POL 602
Prof. Matthew Lebo
Week 2 – September 6 and 8, 2005

Probability Theory I

Although people have been concerned with chance in the context of gambling games since before the dawn of recorded history, the philosophical and mathematical investigation of probability is fairly recent, beginning only in the 1660's. [For an excellent discussion of the early history of probability see Hacking's (1975) *The Emergence of Probability*.]

The early theorists of probability, whose work is now referred to as Classical Theory, were primarily concerned with games of chance involving "fair" devices such as dice, cards and coins.

A important issue for classical theory was the assumption of fairness, or *equiprobability of outcomes*, first justified by appeal to physical features of the gambling tools such as symmetry or evenly-distributed weight, and later formalized by an axiom referred to today as the Principle of Indifference (or the Principle of Insufficient Reason as Keynes put it).

Principle of Indifference: This principle essentially holds that alternatives are to be considered equiprobable if there is no reason to expect or prefer one over the other.

There are two problems, in the modern view, with the principle. The first concerns the circular definition of probability in terms of equiprobablity (i.e. events are assigned probability based on their symmetric probability). In other words, we define how probable certain outcomes are based on assumptions about how probable the outcomes are. Rather tautological.

The second problem is that the Principle of Indifference confounds at least two different senses of prior ignorance (Smithson 1989). On the one hand, we know (given no evidence to the contrary) that all events are equally likely. However, on the other hand, we do not know how likely any of them are.

Example: You are about to throw a standard six-sided die, when the person sitting next to you whispers that the die is loaded (i.e. unbalanced in weight such that some set of sides are favored over others). By the principle of indifference, we should still regard the outcomes as equiprobable, since *we don't know in what way the die is loaded.*

Another way to think about this problem is the assignment of probabilities to outcomes more complex than simple gambling devices. We have a feeling that the outcome of a positive test result from a medical procedure, or that a friend will get married in the next 3 years, are not equiprobable with their converse (a negative result, or unmarried). But how do we assign probabilities to the proper outcome?

Throughout the history of probability theory, there have been several responses to the dilemmas of classical theory. We will examine the two dominant responses: the *relative frequency* or *frequentist* approach, and the *subjective probability* or *Bayesian* [after Rev. Thomas Bayes, some of whose results we will examine in detail later] approach.

Relative Frequency or the Frequentist Approach

Historians credit Venn, Von Mises and Reichenbach as the founders of the frequentist approach.

In simple terms, this approach holds that the probability that a specific outcome obtains can be interpreted to mean the relative frequency with which that outcome would obtain if the process were repeated a large number of times under similar conditions. (note that "large number" and "similar conditions" are rather vague terms in this simple statement of theory).

In the frequentist approach, then, we assign the probability of ½ to getting heads when flipping a fair coin not because of an assumption about equiprobability but because if we were to flip the coin many, many times we expect the proportion of heads to be ½. This is not to say that we expect exactly ½ of any number ($n>1$) of flips to result in heads. In fact, we would be quite surprised if this outcome obtained consistently. This fact emphasizes the importance of a random or stochastic element to our coin flips.

Note also that the frequentist approach can be problematic in some cases, such as the probability of a friend getting married in the next three years mentioned earlier. There is no empirical way (or even a reasonable thought experiment) to have the friend relive the next three years a large number of times for us to determine the relative proportions.

Bayesian Approach

According to the subjective interpretation of probability, the probability that a person assigns to some outcome represents her own judgment of the likelihood that the outcome will occur.

Thus, if you and I have different beliefs, we will assign different *subjective probabilities* to a set of outcomes.

This approach is a powerful way to mathematically incorporate the subjectivity of scientific research into the data analysis process, but it is not without its limitations.

First, it can be difficult in many real world applications to express prior beliefs (or *priors*) in numerical terms that allow for the required calculations. Moreover, it is unlikely that all people or researchers have beliefs about the world that meet the strict requirements of probability theory. Additionally, until very recently, much Bayesian analysis was practically impossible due to computational limitations.

In this course, as in modern political science, the dominant approach to statistical inference is the Frequentist one.

It may be that in the course of our careers this may change, as recent trends in journal publications indicate, so I will try to introduce or discuss Bayesian terminology when I can.

Why might this occur? Because the assumption of the Frequentists about repeated sampling and statistical inference are so frequently broken.

Often, we are not dealing with a sample, but have the whole "universe" of cases. Such as studies of the 50 American States, or of the bills passed in the 108th Congress, or the European democracies. So what exactly are we inferring to?

The most important point to remember is that, with the exception of how probabilities are to be interpreted (and sometimes assigned), both approaches have in common the same mathematical framework of probability regarding two central questions:

1. How are the probabilities of certain events determined given the specified probabilities of each possible outcome of an experiment (i.e. after the frequentist or Bayesian assignment).

2. How are the probabilities of certain events revised or updated after additional relevant information is obtained.

It is these two questions, which require a mathematical theory of probability, that we will concern ourselves with for the remainder of today and all of next class.

The first tool in developing the theory we need is set theory, which you have already reviewed in the math course.

One terminology difference that is fairly standard across presentations of probability theory is that the universal set is now referred to as the *sample space* (S) and is defined as the set of all possible outcomes of an experiment.

Thus the sample space for the roll of a six-sided die is:

$$S = \{1, 2, 3, 4, 5, 6\}$$

I should note also that the textbook uses the bar notation for complements ($\overline{A}$).

Other results from set theory that the text presents that you may not be familiar with are *De Morgan's Laws*:

$$\overline{(A \cap B)} = \overline{A} \cup \overline{B}$$
$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}$$

A few more definitions are in order before proceeding to the axioms of probability.

First, an *experiment* is any process by which an observation is made.

Events are the possible outcomes of an experiment.

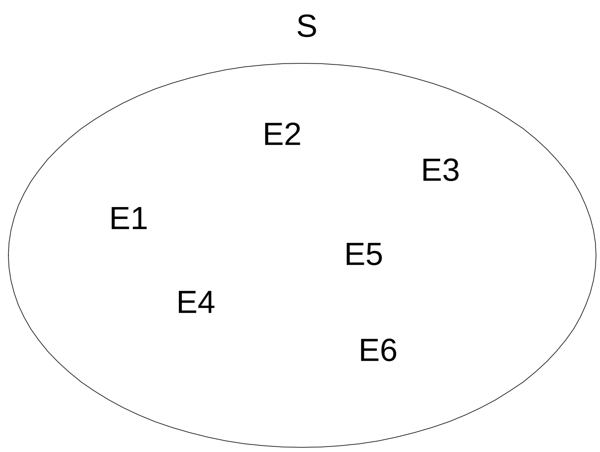
Simple events (E) are those which can not be decomposed.

They correspond to one *sample point*. The sample space, *S*, is the set of all possible sample points.

A *discrete sample space* is one that contains either a finite or denumerably infinite (countable) number of distinct sample points.

Compound events are sets of sample points or unions of simple events.

Example: Venn diagram of the die-tossing experiment



Simple Events:
Rolling a 1 = $\{E_1\}$
Rolling a 2 = $\{E_2\}$ etc…

Compound Events:
Rolling an odd number, event A = $\{E_1, E_3, E_5\}$
Rolling >4 , event B = $\{E_5, E_6\}$

We can now present the axioms of probability. (Note that there are several orders of increasing mathematical sophistication with which these axioms can be presented):

Suppose *S* is a sample space associated with an experiment. To every event *A* in *S* ($A \subset S$), we assign a number *P(A)* [often written Pr(*A*)]called the probability of *A*, such that:

1. $P(A) \geq 0$
2. $P(S) = 1$
3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ for all disjoint events A_1, A_2,…

Thus a probability distribution on a sample space S is a specification of numbers P(A_i) which satisfy axioms 1, 2, and 3.

From these simple statements, all the other properties of the probability function can be derived.

Note that this does not tell us how to assign particular probabilities to simple events (this is the role of classical, frequentist or Bayesian theory).

What are some of the other basic properties of Probabilities? We'll come back to some of these in more detail.

$P(\overline{A}) = 1 - P(A)$ the probability of not-A is equal to 1 minus the probability of A.

$P(\varnothing) = 0$ the probability of an event not in the sample space (the null set) is 0.

$If\ A \subset B, then\, P(A) \leq P(B)$ if A is a component of B, the probability of A is less than or equal to the probability of B.

For any event A, $P(A) \leq 1$.

Another expression of rule 3 above:

Let $A_1, A_2, ..., A_n$ be events defined over S.

$$\text{If: } A_i \cap A_j = \varnothing\ for\, i \neq j, then\, P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i).$$

An example using some of these rules:

Biff decides to apply to two colleges, X and Y. Based on how his friends have fared, he estimates that his probability of being accepted at X is 0.7 and at Y is 0.4. He also suspects that there is a 75% chance that at least one of his applications will be rejected. What is the probability that he gets at least one acceptance?

Let A be the event "School X accepts him" and B, the event that "School Y accepts him."

We are given that P(A) = 0.7, P(B) = 0.4, and P($\overline{A} \cup \overline{B}$)= 0.75.

We are looking for P($A \cup B$).

Which is: P($A \cup B$) = $P(A) + P(B) - P(A \cap B)$.

De Morgan's laws tell us that $\overline{A} \cup \overline{B} = (\overline{A \cap B})$, so

$P(A \cap B) = 1 - P[(\overline{A \cap B})] = 1 - 0.75 = 0.25$

We now have the 3 components we need – Biff's chances are pretty good with an 85% chance of getting in somewhere:

P($A \cup B$) = 0.7 + 0.4 – 0.25 = 0.85.

The *Sample-Point Method*:

A "cookbook" method (see pg. 34):

1. Define the experiment and clearly determine how to describe one simple event.

2. List the simple events associated with the experiment and test each to make certain it cannot be decomposed. This defines sample space S.

3. Assign reasonable probabilities to the sample points in S in accordance with the 1st and 2nd axioms of probability.

4. Define the event of interest, A, as a collection of sample points.

5. Find P(A) by summing the probabilities of sample points in A (axiom 3).

<u>Example (2.1) in the text (pg. 30)</u>
Note that not all sample spaces are discrete (an experiment testing the response time of subjects to stimulus after viewing a negative political ad, for example). These will be discussed in a few weeks. For now, we will focus on methods of calculating the probabilities of events in discrete sample spaces.

We have 5 seemingly identical computer terminals that can be shipped – but two are defective.

If they are chosen randomly, what is the probability that a shipment of 2 will contain no defective terminals.

List the sample space: combinations of terminals that can be either D1, D2, G1, G2, or G3.

There are ten combinations.

A is the event that two are chosen without either being defective.

A is the set of E8 = {G1,G2), E9 = {G1,G3}, E10 ={G2,G3}

If they are each selected randomly the probability of each is 1/10 and the probability of A is 1/10 + 1/10 + 1/10 = 3/10.

Example 2.3 on page 35 – tossing a fair coin 3 times. What is probability of 2/3 heads?

The probability is 3/8.

Here's a digression from the textbook.

For an experiment where there are only two possible outcomes, we can use a shortcut to see how many possible combinations there can be and what the distribution of those combinations can be in terms of the number of sample points in each combination.

Pascal's Triangle

Begin with the number 1 – if we flip a coin 0 times, there is only one outcome, the null set.

1

Add one to the unseen zero to left of one and again to the right of 1 to get the possible combinations of one flip.

```
        1
1               1
```

Then add the numbers in pairs, beginning with the unseen zero to the left of the first one in the second column.

```
        1
   1        1
1       2       1
```

And so on….

```
                        1
                1               1
        1               2               1
1               3               3               1
```

We can read this to see that with 3 flips there are 1+3+3+1 combinations of outcomes. Zero heads will occur 1 time, 1 head 3 times, 2 heads 3 times and 3 heads one time. Back to the sample-point method.

The sample-point method is mechanical but can be tedious and almost impossible to use in the case of large numbers of events or sample points. Accordingly, we borrow elementary results from *combinatorial analysis* or the theory of counting to assist us in analyzing sample spaces.

Tools for Counting Sample Points

The first is the so-called multiplication rule or *mn rule*. With m elements $a_1, a_2, \ldots, a_m$ and n elements $b_1, b_2, \ldots, b_n$, it is possible to form m x n pairs containing one element from each group. (Proof given is by inspection of an *a x b* table). For example, there are 6 times 6 = 36 different ways to roll 2 dice.

This rule can, of course be extended to any number of sets and can thus be used to count the sample points of any experiment that is reducible to a series of steps or parts.

Example—in the coin tossing case above, $m=n=p=2$, therefore the sample space has $2^3=8$ points. That is, 2 possibilities for each of 3 events.

The second tool is the *permutation*.

Consider an experiment in which we select three cards from a deck, one-at-a-time, without replacing the cards already selected (this is called sampling without replacement).

Given a standard 52-card deck, the size of the sample space is given by 52 (52-1) (52-2). There are thus 132,600 permutations or ways to draw three cards. More generally, if we want to find the number of permutations of n elements taken r at a time (so 52, 3 at a time in the example), the formula is:

$$P_r^n = n(n-1)(n-2)\ldots(n-r-1) = \frac{n!}{(n-r)!}$$

Note that this implies sampling without replacement and that $n!$ (read n factorial) is defined as $n(n-1)(n-2)\ldots1$, and that 0! = 1.

Book's definition here: An ordered arrangement of r distinct objects is called a permutation. The number of ways of ordering n distinct objects taken r at a time is designated by the symbol P_r^n. Sometimes this is shown as ${}_nP_r$.

Example: Suppose a club consists of 25 members. How many ways can a president and secretary be chosen (assuming only one office per person)?

Answer: $P_2^{25} = \frac{25!}{(25-2)!} = \frac{25!}{23!} = \frac{(25)(24)(23!)}{23!} = (25)(24) = 600$

Example: How many ways can one arrange six books on a shelf?

Answer: $P_6^6 = \frac{6!}{(6-6)!} = \frac{6!}{0!} = 6! = 720$

Note on sampling *with* replacement: the number of sample points is simply n^r.

For example, think of drawing 1 ball from 10 balls numbered 1..10 in an urn, noting the number and replacing it, a total of 5 times.

What is the probability of getting the sequence 1,2,3,4,5?

The number of sample points is $n^r=10^5$.

With one choice we have a vector of *n* objects.

With two, we have a grid that is n by *n*.

Three, a cube, and so on.

Assuming independent random draws, P(A)=1/10^5.

Obtaining Different Numbers

If $r > n$ it is impossible for all of the selected balls to be different.

So suppose that $r \leq n$.

The number of outcomes in the event E is the number of vectors for which all *r* components are different.

This equals $P_{n,r}$ since the first component x_1 of each vector can have *n* possible values, the second component x_2 can have any of the other $n - 1$ values and so on.

Since S is a simple space containing n^r vectors, the probability *p* that *r* different numbers will be selected is:

$$p = \frac{P_r^n}{n^r} = \frac{n!}{(n-r)!n^r}$$

<u>Example The Birthday Problem:</u>

What is the probability, *p*, that at least two people in a group of *r* [2,365] people will have the same birthday? Assuming no twins, etc.

The sample space is given by 365^r and the possible permutations wherein each person has a different birthday by P_r^{365}. Thus the probability of each person having a different birthday is $\frac{P_r^{365}}{365^r}$ and, by the second axiom of probability, the probability of each person not having a different birthday (i.e. at least two with the same birthday) is given by:

$$1-\frac{P_r^{365}}{365^r}=1-\frac{365!}{(365-r)!365^r}$$

In a class where r = 15, p = .253. There is about a 25% chance that at least two of us share the same birthday.

Combinations: the number of combinations of n objects taken r at a time (or the number of subsets r that can be formed by n objects) is denoted by:

$$nCr\ ,C_r^n\ \text{or}\ \binom{n}{r}$$

The latter is referred to as a *binomial coefficient.*

The number of unordered subsets of size r chosen (without replacement) from n objects is:

$$nCr\ ,C_r^n\ \text{or}\ \binom{n}{r}=\frac{P_r^n}{r!}=\frac{n!}{r!(n-r)!}$$

Example: Suppose a class contains 15 boys and 30 girls and that 10 students are to be selected at random for a special assignment. What is the probability that exactly 3 boys are selected?

Answer: The number of different combinations (assumed to be equally probable) is: $\binom{45}{10}$.

The ways that 3 boys can be selected are $\binom{15}{3}$ and the way 7 girls can be selected is $\binom{30}{7}$.

The answer is thus: $$p=\frac{\binom{15}{3}\binom{30}{7}}{\binom{45}{10}}=.2904$$

Suggested Homework Problems: 1.6, 1.8, 1.11, 1.19, 1.23, *1.30, 2.8, 2.9, 2.16, 2.21, 2.31, 2.35, 2.43, 2.45, 2.55

Probability Theory II

Last time we discussed the philosophy of probability, the axioms of probability and some ways to count sample points (some combinatorics).

Before we move on, I want to introduce one more result from combinatorical theory: the *multinomial coefficient*.

$$N = \binom{n}{n_1 n_2 \ldots n_r}$$

where N is the number of arrangements of *n* items in n_r groups—order doesn't matter.

The number of ways of partitioning N is:

$$N = \binom{n}{n_1 n_2 \ldots n_r} = \frac{n!}{n_1!n_2!\ldots n_r!}$$

Example: How many ways can one place 100 experimental subjects into 1 group of 50 and two groups of 25?

$$N = \binom{100}{50\ 25\ 25} = \left(\frac{100!}{50!25!25!}\right) \simeq 1.2 \times 10^{43}$$

Example II:

A deck of cards contains 13 hearts. The cards are shuffled and dealt among four players, A, B, C, D as they are in bridge so that each player receives 13 cards.

What is the probability *p* that player A gets 6 hearts, B gets 4 hearts, C gets two hearts and D gets one heart?

The total number N of different ways in which the 52 cards can be distributed among the four players so that each player receives 13 cards is:

$$N = \binom{52}{13,13,13,13} = \frac{52!}{(13!)^4}$$

We can assume that each of these ways is equally probable.

Now calculate the number of *M* ways of distributing the cards so that each player gets the specified number of hearts.

The number of different ways they can be distributed and give 6,4,2, and 1 hearts to A,B,C, and D, respectively is:

$$\binom{13}{6,4,2,1} = \frac{13!}{6!4!2!1!}$$

and the number of different ways in which the other 39 cards can then be distributed to the four players so that each will have a total of 13 cards is:

$$\binom{39}{7,9,11,12} = \frac{39!}{7!9!11!12!}$$

$$M = \frac{13!}{6!4!2!1!} * \frac{39!}{7!9!11!12!}$$

And the required probability is:

$$p = \frac{M}{N} = \frac{13!39!(13!)^4}{6!4!2!1!7!9!11!12!52!} = 0.00196$$

So, multinomial coefficients generalize binomial coefficients.

The coefficient $\binom{n}{n_1, \ldots, n_r}$ is the number of ways to partition a set of *n* items into distinguishable subsets of sizes $n_1, \ldots, n_r$ where $n_1, \ldots, n_r = n$.

It is also the number of arrangements of *n* items of *r* different types for which n_i are of type *i for i=1,…,r*.

Conditional Probability

The probability of an event can often be dependent on other random variables.

Another way of saying this is that our priors about the probability of an event may be different if we have additional information.

For example, what is the probability of a certain bill passing the Senate, *given* a Republican majority?

Or what is the probability that G.W. Bush will win a general election given a series of independent variables?

What is the probability that a voter will respond to a telephone survey given that she is White? Black? A Pacific Islander?

These are all examples of *conditional probabilities*.

More formally: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}, P(B) > 0$

where P(A|B) is read as "the probability of A given B," or "A conditional on B," or "A conditioned on B."

Example: imagine an experiment is repeated N times where N is a very large number.

There are two outcomes of interest, A and B, and their frequencies are given by the following table:

	A	$\bar{A}$
B	n_{11}	n_{12}
$\bar{B}$	n_{21}	n_{22}

So $(A \cap B) = n_{11}$ and $P(A \cap B) = \frac{n_{11}}{N}$, etc.

By inspection, $P(A) = \frac{n_{11} + n_{21}}{N}, P(B) = \frac{n_{11} + n_{12}}{N}$.

What is *P(A|B)*?

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{n_{11} / N}{(n_{11} + n_{12}) / N} = \frac{n_{11}}{n_{11} + n_{12}}$$

As another example, given a balanced die, what is the probability of rolling a 1 given that you roll an odd number?

$$P(1 \mid odd) = \frac{P(1 \cap odd)}{P(odd)} = \frac{1/6}{1/2} = \frac{1}{3}$$

[Note: why 1/6? Since "odd" contains the event "roll a 1," their intersection is equal to the probability of rolling a 1:

$A \cap B = A$

What if the probability of event A is not affected by the occurrence of event B:

P(Kerry Wins | New York's average annual rainfall = 7")

In this case, the events are said to be *independent*. More formally:

Two events, A and B, are independent if:

$$P(A \mid B) = A \quad \text{or}$$
$$P(B \mid A) = B \quad \text{or}$$
$$P(A \cap B) = P(A)P(B)$$

Note that if any one of these conditions hold the others will hold as well.

So, showing that one holds (and it is sometimes easier to show one or another) is sufficient proof of independence.

Do not confuse independence with mutually exclusive. The latter means that two things cannot occur simultaneously, implying that outcomes may be related.

Multiplicative Law

We can generalize the third condition for independence with the multiplicative law of probability.

Again given two events A and B,

$$\begin{aligned} P(A \cap B) &= P(A)P(B \mid A) \\ &= P(B)P(A \mid B) \end{aligned}$$

where $P(A \cap B)_{independent} = P(A)P(B)$ is a special case of this law.

Note that this law generalizes further to *k* events:

$$P(A_1 \cap A_2 \cap A_3 \cap ... \cap A_k) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2)...P(A_k \mid A_1 \cap A_2 \cap A_3 \cap ... \cap A_{k-1})$$

We also need laws for the probability of the union of events that are not necessarily disjoint (thus a generalization of the third axiom):

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Proof:

Draw this out as you go…

Assume S={A,B}

By inspection,

$$A \cup B = A \cup (\overline{A} \cap B)$$
$$B = (\overline{A} \cap B) \cup (A \cap B)$$

By Axiom 3,

$$P(A \cup B) = P(A) + P(\overline{A} \cap B)$$
$$P(B) = P(\overline{A} \cap B) + P(A \cap B)$$

Rearranging the second equation:
$$P(\overline{A} \cap B) = P(B) - P(A \cap B)$$

Substitute this into the first probability statement above:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Q.E.D.

This rule can also be extended to more than 2 events. In the case of 3, for example,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Finally, before we move onto some examples, recall also that the 2nd axiom allows us to say:

$$P(A) = 1 - P(\overline{A})$$

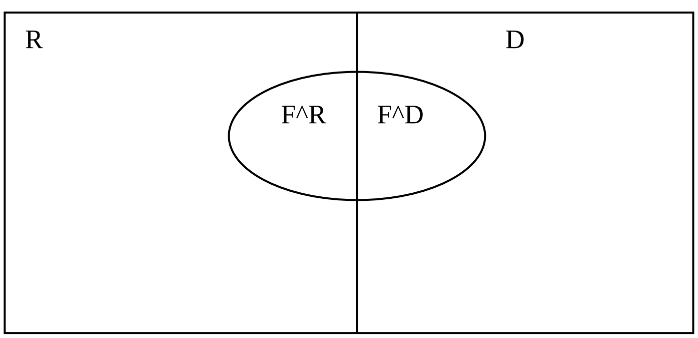
Example (2.17): In a given city, 40% of the voters are Republicans and 60% Dems.

Of the R's, 70% favor a bond issue. Of the D's, 80% are in favor of the issue.

If a voter is selected at random, what is the probability that he favors the bond issue?

Solution: Let F be the event that the issue is favored.

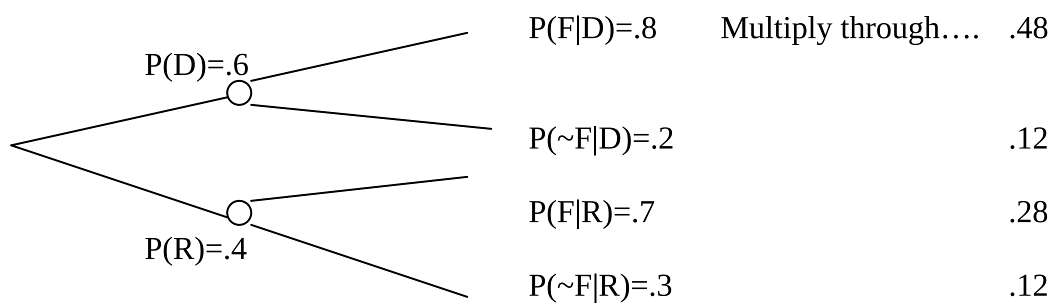
The sample space looks like:



P(R) = .4, P(D) = .6, P(F|R)=.7, P(F|D)=.8

$$P(F) = P(F \cap R) + P(F \cap D)$$
$$P(F) = P(F \mid R)P(R) + P(F \mid D)P(D)$$
$$P(F) = (.7)(.4) + (.8)(.6) = .76$$

This can also be demonstrated using a tree diagram.



And add .48+.28=.76

Example: A man has n keys on a key ring, one of which opens the door to his apartment. Having celebrated a bit too much one evening, he returns home only to find himself unable to distinguish one key from another. Resourceful, he decides to choose a key at random and try it. If it fails to open the door, he will discard it and choose at random one of the remaining n-1 keys, and so on. Clearly, the probability that he gains entrance with the first key he selects is 1/n. Show that the probability the door opens with the *third* key he tries is also 1/n. (Hint: what has to happen before he gets to the third key?)

Let K_i be the event that the ith key tried opens the door, i=1,2,…,n. Then P(door opens first time with 3rd key) =

$$P(\overline{K}_1 \cap \overline{K}_2 \cap K_3) = P(\overline{K}_1) * P(\overline{K}_2 \mid \overline{K}_1) * P(K_3 \mid \overline{K}_1 \cap \overline{K}_2)$$

What is the probability that the first key *doesn't* open the door?

$$1 - \frac{1}{n} = \frac{n}{n} - \frac{1}{n} = \frac{n-1}{n}$$

The probability that the second key opens the door, given that the first didn't is $\frac{1}{n-1}$.

The complement of that is:

$$1 - \frac{1}{n-1} = \frac{n-1}{n-1} - \frac{1}{n-1} = \frac{n-1-1}{n-1} = \frac{n-2}{n-1}$$

And, p that the third key opens the door given that the first two didn't is $\frac{1}{n-2}$.

So, $P(\overline{K}_1 \cap \overline{K}_2 \cap K_3) = P(\overline{K}_1) * P(\overline{K}_2 \mid \overline{K}_1) * P(K_3 \mid \overline{K}_1 \cap \overline{K}_2)$

$$= \quad \frac{n-1}{n} * \frac{n-2}{n-1} * \frac{1}{n-2} = \frac{1}{n}$$

Event Composition Approach

The text generalizes the procedure used to solve this example into what it calls the *event composition approach*, which differs from the sample point approach in that it is not necessary to count all sample points.

1. Define the experiment

2. Visualize sample points

3. Write an equation for the event of interest as a function of known quantities.

4. Apply the multiplicative and additive laws to find the probability of interest.

Example (2.20): The probability of a patient responding to treatment is .9. If 3 patients are treated independently, what is the probability that at least one will respond?
Solution: Let A be the event of interest (at least one responds) and B_n be the event that patient *n* does not respond.

$$\begin{aligned}
P(A) &= 1 - P(\overline{A}) \\
P(A) &= 1 - P(B_1 \cap B_2 \cap B_3) \\
P(A) &= 1 - P(B_1)P(B_2 \mid B_1)P(B_3 \mid B_1 \cap B_2) \\
P(A) &= 1 - P(B_1)P(B_2)P(B_3) \\
P(A) &= 1 - .1^3 = .999
\end{aligned}$$

We can go from line 3 to line 4 because the patients are independent from each other.

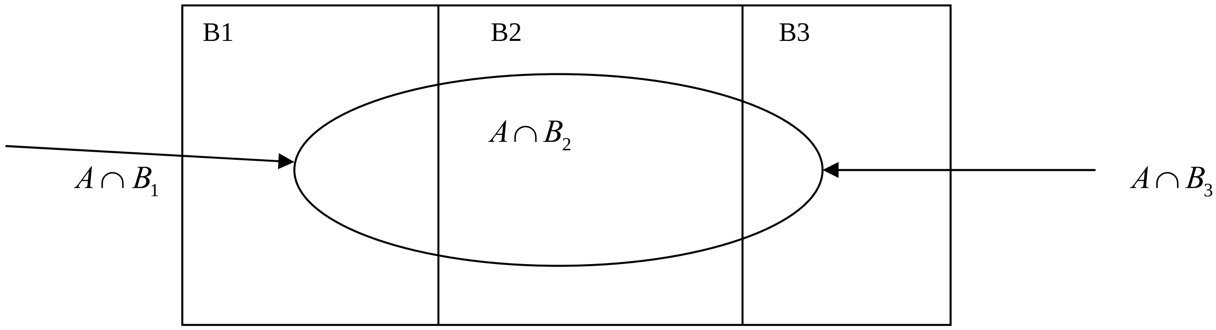
Law of Total Probability

Sometimes the ECA is easier if S is thought of as a union of disjoint subsets. The law of total probability allows for this partitioning:

For a positive integer *k*, let B_1, B_2, …,B_k be defined so that:

1. $S = \bigcup_{n=1}^{k} B_n$
2. $B_i \cap B_j = \varnothing \qquad \forall i \neq j$

Thus { B_1, B_2, …,B_k } forms a partition of S. An event A can thus be decomposed (k=3):



Generally,

$$A = (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_k)$$

(So long at condition 2 above holds.)

This allows for the following result:
if {B_1, B_2, …,B_k} is a partition of S such that $P(B_i)>0$, then

$$P(A) = \sum_{i=1}^{k} P(A \mid B_i)P(B_i)$$

[proof given in text]

Bayes' Rule

The law of total probability allows us to state Bayes' Rule:

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$$

Proof: By the definition of conditional probability:

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)}$$

Rewriting the numerator using the multiplicative law:

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{P(A)}$$

[note this is a common form of presentation for Bayes Rule]

Then, by the law of total probability, the denominator is rewritten:

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i)P(B_i)}$$

Or, in its simplest form: $P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}$

Since P(A) = $P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})$

We can expand this to: $P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})}$

Bayes' Rule allows us to formalize the idea of "updating" or revising a belief about a probability given new evidence. In the language of Bayesian statistics, P(B_j) is called the prior probability and P(B_j|A) is a posterior probability.

Example: Suppose that a given medical test is 90% reliable (i.e. if you have the disease there is .9 chance of a proper positive response and if you don't have the disease there is a .1 chance you will have a false positive). Data from a large sample yields that 1/10,000 in the population actually have the disease [prior].

If you test positive, what is the probability that you have the disease [posterior]?
Solution: if A = you test positive; B= you have the disease

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})}$$
$$P(B \mid A) = \frac{.9(.0001)}{.9(.0001) + .1(.9999)} = .0009 = \frac{9}{10000}$$

Another way to think about this is 1/10000 people have the disease, but the test gives a false positive result for 1/10 people who take it. So, the number of positive results is about 1000 times greater that the number of people with the disease.

Example II

Recently the U.S. Senate Committee on Labor and Public Welfare investigated the feasibility of setting up a national screening program to detect child abuse. A team of consultants estimated the following probabilities: (1) 1 child in 90 is abused, (2) a physician can detect an abused child 90% of the time, and (3) a screening program would incorrectly label 3% of all nonabused children as abused. What is the probability that a child is actually abused given that the screening program diagnoses him as such? How does this probability change if the incidence of abuse is 1 in 1000? 1 in 50?

Let *B* denote the event that the program diagnoses the child as abused, and let A be the event that the child *is* abused. Then *P(A)* = 1/90, *P(B|A)*=0.90, and $P(B\mid\bar{A})$ =0.03, so

$$P(A\mid B)=\frac{(0.90)(1/90)}{(0.90)(1/90)+(0.03)(89/90)}=0.25$$

If *P(A)*=1/1000, *P(A|B)*=0.029; if *P(A)*=1/50, *P(A|B)*=0.38.

Two final definitions that we can now state with more precision:
1. A *random variable* is a real-valued function for which the domain is a sample space. If y is an observed value of random variable Y, then P(Y=y) is the sum of the probabilities of the sample points assigned value y.

2. A *random sample* is a sample selected such that each of the $\binom{N}{n}$ possible samples has an equal probability of selection.

Practice Homework Problems: 2.57, 2.58, 2.70, 2.77, 2.78, 2.96, 2.101, 2.103, 2.109, 2.110, 2.112, 2.114.