Elements of Probability Theory I

Trials, Probability Experiments, Outcomes, and Sample Spaces

A simple probability experiment, experiment, is a pair; an action along with an observation of an uncertain but unique result of the action. The result of the experiment is called an outcome. A sample space of an experiment is the set of all possible outcomes of the experiment.

If the action is taken to be tossing a coin, and we agree to observe to the result of the coin face landing up being a H or a T, then the sample space is S = {H, T}. However, if we agree to observe the time it takes the coin to land, after being thrown into the air with random force, then we are taking about a different experiment, and the sample space is the set of real numbers.

It is assumed that all members of a sample space represent unique results, so that they are disjoint, and that all the members represent the entire collect of results, so that they are exhaustive. So a sample space S must include results which can occur, however improbable – like the coin landing on its side, it being eaten by a cat on the way down, the environment of the experiment being swept away by a hurricane, etc. We generally group these red herring results into a single result that the experiment wasn't completed, and do not consider it a result.

A repetition of the same simple experiment, called a trail, gives a set of outcomes that can be observed as ordered or not ordered by the number of the trail. This repetition is actually a new experiment, called a compound experiment. If the observation of this new experiment is to regard the order, then the outcomes of it are n-tuples of the random variable that is defined in the simple experiment, and the sample space is the set of all such n-tuples.

For example, if we perform the <coin toss, observe the coin face up> experiment twice, the sample space S is {HH, HT, TH, TT}. Note that this is the same experiment as flipping two coins, essentially. If we do it 3 times, we get {HHH, HHT, HTT, TTT, THT, TTH}.

If the observation does not include order, then it is a statistical set, a set where elements can be listed more than once, given each element a frequency of occurrence. This sample space is generally in the Reals. This set can also be viewed as a set of ordered pairs, the first component being the outcome, and the second, the frequency of the outcome.

In statistics, such compound experiments are performed and the observations are not ordered. The performing of each trial is call sampling, and the aggregated outcome, a statistical subset of R, is called a sample. Generally, sampling without replacement is performed, as replacing an element of the sample could result in it being repeatedly selected, and thus biasing the random process of the simple experience.

Example

The experiment is to select a distinct person in the class, and on each trail agree to not select one that we have previously selected, and measure his/her height. If we repeat the experiment 10 times, instead of the observation of a 10 tuple made from the individual trails, we think of each trial as a measurement, and observe the aggregate outcome of all the trails as a set of statistical data in the sample space R, something like {160, 145, 150, 150, 148, 187, 125, 125, 144, 150}.

These experiments produce sample spaces that are made from just one variable, they are uni-variate.

Bi-variate Sample Spaces

If we perform an experiment and make two different observations, giving use two different sample spaces, then a new sample space composed by the Cartesian product of the two sample spaces is a new sample space and is said to be bi-variant. Etc. for multivariate.

For example, if in the <coin toss, observe the coin face up> we add the time it takes for the coin to hit the ground, we get <coin toss, observe the coin face up, observe the time in the air> and do the experiment 2 times, the sample space could look like {(H,3), (T,3.2)}.

Now suppose the experiment is selecting a person, one descriptor is the person's height and the other is the person weight, and we do it 50 times in a population of some good size, like 1000. The set of outcomes could look like {(160,93), (142,78), ….}. This data can be plotted and checked for associations, like correlation.

The rest of this chapter only deals with univariate sample spaces.

Univariate Sample Spaces

There are two interesting things about univariate sample spaces; each element's description and the number of elements. This is because in elementary probability theory, to find the probability of the occurrence of an element, we must be able to identify it. The probability of the occurrence is then just the number of occurrences divided by the total number of elements in the sample space, as long as each element has the same probability of occurrence. For instance, suppose we have a sac with 2 green marbles and 4 red marbles. Thus we have the descriptions of the elements, and we have how many – 6. So, the probability of choosing a green marble at random from the sac is the # of occurrences of green marbles / number of marbles in the sac. So it's 2/6 = 1/3.

Examples

1. Experiment: Flip a fair coin and observe the characteristic of the side landing up.
Sample Space: S = {H, T} H and T occur once, so the probability of either is ½
(a fair coin means H and T are equally probable).

2. Experiment: Select a student in your class and observe the color of his or her hair.
Sample Space: {red, black, brown, blond, green, ...} Hopefully, the probability a student has green hair is 0.

3. Experiment: Roll a die and observe the number on the side landing up.
Sample Space: S = {1, 2, 3, 4, 5, 6}. The probability of each of 1,..,6 is 1/6 as long as the die are fair.

4. Experiment: Roll two dice and observe the numbers on the two sides landing up.
Sample Space: S = the set with all the elements listed in the table below

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

The probability of any one of these 36 is 1/36, assuming fair dice. What's the probability of a 3 being rolled. There are 11 of these pairs in which a 3 appears, so the probability of a 3 is 11/36.

5. Experiment: Roll two dice and observe the sum of the numbers on the two sides landing up. Sample Space: S = {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

What is the probability of a 5. It's not 1/12, because the probability of each of 2,…,12 is not the same. Unless you know about the sample space in 4., you can't determine this probability. Once you recognize that you can go back to the sample space in 4., then it is easy to count the number of pairs that give 5, and the probability is 4/36.

6. Experiment: Choose 2 numbered marbles (without regard to order) at random from a bag of 6 numbered marbles and observe their numbers
Sample Space: The set of all collections of 2 marbles chosen from 6 marbles;
{m_1m_2, m_1m_3, …., m_1m_6, …., m_5m_6}

|S| = C(6, 2) = 15

C(N,R)

Probability Function P

(1) A probability function is a function P that assigns a number $P(s_i)$ to each outcome s_i in a sample space $\{s_1, s_2, . . . , s_n\}$ of an experiment in such a way that

(a) $0 \le P(s_i) \le 1$
(b) $P(s_1) + P(s_2) + . . . + P(s_n) = 1$

In no way are probability functions unique. However, we generally speak of the one probability function that is a representation of the relative likelihood of the occurrence of any of the outcomes. For example, if we assigned the probability P to the set $\{s_1, s_2, s_3\}$ by $P(s_1) = ¼$, $P(s_2) = ¼$, and $P(s_3) = ½$, then we are representing that the s_1, s_2 are equally likely, and both are half as likely as s_3. Each s_i might be the outcome that runner i wins a race. However, if a bag contains 3 socks, and s_i represents the outcome of randomly picking out sock i from the bag, then P is not representative of the likelihood of the occurrence of the s_i. What probability function is representative?

What is the probability function for the sample space in 5. above. It's

x	2	3	4	5	6	7	8	9	10	11	12
P(x)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Note that P(2 or 3) = P(2) + P(3) because the basic elements of the sample space, the observations, are disjoint - have nothing in common.

The assignment of probabilities is often performed by looking at historical data and addressing the relative frequency of occurrences. However, it is equally obtained by observation of the laws of nature which might govern the action part of the experiment. If a coin is manufactured correctly, when we flip it, it will land H with probability .5. We can ascertain that from the laws of physics. And note that if we find a disagreement in that when we flip it 1000 times, heads comes up only 400 times, then we can assume that the coin is biased, and not made as we thought. Thus the two ideas of probability must overlap. Another popular type of probability assignment is subject assignment. This is done by simply making a guess at the number. For instance, the probability of life on Mars is .09. This type is a belief, and is not generally verifiable to be in line with the first two accepted methods. There are many alternative points of view on this, and I refer you to the document "Symmetry Is the Very Guide of Life" posted in the class folder on my website for a fairly involved read about all the different types - • Actual and hypothetical frequentism • Classical probabilities, with finite sample spaces • Fully constrained logical probability • Symmetry-based propensities • Constrained subjectivism • Classical probabilities in infinite sample spaces • Less constrained logical probability • Non-symmetry-based propensities • Unconstrained subjectivism.

As said before, an experiment can be thought of as having two parts, the primary or action part, which is to do something, and the secondary part, which is to perform an observation of the result of primary part. The characteristic of the observation performed in the second part, the thing that we are measuring, is called a **random variable (rv)**. It is so termed because it can take on the random values of the outcome of the experiment.

In the experiment of flipping a coin, the primary part is the flipping of it. There many secondary parts, or observations that can be made. We generally just observe the face of the coin landing up, the rv is the characteristic of the side that it lands on and can take on the values H or T. (If we observe the number of times it flips in the air before it lands, the rv can take on the number 1,2,3, …. If we observe the loudness of the impact of its landing, the rv can take on a number of decibels. If we observe the amount of time it takes from it being thrown until coming to rest, the rv can take on values in the set of real numbers. A random variable can be discrete or continuous, independent of the primary part of the experiment. It is discrete if the values it can take on are in the set on integers, otherwise it is continuous.)

Instead of referring to a random variable as rv, let's make it simpler and refer to it as X.

We generally want to know the probability of an **event**, E, a subset of the sample space. In an experiment, we think of the random action part of the experiment as defining the event. So, if we draw a marble from our sac containing 2 green marbles and 4 red marbles, we can ask about the probability of drawing a green marble. We think of the thing being drawn as a random variable, X, that describes the event. So we ask about P(X = green marble), the event, E, is the subset of the sample space that contains green marbles.

Given a probability function, P, we can obtain the probability of an event E by adding together the probabilities of the outcomes in E. In general, $P(E) = P(s_1) + P(s_2) + …. + P(s_k)$ where s_1, s_2, … s_k make up the event. However, when all the outcomes of S are equally likely, $P(s_i) = 1/|S|$ so that $P(E) = P(s_1) + P(s_2) + …. + P(s_k) = K/|S| = |E|/|S|$. Getting back to the sac with 2 green marbles and 4 red marbles, it is clear that all the outcomes are equally likely, so that P(X = green marble) = 2/6.

What about the sac of six labeled marbles. Let's draw 2 marbles. It is clear that there are C(6,2) =15 elements in the sample space. If I choose one of the elements, $P(X = \{m_i, m_j\})$ = 1/15 for any i,j = 1,..,6, i≠j. What is $P(X=\{m_{\mathbf{1}}, m_j\})$ for some j = 1,…,6. That is, what is the probability that I have drawn the marble labeled 1. The event E is { $\{m_1, m_2\}$, $\{m_1, m_3\}$, $\{m_1, m_4\}$, $\{m_1, m_5\}$, $\{m_1, m_6\}$ }; so the probability is 5/15 = 1/3.

Suppose a die is loaded so that the probability of a 6 showing on top is .2, and the probability of showing any other number is .16. If I roll the die, what is P(X = 4). It's .16. What about P(X = 3 or 4), it's .16 + .16 = .32; it is not 2/6 = .3333….

In the fare coin toss experiment, the event of the coin showing heads is given by P(X = "H") = ½.

In the case of observing the sum of the numbers facing up on the roll of two fair die, the event that a 7 is rolled is X = 7. P(X = 7) = |E|/|S| = 6/36 = 1/6

Experiment: Draw a hand of two cards from a deck of 52.
Event H: Both cards are diamonds.
H is the set of all hands of 2 cards chosen from 52 such that both cards are diamonds. There are 13 diamonds, so |H| = C(13,2) = 13*12/2 = 78. P(X = H) = 78 / C(52,2) = 78 / 26*51 = .0588 approximately.

Your broker recommends four companies. Unbeknownst to you, two of the four happen to be duds. You invest in two of them at random. Find the probability that:

(a) you have chosen the two losers
(b) you have chosen the two winners
(c) you have chosen one of each

At first, one thinks of {WW, WL, LW, LL} as the sample space. However, it is not.
Let the companies be C_1, C_2, C_3, C_4. Then, wlog, C_1 and C_2 are the winners, and C_3 and C_4 are the losers. There are C(4,2) elements in the sample space, namely,
$\{C_1, C_2\}$ $\{C_1, C_3\}$ $\{C_1, C_4\}$ $\{C_2, C_3\}$ $\{C_2, C_4\}$ $\{C_3, C_4\}$, and now you get

(a) you have chosen $\{C_3, C_4\}$ with probability 1/6
(b) you have chosen $\{C_1, C_2\}$ with probability 1/6
(c) you have chosen one of each one of $\{C_1, C_3\}$ $\{C_1, C_4\}$ $\{C_2, C_3\}$ $\{C_2, C_4\}$ with probability 4/6 = 2/3

Example
In the card game poker, a hand consists of a set of five cards drawn from a standard deck of 52. Note that each hand is equally likely. A full house is a hand consisting of three cards of one denomination ("three of a kind"—e.g. three 10s) and two of another ("two of a kind"—e.g. two Queens). Here is an example of a full house: 10C , 10D, 10S, QH, QC.

(a) How many different poker hands are there?
(b) How many different full houses are there that contain three 10s and two Queens?
(c) How many different full houses are there altogether? What's the probability of a full house?

Solution
(a) Since the order of the cards doesn't matter, we simply need to know the number of ways of choosing a set of 5 cards out of 52, which is C(52, 5) = 2,598,960 hands.

(b) The 3 10s can be chosen from the 4 10s available C(4,3) = 4 ways.
For each of those ways, the 2 Qs can be chosen C(4,2) = 6 ways. Since there are no other cards to be chosen, that's it, and there are 4*6 = 24 such hands.

(c)
Step 1: Choose a denomination for the three of a kind; 13 choices.
Step 2: Choose 3 cards of that denomination. Since there are 4 cards of each denomination (one for each suit), we get C(4, 3) = 4 choices.
Step 3: Choose a different denomination for the two of a kind. There are only 12 denominations left, so we have 12 choices.
Step 4: Choose 2 of that denomination; C(4, 2) = 6 choices.
Thus there are a total of 13 * 4 * 12 * 6 = 3744 possible full houses.

Homework S1

1. Two coins are tossed. What is the probability that the result shows at most one tail.
2. Two indistinguishable dice are rolled. What is the probability that the numbers add to 5.
3. Motor Vehicle Safety: The following table shows crashworthiness ratings for 10 small SUVs. (3=Good, 2=Acceptable, 1=Marginal, 0=Poor)

Frontal Crash Test Rating	3	2	1	0
Frequency	1	4	4	1

   a. What is the probability that a randomly selected SUV will have a crashworthiness of 3?
   b. What is the probability that a randomly selected SUV will have a crashworthiness of acceptable of better?
   c. What is the probability that a randomly selected SUV will have a crashworthiness of marginal or acceptable?

4. A poker hand consists of five cards from a standard deck of 52.

   a. How many hands are there containing two pairs (two of one denomination, two of another denomination, and one of a third)
   b. What's the probability of getting 2 tens and 2 queens?
   c. What's the probability of getting the ten of hearts, the ten of spades, the queen of diamonds, and the queen of hearts?
   d. What's the probability of getting three of a kind (three of one denomination, one of another denomination, and one of another denomination)
   e. What's the probability of getting 3 aces.

Combining Events
An event may often be described in terms of other events, using set operations.
Let E and F be events in S. Recall that E is a certain subset of S and that its elements are described by an English statement, or by a listing of its elements.

The negation of E, written E′, is the event that E does not occur. This means that the outcome of the experiment is not in E, but it is still in S, it is in the compliment of E (in S). Whenever there is an English statement describing E, E′ is described by negating it.

So if the event E is described by "An even number", E′ is described by "An odd number". If S = {5,6,7,8,9,10}, then E′ = {5,7,9}.

What about the events "E or F" and "E and F".

A simple example: the experiment is throwing a single die. Let E be the event that the outcome is a 5, and let F be the event that the outcome is an even number. Thus,

E = {5}, F = {2, 4, 6}. So, E or F = {5, 2, 4, 6}.

In other words, E or F is the event that the outcome is either a 5 or an even number.

Whenever the is an English statement describing E and F, the event E or F is described by the statement obtained by inserting "or" between the statement for E and the statement for F. In a like manner, the event "E and F" described by the statement obtained by inserting "and" between the statement for E and the statement for F.

If we do not have an English description for E or F, we rely on the set theoretic notation for the two sets E and F to make the compliment, union, or intersection. They are defined as E′, E Ų F, and E Π F, in which case two sets E and F are compliment, joined, or intersected as if they were described in English. The definitions are simply:

E′ ={ s | not s ϵ E; s ϵ S }, E Ų F ={ s | s ϵ E or s ϵ F; s ϵ S }, E Π F ={ s | s ϵ E and s ϵ F; s ϵ S}

Example
The following table shows sales of recreational boats in the U.S. during the period 1999–2001.

	Motor boats	Jet skis	Sailboats	Total
1999	330,000	100,000	20,000	450,000
2000	340,000	100,000	20,000	460,000
2001	310,000	90,000	30,000	430,000
Total	980,000	290,000	70,000	1,340,000

The experiment is that a recreational boat is selected at random from those in the table. Let E be the event that the boat was a motor boat, let F be the event that the boat was purchased in 2001, and let G be the event that the boat was a sailboat. Find the probabilities of the following events:

(a) E (b) F (c) E Π F (d) G' (e) E Ų F'

Compliments and their probabilities

Let E and F be any two events in a non-empty space of outcomes S. Recall that the probability of the P(E) is the sum of the probabilities of the outcomes in E, and P(F) is the sum of the probabilities of the outcomes in F.

Remember also that the outcomes in S, the s_i, need not have equal probabilities; however, they are disjoint and the sum of the $P(s_i)$ must be 1. Also, remember that in general $P(E) \neq N(E)/N(S)$, but instead P(E) = sum of the probabilities of the outcomes, s_i, in E. Because of this, the probability of the union of any number of disjoint events equals the sum of the probabilities of the individual events.

Now it's easy to see that S = E U E', and E, E' disjoint. So
$1 = P(S) = P(EUE') = P(E) + P(E')$

$P(E') = 1 - P(E)$.

In class exercise

You roll a pair of dice and observe the sum of the two top numbers. Describe the event that the sum of the numbers is not 6. What is its probability?

Conditional Events and their probabilities

One of the most important concepts in probability study is that of conditional events. Given two events E and F, in a sample space S, the conditional event E|F is a statement describing the event that E occurs given that the event F occurs.

Example

If S = {1,2,….,10}, and F = event that the outcome is ≤7, and E = event that the outcome is even. Then E|F is the statement that E occurs given that F occurs, so E|F is the statement that the outcome is even given that the outcome is ≤7.

S = {*1*,***2***,*3*,***4***,*5*,***6***,*7*,8,9,10}   E = {*1*,***2***,*3*,***4***,*5*,***6***,*7*,**8**,9,**10**}   F = {*1*,***2***,*3*,***4***,*5*,***6***,*7*}

It is common to confuse the intersection of two events with a conditional involving the two events. In the above example, the intersection of E and F is H = {2,4,6}. So P(E Π F) = 3/10. The probability of E|F is computed by observing the occurrence of E given that F occurs. If F occurs, then P(E) = 3/7.

Example

There are four people being considered for the CEO position at Nortol Tech. Three of the applicants are over 60 years old and two are women, one of the women is over 60.

a. What is the probability that a candidate is over 60 and a female?
b. Given that the candidate is male, what is the probability the he is over 60?
c. Given that the person is over 60, what is the probability that he is male?
d. Given that the person is over 60, what is the probability that he is female?

Probabilities of Intersections

Unlike observations, two events can, and often do, share common observations. These observations are in the intersection of the two events.

Evidentially, from the above discussion on conditional probabilities, the probability of the intersection of E and F can be computed by looking at the actual outcomes in the intersection, of by

P(E Π F) = P(E|F) * P(F) = P(F|E) * P(E)

P(E Π F) is called the joint probability of E and F.

If P(E Π F) = P(E) * P(F) then the events E and F are said to be independent. This is **because** if

P(E Π F) = P(E) * P(F) = P(E|F) * P(F), so **P(E) = P(E|F)**

and

P(E Π F) = P(E) * P(F) = P(F|E) * P(E), so **P(F) = P(F|E)**

Probabilities of Unions

For two sets, E and F then set $E \sim F = \{s \mid s \in E \text{ and not } s \in F\} = E \sqcap F'$

Note that for two events E and F in a space, S,
$P(E \sim F) = P(E \sqcap F') = P(E) - P(E \sqcap F)$

It's easy to see that
$E \cup F = (E \sim F) \cup (F \sim E) \cup (E \sqcap F)$.

The RHS is the union of disjoint sets in a sample space. So,
$P(E \cup F) = P(E \sim F) + P(F \sim E) + P(E \sqcap F) = P(E) - P(E \sqcap F) + P(F) - P(E \sqcap F) + P(E \sqcap F)$
$= P(E) + P(F) - P(E \sqcap F)$

If $P(E \sqcap F) = 0$, the two events E and F are said to be mutually exclusive, meaning they contain no common outcomes.

One would think that ME Implies ID.

$0 = P(E \sqcap F) = P(E) * P(F) = P(E|F) * P(F)$ so that $P(E|F)$ must be 0 or $P(F) =0$

What does $P(E|F) =0$ imply. It implies that E does depend on F, and if F occurs, E doesn't. The other way around, does ID imply ME. No, it just says

$P(E \cup F) = P(E) + P(F) - P(E \sqcap F) = P(E) + P(F) - P(E)*(F)$

A coin is tossed three times and the sequence of heads and tails is recorded. Decide whether the following pairs of events are mutually exclusive, independent.
(a) A: the first toss shows a head, B: the second toss shows a tail
(b) A: all three tosses land the same way up, B: one toss shows heads and the other two show tails.
(c) A: all three tosses land heads up, B: one toss shows heads.
(d) A: the first two land heads up B: the last one lands heads up

Solution:
(a) mutually exclusive and independent
(b) mutually exclusive, not independent
(c) not mutually exclusive, not independent
(d) not mutually exclusive and independent

Your broker recommends four companies. Unbeknownst to you, two of the four happen to be duds. You invest in two of them at random. Find the probability that:

(a) you have chosen the two losers
(b) you have chosen the two winners
(c) you have chosen one of each

We've seen this before. Now we solve (b) a different way than before. Let W represent a win.

P(W and W) = P(W on 1st choice and W on 2nd choice)
= P(W on 2nd choice | W on the 1st choice) * P(W on the 1st choice)
P(W on the 1st choice) = 2/4 = 1/2
If we get a W on the 1st choice, then there is 1 winner left and 2 losers, so
P(W on 2nd choice | W on the 1st choice) = 1/3
Therefore:
P(two winners) = P(W and W) = 1/2 * 1/3 = 1/6
same answer as we got before.

Example
The astrology software package Turbo Kismet works by first generating random number sequences, and then interpreting them using numerology. When I ran it yesterday, it informed me that there was a 1/3 probability that I would meet a tall dark stranger this month, a 2/3 probability that I would travel within the next month, and a 1/6 probability that I would meet a tall dark stranger on my travels this month. What is the probability that I will either meet a tall dark stranger or that I will travel this month?

One can interpret "meeting a tall dark stranger on my travels this month" as being "meeting a tall dark stranger given that I travel this month", so it should be clarified that what is meant is that there is a 1/6 probability that I both travel this month and meet a tall dark stranger. Then the solution of the problem is an easy application of
P(E) + P(F) - P(E Π F) = 1/3 + 2/3 – 1/6 = 2/3.

Homework S2

Suppose two dice (one red, one green) are rolled. The following are events

A: the red die shows 1

B: the numbers add to 4

C: at least one of the numbers is 1

D: the numbers do not add to 11.

In Exercises 1–4, express the given event in symbols E and F

1. The red die shows 1 and the numbers add to 4.
2. The numbers do not add to 4 but they do add to 11.
3. Either the numbers add to 11 or the red die shows a 1.
4. At least one of the numbers is 1 or the numbers add to 4.

Let W be the event that you will use the web site tonight, let I be the event that your math grade will improve, and let E be the event that you will use the web site every night. In Exercises 5–8, express the given event in symbols.

5. You will use the web site tonight and your math grade will improve.
6. Either you will use the web site every night, or your math grade will not improve.
7. Your math grade will not improve even though you use the web site every night.
8. You will either use the web site tonight with no grade improvement, or every night with grade improvement.

9. Complete the following. Two events E and F are mutually exclusive if their intersection is ______.

10. If E and F are events, then $(E \cup F)'$ is the event that ________.

Publishing Exercises 11–15 are based on the following table, which shows the results of a survey of 100 authors by a publishing company.

	New Authors	Established Authors	Total
Successful	5	25	30
Unsuccessful	15	55	70
Total	20	80	100

Compute the following probabilities of the given events

11. An author is established and successful
12. An author is a new author
13. An author is unsuccessful
14. An unsuccessful author is established
15. A new author is unsuccessful

16. A pharmaceutical company is running trials on a new test for anabolic steroids. The company uses the test on 400 athletes known to be using steroids and 200 athletes known not to be using steroids. Of those using steroids, the test is positive for 390 and negative for 10. Of those not using steroids, the test is positive for 10 and negative for 190. What is the estimated probability of a false negative result (the probability that an athlete using steroids will test negative)? What is the probability of a false positive result (the probability that an athlete not using steroids will test positive)?

17. Tony has had a "losing streak" at the casino—the chances of winning the game he is playing are 40%, but he has lost 5 times in a row. Tony argues that, since he should have won 2 times, the game must obviously be "rigged." Comment on his reasoning.

18. In 1999, the probability that a consumer would shop for holiday gifts at a discount department store was .80, and the probability that a consumer would shop for holiday gifts from catalogs was .42. Assuming that 90% of consumers shopped from one or the other, what percentage of them did both?

19. In 2001, 6.1% of all U.S. households were connected to the Internet via cable, while 2.7% of them were connected to the internet through DSL. What percentage of U.S. households did not have high-speed (cable or DSL) connection to the Internet? (Assume that the percentage of households with both cable and DSL access is negligible.)

20. In 2000 the top 100 chain restaurants in the U.S. owned a total of approximately 130,000 outlets. Of these, the three largest (in numbers of outlets) were McDonalds, Subway, and Burger King, owning between them 26% of all of the outlets. The two hamburger companies, McDonalds and Burger King, together owned approximately 16% of all outlets, while the two largest, McDonalds and Subway, together owned 19% of the outlets. What was the probability that a randomly chosen restaurant was a McDonalds?

21. A manufacturer of an electric kitchen utensil conducted a survey of consumer complaints. The results are summarized in the following table:

	Reason for Complaint			
	Electrical	Mechanical	Appearance	Totals
During Guarantee Period	18%	13%	32%	63%
After Guarantee Period	12%	22%	3%	37%
Totals	30%	35%	35%	100%

(a) Calculate the probability that a customer complains about appearance (dents, scratches, etc.) given that the complaint occurred during the guarantee time.
(b) Calculate the probability that a customer complains about appearance.

Publishing Exercises 22–27 are based on the following table, which shows the results of a survey of 100 authors by a publishing company.

	New Authors	Established Authors	Total
Successful	5	25	30
Unsuccessful	15	55	70
Total	20	80	100

Compute the following conditional probabilities:

22. That an author is established, given that she is successful
23. That an author is successful, given that he is established
24. That an author is unsuccessful, given that she is established
25. That an author is established, given that he is unsuccessful
26. That an unsuccessful author is established
27. That an established author is successful

28. I choose 2 marbles at random and without replacement from a sac containing 6 marbles labeled 1 to 6. Show that calculating the probability using combined events, M1 on the first draw (E), or M1 on the second draw(F), leads to the same probability we got before (=1/3). Careful here, the P(M1 on the first draw and M1 on the second draw) = 0, but $1/6 + 1/5 \neq 1/3$. Use the fact that $P(F) = P(F \cap E) + P(F \cap E')$, and then $P(F \cap E') = P(F|E')*P(E')$

Bayesian Analysis

"De Finetti did not accept P(A Π B)/P(B) as the definition of the conditional probability PB(A).. But he started, without quibble, with an assumption that originated with the measure-theorists: the giver B and bearer A of a conditional probability PB(A) are both timeless events—subsets of a sample space. Here is his definition of a person's subjective conditional probability PB(A): it is the price the person assigns to a contract that pays $1 if A and B both happens, pays $0 if B happens but A does not, and is cancelled (the payment is returned) if B does not happen. This definition applies even if P(B) = 0, but if P(B) > 0, then the person must make PB(A) equal to P(A Π B)/P(B) if he wants to avoid incoherence—i.e., if he wants to avoid setting prices that would enable an opponent to make money from him for certain." Ŧ

In subjective probability, we can throw away the notion of a set; however, if whatever we speak of can be interpreted using sets, then we want the same arguments to be valid using sets. So, it's reasonable to pull many of the ideas from sets into the subjective case. Here, in subjective probability, the definition of an event is:

An event is a logical entity which can assume the two values, true or false (that is: it did occur or it will happen, respectively, it did not occur or it will not happen).

Let E and F be events with P(F) ≠0.
P(E|F) = P(E Π F) / P(F). E Π F and E|F are often confused, mostly by the speaker of them. Suppose that E is the event that it rains tomorrow and F is the event that it is cloudy tomorrow. What's the difference between the events E Π F and E|F? In words,
E Π F is the event that it rains tomorrow and that it is cloudy tomorrow
E|F is the event that it rains tomorrow given that it is cloudy tomorrow
P(E Π F) is the portion of times it both rains and is cloudy
P(E|F) is the portion of times that it rains given that it's cloudy
= the portion of times it both rains and is cloudy / portion of times it is cloudy
F = (E Π F) U (E' Π F); P(F) = P(E Π F) + P(E' Π F)

This "portion of times" notion gives the set theoretic equivalence of the subjective mood. It can be used when the meaning of the subjective statements escape us.

Note that P(E|F) ≥ P(E Π F) and also that if E only occurs when F occurs, and P(F)≠0, P(E|F) = P(E Π F) / P(F) = P(E) / P(F).

Also, P(E|F) + P(E'|F) = 1. (P(E Π F) + P(E' Π F) = P(E Π F Π E' Π F) = P(F))

Ŧ The Notion of Event in Probability and Causality, Situating myself relative to Bruno de Finetti
Glenn Shafer, gshafer@andromeda.rutgers.edu

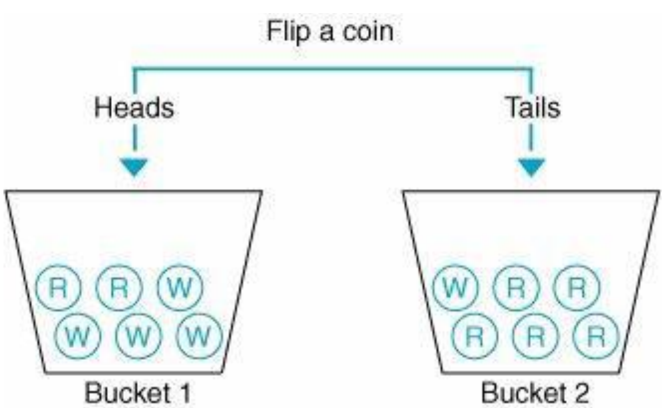
Bayes
Thomas Bayes gave a special case involving continuous prior and posterior probability distributions and discrete probability distributions of data, but in its simplest setting involving only discrete distributions, Bayes' theorem relates the conditional and marginal probabilities of events E and F, where F has a non-vanishing probability:

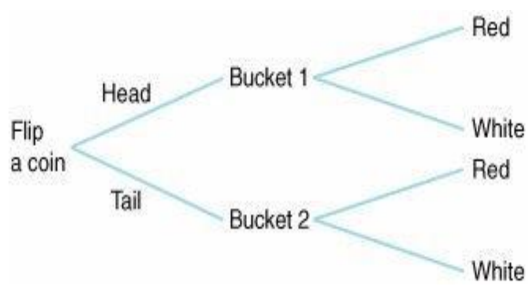
Each term in Bayes' theorem has a conventional name:
P(E) is the <u>prior</u> probability or marginal probability of E. It is "prior" in the sense that it does not take into account any information about F.
P(E|F) is called the <u>posterior</u> probability because it is derived from or depends upon the specified value of F.
P(F|E) is called the <u>likelihood</u>.

Bayes' theorem in its simplest form gives a mathematical representation of how the conditional probability of event E given F is related to the converse conditional probability of F given E.

You have already seen Bayes's Theorem, it's simply
P(E Π F) = P(E|F) P(F) = P(F|E) P(E) written as P(E|F) = P(F|E) P(E) / P(F)

Example -- USING A BAYSEIAN TREE
A two part experiment. Flip a fair coin. If heads lands up, draw a ball from unr1, if tails lands up, draw a ball from urn2. What is the probability that the ball drawn is from urn1 given that it is red?





We want P(H|R). We know P(H) = 1/2. P(R|H) = 2/6 = 1/3 , so P(R.H) = 1/2 * 1/3 = 1/6.
Also, P(T) = 1/2 and P(R|T) = 5/6, so P(R.T) = 1/2 * 5/6 = 5/12
P(H|R) = P(R.H) / P(R) = P(R.H) / (P(R.H) + P(R.T)) = 1/6 / (1/6 + 5/12) = 2/7.
(“.” represents “and”)

Example
A production manager for a manufacturing firm is supervising the machine setup for the production of a product. The machine operator sets up the machine. If the machine is set up correctly, there is a 10% chance that an item produced on the machine will be defective; if the machine is set up incorrectly, there is a 40% chance that an item will be defective. The production manager knows from past experience that there is a .50 probability that a machine will be set up correctly or incorrectly by an operator. In order to reduce the chance that an item produced on the machine will be defective, the manager has decided that the operator should produce a sample item. The manager wants to know the probability that the machine has been set up incorrectly if the sample item turns out to be defective.

The probabilities given in this problem statement can be summarized as follows:

$$P(\text{C}) = .50 \qquad P(\text{D}|\text{C}) = .10$$
$$P(\text{IC}) = .50 \qquad P(\text{D}|\text{IC}) = .40$$

Where C = correct; IC = incorrect; D = defective

The posterior probability for our example is the conditional probability that the machine has been set up incorrectly, given that the sample item proves to be defective, or P(IC|D). In Bayesian analysis, once we are given the initial marginal and conditional probabilities, we can compute the posterior probability by using Bayes's rule, as follows:

$$\begin{aligned} P(\text{IC} \mid \text{D}) &= \frac{P(\text{D}|\text{IC})P(\text{IC})}{P(\text{D}|\text{IC})P(\text{IC}) + P(\text{D}|\text{C})P(\text{C})} \\ &= \frac{(.40)(.50)}{(.40)(.50) + (.10)(.50)} \\ &= .80 \end{aligned}$$

Previously, the manager knew that there was a 50% chance that the machine was set up incorrectly. Now, after producing and testing a sample item, the manager knows that if it is defective, there is a .80 probability that the machine was set up incorrectly. Thus, by gathering some additional information, the manager can revise the estimate of the probability that the machine was set up correctly. This will obviously improve decision making by allowing the manager to make a more informed decision about whether to have the machine set up again by the same employee.

Homework S3

1. A service station owner sells Goodroad tires, which are ordered from a local tire distributor. The distributor receives tires from two plants, A and B. When the owner of the service station receives an order from the distributor, there is a .50 probability that the order consists of tires from plant A or plant B. However, the distributor will not tell the owner which plant the tires come from. The owner knows that 20% of all tires produced at plant A are defective, whereas only 10% of the tires produced at plant B are defective. When an order arrives at the station, the owner is allowed to inspect it briefly. The owner takes this opportunity to inspect one tire to see if it is defective. If the owner believes the tire came from plant A, the order will be sent back. Using Bayes's rule, determine the posterior probability that a tire is from plant A, given that the owner finds that it is defective.

2. A large research hospital has accumulated statistical data on its patients for an extended period. Researchers have determined that patients who are smokers have an 18% chance of contracting a serious illness such as heart disease, cancer, or emphysema, whereas there is only a .06 probability that a nonsmoker will contract a serious illness. From hospital records, the researchers know that 23% of all hospital patients are smokers, while 77% are nonsmokers. For planning purposes, the hospital physician staff would like to know the probability that a given patient is a smoker if the patient has a serious illness.

3. Two law firms in a community handle all the cases dealing with consumer suits against companies in the area. The Abercrombie firm takes 40% of all suits, and the Olson firm handles the other 60%. The Abercrombie firm wins 70% of its cases, and the Olson firm wins 60% of its cases. Determine the probability that the Olson firm handled a particular case, given that the case was won.

4. A metropolitan school system consists of three districts, north, south, and central. The north district contains 25% of all students, the south district contains 40%, and the central district contains 35%. A minimum-competency test was given to all students; 10% of the north district students failed, 15% of the south district students failed, and 5% of the central district students failed. What is the probability that a student selected at random failed the test?

5. In a certain country, during the winter, when it is cloudy in the morning it rains with probability ¼ and when it's not cloudy in the morning it doesn't rain with probability ¼ . Historically, the probability of rain in a given day is .2. When it's cloudy in the morning, Mr. P flips a fair coin to determine if he will take his umbrella. If heads comes up, he takes it, otherwise he doesn't. What is the probability that Mr. P has his umbrella, given that it rains?

6. In a poker hand, John has a very strong hand and bets 5 dollars. The probability that Mary has a better hand is .04. If Mary had a better hand she would raise with probability .9, but with a poorer hand she would only raise with probability .1.
If Mary raises, what is the probability that she has a better hand than John does?

7. You've been kidnapped by the Alkida. You are given two urns and fifty balls. Half of the balls are white and half are black. You are asked to distribute the balls in the urns with no restriction placed on the number of either type in an urn. You must distribute all the balls into the urns. Soon, you will be blindfolded and asked to draw a ball from an urn. If you draw a white ball, you can go free; if you draw a black ball, you will be killed. How should you distribute the balls in the urns? Justify your answer.

Discrete probability distribution functions

A discrete probability distribution function, f_X, is a function which associates a random variable X with a value between 0 and 1, inclusive. The notation used is $f_X(x) = P(X = x) = z$ to say that P(x) = z. The idea is that the function f_X is a picture of all of the individual probabilities. The probabilities must add to 1 as usual, and be non-negative:

$P(X = x) \geq 0$, and $\sum_x P(X = x) = 1$.

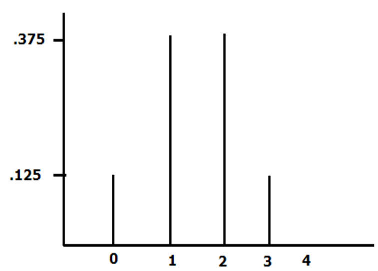
Example 1

Let the random variable X be the number of heads that come up when a coin is tossed three times—we obtain

the event that X = 0 is {TTT}	$P(x=0) = 1/8 = 0.125$
the event that X = 1 is {HTT, THT, TTH}	$P(x=1) = 3/8 = 0.375$
the event that X = 2 is {HHT, HTH, THH}	$P(x=2) = 3/8 = 0.375$
the event that X = 3 is {HHH}	$P(x=3) = 1/8 = 0.125$
the event that X = 4 is Ø	$P(x=4) = 0$

Picturing the Discrete Probability Distribution Function:

The most common way to represent the distribution is via a graph, such as the following for the above example.



Right away, from the picture, we can tell that the likelihood of a 1 or 2 is equal and about 3 times more likely than a 0 or 1, which are also equal likely.

There are two types of distribution functions corresponding to the two different types of random variables, discrete and continuous. The first type is a discrete distribution function, whose domain is a discrete set of real numbers – a set with gaps between the elements. The second type is a continuous distribution function, whose domain is a non-discrete set of real numbers.

The above example is an example of a discrete probability distribution function.

We have assumed that the random variable takes on real number values. In the case that it doesn't, then the sample space must be mapped into a new sample space of real values, the map preserving the set theoretic properties of the domain sample space. This mapping can then be taken as the random variable of the experiment, in that the outcomes of the experiment are in the range of it, and this range gives the domain of the distribution function.

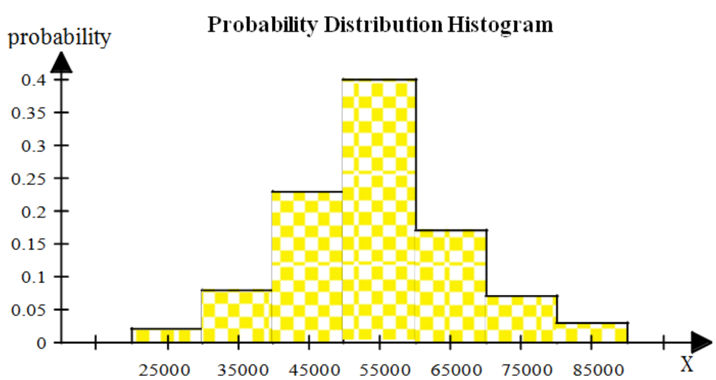
In general, sample data collected by experimentation gives us what's called a custom distribution. It is called custom because the data may not "fit" into a common, well known probability distribution. Inferential statistics concerns, in part, finding a probability distribution in which the sample distribution fits. That is, given a sample, is there a well-known probability distribution that describes the population and therefore, the sample. Intuitively, you should think that the best method of collecting data for fitting is to collect many samples, and do so in a proper, random way. There are many sampling techniques, and simple methods for determining the number of elements to include in the sample, to assure to any confidence factor, a good fit. Once this is done, the custom distribution might be found to be a common distribution. If not, it can be used just as is, empirically. The use may be to infer, to describe, or to obtain probabilities of occurrences of certain (other) samples.

Example:
An experiment consists of repeatedly sampling clusters of 100 attorneys in Ohio, where we're sure that the clusters cover all categories of attorneys in Ohio. Let X be the sample mean income range and put these X into classes of 10,000 starting with 20,000. Although the actual sampling random variable is continuous, using classes (income brackets) allows us to approximate it by a discrete random variable.

Suppose the following table summarizes the sampling information.

Class	25000	35000	45000	55000	65000	75000	85000
Frequency	2	8	23	40	17	7	3

From the table we see that the mean value was between 20,000 and 30,000 two times; it was between 30,000 and 40,000 eight times, etc.



It's fairly obvious that this distribution shouldn't fit a common one, it could be thought of as the "Ohio attorney salary distribution, 2010". Statistical analysis could be used to infer the mean and standard deviation of the actual population from the sample information. The inferred population distribution will have at least two population parameters, a mean and a standard deviation. A probabilist would use these population parameters in computing probabilities involving other samples of the population.

The probabilist works with samples of populations with known parameters to make inferences about the samples, just as we did in the previous sections of this chapter. The statistician works with samples of populations of unknown parameters to make inferences about the populations.

Mean (Expected Value) and Standard Deviation of a Random Variable
In the example above, let's suppose that the population exactly conforms to the sample so that we can compute the population mean and standard deviation from the sample data. To get the mean of the distribution, we add the outcomes and divide by the number of them:

2 attorneys @ $25,000 .... $50,000
8 attorneys @ $35,000 .... $280,000
23 attorneys @ $45,000 .... $1,035,000
40 attorneys @ $55,000 .... $2,200,000
17 attorneys @ $65,000 .... $1,105,000
7 attorneys @ $75,000 .... $525,000
3 attorneys @ $85,000 .... $255,000
$5,450,000

So the mean is $5,450,000/100 = $54,500

Method 2 (Using the probability distribution) Since we have multiplied each class midpoint, the x value, by its frequency and then divided the sum of these by the sum of the frequencies, we might as well have just multiplied each x value by its probability, and then added. This would result in the same answer:

The expected value or mean of a discrete distribution function $f_X(x) = P(X = x)$ is

$$E(X) = \sum_x x\,P(X = x)$$

= 25,000 *.02 + 35,000 * .08 + …. + 85,000 * .03

The same argument holds for computing the variance so,

$$\text{Var}(X) = \text{E}((X - E(X))^2) = \sum_x (x - E(X))^2\ \text{P}(x)$$

$$= (25{,}000 - 54{,}500)^2 * .02 + (35{,}000 - 54{,}500)^2 * .08 + \ldots + (85{,}000 - 54{,}500)^2 * .03$$

The expectation of X, $E(X)$ is generally denote by "μ", and the standard deviation = $\sqrt{\text{Var}(X)}$ is denoted by "σ", and sometimes, "sd" (standard deviation).

Note that the variance of X is the expected value (mean) of the squares of the differences of the x values and the mean. And then we take the square root of that to get the standard deviation. A much better method is to take the absolute values of the differences of the x values and the mean, without squaring. This method gives the Mean Absolute Deviation, always less than the standard deviation and a more intuitive measure of dispersion. It is also more accurate when there are errors in the measurements of the sample data. The reason it is not used in the theoretical development of statistics and probability is because the standard deviation defined as the squares of the differences mathematically makes the rest of the theory harmonious. The absolute value function is too awkward to fit in theoretically, however, using it actually gives better results, and many statisticians do actually use it.

See: Revisiting a 90-year-old debate: the advantages of the mean deviation, Stephen Gorard, Department of Educational Studies, University of York, Paper presented at the British Educational Research Association Annual Conference, University of Manchester, 16-18 September 2004

The above distribution is important only to someone studying the salaries of attorneys. It is not general enough to be useful in a wide range of applications. I should also point out that a statistician looks at small samples as being a bit different from the populations. The mean of a sample is calculated the same as the mean of a population, but the sample variance is calculated using a slightly different method, namely $\sum_x fx(x - E(X))^2$/ N-1, where fx is the frequency of x, and N is the sum of all the frequencies. It turns out for small values of N this is a closer estimate of the true sample variance. The statistician always differentiates a sample variance from a population variance by denoting the former as “s” and the latter as “σ”, and also differentiates the sample mean from the population mean by denoting the former as “ $\bar{x}$ ” and the latter as “μ” even though they are computed in the same manner.

The common and important distributions

There are many well know, standard probability distributions. To name a few most popular ones, there is the discrete Binomial, Uniform, and Poisson; and there is the continuous Uniform, Normal, and Exponential.

Distributions are really more than one distribution, they are families of distributions. They take on different forms given the values of their parameters. (Pronounced par-ram-it-ters, with emphasis on “ram” and not on “met”). For instance, the binomial has 2 parameters, the number of trials and the probability of a success on each trial (which does not change from trail to trail). So for each pair, N, a natural number, and P, a real number between 0 and 1, we get a binomial distribution. The distribution in example 1 is a binomial distribution with N=3 and P=.5.

Discrete Distributions	Parameters	Domain of f_X
Binomial	N,P	0,…,N
Uniform	μ=1/(b-a)	I
Poisson	μ	I+

Continuous Distributions	Parameters	Domain of F_X
Uniform	μ=(a+b)/2	in R
Normal	μ, σ	R
Exponential	μ	R+

The domain is where a particular distribution may be defined. Here I+ is the positive Integers and R+ the positive Reals. The capital F_x refers to the fact that instead of a distribution function, a predominant function used for continuous distributions is the probability density function (pdf), denoted by a capital F. $F_x(x)$ is the area under the pdf, from X = -∞ to x. This is necessary because for continuous function, $f_x(x)$ = P(X=x) is 0, due to the non-discrete nature of the domain.

μ (pronounced mew) is generally used to denote the mean of the distribution and σ (sigma) is generally used to denote its standard deviation. For the most part, distributions are defined by their means and standard deviations. Even the binomial is defined that way, since any pair of appropriate N, and P, give a unique μ and δ.

What is the mean and variance of the binomial distribution?

$\mu = np$, $\sigma^2 = np(1-p)$

You can see the derivation for this in any good book on probability.

The common distributions and their functions have names, and the names are shared with the random variables on which they are defined.

Example

The Discrete Uniform Distribution

$f_X(x) = P(X = x) = 1/C$ for some constant C, and defined over some set of C equally likely outcomes that are real numbers. The variable X is said to be uniformly distributed. f_X is a uniform distribution function.

Example

I have 10 marbles label 1, …., 10 in a sac. The experiment is to choose a marble from the sac and observe its label. The random variable is "the marble is labeled with" and it is uniformly distributed with, for x=1,…,10, P(X = x) = 1/10. The label of the marble is uniformly distributed. $f_X(x) = 1/10$ is a uniform distribution function.

The discrete uniform distribution is important because it is often chosen as the default distribution when a true distribution for the random variable is unknown. For example, if you went fishing in a pound stocked with fish, you would use a discrete uniform random distribution to calculate the probability of you catching a fish, even though you may know that there are other factors involved in the experiment, like the depth of the water you are fishing in, the time of day, the type of bait you are using, the number of times you cast out, etc. That is, you may assume that catching a fish is only dependent on how man fish there are in the pond - independent of the time, place, method, to some degree.

Most of the well-known distributions have evolved historically from performing experiments. For example, the uniform distribution evolved as a description of the probabilities of events in sampling from a set of elements whose probability of occurrence is equal. The binomial distribution involves a certain type of experiment, called a binomial experiment, in which there are trials.

A binomial experiment is defined by the following conditions:

1. There are n "trials" where n is determined in advance and is not a random value.
2. There are two possible outcomes on each trial, called "success" and "failure".
3. The outcomes are independent from one trial to the next.
4. The probability of a "success" remains the same from one trial to the next.

Generally speaking, the probability of a success is denoted by “p” and that of failure, by “q”; and p = 1- q.

The trials in a binomial experiment are termed Bernoulli Trials, after Jakob Bernoulli (1654–1705).

A **binomial random variable** is defined as *X* = number of successes in the *n* trials of a binomial experiment, with probability p of a success on each trial.
$f_X(x) = P(X=x) = C(n,x)p^x q^{n-x}$

The number of heads in three tosses of a fair coin, the number of girls in six independent births, and the number of men who are six feet tall or taller in a random sample of ten adult men from a large population are all examples of binomial random variables. The number of fish you may catch in that pound stocked with (a large number of) fish is also a binomial random variable.

Random Circumstance	Random Variable	Success	Failure	*n*	*p*
(1) Toss three fair coins	*X* = number of heads	Head	Tail	3	1/2
(2) Roll a die eight times	*X* = number of 4s and 6s	4, 6	1, 2, 3, 5	8	2/6 = 1/3
(3) Randomly sample 1000 U.S. adults	*X* = number who have seen a UFO	Seen UFO	Have not seen UFO	1000	Proportion of all adults who have seen a UFO
(4) Roll two dice once	*X* = number of times sum is 7	Sum is 7	Sum not 7	1	6/36 = 1/6

Sample surveys can produce a binomial random variable when we count how many individuals in the sample have a particular opinion or trait (see the third example in the table above). A “trial” is one sampled individual. A “success” is that the individual has the opinion or trait. The probability of “success” is the proportion in the population who have the opinion or trait. If a random sample is taken without replacement from a large population, the conditions of a binomial experiment are considered to be met, although the probability of a “success” actually changes slightly from one trial to the next as each sampled individual is removed from eligibility.

Sampling without replacement from a small population does not produce a binomial random variable. Suppose a class consists of ten boys and ten girls. Five children are randomly selected to be in a play. Let *X* be the number of girls selected. Notice that *X* is *not* a binomial random variable because the probability that a girl is selected each time depends on who is already in the sample and who is left, violating the condition that the probability of “success” must remain the same on each trial. *X* is an example of a *hypergeometric random variable.*

Any individual random circumstance can be treated as a binomial experiment with $n = 1$ and p = probability of a particular outcome. In this case, the random variable X is either 0 or 1, and the random variable may also be called a Bernoulli random variable. The fourth example in the table above illustrates such a variable.

More examples of Binomial Random Variables

1. Roll a die 10 times and let X be the number of times you roll a six.

2. Provide a property with flood insurance for 20 years; let X be the number of years, during the 20-year period, in which the property experiences a flooded. (Assume that flooding is independent from one year to the next.)

3. 60% of all bond funds will depreciate next year, and you randomly select 4 from a very large number of possible choices; X is the number of bond funds you hold that will depreciate next year. (X is approximately binomial. Since the number of bond funds is extremely large, choosing a "loser" (a fund that will depreciate next year) does not significantly deplete the pool of "losers," and so the probability that the next fund you choose will be a "loser," is hardly affected. Hence we can think of X as being a binomial variable.)

Example
Probability Distribution of a Binomial Random Variable

Suppose that we have a possibly unfair coin, whose probability of heads is p and whose probability of tails is q = 1-p.

(a) Let X be the number of heads you get in a sequence of 5 flips. Find P(X = 2)
(b) Let X be the number of heads you get in a sequence of n flips. Find P(X = x).

Solution

(a) Call the flips F_1 F_2 F_3 F_4 F_5. We want the probability that exactly two flips are heads. There are exactly C(5,2) pairs i,j such that F_i and F_j are both heads that account for all the sequences of flips with exactly 2 heads. Each sequence of 5 flips with exactly 2 head flips has tails in the remaining 3 flips. Since the flips are independent, the probability of one such sequence of flips is the product of the probabilities of the individual flips, which is p*p for the heads, and q*q*q for the tails. So each flip sequence has probability p^2q^3. And there are C(5,2) mutually exclusive sequences with 2 heads and 3 tails, so we can add the probabilities of each sequence to get the probability of the event that we get any configuration of 2 heads and 3 tails, giving us C(5,2) p^2q^3. (We've used the basic definition that the probability of the event is the sum of the probabilities of the outcomes that make the event happen.)

(b) There is nothing special about 5 and 2. The general formula for n flips is
$P(X=x) = C(n,x)p^xq^{n-x}$.

Questions

100 customers must select a preference among three sodas: your company's new Hyper Cola and the two competitors (you know what they are...). Success, of course, means selecting Hyper Cola. Is this binomial?

You select 3 bonds from 10 recommended ones. Unbeknownst to you, 8 of them will go up, and three are stones. x is the number of winners you select. Is this binomial?

You select 3 bonds from a large number of recommended ones. Unbeknownst to you, 80% of them will go up, and 20% are stones. x is the number of winners you select. Is this binomial?

Example
Will You Still Need Me When I'm 64?
The probability that a randomly chosen person in the US is 65 or older is approximately 0.2.

(a) What is the probability that, in a randomly selected sample of 6 people, exactly 4 of them are 65 or older?
(b) If X is the number of people of age 65 or older in a sample of 6, construct the probability distribution of X and plot its graph.
(c) Compute $P(X \leq 2)$.
(d) Compute $P(X \geq 2)$.

Solution

(a) The experiment is a sequence of Bernoulli trials; in each trial we select a person and ascertain his age. If we take “success” to mean selection of a person 65 or older, the probability distribution is

$P(X=x) = C(n,x)p^x q^{n-x}$

Where n = number of trials = 6,
p = probability of success = 0.2, and
q = probability of failure = 0.8.

So, P(X = 4) = C(6,4)(0.2)4(0.8)2 = 15 * 0.0016 * 0.64 = 0.01536

(b) We have already computed P(X = 4). Here are all the calculations.
P(X = 0) = C(6,0)(0.2)0(0.8)6 = 1*1*0.262144 = 0.262144
P(X = 1) = C(6,1)(0.2)1(0.8)5 = 6*0.2*0.32768 = 0.393216
P(X = 2) = C(6,2)(0.2)2(0.8)4 = 15*0.04*0.4096 = 0.24576
P(X = 3) = C(6,3)(0.2)3(0.8)3 = 20*.008*0.512 = 0.08192
P(X = 4) = C(6,4)(0.2)4(0.8)2 = 15*0.0016*0.64 = 0.01536
P(X = 5) = C(6,5)(0.2)5(0.8)1 = 6*0.00032*0.8 = 0.001536
P(X = 6) = C(6,6)(0.2)6(0.8)0 = 1*0.000064*1 = 0.000064

The probability distribution is:

x	0	1	2	3	4	5	6
P(X=x)	0.262144	0.393216	0.24576	0.08192	0.01536	0.001536	0.000064

(c) P(X ≤ 2), the probability that the number of people selected who are at least 65 years old is either 0, 1, or 2, is the union of these events, and is thus the sum of the three probabilities, P(X ≤ 2) = P(X = 0) + P(X = 1) + P(X = 2)
= 0.262144 + 0.393216 + 0.24576 = 0.90112.

(d) To compute P(X ≥ 2), we could compute the sum
P(X ≥ 2) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6), but it is far easier to compute the probability of the complement of the event,
P(X < 2) = P(X = 0) + P(X = 1) = 0.262144 + 0.393216 = 0.65536,
and then subtract the answer from 1:
P(X ≥ 2) = 1 - P(X < 2) = 1 - 0.65536 = 0.34464.

Example

(a) Your manufacturing plant produces 10% defective airbags. If the next 5 airbags are tested, find the probability that three of them are defective.
(b) Compute the probability distribution (that is, find p(0), p(1), ..., p(5)), graph them, and locate μ and δ on the graph.
(c) What fraction of the outcomes will fall within 2 standard deviations of the mean?

Answer to part (b)

$$P(X=x) = \binom{5}{x}(0.1)^x(0.9)^{5-x}.$$

Thus: $P(X=0) = \binom{5}{0}(0.1)^0(0.9)^{5-0} = 0.59049$

$$P(X=1) = \binom{5}{1}(0.1)^1(0.9)^{5-1} = 0.32805$$

$$P(X=2) = \binom{5}{2}(0.1)^2(0.9)^{5-2} = 0.07290$$

$$P(X=3) = \binom{5}{3}(0.1)^3(0.9)^{5-3} = 0.0081$$

$$P(X=4) = \binom{5}{4}(0.1)^4(0.9)^{5-4} = 0.00045$$

$$P(X=5) = \binom{5}{5}(0.1)^5(0.9)^{5-5} = 0.00001$$

Answer to (c) We calculate $\mu = 0.5$, and $\sigma = 0.67$. Thus, the interval is

$$[\mu - 2\sigma, \mu + 2\sigma] = [-0.84, 1.84]$$

These are values of x, and the interval includes $x = 0$ *and* 1. Since

$P(X=0 \text{ or } X=1) = 0.59049 + 0.32805 = 0.9185$, we conclude that at least 91.85% of the outcomes will be within 2 standard deviations of the mean.

Homework S4

1. A manufacturing company has 10 machines in continuous operation during a workday. The probability that an individual machine will break down during the day is .10. Determine the probability that during any given day 3 machines will break down.

2. A polling firm is taking a survey regarding a proposed new law. Of the voters polled, 30% are in favor of the law. If 10 people are surveyed, what is the probability that 4 will indicate that they are opposed to the passage of the new law?

3. An automobile manufacturer has discovered that 20% of all the transmissions it installed in a particular style of truck one year are defective. It has contacted the owners of these vehicles and asked them to return their trucks to the dealer to check the transmission. The Friendly Auto Mart sold seven of these trucks and has two of the new transmissions in stock. What is the probability that the auto dealer will need to order more new transmissions?

4. A new county hospital is attempting to determine whether it needs to add a particular specialist to its staff. Five percent of the general hospital population in the county contracts the illness the specialist would treat. If 12 patients check into the hospital in a day, what is the probability that 4 or more will have the illness?

5. A retail outlet receives radios from three electrical appliance companies. The outlet receives 20% of its radios from A, 40% from B, and 40% from C. The probability of receiving a defective radio from A is .01; from B, .02; and from C, .08. What is the probability that a defective radio returned to the retail store came from company B?

6. A metropolitan school system consists of three districts, north, south, and central. The north district contains 25% of all students, the south district contains 40%, and the central district contains 35%. A minimum-competency test was given to all students; 10% of the north district students failed, 15% of the south district students failed, and 5% of the central district students failed. What is the probability that a student selected at random failed the test?

7. The Ramshead Pub sells a large quantity of beer every Saturday. From past sales records, the pub has determined the following probabilities for sales:

Barrels	Probability
6	.10
7	.20
8	.40
9	.25
10	.05

Compute the expected number of barrels that will be sold on Saturday.

8. A market in Boston orders oranges from Florida. The oranges are shipped to Boston from Florida by railroad, truck, or airplane; an order can take 1, 2, 3, or 4 days to arrive in Boston once it is placed. The following probabilities have been assigned to the number of days it takes to receive an order once it is placed (referred to as lead time):

Lead Time	Probability
1	.20
2	.50
3	.20
4	.10

Compute the expected number of days it takes to receive an order and the standard deviation.

9. An investor is considering two investments, an office building and bonds. The possible returns from each investment and their probabilities are as follows:

Office Building		Bonds	
Return	Probability	Return	Probability
$50,000	.30	$30,000	.60
60,000	.20	40,000	.40
80,000	.10		1.00
10,000	.30		
0	.10		
	1.00		

Using expected value and standard deviation as a basis for comparison, discuss which of the two investments should be selected.

REVIEW A

1. Four people are to be arranged in a row to have their picture taken. In how many ways can this be done?

2. In how many ways can we choose five people from a group of ten to form a committee?

3. How many seven-element subsets are there in a set of nine elements?

4. Charles claims that he can distinguish between beer and ale 75 percent of the time. Ruth bets that he cannot and, in fact, just guesses. To settle this, a bet is made: Charles is to be given ten small glasses, each having been filled with beer or ale, chosen by tossing a fair coin. He wins the bet if he gets seven or more correct.
   a. Find the probability that Charles wins if he has the ability that he claims.
   b. Find the probability that Ruth wins if Charles is guessing.

5. A lady wishes to color her fingernails on one hand using at most two of the colors red, yellow, and blue. How many ways can she do this?

6. A coin is tossed three times. What is the probability that exactly two heads occur, given that
   (a) the first outcome was a head?
   (b) the first outcome was a tail?
   (c) the first two outcomes were heads?
   (d) the first two outcomes were tails?
   (e) the first outcome was a head and the third outcome was a head?

7. A die is rolled twice. What is the probability that the sum of the faces is greater than 7, given that
   (a) the first outcome was a 4?
   (b) the first outcome was greater than 3?
   (c) the first outcome was a 1?
   (d) the first outcome was less than 5?

8. A card is drawn at random from a deck of cards. What is the probability that
   (a) it is a heart, given that it is red?
   (b) it is higher than a 10, given that it is a heart? (Interpret J, Q, K, A as 11, 12, 13, 14.)
   (c) it is a jack, given that it is red?

REVIEW B

1) On a certain Psychology exam, 20% of the students earned a score of 90 or above. It was also true that 20% of the male students earned a score of 90 or above. Are the events "earning a score of 90 or above" and "being male" independent?

2) The employees of a company were surveyed on questions regarding their educational background and marital status. Of 600 employees, 400 had college degrees, 100 were single, and 60 were single college graduates. What is the probability that an employee of the company is single or has a college degree?

3) The table below displays the probabilities for each of the six outcomes when rolling a particular unfair die. Suppose that the die is rolled once. Let A be the event that the number rolled is less than 4 and let B be the event that the number rolled is odd. Find P(A|B).

Outcome	1	2	3	4	5	6
Probability	.1	.1	.1	.2	.2	.3

4) An urn contains 10 balls, 6 red and 4 black. I select 4 balls from the urn at random, one by one, and put them aside. What is the probability that I have selected all black balls.

5) Urn 1 contains 10 balls, 6 red and 4 black. Urn 2 contains 12 balls, 6 red and 6 black. I draw 3 balls from the urns, choosing an urn with probability ½ at each drawing. What is the probability that I draw 3 red balls.

6) A dice game involves rolling three dice and betting on one of the six numbers that are on the dice. The game costs $7 to play, and you win if the number you bet on appears on any of the dice. The distribution of the outcomes of the game (including the profit) is shown below:

Number of dice with your number	Profit	Probability
0	-7	125/216
1	7	75/216
2	9	15/216
3	21	1/216

What is your expected profit from playing this game?