MAT 105 - Summary of Lecture 4
Thursday, January 14, 2016

Probability theory is nothing but common sense reduced to calculation. (Pierre Laplace)

1 Basics of probability

People say that probability and statistics are the two sides of the same coin. Why is that? Consider a bin with balls that we want to study. If we already know that the bin has 20 red balls, 10 blue and 30 green, probability ideas help us answer questions like "what are the odds of choosing a blue ball, when picking a ball at random?" Intuitively, I have 10 out of 60 chance of picking a blue ball, so this probability is $1/6$. What if I know there are balls in the urn, but I do not know what kind of balls there are? Then we can sample from the bin and make decisions about the contents based on the sample. Statistics provides a framework for sampling and analyzing the sample data.

Let us revisit the example from last week, where we wanted to draw conclusions about the height of people in the US. If we know all heights of the population, we can compute **exactly** the odds of a person picked at random to be 5.6 ft, by using probability. However, if we do not know all heights, we can use a *good random sample* to compute various parameters (such as mean, variance) which help us **approximate** the odds of a person to be around 5.6 ft tall.

Many times we can recognize the distribution of the data, based on knowledge about the experiment. In that case, we use probability to study the data. Other times, we can only approximate the distribution, using statistics ideas, and draw conclusions based on the approximation. In order to discuss sampling theory and statistical inference, we will also need to use some ideas from probability. Thus, the focus of our next few weeks will be to study probability and we will return to statistics later on in the semester.

Def: An **experiment** is an activity or a procedure that leads to *distinct* and well-defined possibilities called **outcomes**. The set of all outcomes forms the **sample space**, which we denote by **S** (some books use Ω for the sample space). If S is finite, let $|S|$ denote the number of outcomes in the sample space. We would like to find odds (probabilities) for various "events" of interest. So what is an "event"? It is a statement about the outcome of the experiment.
If the sample space is finite and each outcome is *equally likely*, we compute the probability of an event E by

$$P(E) = \frac{\#\text{ outcomes in } E}{\#\text{ outcomes in } S}.$$

Examples:

1. Roll a six-sided fair die (this is our experiment!). Then $S = \{1, 2, 3, 4, 5, 6\}$ and $|S| = 6$. Let $E =$"the die lands on 6" be the event of interest. There is one outcome from the sample space that satisfies the statement E. We also say that "6 is in E". We can find the probability of this event by

$$P(E) = \frac{\#\text{ outcomes in } E}{\#\text{ outcomes in } S} = \frac{1}{6}.$$

2. Flip two fair coins. Then $S = \{HH, HT, TH, TT\}$ and $|S| = 4$. Let Let $E =$"both coins land on heads" be the event of interest. Then $E = \{HH\}$.

$$P(E) = \frac{\#\text{ outcomes in } E}{\#\text{ outcomes in } S} = \frac{1}{4}.$$

We observe that the main question in finding probabilities involve being able to count the number of outcomes satisfying a certain property, or counting the number of outcomes in the sample space. The main counting ideas are illustrated in the rest of the lecture.

2 Counting

Example 1: I have 5 bananas, 4 oranges and 3 apples. In how many ways can I pick a fruit?
I pick one banana OR one orange OR one apple. I can do this in $5 + 4 + 3 = 12$ ways. Note that we assume that fruits are distinct (fruits of the same type are different).

Sum Rule: If there are k types and $n_1, n_2, \ldots, n_k$ objects of type $1, 2, \ldots, k$ respectively, there are

$$n_1 + n_2 + \cdots + n_k$$

ways to pick one object from the set.

Example 2: I form a game team. I need a programmer, a physics person, a producer, an artist and a designer. My team must include one of each. There are 10 programmers, 2 physics people, 7 producers, 5 artists, and 8 designers who are willing to be on my team. In how many ways can I pick the team?
I first pick the programmer (10 ways), AND then I pick the physics person (2 ways) AND then I pick the producer (7 choices) AND then the artist (5 choices) AND lastly the designer (8 choices). There are $10 \times 2 \times 7 \times 5 \times 8 = 5600$ possible teams I can create! This follows the following rule:

Multiplication Rule: If there are k types and $n_1, n_2, \ldots, n_k$ objects of type $1, 2, \ldots, k$ respectively, there are

$$n_1 \times n_2 \times \cdots \times n_k$$

ways to pick one object from each set.
Remark: to distinguish between when to use the sum rule and when to use the multiplication rule, think that for sum you pick A OR B, but for multiplication you pick A AND B.

Example 3: In how many ways can I order 4 distinct dice?
I have 4 choices for the first die, AND then I have 3 choices left for the second die AND now 2 choices for third die AND one choice left for the last die. There are $4 \times 3 \times 2 \times 1 = 4!$ ways to order the dice.

Factorial: $n! = n(n-1)(n-2) \cdots 2 \times 1$ represents the number of ways to order n objects.
Some properties of factorials:

- $0! = 1$ by definition.
- $(n+1)! = (n+1) \times n!$

Example 4: There are 47 students in the class. In how many ways can I pick 4 representatives to serve as president, vice-president, secretary, treasurer for a club I want to organize?
I pick one president out of 47, AND one vice-president out of 46 students remaining, AND a secretary out of 45 AND lastly a treasurer out of 44. There are $47 \times 46 \times 45 \times 44 = \frac{47!}{43!}$ possible committees I can form.

Permutations: Permutations of n objects, taken k at a time count the number of ways to pick an *ordered* set of k objects out of n. It is defined as

$$P(n,k) = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

Example 5: There are 47 students in the class. In how many ways can I pick 4 class representatives, without specific jobs?
We might think that there should be $P(47,4)$, just as in Example 4, but we notice that we would be over-counting by a *factor* of 4! which represents the number of different ways in which the 4 representatives can be assigned 4 distinct jobs. Thus, there are $\frac{P(47,4)}{4!} = \frac{47!}{(43!)(4!)} = \binom{47}{4}$ ways to pick 4 representatives.

Combinations: Combinations of n objects, taken k at a time count the number of ways to pick an *unordered* set of k objects out of n. It is defined as

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Remark: we use permutations when the desired set is ordered, and when the objects are distinct; we use combinations when the desired set is unordered, and when the objects are not distinct.

Example 6: I roll 8D6.

(a) In how many ways can we get two 2's and six 6's? Imagine there are 8 slots to be filled with 2's or 6's. We choose two slots where we place the 2's in $\binom{8}{2} = \frac{8!}{2!6!} = 28$ ways. The remaining slots get filled with 6's.

(b) In how many ways can we get two 2's, three 3's and three 5'2? First, we choose two slots where to place the 2's in $\binom{8}{2} = \frac{8!}{2!6!} = 28$ ways. We are left with 6 empty slots and we choose three of them where to place the 3's in $\binom{6}{3} = \frac{6!}{3!3!} = 20$ ways. The 5's get placed in the remaining 3 slots. Thus, there are

$$\binom{8}{2} \times \binom{6}{3} \times \binom{3}{3} = \frac{8!}{2!6!} \times \frac{6!}{3!3!} \times \frac{3!}{3!0!} = \frac{8!}{2!3!3!} = \binom{8}{2,3,3} = 560$$

ways to get such an outcome.

Multinomial: The multinomial $\binom{n}{n_1, n_2, \ldots, n_k}$ counts the number of ways in which n objects can be placed in k bins of sizes $n_1, n_2, \cdots, n_k$ with $n_1 + n_2 + \cdots + n_k = n$ and is defined as

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1! \times n_2! \times \cdots \times n_k!}.$$

Remarks:

- Note that each element is placed in some bin.
- If there are only two bins, then $\binom{n}{n_1, n_2} = \frac{n!}{n_1! \times n_2!} = \binom{n}{n_1}$, the combination of n objects taken n_1 at a time, which is also called a **binomial**. We will discuss the binomial and the related binomial distribution throughout the semester.
- We can interpret the multinomial as first placing n_1 objects out of n in bin 1, then placing n_2 objects out of $n - n_1$ in bin 2, etc, n_k out of $n - n_1 - n_2 - \cdots - n_{k-1} = n_k$ in bin k. This gives

$$\begin{aligned}\binom{n}{n_1, n_2, \ldots, n_k} &= \binom{n}{n_1} \times \binom{n - n_1}{n_2} \times \binom{n - n_1 - n_2}{n_2} \times \cdots \times \binom{n_k}{n_k} \\ &= \frac{n!}{n_1!(n-n_1)!} \times \frac{(n-n_1)!}{n_2!(n-n_1-n_2)!} \times \frac{(n-n_1-n_2)!}{n_3!(n-n_1-n_2-n_3)!} \times \cdots \times \frac{n_k!}{n_k!0!} \\ &= \frac{n!}{n_1! \times n_2! \times \cdots \times n_k!},\end{aligned}$$

  since everything else cancels and $0! = 1$.

More examples:

1. I have 10 games.
   (a) In how many ways can I order them? In 10! ways.
   (b) In how many ways can I choose 3 games to give away? In $\binom{10}{3} = \frac{10!}{3!7!}$ ways, since the set of 3 needs not be ordered.
   (c) In how many ways can I choose 3 games to give to Al, Bob and Cate? In $P(10, 3) = \frac{10!}{7!}$ ways, since the set of 3 needs to be placed in order: Al, Bob, Cate.
   (d) For each game, I decide to keep it, donate it or trash it. How many possible outcomes are there? There are 3 possible outcomes for each game, so using the Multiplication Rule, there are 3^{10} possible outcomes.
2. Flip a fair coin twice. What is the probability that we get one H and one T?

   We might be tempted to say that there are three possible outcomes: 2H, 2T or 1H+1T, so the probability is 1/3. But that would be wrong because these outcomes are not equally likely! In fact, 1H+1T is twice as likely as 2H.

   A more accurate count is obtained by looking at the outcome of each flip. Let S denote the sample space as usual, and let E be the event that we get one H and one T. Then

$$S = \{HH, HT, TH, TT\}, \quad E = \{HT, TH\} \quad \Rightarrow \quad P(E) = \frac{|E|}{|S|} = \frac{2}{4} = \frac{1}{2}.$$

   Another way to reason about this problem is to say that it does not matter what the first coin lands on, but the second coin has to NOT match it, which happens with probability 1/2. This argument is correct and intuitive, but uses *independence* between flips, which we need to justify. We will discuss it in later lectures.

3. Roll two six-sided dice, of different colors. The sample space is the set of all pairs (i, j) with i, j both ranging from 1 to 6. That is,

$$S = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}.$$

Thus, $|S| = 6 \times 6 = 36$. Let E be the event $E =$ "both dice land on 6". Then only the outcome $(6, 6)$ is in E, so

$$P(E) = \frac{|E|}{|S|} = \frac{1}{36}.$$

Let $F =$"at least one die lands on 6". Looking in the table above, we count 11 outcomes satisfying this event, so

$$P(F) = \frac{|F|}{|S|} = \frac{11}{36}.$$

Let $G =$"the sum of the dice is 7". Then $G = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, so

$$P(G) = \frac{|G|}{|S|} = \frac{6}{36}.$$

Let $H =$"the sum of the dice is 9". Then $H = \{(3, 6), (4, 5), (5, 4), (6, 3)\}$, so

$$P(H) = \frac{|H|}{|S|} = \frac{4}{36}.$$

We can intuitively say that $P\{\text{sum is } 9\} < P\{\text{sum is } 7\}$ even without direct computation, by noting that you can make 9 less often than making 7. Our intuition is correct!

4. On my shelf, I have 10 Math books, 3 CS books and 5 Physics books. In how many ways can I

   (a) order the books on the shelf?

   Since there are 18 books in total, there are 18! ways to order them.

   (b) pick 4 Math books?

   In $\binom{10}{4} = \frac{10!}{4!6!}$ ways, since I choose only out of 10 possible Math books and the order in which I pick them does not matter.

   (c) pick one book of each type?

   By the Multiplication Rule, in $10 * 3 * 5 = 150$ ways.

   (d) pick one book from the set?

   In $10 + 3 + 5 = 18$ ways, using the Sum Rule.