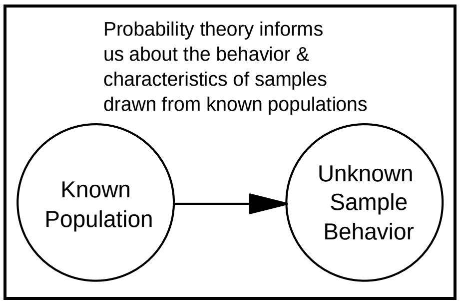
Lecture: PROBABILITY THEORY

1 Introduction

Now, we will begin our study of probability theory, the second major component of statistical analysis. As I mentioned at the start of the course, statistical analysis often has to concern itself with uncovering information about an unknown population based on information gleaned from a **sample.** Since, by definition, a sample does not provide a complete and comprehensive view of the population under study, we try to use information about the sample to *reduce the uncertainty surrounding our understanding of the unknown population.* Probability theory is absolutely fundamental in this enterprise. It provides a set of conclusions about randomness and the unknown world that will allow us to reduce our uncertainty about the population in question.



That is, we will be attempting to provide answers to questions of the following type:

<Suppose we have a die which is perfectly balanced; how often can we expect an ace to turn up in repeated throws (samples)?>

Or,

<Suppose we know that the weight of all women aged 20 in the U.S. averages 123 lbs and is distributed normally with 15 lbs; how often can we expect the sample mean of a random sample of size 20 to be within 15 lbs of the population mean?>

Note, that in none of the following development of probability theory will I refer to a given sample and ask questions about the population from which it comes; that is a question of **statistical inference** to which we will turn after we're done with probability theory. What we're asking now are guestions about the features of samples drawn from *known populations*. Later, we'll use our accumulated information developed in this part of the course to turn around and ask questions of *unknown populations*.

2 Defining Probability

In addition to the many formal applications of probability theory, the concept of probability enters our everyday life and conversation. We often hear and use such expressions as: "It probably will rain tomorrow afternoon"; "It is very likely that the plane will arrive late"; or "The chances are good that he will be able to join us for dinner this evening." Each of these expressions is based on the concept of the probability, or the likelihood, that some specific event will occur.

Despite the fact that the concept of probability is such a common and natural part of our experience, no single scientific interpretation of the term probability is accepted by all statisticians, philosophers, and other authorities. Through the years, each interpretation of probability that has been proposed by some authorities has been criticized by others. Indeed, the true meaning of probability is still a highly controversial subject and is involved in many current philosophical discussions pertaining to the foundations of statistics. Two different interpretations of probability will be described here. Each of these interpretations can be very useful in applying probability theory to practical problems.

2.1. The Frequency Interpretation of Probability

In many problems, the probability that some specific outcome of a process will be obtained can be interpreted to mean the *relative frequency* with which that outcome would be obtained if the process were repeated a large number of times under similar conditions. For example, the probability of obtaining a head when a coin is tossed is considered to be ½ because the relative frequency of heads should be approximately ½ when the coin is tossed a large number of times under similar conditions. In other words, it is assumed that the proportion of tosses on which a head is obtained would be approximately ½.

Of course, the conditions mentioned in this example are too vague to serve as the basis for a scientific definition of probability. First, a "large number" of tosses of the coin is specified, but there is no definite indication of an actual number that would be considered large enough. Second, it is stated that the coin should be tossed each time "under similar conditions," but these conditions are not described precisely. The conditions under which the coin is tossed must not be completely identical for each toss because the outcomes would then be the same, and there would be either all heads or all tails. In fact, a skilled person can toss a coin into the air repeatedly and catch it in such a way that a head is obtained on almost every toss. Hence, the tosses must not be completely controlled but must have some "random" features.

Furthermore, it is stated that the relative frequency of heads should be "approximately ½," but no limit is specified for the permissible variation from ½. If a coin were tossed 1,000,000 times, we would not expect to obtain exactly 500,000 heads. Indeed, we would be extremely surprised if we obtained exactly 500,000 heads. On the other hand, neither would we expect the number of heads to be very far from 500,000. It would be desirable to be able to make a precise statement of the likelihoods of the different possible numbers of heads, but these likelihoods would of necessity depend on the very concept of probability that we are trying to define.

Another shortcoming of the frequency interpretation of probability is that it applies only to a problem in which there can be, at least in principle, a large number of similar repetitions of a certain process. Many important problems are not of this type. For example, the frequency interpretation of probability cannot be applied directly to the probability that a specific acquaintance will get married within the next two years or to the probability that a particular medical research project will lead to the development of a new treatment for a certain disease within a specified period of time.

According to the subjective, or personal, interpretation of probability, the probability that a person assigns to a possible outcome of some process represents her own judgment of the likelihood that the outcome will be obtained. This judgment will be based on each person's beliefs and information about the process. Another person, who may have different beliefs or different information, may assign a different probability to the same outcome. For this reason, it is appropriate to speak of a certain person's *subjective probability* of an outcome, rather than to speak of the *true probability* of that outcome.

2.2. The Subjective Interpretation of Probability

As an illustration of this interpretation, suppose that a coin is to be tossed once. A person with no special information about the coin or the way in which it is tossed might regard a head and a tail to be equally likely outcomes. That person would then assign a subjective probability of ½ to the possibility of obtaining a head. The person who is actually tossing the coin, however, might feel that a head is much more likely to be obtained than a tail. In order that people in general may be able to assign subjective probabilities to the outcomes, they must express the strength of their belief in numerical terms. Suppose, for example, that they regard the likelihood of obtaining a head to be the same as the likelihood of obtaining a red card when one card is chosen from a wellshuffled deck containing four red cards and one black card. Because those people would assign a probability of 4/5 to the possibility of obtaining a red card, they should also assign a probability of 4/5 to the possibility of obtaining a head when the coin is tossed.

This subjective interpretation of probability can be formalized. In general, if people's judgments of the relative likelihoods of various combinations of outcomes satisfy certain conditions of consistency, then it can be shown that their subjective probabilities of the different possible events can be uniquely determined.

Difficulties: However, there are two difficulties with the subjective interpretation. First, the requirement that a person's judgments of the relative likelihoods of an infinite number of events be completely consistent and free from contradictions does not seem to be humanly attainable. Second, the subjective interpretation provides no "objective" basis for two or more scientists working together to reach a common evaluation of the state of knowledge in some scientific area of common interest.

Benefits: On the other hand, recognition of the subjective interpretation of probability has the salutary effect of emphasizing some of the subjective aspects of science. A particular scientist's evaluation of the probability of some uncertain outcome must ultimately be that person's own evaluation based on all the evidence available. This evaluation may well be based in part on the frequency interpretation of probability, since the scientist may take into account the relative frequency of occurrence of this outcome or similar outcomes in the past. It may also be based in part on the classical interpretation of probability, since the scientist may take into account the total number of possible outcomes that are considered equally likely to occur. Nevertheless, the final assignment of numerical probabilities is the responsibility of the scientist herself.

The subjective nature of science is also revealed in the actual problem that a particular scientist chooses to study from the class of problems that might have been chosen, in the experiments that are selected in carrying out this study, and in the conclusions drawn from the experimental data. The mathematical theory of probability and statistics can play an important part in these choices, decisions, and conclusions.

NOTE: The Theory of Probability Does Not Depend on Interpretation. The mathematical theory of probability we will develop is developed without regard to the controversy surrounding the different interpretations of the term probability. This theory is correct and can be usefully applied, regardless of which interpretation of probability is used in a particular problem. For most of what we'll do, the difference between subjective and objective probability concept does not make a difference in our analysis of a problem. The book takes a frequentist approach to the problem.

2.3. Some Differences between the Frequentist and Subjective Views

How about the following **frequentist** definition: Assume that we're betting *against* rolling a one when throwing a die. Now, intuitively we expect each of the six numbers on a die to be equally probable, Right?

What therefore is the probability of rolling an ace?> 1/6, *provided it is an honest die*. If we're not sure it's an honest die, the probability of getting an ace might be different.

In fact, if we rolled that die many times we would expect to get a better and better idea of whether that die is honest or not, right?

Let's say that there are six possible **events** each time we roll the die, *el, e2, e3, e4, e5, e6*, Then *after many throws* we expect the probability of throwing an ace *(e1)* to be

Frequentist Definition of Probability:

$\Pr(e_1) = \lim_{n \to \infty} \left(\frac{n_1}{n}\right)$ where n is the number of tosses and n_1 is the number of aces (ones).

In other words as the number of throws gets very large, the ratio of number of aces (n_1) to the number of throws n (the relative frequency of aces) is the probability of e_1.

Seems reasonable doesn't it? We'll just define the probability of an event e_1 as the long-run relative frequency of that event in repeated trials.

Note something important here: Under this definition of probability *only a random variable can have a probability and an associated probability distribution.* For example, the *population* mean weight of all 20 year old women is not a random variable. It exists and does not vary at a given point in time. Sample means will vary all over the place because they're functions of random drawings made out of a very large population.

Consequently, under a frequentist notion of probability, only sample means, standard deviations, etc. can have probability distributions. Population means, standard deviations, variances, etc do not; either the population mean is a given value (Pr = 1) or it is not (Pr = 0). <Does this seem reasonable to you?>

Subjectivists believe something different: Probability generally is subjective; it represents our opinions about unknown events whether we have new data or not. If you believe probability is subjective, then you see noting wrong in saying something like, "The probability that the average weight of 20 -year old females is between 115 and 125 lbs is 0.65." You are constructing your own probability distribution around the population mean; frequentists would say that this is incorrect -- you cannot associate probability with a population parameter, only with a chance event, like a sample mean. This is one element in the current hot debate between two schools of statisticians, the Bayesians (challengers, wearing black trunks) and the Classicists (champion in in white).

Mathemeticians take another approach completely: They make some basic and reasonable assumptions *about how a probability measure should behave and then they use these assumptions (postulates, axioms) to deduce* *many interesting characteristics of samples drawn from known populations.* We'll sketch how this is done in a minute, but first some terminology from set theory.

3 The Probability Model

3.1. The Experiment and Outcomes of Experiment

The Experiment: when we talk about an experiment we mean any situation capable of replication under essentially stable circumstances e.g., flipping a coin, timing the length of time it takes a rat to travel a maze, measuring the a speed of a pitcher's fast ball, measuring the failure rate of a space craft electronic component.>

Outcomes of an experiment: There may be a finite number of outcomes to our experiment. E.g. in flipping a coin there are only 2 outcomes - Rolling one die there are 6 outcomes

< How many possible outcomes when 2 dice are rolled?> 36(=6x6)

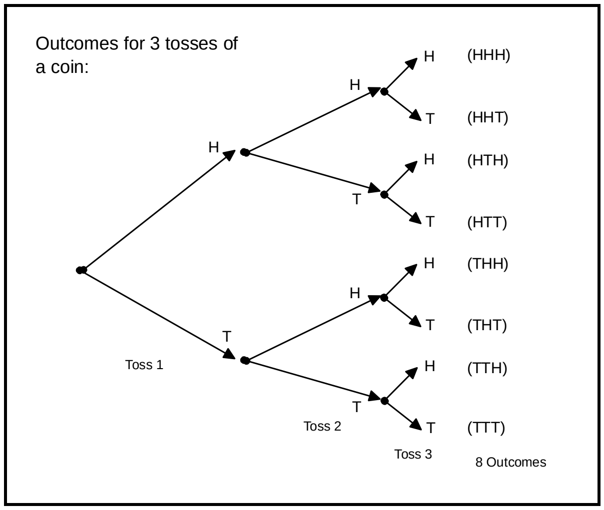
<When 3 dice are rolled?> (6x6x6=) 216

The number of outcomes to an experiment may be infinite:<In timing a rat's trip through a maze how many possible outcomes?> Potentially an infinite number, depending on how closely *time* is measured."

3.2. Determining Outcomes: Counting, Permutations & Combinations

Often it is very difficult to determine just what all the possible outcomes of an experiment are. For example, suppose we toss a coin three times: <How many different possible outcomes are there to this experiment?>

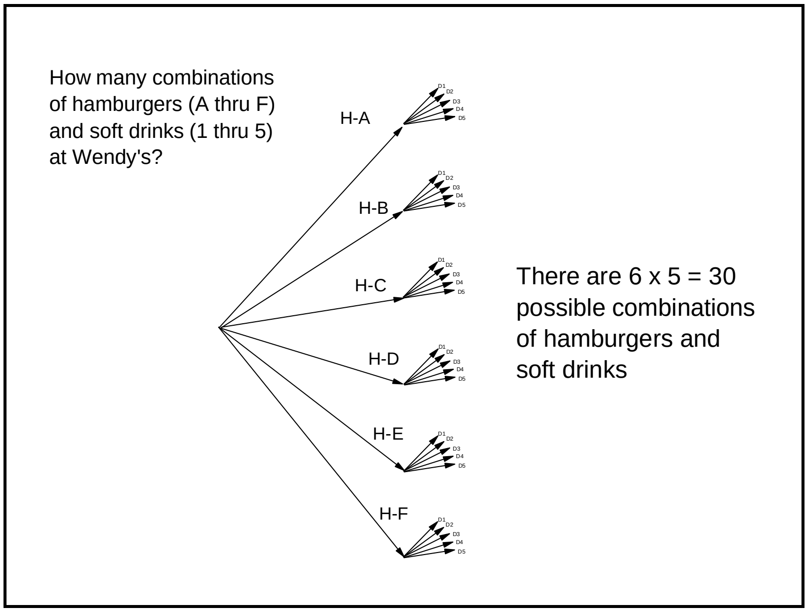
Perhaps the easiest way to figure this out is with a tree diagram **{Next Slide}** :



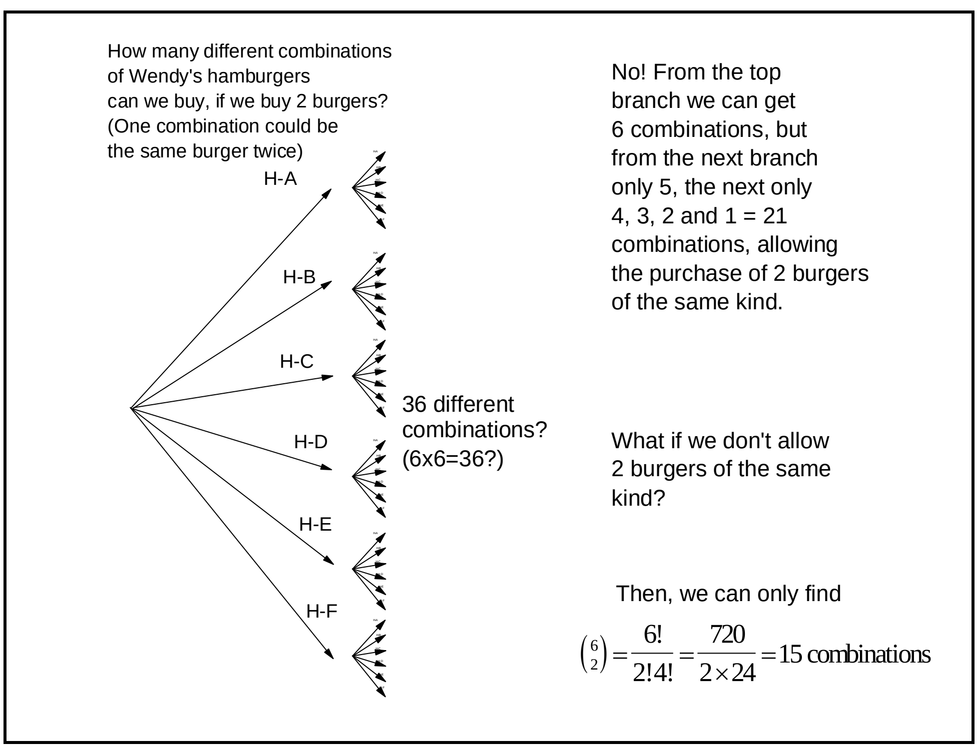
Mathematically, we have 2 outcomes for the first throw, 2 for second and 2 for the third ... consequently we have $2^3 = 8$ possible outcomes since either outcome can occur in second- and third tosses no matter what happens on the first (This is the **exponential counting rule**).

<Is each outcome equally likely?> only if coin is "fair"

<What if you were to go to Wendy's and order one of their 6 different types of hamburgers and one of 5 different drinks … assuming that an outcome is a *combination* of hamburger and drink, how many possible combinations are possible 6 x 5 = 30 outcomes **{Next Slide}**



<Finally, suppose out of the 6 different hamburgers at Wendy's, I wanted to buy two hamburgers. How many different combinations of hamburgers could I buy?> **{Next Slide}**



Now, let's present some computing formulas that will help us to determing the number of outcomes in various experiments **{Next Slide}** :

Number of Permutations of *n* things taken *r* at a time

$$_nP_r = \frac{n!}{(n-r)!} = n \times (n-1) \times (n-2) \times \ldots \times (n-r+1)$$

Number of Combinations of *n* things taken *r* at a time

$$_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{_nP_r}{r!}$$

<Suppose there are 21 candidates for the student legislature in the upcoming elections with 5 to be elected; how many possible outcomes to the election (assume the top 5 vote getters are elected)>

21 x 20 x, 19 x 18 x 17 = 2,441,880 possible, outcomes , ordered quintuples, $_{21}P_r$

<Assume further that the Blue Sky party is running 5 candidates. How many of these outcomes are favorable to the event that all of the blue-skyers are -elected? That is, how many different ways can these 5 candidates arrange themselves?)> 120 (= 5!)

<If each outcome is equally likely, what is the probability that all 5 blue-skyers will be elected?> 120/2,441,880 = 0:0000491 = $\frac{5!}{_nP_r}$

In this example, what we're about to call the **sample space** contains 2,441,880 outcomes of which only 120 are favorable to the event "clean-sweep by B1ue-Sky "

Note, that often we're not interested strictly in different outcomes: for example if we're interested in a clean sweep by the blue-sky party, we are interested only in those outcomes in which all those elected are members of blue-sky…

But this is a special case of the general question, "How many different ways can we choose a group -of five persons WITHOUT REGARD T0 THE ORDER IN WHICH THE TOP FIVE FINISH? In our example, the answer is,

$${}_nC_r = \binom{n}{r} = \frac{21!}{5!(21-5)!} = 20{,}349 \text{ distinct combinations}$$

of 5 candidates only one of which contains all members of Blue-Sky

$\frac{1}{20349} = 0.0000491$ as before.

What we've been doing is undertaking a brief review of **permutations** and **combinations** as a way of determining the number of possible outcomes to a given experiment.

3.3. Sample Space

Sample space: The set of all possible outcomes of an experiment.

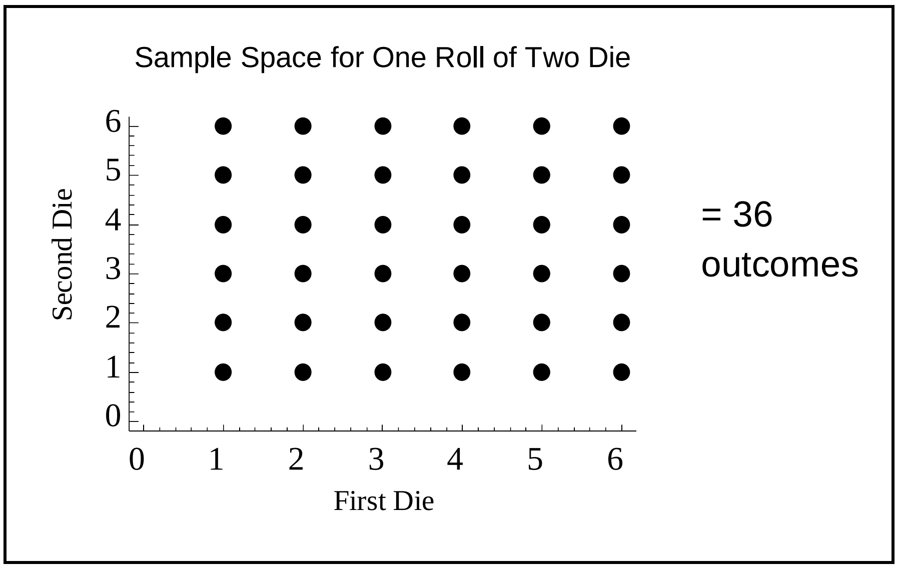
A sample space is **discrete** if all possible outcomes can be identified and counted Eg, the sample space for a coin flip contains the following possible outcomes: {H;T} if we flip the coin twice the set of all possible outcomes is {H,H;   H,T;   T,H;   T,T}

<What if we flip a coin 3 times…what is the set of all possible outcomes?> 8 outcomes:

{H,H,H;   H,H,T;   H,T,T;   T,T,T;   T,H,T;   TH,H;   T,T,H; H,T,H}

<What is the sample space for one roll of a die?> { 1,2,3,4,5,6}

<What is the sample space for one roll of two dice?> **{Next Slide}**



A **continuous sample space** exists if the number of possible outcomes is infinite and uncountable.

Eg: the exact weight of a 20 year old college student, the 100 yard dash time of a runner.

Whether or not a sample space is continuous or discrete often depends on the level of measurement of the outcome; e.g., if we measure weight to the nearest pound, then our sample space is discrete.

Often when working with continuous sample spaces we may want to group our measures in order to produce a discrete sample space.

In defining a sample space of an experiment, *we must be sure that it is not possible for two or more outcomes in the same replication* of the experiment. If one and only one am outcome is possible in each experiment then the outcomes are mutually exclusive.

We can call each mutually exclusive outcome an **elementary (or simple) event** <this definition is not in the book, but remember it anyway>...

Sample space contains all elementary events (or all possible mutually exclusive outcomes)

The set of elementary events must be exhaustive also, that is it must contain all possible outcomes.Thus, the set of elementary events, *s*, that comprises a sample space is both exhaustive and mutually exclusive.

3.4. Random variables:

Given a sample space and the set of elementary events, we often want to determine the probability of one of the elementary events. Or, we may want to consider the probability of *one or more* of these elementary events:

For example, in matching pennies we may want to determine the probability of having either both heads or both tails. There are 4 outcomes (elementary events) to the experiment only two of which meet our criterion for a success.

Likewise, if we want to determine the probability that one roll of a die will give a number less than three, we must include rolls of *1* and *2*. Therefore it is useful to have a standard way to define the relevant outcomes (elementary events) of any given experiment (sample space).

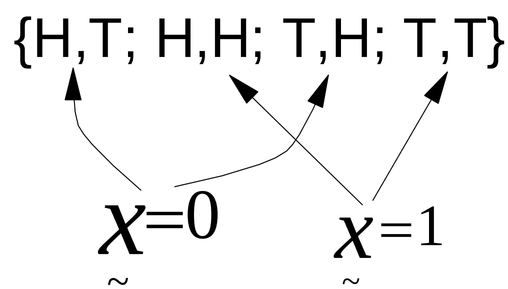
3.4.1 A Random variable

A random variable is a well-defined rule for making the assignment of a numerical value to any outcome of the experiment. Example: To define the outcomes of a simple coin toss we define the Random Variable < $\underset{\sim}{x}$ (Notation)>**{Next Slide}**

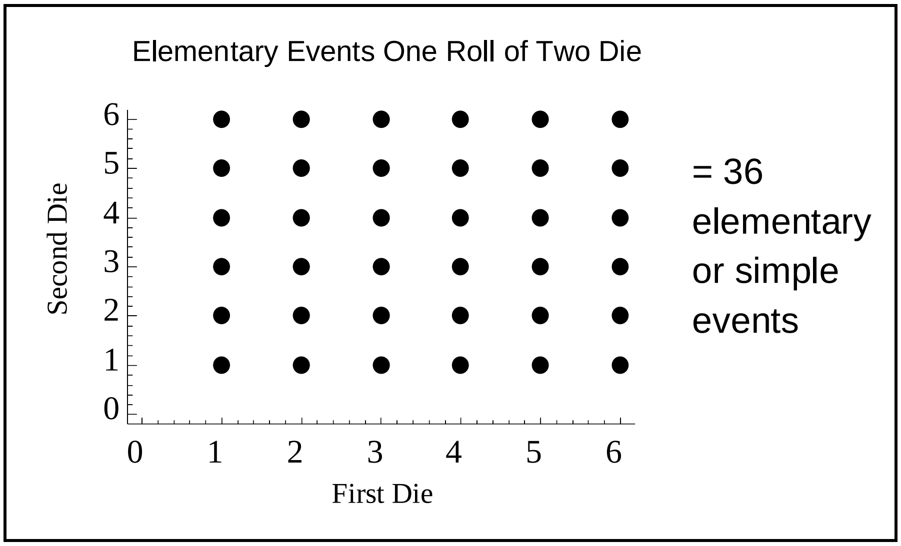
$$\underset{\sim}{x} = \begin{cases} 1 \text{ if outcome = H} \\ 0 \text{ if outcome = T} \end{cases}$$

Here's another example: Matching pennies **{Next Slide}**

$$\underset{\sim}{x} = \begin{cases} 1 \text{ if there is a match} \\ 0 \text{ if there is no match} \end{cases}$$ **{Next Slide}**



What about rolling 2 dice? The elementary events are: **{Next Slide}**



What is the probability of any one of these elementary events? ***1/36.***

Now, suppose that we want to define a Random Variable so that it takes on the values of the sum of the two dice.

Define $\underset{\sim}{x}$= Roll#1 + Roll#2

What are the possible values that our Random Variable can take? How many elementary events (outcomes) go with each value of *x*?

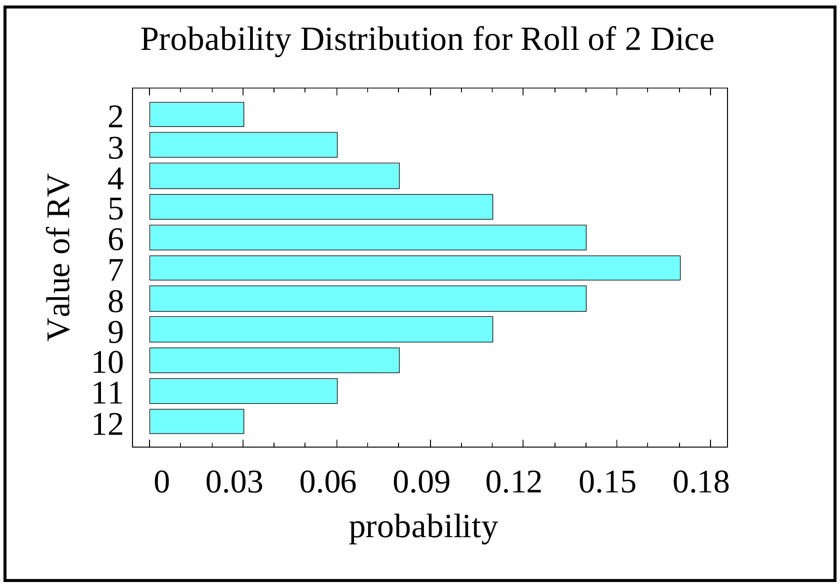
$\underset{\sim}{x}$	Number of Elementary events
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

Now, intuitively what is the probability

$\underset{\sim}{x}$= ?	Probability
1	0
2	1/36
7	6/36

etc.

A **probability distribution of a Random Variable** gives, for each value of the R.V. the probability that a *given value* will be obtained - for a discrete distribution, the probability of each value of *x* on our dice rolling experiment looks like **{Next Slide}**



I'll define **probability** more precisely in a moment.

To Repeat: To construct a probability model we undertake the following steps:

1. **Define the experiment** and recognize its set of *mutually* exclusive and exhaustive outcomes (the **sample space**)
2. Define a **Random Variable** which assigns a number to EACH outcome in the sample space.
3. Determine the probability of each value of the R.V. to obtain a **probability distribution.**

3.5. Events & set notation:

The **sample space** is the set of all possible outcomes of an experiment. <What do you think is the probability of the sample space?> **One!!**

An **Elementary Event** is another name for each possible outcome in the sample space. In a very real sense, our elementary events are the "atoms" of the theory of statistics, or of **probability theory** as it is also called.

<What is the probability of each elementary event in our dice rolling experiment.> **1/36**

Therefore, more rigorously we can define a **sample space** as a set of *elementary events* that are *mutually exclusive* and *exhaustive*.

Definition of **probability:**

We define the **probability** of an elementary event as a number between zero and one such that the sum of the probabilities over the sample space is one. To each elementary event, e_i, we assign a number between zero and one, call it p_i. We can write this as: **{Next Slide}**

$$\begin{array}{llllll} S = \{ e_1, & e_2, & \ldots & \ldots, & e_k \} & \\ \quad\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \text{for a sample space having } k \text{ elementary events,} \\ \quad\ p_1, & p_2, & \ldots & \ldots, & p_k \} & \end{array}$$

or, we can write this as: **{Next Slide}**

$$e_1 \to pr(e_1) = p_1$$
$$e_2 \to pr(e_2) = p_2$$
$$e_3 \to pr(e_3) = p_3$$
$$\cdots$$
$$e_k \to pr(e_k) = p_k$$

Where the expression pr(e_2) means “assign a specific number between zero and one to the elementary event that is designated in the argument,” e_2 in this case; and the value given by that assignment is p_2 . The notation reinforces the idea that we are assigning a number to each elementary event.

Therefore, a **probability distribution**, as its name implies is a statement of how probabilities are distributed over the elementary events in the sample space.

<In any experiment must the probability of each elementary event always be equal to the probability of all other elementary events?> *No,* take for example an unfair coin that has

P(HEAD) = .7

P(TAIL) = . 3

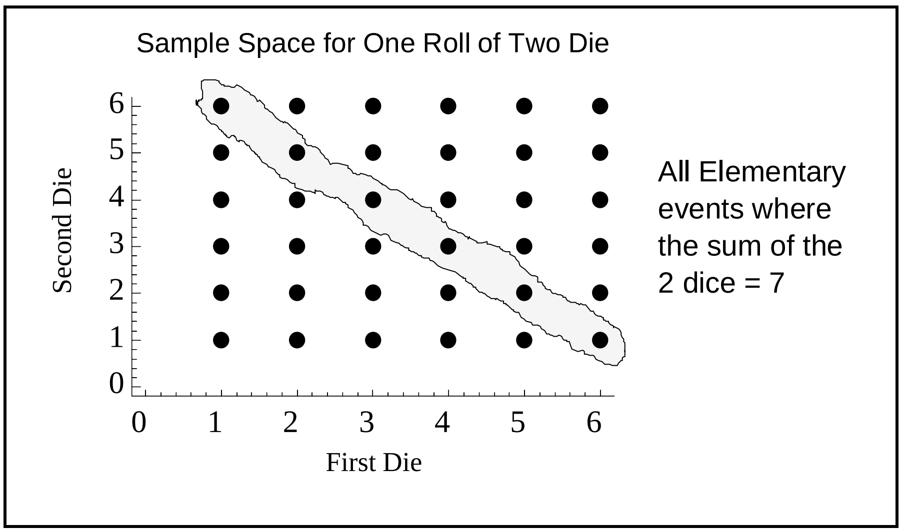
3.5.1 Compound events

When we talk about the probability that *at least one of a collection of elementary events* occurs, we are talking about the probability of a **compound event.** When the compound event is made up of elementary events, then it turns out that we can add up the individual probabilities of the elementary events:

$$pr(e_i,\ \text{or}\ e_j,\ \text{or}\ e_k, \ldots,\ \text{or}\ e_m) = p_i + p_j + p_k + \cdots + p_m.$$

An **Event** is a subset of the sample space containing *zero or more of the elementary events.*

<What about the event { Sum of two dice =7}? that is the collection of elementary events along the diagonal of our model. **{Next Slide}**



Properties of our Probability Measure:

1. $P(\text{Sample Space}) = 1.$
2. $0 \leq P(\text{Any Event in S}) \leq 1.$

3.5.2 Set Notation And Venn diagrams

A sample space is a set of all possible Elementary Events {sample space elements}; however, when the number of elements in a set is too large to list, we define the sample space as a set with a rule for inclusion in the set.

For example, the toss of two coins leads to a sample space with four possible elementary events: {HT,HH, TH,TT} ; however, sometimes the sample space is just too large to list all the elementary events, so we resort to the following kinds of identification of set members:

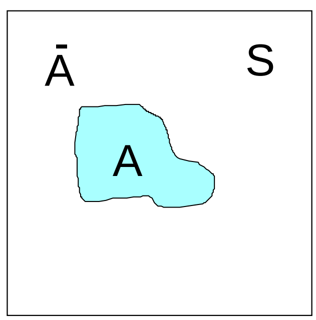
$C = \left\{ \underset{\sim}{x} \,|\, 0 \leq \underset{\sim}{x} \leq \infty \right\}$ represents all numbers between zero and infinity.

$D = \left\{ \underset{\sim}{x} \,|\, 0 \leq \underset{\sim}{x} \leq 300,000 \text{ and } \underset{\sim}{x} = \text{integer} \right\}$ represents all integers between 0 and 300,000.

The **complement** of an event is that subset of a sample space containing NO elements of the event

Given a sample space S: Let A = Event and $\bar{A} = \text{complement of the event}$

{Next Slide}

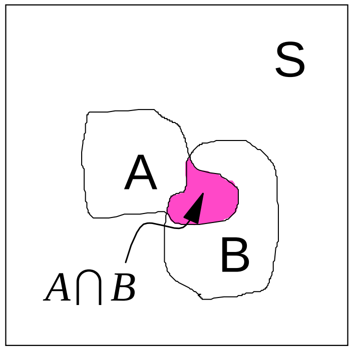


- The sample space, S, contains all the elementary events
- The Event A contains all elementary events corresponding to an inclusion rule.
- All other elementary events are in the complement A-bar.

$\bar{A} = S - A$ = {all sample points not in A}

3.5.1 Intersection of two events:

Consists of all outcomes common to two events. **{Next Slide}**



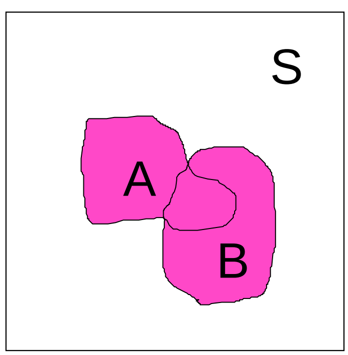
- The intersection consists of all outcomes common to two events.

The **union** of two events consists of all elementary events that are members of either event:

$$A \cup B = \{\underset{\sim}{x} | x \in A \; or \; x \in B\}$$

{Next Slide}

$$A \cup B = \{\underset{\sim}{x} | x \in A \; or \; x \in B\}$$

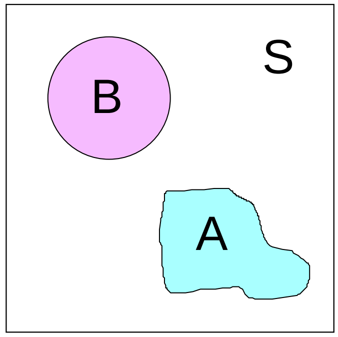


- The Union of two events A & B consists of all elementary events contained in <u>either</u> A or B.

3.6. The Rules of Probability

3.6.1 Probability of the Union of mutually exclusive events (Special addition Rule)
{Next Slide}

$$\Pr(A \cup B) = P(A) + P(B) = \Pr(A \text{ or } B)$$



A and B are two mutually exclusive events (no elementary events in common)

The Special Additon Rule: If event A and event B are mutually exclusive, then,

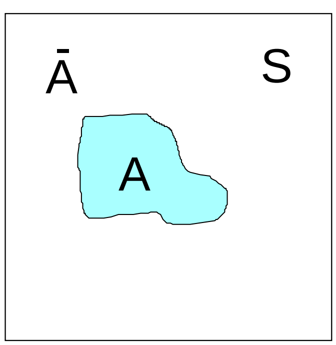
$$\Pr(A \cup B) = P(A) + P(B) = \Pr(A \text{ or } B)$$

or, more generally, if events A, B, C, ... are mutually exclusive then

$$\Pr(A \cup B \cup C \cup \ldots) = P(A) + P(B) + P(C) + \ldots = \Pr(A \text{ or } B \text{ or } C \text{ or } \ldots)$$

3.6.2 The Probability of the Union of Complementary Events: ***{Next Slide}***

$$\begin{aligned} \Pr\{A \cup \bar{A}\} &= \Pr(A) + \Pr(\bar{A}) \\ &= \Pr(A) + 1 - \Pr(A) \\ &= 1. \end{aligned}$$



- The Probability of the union of complementary events, A & A-bar is one!

The **Complementation Rule** states that for any event A, the probability of event A is 1 minus the probability of "not A":

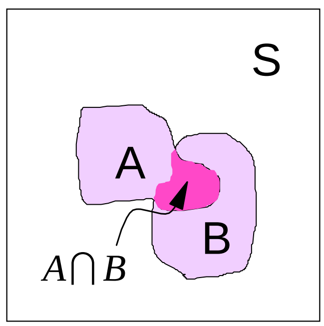
$$P(A) = 1 - P(\overline{A}).$$

3.6.3 The General Addition Rule

If A and B are any two events, then: **{Next Slide}**

$$\Pr(A \cup B) = P(A) + P(B) - P(A \cap B) = \Pr(A \text{ or } B)$$

$$\Pr(A \cup B) = P(A) + P(B) - P(A \cap B) = \Pr(A \text{ or } B)$$



The **General Addition Rule** requires the subtraction of the probability of Intersecting events.

or, more generally

$$\Pr(A \cup B \cup C \cup \ldots) = P(A) + P(B) + P(C) + \ldots - P(A \cap B \cap C \cap \ldots) = \Pr(A \text{ or } B \text{ or } C \text{ or } \ldots)$$

3.6.4 The Probability of the Intersection of Complementary Events

The intersection of two events consists of all those elementary events that are common to both events. Complementary events have no elementary events in common. Therefore,

$$\Pr\{A \cap \bar{A}\} = 0.$$

3.7. Contingency Tables & Conditional Probability

Suppose a drug company wants to test the effectiveness of a new anti-allergy drug they wish to put on the market to compete with, Claritin. They propose the following experiment:

1. First, sample 100 people with allergy problems. <How do we describe the population from which the sample is drawn?> The population of allergy sufferers.
2. Second, Randomly select 20 of the 100 to receive a placebo, the other 80 to receive the experimental drug.
3. Third, Administer the drug and check after 24 hours to see if the individual reacts after being exposed to ragweed pollen.

Here are the results expressed in a table: **{Next Slide}**

	Received the Drug (D)	Received the Placebo $(\overline{D})$	Total
Allergic Symptoms (A)	16	8	24
No Allergic Symptoms $(\overline{A})$	64	12	76
Total	80	20	100

The colored interior boxes are called **cells**. They show the number of observations exhibiting each combination of the two events. The **total** column shows the number of obsservations showing both types of allergic reaction, and the **total** row shows the number of observations receiving each kind of treatment.

The table itself is called a **contingency table** and it shows the frequency distribution for **bivariate data.** This kind of table can also be called a **two-way table.**

<Now, suppose we were to randomly select one person from our group:

1. <What is the probability that any one person will have been selected? > [1/100]

2. <What is the probability that the person selected will have received the drug *and* shown allergic symptoms?> Only 16 chances out of 100. This is the probability of the intersection of two events:

$$\therefore\ P(D \cap A) = \tfrac{16}{100} = 0.16.$$

3. <What is the probability that any person chosen will have received the drug.> 80 chances out of 100.

$$\therefore\ P(D) = \tfrac{80}{100} = 0.80.$$

4. <What is the probability that a randomly selected respondent will have exhibited allergic symptoms?>

$$\therefore\ P(A) = \tfrac{16+8}{100} = 0.24.$$

Let's create another table that shows, in each cell, the probability that a particular event occurs:

{Next Slide}

	Received the Drug (D)	Received the Placebo $(\overline{D})$	$P(A)$ & $P(\overline{A})$
Allergic Symptoms (A)	$P(A \cap D) = 0.16$	$P(A \cap \overline{D}) = 0.08$	$P(A) = 0.24$
No Allergic Symptoms $(\overline{A})$	$P(\overline{A} \cap D) = 0.64$	$P(\overline{A} \cap \overline{D}) = 0.12$	$P(\overline{A}) = 0.76$
$P(D)$ & $P(\overline{D})$	$P(D) = 0.80$	$P(\overline{D}) = 0.20$	1.00

{Next Slide}

	Received the Drug D	Received the Placebo $\overline{D}$	$P(A)\& P(\overline{A})$
Allergic Symptoms A	$P(A \cap D) = 0.16$	$P(A \cap \overline{D}) = 0.08$	$P(A) = 0.24$
No Allergic Symptoms $\overline{A}$	$P(\overline{A} \cap D) = 0.64$	$P(\overline{A} \cap \overline{D}) = 0.12$	$P(\overline{A}) = 0.76$
$P(D)\& P(\overline{D})$	$P(D) = 0.80$	$P(\overline{D}) = 0.20$	1.0

3.7.1 Joint & Marginal Probabilities

The probabilities listed in this table have names. In the cells, they are called **joint probabilities** and on the margins they are called **marginal probabilities. {Next Slide}**

Joint Probabilities

Marginal Probabilities

	Received the Drug D	Received the Placebo $\overline{D}$	$P(A)\& P(\overline{A})$
Allergic Symptoms A	$P(A \cap D) = 0.16$	$P(A \cap \overline{D}) = 0.08$	$P(A) = 0.24$
No Allergic Symptoms $\overline{A}$	$P(\overline{A} \cap D) = 0.64$	$P(\overline{A} \cap \overline{D}) = 0.12$	$P(\overline{A}) = 0.76$
$P(D)\& P(\overline{D})$	$P(D) = 0.80$	$P(\overline{D}) = 0.20$	1.0

Joint probabilites show the probability of two events happening together, while marginal probabilities show the overall probability of a single event happening, irrespective of what's happened with the other event.

We can use the joint and marginal probabilities to develop the notion of **conditional probability,** an extremely important type of probability.

<What if we chose randomly *someone who **we know** had received the drug*? What is the probability that the person would have exhibited allergic symptoms?>

$$P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{\frac{16}{100}}{\frac{80}{100}} = \frac{16}{80} = 0.20.$$

That is, *the probability that a person has an allergic reaction* ***conditional upon his already having received the drug*** *is equal to the quotient of the joint probability and the marginal probability of having received the drug.*

Therefore, the general formula for the **conditional probability** $P(A|D)$, can be written:

$$P(A|D) = \frac{P(D \cap A)}{P(D)}.$$

NB: The conditional probability involves the use of *additional information related to the second event;* consequently, the conditional probability of an event will often be higher than the event's unconditional probability, assuming the two events *A* and *D* are related to each other.

Let's state more formally the **conditional probability rule: {Next Slide}**

> **The Conditional Probability Rule:**
> If A and B are any two events, then
>
> $$P(B|A) = \frac{P(A \& B)}{P(A)}.$$
>
> In words, for any two events the conditional probability that one event occurs given that the other even has occurred equals the joint probability of the two events divided by the probability of the given event.

Notice that I changed the notation to be consistent with other rules.

<Now, what is the probability of showing allergic symptoms *if the respondent did not get the new drug?*>

$$P\left(A|\overline{D}\right) = \frac{P(\overline{D} \cap A)}{P(\overline{D})}.$$

But since D and $\overline{D}$ are complements we know that $P\left(\overline{D}\right) = 1 - P(D) = 0.20.$

Our experimental results showed that $P(\overline{D} \cap A) = \frac{8}{100} = 0.08.$

$\therefore\ P\left(A|\overline{D}\right) = \frac{0.08}{0.2} = 0.4.$

<Is the drug effective?> Yes it appears so. The probability of having an allergic reaction is cut in half.

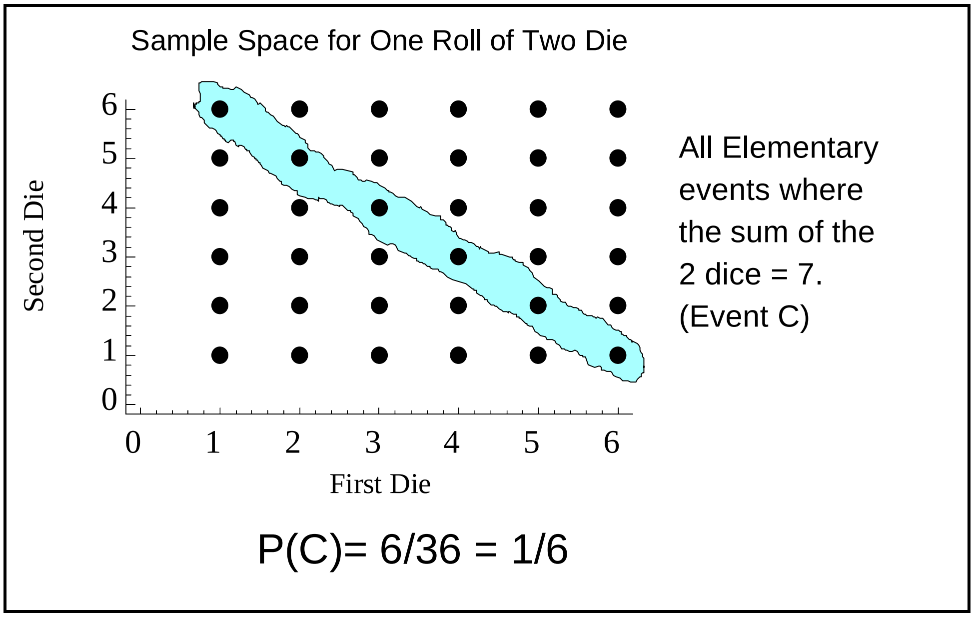
$$P\left(A|\overline{D}\right) = 0.4$$
$$P(A|D) = 0.2$$

The conditional probability shows us something very important: It shows us how the probability of an event is affected by the occurrence, or non-occurrence, of another event. You can see how important this kind of finding would be if you were testing a new drug, or trying to assess the effectiveness of an economic policy, etc.

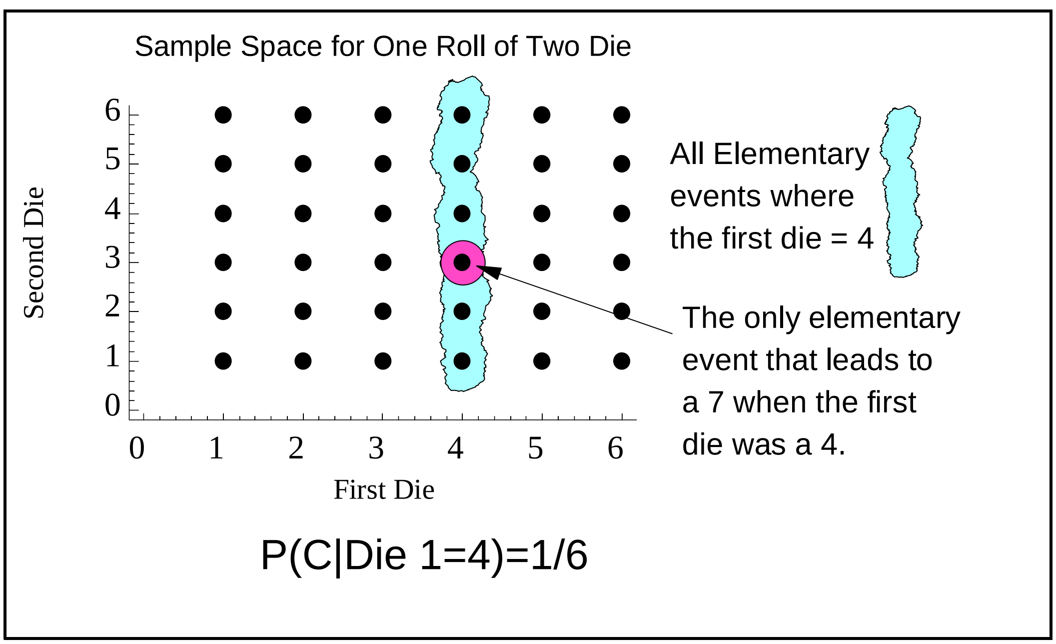
Now, let's try another experiment: Suppose we want to know the probability of rolling 2 dice and getting a 7:

Let $C =$ we got a 7
$\overline{C} =$ we did not get a 7

<What is *P(C)*?> ***P(C) = 6/36 = 1/6*** **{Next Slide}**



<What if we know that the first die was a 4?> What is the probability that we roll a 3 on the second die to get our 7?> **{Next Slide}**



It looks as though knowing that the first die was a "4" was no help in predicting the probability of getting a "7". <Why Not?>

Well, it might be because the two rolls of the dice are **independent** of each other.

Let's look at the joint distribution:

	1	2	3	4	5	6	
1	0.028	0.028	0.028	0.028	0.028	0.028	0.167
2	0.028	0.028	0.028	0.028	0.028	0.028	0.167
3	0.028	0.028	0.028	0.028	0.028	0.028	0.167
4	0.028	0.028	0.028	0.028	0.028	0.028	0.167
5	0.028	0.028	0.028	0.028	0.028	0.028	0.167
6	0.028	0.028	0.028	0.028	0.028	0.028	0.167
	0.167	0.167	0.167	0.167	0.167	0.167	1.000

{Next Slide}

Ways to Roll a 7

Ways to roll a 3 when Die 1=4

First Die (A)

Second Die (B)

	1	2	3	4	5	6	
1	0.028	0.028	0.028	0.028	0.028	0.028	0.167
2	0.028	0.028	0.028	0.028	0.028	0.028	0.167
3	0.028	0.028	0.028	0.028	0.028	0.028	0.167
4	0.028	0.028	0.028	0.028	0.028	0.028	0.167
5	0.028	0.028	0.028	0.028	0.028	0.028	0.167
6	0.028	0.028	0.028	0.028	0.028	0.028	0.167
	0.167	0.167	0.167	0.167	0.167	0.167	1.000

$$P(B|A) = \frac{P(A \& B)}{P(A)} = \frac{P(B=3 \cap A=4)}{P(A=4)} = \frac{0.028}{0.167} = 0.167 = P(B=3) = P(C=7).$$

And we can see that, since the probability of rolling a 3 with the second die is the same, no matter what the result with the first die, knowing the roll of the first die gives us no useful information to update our probability of getting a 7. **But ... What if we wanted to know the probability of rolling a 12 and the first die = 1? Then the first roll gives useful information, and the probability changes to zero!** We'll come back to this notion of **independent events** in a minute. First, however, let's use the **conditional probability rule** to define another probability rule.

3.8. More Probability Rules

3.8.1 The General Multiplication Rule

Because we define conditional probability as:

$P(A|B) = \frac{P(B\&A)}{P(B)}$, we can write the **general multiplication rule** as follows:

If A and B are any two events then **{Next Slide}**

$$P(B\&A) = P(B) \bullet P(A|B).$$

In words, for any two events, their joint probability equals the probability that one of the events occurs times the conditional probability of the other even, given that event.

The conditional probability rule and the general multiplication rule are simply variations of each other.

3.8.2 Statistical Independence

One of the most important concepts in probability is that of **statistical independence** of events.

For two events, statistical independence, or more simply, **independence**, is defined as follows: **{Next Slide}**

Independent Events

Event A is said to be **independent** of event B if the occurrence of event B does not affect the probability that event A occurs. In symbols,

$$P(A \mid B) = P(A).$$

In words, knowing whether event B has occurred provides no probabilistic information about the occurrence of event A.

So, when two events are independent the following must hold:

3.8.3 The Special Multiplication Rule

Now, we're in a position to produce a special version of our general multiplication rule for two independent events. The general multiplication rule says

{Next Slide}

Independence Implies

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) = P(A|S)$$

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = P(B) = P(B|S)$$

Rearranging these two equations gives the General Multiplication Rule:

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

which implies, since

$$P(A \cap B) = P(B \cap A) \Rightarrow P(A|B)P(B) = P(B|A)P(A)$$

that, if A and B are independent events, then from the definition of independence:

$$P(A \cap B) = P(A|B)P(B) = P(A) \bullet P(B) \text{ or,}$$
$$P(B \cap A) = P(B|A)P(A) = P(B) \bullet P(A)$$

{Next Slide}

The Special Multiplication Rule (for Two Independent Events)

If *A* and *B* are independent events, then

$$P(B \,\&\, A) = P(B) \bullet P(A),$$

and conversely, if

$$P(B \,\&\, A) = P(B) \bullet P(A),$$

then *A* and *B* are independent events. In words, two events are independent if and only if their joint probability equals the product of their marginal probabilities.

<So, can independent events be *mutually exclusive*?> No, because the probability of one event

depends upon the occurrence, or nonoccurrence of the other.

Therefore, mutually exclusive events cannot be independent.

3.8.4 Example Problem: Smoking Study

An epidemiological study of smoking found that for three different age groups:

less than age 30 (<30)

ages 30-50

greater than 50 (>50

that half those under 30 were found to smoke. <How do I write this probability?>

$$P(Sm| < 30) = 0.5$$

a) If $P(< 30) = 0.5$ find the *joint probability that a person is under 30 <u>and</u> smokes.*

<What, excactly, in probability terms are we trying to find?>

$$P(Sm \cap < 30).$$

<What information do we have?>

$$P(Sm| < 30) = 0.5$$
$$P(< 30) = 0.5$$

<How do we use this information?> Use the formula for the

intersection of two events (or the General Multiplication Rule)

$$P(Sm \cap < 30) = P(Sm| < 30)P(< 30)$$
$$P(Sm \cap < 30) = 0.5 \times 0.5 = 0.25.$$

b) <Now, are the probabilities of smoking and being under age 30 independent or dependent?>

We don't have enough information to tell.

However, if I gave you the following information, what would you say, dependent or independent?>

$$P(Sm| < 30) = 0.5$$
$$P(Sm|30-50) = 0.5$$
$$P(Sm| > 50) = 0.25$$

Age and smoking are dependent events because the conditional probabilities vary by age.

c) Now, if the probability of a person's being in a particular age group are:

$$P(< 30) = 0.50$$
$$P(30-50) = 0.25$$
$$P(> 50) = 0.25$$

Find, *P(Sm)*, the unconditional probability that a person smokes if randomly selected from the population.

<What are we trying to find here?> The **marginal probability of *Sm***. Since the age groups are mutually exclusive, we have:

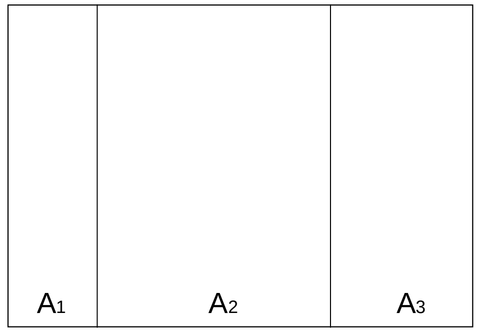
$$\begin{aligned} P(Sm) &= P(Sm \cap < 30) + P(Sm \cap 30-50) + P(Sm \cap > 50) \\ &= P(Sm| < 30) \bullet P(< 30) + P(Sm|30-50) \bullet P(30-50) + P(Sm| > 50) \bullet P(> 50) \\ &= 0.5 \times 0.5 + 0.5 \times 0.25 + 0.25 \times 0.5 \\ P(Sm) &= 0.25 + 0.13 + 0.06 = 0.44 \end{aligned}$$

The **marginal probability** of an event is the sum of the intersections of that event with all possible states of another event. This is an example of the **Total Probability Rule**, which I'll now discuss.

3.8.5 The Rule of Total Probability

Define: **Exhaustive Events** are exhaustive because at least one of them must occur. E.g., Governor's at a conference must be either Republican, Democrat or Independent. Exhaustive events are also mutually exclusive, since only one category can occur at a time.

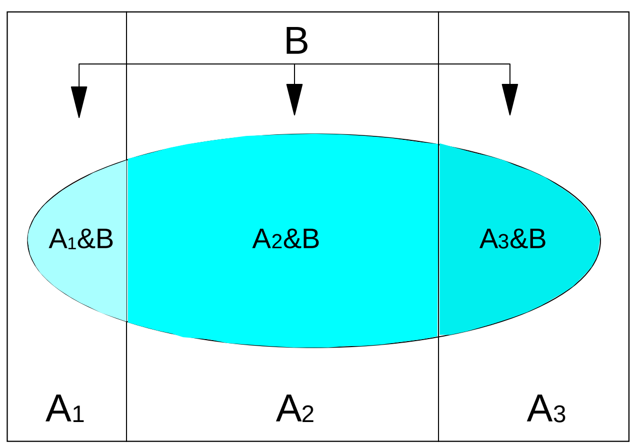
Look at the following sample space, totally occupied by an event *A* which can take on only one of three values:



Sample space, S, is completely filled by event A.

{Boone: Simply click on the above slide in the presentation, and the following object will materialize}

Now, let's add a second event, *B*, that can occur in conjunction with any of the three values of *A* **{Next Slide}** *:*



Sample space, S, is completely filled by event A.

Now, *B* can occur in conjunction with one, and only one of the events *A1, A2, or A3.* So, the probability of *B* is:

$$P(B) = P(A_1 \& B) + P(A_2 \& B) + P(A_3 \& B).$$

But, the general multiplication rule allows us to substitute for each term on the right side of this equation:

$$P(B) = P(A_1) \bullet P(B|A_1) + P(A_2) \bullet P(B|A_2) + P(A_3) \bullet P(B|A_3)$$

Or, we can write:

{Next Slide}

The Rule of Total Probability
Suppose that Events $A_1, A_2, \ldots, A_k$ are mutually exclusive and exhaustive; that is, exactly one of the events must occur. Then, for any event *B,*

$$P(B) = \sum_{j=1}^{k} P(A_j \cap B) = \sum_{j=1}^{k} P(A_j) \bullet P(B \mid A_j).$$

3.9. Bayes's Rule

3.9.1 Bayes's Theorem: First Pass

Using the probability notions that we've just developed, we can represent the probability that two events *A* and *B* occur simultaneously in the following manner:

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$$

Then, using the last two terms of this equality and rearranging we can express the probability that *B* occurs, given that *A* occurred in terms of the probability that *A* occurs given *B:*

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

This is a very simple version of **Bayes's Rule (or Theorem).** This theorem, which follows immediately from our Rules (or Axioms) of probability is enormously important to that group of statisticians who call themselves "Bayesians".

Bayesians argue as follows: Before we go out to collect data on a phenomenon, we usually have a prior opinion (subjective) as to what the value of that phenomenon is.

This prior opinion is represented by *P(A)*. After we have actually collected a sample, we have to modify our opinions in light of sample evidence, *right?*

Bayes Rule give us a method by which we can consistently modify (update) our prior opinion.

We'll come back to this discussion after we develop a more complete understanding of Bayes Rule.

3.9.2 Bayes's Rule Derived

Using the rule of total probability we can derive Bayes's rule. For simplicity, let's consider three events A_1, A_2, and A_3, that are mutually exclusive and exhaustive and let B be any event. For Bayes's rule we assume that the following probabilities are known: **{Next Slide}**

$$P(A_1) \quad P(A_2) \quad P(A_3)$$
$$P(B|A_1) \quad P(B|A_2) \quad P(B|A_3)$$

Our goal is to use these known probabilities to find the following probabilities:

{Next Slide}

$$P(A_1|B) \quad P(A_2|B) \quad P(A_3|B)$$

From the conditional probability rule, we know that we can express each of these conditional probabilities as: **{Next Slide}**

From the Conditional Probability Rule, we know that:

$$P(A_i \mid B) = \frac{P(B \,\&\, A_i)}{P(B)} \text{ for } i = 1, 3.$$

Next, let's use what we've established through our previous work to alter this equation. Apply the general multiplication rule to the numerator:

{Next Slide}

Next, apply the General Multiplication Rule to the Numerator:

$$P(A_i \mid B) = \frac{P(B \,\&\, A_i)}{P(B)} \; i = 1, 3.$$

That gives us:

Prior probability

$$P(A_i \mid B) = \frac{P(A_i) \bullet P(B \mid A_i)}{P(B)} \; i = 1, 3.$$

Posterior probability

{Next Slide}

Finally, apply the rule of Total Probability in the denominator to give us **Bayes's Rule:**

$$P\left(A_i \mid B\right) = \frac{P\left(A_i\right) \bullet P\left(B \mid A_i\right)}{\sum_{j=1}^{k} P\left(A_j\right) \bullet P\left(B \mid A_j\right)} \quad i = 1,3$$

Bayes's Rule turns out to be an extraordinarily useful probability theorem. It probably doesn't mean a thing to you, right now, but I hope you'll see its utility shortly.

3.9.3 An example using Bayes's Rule (This is from the Mendenhall text!)

The Nickel and Dime department store is considering adopting a new credit management policy in an attempt to reduce the number of credit customers who default on their payments. The credit manager has suggested that in the future credit should be discontinued to any customer who has twice been a week or more late with his monthly installment payment. He supports his claim by noting that past credit records show that 90% of all those defaulting on their payments were late with at least two monthly payments.

Suppose from our own investigation we have found that two per cent of all credit customers actually default on their payments and that 45% those who have not defaulted have had at least two "late" monthly payments.

Find the probability that a customer with two or more late payments will actually default on his payments and, in light of this probability, criticize the credit manager's plan.

Define: L -- a credit customer is two or more weeks late with at least two monthly payments,

D: a credit customer defaults on his payments. $\overline{D}$ is the complement of D.

<What do we wish to find out?> The probability that a customer who has paid late at least twice will default.

<How do we represent this symbollically?> $P(D|L)$

<What information do we have to work on?>

$$
\begin{aligned}
&P(L|D) = 0.9 \\
&P(D) = 0.02 \text{ (the prior)} \\
&P\left(L|\overline{D}\right) = 0.45
\end{aligned}
$$

Now, we can use Baye's Theorem here:

$$
\begin{aligned}
P(D|L) &= \frac{P(L|D) \bullet P(D)}{P(L)} \\
&= \frac{P(L|D) \bullet P(D)}{P(L|D) \bullet P(D) + P\left(L|\overline{D}\right) \bullet P\left(\overline{D}\right)} \\
&= \frac{(0.9 \times 0.02)}{(0.9 \times 0.02) + (0.45 \times 0.98)} \\
&= \frac{0.018}{0.018 + 0.441} = 0.0392
\end{aligned}
$$

The probability that a late payer will default is 0.0392.

<Would you say that the credit manager's idea is a good one?> No, for every 25 late payers only one, on the average, will default, and 24 good credit customers will be lost.

3.9.4 Another Example of Bayes's Rule (Another Smoking Study)

According to the American Lung association 7.0% of the population has lung disease. Of those having lung disease, 90% are smokers; of those not having lung disease, 25.3% are smokers. Determine the probability that a randomly selected smoker has lung disease.

What information do we have available:

- Let *Sm* indicate that a randomly selected person is a smoker.
- Let L_1 represent the event that the person selected has no lung disease.
- Let L_2 represent the event that the person selected has lung disease.

Note that L_1 and L_2 are complementary, which implies that they are mutually exclusive and exhaustive.

Now, what (in symbols) do we want to know?

$$P(L_2|Sm)$$

What information do we have:

$$P(L_2) = 0.07$$
$$P(Sm|L_2) = 0.90$$
$$P(Sm|L_1) = 0.253$$

Now, write an expression (symbolically) that expresses what we want to know in terms of what we know:

$$P(L_2|Sm) = \frac{P(Sm|L_2) \bullet P(L_2)}{P(Sm|L_1) \bullet P(L_1) + P(Sm|L_2) \bullet P(L_2)} = \frac{0.9 \times 0.07}{(0.9 \times 0.07) + (0.253 \times 0.930)} = 0.211.$$

Wow, look at this: If we know that a person is a smoker our estimate of the probability that he/she has lung disease triples! The additional evidence of his/her smoking means that the probability of lung disease changes from fairly rare (7%) to not unexpected (21.1%)

Bayes's Rule allows us to reduce our uncertainty about an event by (in this case) a considerable amount.

3.9.5 Another Bayes's Theorem Problem (Unemployment)

In a population of workers suppose 40% are grade school graduates, 50% are high school graduates, and 10% are college graduates. Among the grade school graudates, 10% are unemployed, among the high school graduates 5% are unemployed and among the college graduates 2% are unemployed.

If a worker is chosen at random and found to be unemployed, what is the probability that he is a college graduate?

Information:

$$\left.\begin{array}{l} P(G) = 0.4 \\ P(H) = 0.5 \\ P(C) = 0.1 \end{array}\right\} \text{ mutually exclusive(this is a prior distributon on education)}$$

Also:

$$P(U|G) = 0.1$$
$$P(U|H) = 0.05$$
$$P(U|C) = 0.02$$

<What do we want to find out?> $P(C|U)$

<What theorem can we apply?> Bayes's!

$$P(C|U) = \frac{P(U|C) \bullet P(C)}{P(U|G) \bullet P(G) + P(U|H) \bullet P(H) + P(U|C) \bullet P(C)} = \frac{0.02 \times 0.1}{(0.1 \times 0.4) + (0.05 \times 0.5) + (0.02 \times 0.1)} = \frac{0.002}{0.067} = 0.03.$$

So, our prior probability of finding a college graduate started out at 10%, but after finding out that the person is unemployed, the probability that he is a college graduate drops to 3%.