B \cap \bar{A} \mid C)$$

3. In the OJ Simpson trial it was claimed by the defense that the fact that Simpson battered his wife should be excluded from the evidence as the fraction of wife batterers that murder their wives is 1/2000 (Dershowitz Lemma). In this problem we will condition this probability on the event that Nicole is dead and see that we obtain a somewhat different result.

   We have the following statistical information (all for 1994):

- number of women = 125,000,000
- number of women murdered in 1994 = 6000
- number of women murdered by spouse in 1994 = 1500
- number of women with spouse in 1994 = 50M
- number of women battered by spouse in 1994 = 3M

Being murdered counts as being battered. Also, we assume that the event of being murdered by somebody else is not influenced by whether the woman was being battered or not.

Define the events

S = Woman is murdered by spouse
M = Woman is murdered
B = Woman is battered by spouse

Use the given information to evaluate the following

(a) $Pr(S \mid B)$, i.e., the probability that a battered woman is murdered by her spouse.

(b) $Pr(S \mid M)$, i.e., probability that a murdered woman was done in by her spouse.

(c) $Pr(S \mid B \cap M)$, i.e., probability that a battered, murdered woman is done in by her spouse.

What can you conclude about Dershowitz' argument that the battery of a spouse should not be admitted as evidence because of his lemma ?