Massachusetts Institute of Technology
6.042J/18.062J: Mathematics for Computer Science
Professor David Karger

Lecture 21
May 1, 1997

Notes 21

1 Probability and Counting

The rules for probabilities that we saw last time and will see again today are analogous to the rules we have already seen for counting.

This is most obvious when dealing with an equiprobable distribution,

Definition 1.1 A sample space S is said to be uniform (or equiprobable) if $Pr(s)$ is the same for all $s \in S$. In particular, $Pr(s) = \frac{1}{|S|}$.

Theorem 1.2 If S is uniform and $A \subseteq S$, then $Pr(A) = |A|/|S|$.

Proof. $Pr(A) = \sum_{a \in A} Pr(a) = \sum_{a \in A} 1/|S| = |A|/|S|$ ∎

Definition 1.3 The complement of the event A is $S - A$. It will be denoted by $\bar{A}$.

Theorem 1.4

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

Proof. Obvious. ∎

Although equiprobably is easiest, these rules are actually true for any distribution.

One explanation: "many worlds" model. Probabilities are measuring fraction of worlds where event happened.

1.1 Sum Rules

Sum rule for counting: if A and B are disjoint, the $|A \cup B| = |A| + |B|$.

Divide both sides by $|S|$: $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.

Monotonicity rule for counting: $A \subseteq B \Rightarrow |A| \leq |B|$.

Divide both sides by $|S|$: $A \subseteq B \Rightarrow \Pr(A) \leq \Pr(B)$.

The sum rule can be extended to nondisjoint sets.

Inclusion-Exclusion rule for counting: $|A \cup B| = |A| + |B| - |A \cap B|$.

Divide both sides by $|S|$: $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

Like for counting, the inclusion-exclusion rule can be used to do estimates:

Corollary 1.5 (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 $Pr(A \cap B) = 0$. By induction, Boole's Inequality can be easily extended to n events.

Theorem 1.6 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. 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:

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

■

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 give bounds on this probability. Using Theorem 1.6, 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 use the subset rule 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 just the complement of the event that a connection is made incorrectly; so the probability that the chip gets wired correctly in this example is between .9 and .99.

1.2 Case Trees and Conditional Probability

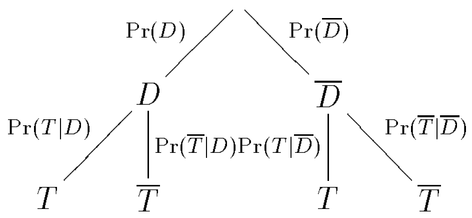
We analyzed by case trees to do counting.

It's also the way we studied probability for Monty Hall.

Related to conditional probability we studied last time:

Recall: $\Pr(A \mid B) = \Pr(A \cap B)/\Pr(B)$

Baby story: $\Pr(D) = 10^{-4}$, $\Pr(T \mid \overline{D}) = 10^{-2}$. Gives $\Pr(D \mid T) = .01$.



Value on edges are conditional probabilities, conditioning on choices made before.

Our rule that probability of a node is product of probability on path is the rule

$$\Pr(A_1 \cap A_2 \cap A_3 \cdots) = \Pr(A_1)\Pr(A_2 \mid A_1)\Pr(A_3 \mid A_1 \cap A_2) \cdots$$

that was proved last time.

What about our rule for adding up probabilities? It corresponds to the following theorem:

Theorem 1.7 If $A_1, \ldots, A_n$ is a partition of the sample space, then

$$\Pr(B) = \sum \Pr(B \mid A_i) \Pr(A_i)$$

2 Independence and the Product Rule

Now we turn to the probabilistic analogue of the product rule. Remember our rule that $|A \times B| = |A| \cdot |B|$? Creep up on analogue.

Definition 2.1 An event A is independent of an event B if

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

In other words, knowing that B occurs doesn't change the probability that A occurs.

For example, if we flip two coins, then the result of coin 1 is unrelated to that of coin two, thus the event "coin 1 is head" is independent of the event "coin two is head".

Independent events are particularly nice to work with. This is because there is a very simple rule to compute the probability of their intersection.

Theorem 2.2 If A is independent of B, then

$$Pr(A \cap B) = Pr(A) \cdot Pr(B)$$

Proof.

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

■

Note that the product rule is symmetric in A and B; it is not by chance.

Theorem 2.3 If A is independent of B, then B is independent of A. In other words, independence is a symmetric relation.

Proof.

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

■

Thus it is appropriate to say that A and B are "independent of each other". This makes it possible to define mutual independence of more than two events.

Definition 2.4 Events $A_1, A_2, \cdots, A_n$ are mutually independent, if for all i, $1 \leq i \leq n$ and for all $J \subseteq \{1, \ldots, n\} - \{i\}$ we have

$$Pr(A_i | \bigcap_{j \in J} A_j) = Pr(A_i)$$

In other words, a collection of events is independent, if each of them is independent of the intersection of any others. Equivalently, the above events are mutually indenpendent, if for all $J \subseteq \{1, \ldots, n\}$, $\Pi_{i \in J} Pr(A_i) = Pr\left(\cap_{i \in J} A_i\right)$.

For example, if we toss n coins and A_i is the event that the i-th coins is head, then these events are mutually independent, as $Pr(A_i) = 1/2$ and this is true regardless of the results of other coins.

The following example is a little bit tricky. We flip three fair coins. Let

- A_1=coin 1 matches coin 2
- A_2=coin 2 matches coin 3
- A_3=coin 3 matches coin 1

Let's see if these events are mutually independent. Clearly, $P(A_i) = 4/8 = 1/2$ for all i. Similarly, $P(A_i \cap A_j) = 2/8 = 1/4$ for any i and j, if $i \neq j$, as such an event occurs iff all three coins are head or all of them are tail. Thus $P(A_i) \cdot P(A_j) = P(A_i \cap A_j)$. However, these events are NOT mutually independent, as $P(A_1 \cap A_2 \cap A_3) = 1/4 \neq P(A_1)P(A_2)P(A_3) = 1/8$. This illustrates how important it is to check all subsets when looking for mutual independence.

A weaker condition on independence is the following:

Definition 2.5 Events $A_1, A_2, \cdots, A_n$ are pairwise independent, if for all i and j, $i \neq j$, A_i and A_j are independent.

Thus mutual independence implies pairwise independence, but the converse is not true, as the previous example shows.

3 The Birthday Problem

Say there are m=80 people in the classroom. What is the probablility of the event A that there are two of them who have the same birthday?

There are N=365 days in a year. Therefore, we can describe the data in a vector of length m, say $(b_1, b_2, \cdots, b_m)$ so that b_i denotes the birthday of the i-th person. Then we have N choices for each coordinate. The set of these vectors will be our *sample space*.

In the sequel we assume that our sample space is uniform, that is, the birthdays of the people in the classroom are independent from each other. This means that for any $s \in S$, we have $Pr(s) = 1/N^m$ by the previous theorem.

Is this assumption reasonable?

Now we are computing the number of sample points (vectors) with all different values to get the probability of the event $\bar{A}$. This is easy: we have N choices for the first person's birthday, N-1 for the second person's birthday, and so on, finally we have N-m+1 choices for the last person's birthday. Thus

$$Pr(\bar{A}) = \frac{N(N-1)\cdots(N-m+1)}{N^m}$$

Note that this was phrased combinatorially. Can also phrase it using product rule for conditional probabilities. Prob first birthday unique is 1. Prob. second is unique given first is is 364/365. And so on.

This answers our question, but we are interested to know how large this number is. Sums are usually easier to deal with than products, thus we apply the well known inequality $1-x < e^{-x}$ with $x = i/N$. This yields:

$$Pr(\bar{A}) = \Pi_{i=1}^{m-1}(1 - \frac{i}{N}) \leq \Pi_{i=1}^{m-1} e^{-i/N} =$$

$$= e^{-\sum_i^{m-1} \frac{i}{N}} = e^{-\frac{(m-1)m}{2N}}$$

To get a lower bound, use the fact that $1 - x > e^{-x-2x^2/3}$ for $0 \leq x \leq 1/3$. This yields

$$Pr(\bar{A}) \geq \Pi_{i=1}^{m-1} e^{-i/N-2i^2/3N^2} = e^{-\sum_{i=1}^{m-1} i/N+2i^2/3N^2} \geq$$

$$\geq e^{-\frac{m(m-1)}{2N}+\frac{2m^3}{9N^2}}$$

If we use our upper bound in the case of N=365, m=80, then it follows that there is only 0.02 percent chance not to have two people in the classroom who have the same birthday. This is very surprising, as 80 is not even one fourth of 365.