MthStat 465, Spring 2005, Lecture Number 9
Combinatorics continued; Geometric Probability; Conditional Probability and Independence.

The day's lecture wrapped up combinatorics, introduced geometric probability, and then used geometric probability to motivate the idea of conditional probability and independence.

1. Combinatorics (conclusion)

It is important to use diagrams and other visual aids to solve elementary combinatorial problems. Here are some examples.

The following problem was suggested by one of Henry Kranendonk's examples from the preceding lecture. Fifteen raisins are to be distributed among the five vertices of an irregular pentagon. If at least one raisin is to be placed at each vertex, how many arrangements are there? If we let x_i be the number of raisins at vertex i, then we have $x_1 + x_2 + x_3 + x_4 + x_5 = 15$ and each x_i is a positive integer. If we let R represent a raisin, we can line the raisins up is a row:

$$R\ R\ R\ R\ R\ R\ R\ R\ R\ R\ R\ R\ R\ R\ R$$

There is a one-to-one correspondence between solutions of our equation, $(x_1, x_2, x_3, x_4, x_5)$ and the placement of 4 vertical dividers, $|$, between the R's. For example,

$$R\,|\,R\ R\,|\,R\ R\ R\ R\,|\,R\,|\,R\ R\ R\ R\ R\ R\ R$$

corresponds to $(x_1, x_2, x_3, x_4, x_5) = (1, 2, 4, 1, 7)$. Since there are 14 positions and we must select 4 of them for the dividers, there are ${}_{14}C_4 = 1001$ possible solutions.

If we change the problem slightly by requiring that the x_i be non-negative, then we can solve it as follows. We have to arrange 15 R's and 4 $|$'s in some order. Each of these orderings indicates as solution. What is new here is that you can have two $|$'s next to one another. This signifies that the corresponding variable is 0. For example,

$$R\,|\,R\ R\,\|;R\ R\ R\ R\,|\,|\,R\ R\ R\ R\ R\ R\ R\ R$$

corresponds to $(x_1, x_2, x_3, x_4, x_5) = (1, 2, 4, 0, 8)$. Now we must select 4 positions out of 19 for the $|$'s, so there are ${}_{19}C_4 = 3876$ solutions.

On a related note, we were reminded that games of chance are generally modeled using classical probability, so it is important to be able to count. As an example, there are ${}_{52}C_5$ sets of 5 cards drawn from a standard deck of playing cards. There are $13 \times 12 \times_4 C_3 \times_4 C_2 = 3744$ ways to get three of one type of card and 2 of another. This is an application of the multiplication principle. There are 13×12 ways to select the value to get three of and the value to get two of. One example is 3 queens and 2 kings. There are 4 ways to get 3 queens (just decide which queen you don't want!) and 6 ways to pick the 2 kings. The probability of such a hand is 6/4165.

Lastly, the Pascal's triangle relation was discussed. This says that

$$_{N+1}C_{k+1} =_N C_{k+1} +_N C_{k+1}.$$

Although this identity can be proven algebraically, it can be proved directly as well. Imagine $N + 1$ balls numbered 1 to N+1. When I select $k + 1$ balls either I select ball N+1 or not. There are ${}_N C_{k+1}$ selections of $k + 1$ balls that do not include ball N+1, and there are ${}_N C_k$ selections that include ball N+1. The latter is true because since ball N+1 must be selected, there are now N balls to select the remaining k balls from. This idea was extended to selecting from a set of objects of two types, where we were to get a objects of one type and b objects of the other type.

2. Geometric Probability

Geometric probability refers to defining the probability of an object as proportional to its size. For example, if the sample space is a rectangle of area 12 and an event is a disk inside the rectangle of area π, then the probability of the disk is $\pi/12$. When you draw Venn diagrams to investigate properties of events, you are already thinking of their probabilities geometrically, as you naturally draw bigger sets for set of higher probability.

3. Conditional Probability and Independence

The issue here is to give mathematical meaning to the quantity "The probability of the event A given that the event B has been observed." Let us denote this quantity by $\Pr(A|B)$. For example, what is the probability of two pairs given that all five cards are black. The idea is that the sample space has been reduced from S to B.

A reasonable mathematical formulation can be deduced from examples in geometric probability with sets in a plane. Let S denote the original sample space. For any event F we have $\Pr(A) = \text{Area}(A)/\text{Area}(S)$.

Since we are trying to give meaning to $\Pr(A|B)$, we are only concerned with outcomes that are already in B. This means that A is no longer of interest, and that we should consider $A \cap B$ instead, the elements of A that are also in B. The probability we want is $\text{Area}(A \cap B)/\text{Area}(B)$ provided the area of B is positive. However, it is the case that

$$\frac{\text{Area}(A \cap B)}{\text{Area}(B)} = \frac{\text{Area}(A \cap B)/\text{Area}(S)}{\text{Area}(B)/\text{Area}(S)} = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Based on this type of example, we make the following general definition: If $\Pr(B) > 0$ then $\Pr(A|B) := \Pr(A \cap B)/\Pr(B)$.

Now, it is also clear that saying that the events A and B are independent should mean that $\Pr(A|B) = \Pr(A)$ and $\Pr(B|A) = \Pr(B)$. However, this requires two statements and that neither event have probability 0. Each statement implies that $\Pr(A \cap B) = \Pr(A)\Pr(B)$, a single statement that is symmetric in the events and has no restrictions on the probabilities of A or B. For that reason we make the definition that A and B are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$ and then note that if A and B are independent and $\Pr(B) > 0$, then $\Pr(A|B) = \Pr(A)$, as desired.

It is important to note that independence is a function of the choice of Pr in your model. You can sometimes choose Pr in such a way to force certain events to be independent, but then other events may be dependent. If Pr is chosen first, then independence must be verified for any pair of events.