FORWARD

Alice, age 21, had a single sexual encounter with a man she met on a ski trip. A few months later she became concerned that she might have become HIV infected. With some hesitation, she had a test for HIV and it came back positive. Upset, she told her friend Jan about the results of the test. “Don’t be too worried," said Jan “the odds are still 24 to 1 against you having HIV”. She has the test repeated and this second test is positive. She tells Jan of the result of the second test. Jan tells her that now the odds are about 3 to1 of her being infected. Is Jan right in her claims about the odds?

Experience shows that not everyone who reserves space on an airline flight actually shows up for the flight. For this reason airlines overbook flights (because the loose money if they fly with empty seats). What is the chance that you will actually get a seat on a flight that you book?

In the television game show “Let's Make a Deal” a contestant is asked to pick one of three doors. Behind one of the doors is a substantial prize and there is nothing behind the other two doors. The host of the show then opens one of the remaining two doors to show nothing is behind it. He the then offers the contestant the opportunity to switch the door the contestant picked with the remaining door. The problem of determining the chance for the contestant to win the prize if the contestant switched gained some popularity in 1990 because of its appearance in the “Ask Merilyn” column of Parade Magazine. Many people (including some mathematicians) wrote in with arguments for various possibilities such as 1/3, 1/2, and 2/3. What is the correct chance?

Martha is a 58-year-old woman who entered menopause about six years ago. She has heard about HRT (Hormone Replacement Therapy) but is somewhat hesitant to go on it because she has heard both pro and con on its benefits. She seeks advice from her doctor. She tells Martha that HRT can substantially reduce the rate of progression of osteoporosis (a loss of bone) that occurs in all women after menopause). She also informs Martha that recent studies show that HRT can cut the risk of cardiovascular disease by 50%. However, some studies show that HRT can increase the risk of breast cancer. These results are mixed. Some studies show that HRT has no effect on the risk of breast cancer while others show it may double that risk. If you were the doctor what would you advice Martha to do?

What do mutations on a chromosome, cosmic rays, and raisons in a cake have in common?

Why do the gambling casinos make so much money? Why do all casinos have a house limit on the size of the bet you can make at any one play?

Questions of the above type are answered by use of a branch of mathematics known as probability theory. Probability theory enables us to use calculation rather than speculation to answer such questions. In this course, we will investigate some of the basic concepts of this subject. You will learn enough to supply the answers to questions the above type. We will not attempt to cover the subject in a comprehensive manner. Instead, we will focus on just a few very basic ideas of widespread use. These ideas will be then be used to solve some real problems that arise primarily in biology and medicine.

TABLE OF CONTENTS

CHAPTER 1
BASIC PROBABILITY

1 RANDOM VARIABLES

Consider dropping a stone from a 100 foot height in a vacuum. Let T be the time the stone hits the ground measured to the nearest half second. If we drop the same stone 50 times, or if we drop 50 different stones once each, we will always find that they hit the ground at T = $\sqrt{100/16}$ sec. = 2.5 sec.

The variable T is an example of a deterministic variable. A deterministic variable is one that always has the same value if it is observed under identical conditions

Now consider measuring the systolic blood pressure, X, of an individual. In Table 1 we summarize the result of the measurement of the systolic blood pressure of 880 persons.

TABLE 1-1

blood pressure	Count	Percent
85 To 90	2	0.23
90 To 95	6	0.68
95 To 100	22	2.50
100 To 105	38	4.32
105 To 110	79	8.98
110 To 115	136	15.45
115 To 120	169	19.20
120 To 125	164	18.64
125 To 130	131	14.89
130 To 135	82	9.32
135 To 140	32	3.64
140 To 145	15	1.70
145 To 150	4	0.45

As can be seen, unlike the time T, there is considerable variability in the values that the blood pressure X can assume. A variable such as X is called a random variable, and the experiment of observing a value of X is a random experiment. Unlike a deterministic variable, the observation of a random variable under identical conditions does not always yield the same value. For a deterministic variable we can predict in advance of observing it what its value will be. For a random variable no such precise prediction is possible.

At first it might seem that we could say nothing about the values that a random variable can take (other than perhaps trivia such as that it must have a non-negative value if is known to represent a non-negative quantity, etc.). But that is not the case. Based on the evidence in Table 1-1 it seems that it is much more likely a blood pressure will be between 120 and125 than between 140 and 145. So, although we cannot precisely predict the exact value of X, we might be able to rather accurately predict how likely it is that the value of X will fall in some specified subset A of the line.

We make precise this idea of how likely it is that the random variable X has a value the set A by the concept of the distribution of X. Suppose that for each subset of numbers A we are given a number that is between 0 and 1 called the probability that X has a value in the set A. We denote that number by $P(X \leq A)$. [Read $P(X \in A)$ as "probability X has a value in A"]. The distribution of X is the collection of all the probabilities $P(X \in A)$ as A runs over all subsets of the line. Since there are infinitely many subsets we clearly cannot list all of these probabilities. Instead, we give a rule for how to compute each of the probabilities. Often the rule itself is called the distribution of A.

2 DISCRETE RANDOM VARIABLES

2.1 PROBABILITY FUNCTION

The simplest kind of random variables are those that can have only finitely many different values. These are called discrete random variables. Call the values that X can have x_1, x_2, $\ldots x_n$. Let $P(X = x_i)$ be the probability that X has the value x_i. The collection of numbers, $\{P(X = x_1), \ldots, P(X = x_n)\}$, is called the probability function of X. When n is small we often give this as a table:

X	x_1	…	x_n
P(X=x)	$P(X=x_1)$	…	$P(X=x_n)$

Mathematically, The probabilities $P(X_i = x_i)$ can be any numbers that satisfy the following two requirements:

(i) $0 \le P(X_i = x_i) \le 1$,

(ii) $P(X = x_1)+\ldots+ P(X = x_n) = 1$.

COMMENTS

P(…) is to always be read as probability of what's ever in (…), e.g. P(X<3) is read "probability that X<3".

It is customary to denote random variables by capital letters such as X, Y, etc. A generic value of a random variable is denoted by the lower case letter of the capital letter used to denote the random variable. For example, if X denotes the random variable, then a value of X is denoted by x. Consequently, P(X=x) denotes the probability that X has the value x.

2.2 Computation of $P(X\in A)$

For a discrete random variable X we compute $P(X\in A)$ by the rule:

$P(X\in A)$ = sum of all $P(X_i = x_i)$ for x_i in A. (1.1)

Symbolically we write "sum of all $P(X_i = x_i)$ with x_i in A" as. $\sum_{x\in A} P(X = x)$. [Σ is universally used in mathematics to denote sum]. Thus we write the above as

$$P(\in A) = \sum_{x\in A} P(X = x)$$

EXAMPLE 1.1

Suppose X is a random variable having the following probability function:

x	-2	-1	1	2	3	5	6	8
P(X=x)	0.1	0.05	0.1	0.2	0.1	0.05	0.15	0.25

Find (a) $P(X> 2)$, (b) $P(X\ge 2)$, (c) $P(-1< X \le 3)$, (d) $P(X> 3)$

SOLUTION

(a) $P(X> 2) = P(X = 3) + P(X = 5) + P(X = 6) + P(X = 8) = 0.1+ 0.05 + 0.15 + 0.25$.

= 0.55.

(b) $P(X \geq 2) = 0.2 + 0.1 + 0.05 + 0.15 + 0.25 = 0.75$.

(c) $P(-1 < X \leq 3) = 0.1+0.2+0.1 = 0.4$.

(d) $P(X > 3) = 0.05 + 0.15 + 0.25 = 0.45$

EXAMPLE 1.2

Suppose X has probability function

x	0	1	2	3
P(X=x)	0.2		0.3	0.1

Find $P(X = 1)$.

SOLUTION

The sum of the probabilities must be 1.Therefore $P(X=1) = 1 – (0.2+0.3+0.1) = 0.4$

2.3 INDICATOR VARIABLE

One of the simplest discrete random variables is an indicator (also called a Bernoulli) random variable. An indicator random variable X can only have the values 1 and 0. Since $P(X = 1) + P(X = 0) = 1$, an indicator random variable is specified by giving the probability $p = P(X = 1)$. The probability p is called the success probability. Indicator variables are used to model the occurrence (X = 1) or non-occurrence (X = 0) of some specified event, e.g. head or tail on a coin toss, gender, presence or absence of a risk factor, etc. Often we will refer to an experiment whose outcome is modeled by an indicator random variable as a coin flip and success as a head. We will see many applications of indicator variables in later material.

3 JOINT DISCRETE DISTRIBUTIONS

There is not much we can do with just a single random variable. In most cases we have several random variables we want to consider simultaneously. For example, we might want to consider how blood factors such as rh, blood type, depend on race and sex. In this case we would be interested in looking at 4 random variables simultaneously, viz. age, sex, rh, and blood type.

3.1 RANDOM PAIRS

Suppose X and Y are two discrete random variables. To consider them simultaneously we use the random pair (X,Y). If X has values $x_1, \ldots, x_m$ and Y has values $y_1,\ldots,y_n$ then the random pair (X,Y) has the nm pairs (x_i,y_j), i = 1,…,m, j =1,…,n as values. The probability function for the random pair (X,Y) are the nm probabilities $P((X,Y) = (x_i,y_j))$. Mathematically, these nm probabilities can be any numbers between 0 and 1 that sum to 1.

If n and m are small we often write this probability function as a two-way table.

Table 1-2Table 1-3

		y			
		y_1	y_2	…	y_n
x	x_1	$P(X,Y) = (x_1,y_1))$.	$P(X,Y) = (x_1,y_2))$.	…	$P(X,Y) = (x_1,y_n))$.
	x_2	$P(X,Y) = (x_2,y_1))$.	$P(X,Y) = (x_2,y_2))$.	…	$P(X,Y) = (x_2,y_n))$.
	.				
	.	.	.	…	.
	.	.	.	.	.
	x_m	$P(X,Y) = (x_m,y_1))$.	$P(X,Y) = (x_m,y_2))$.		$P(X,Y) = (x_m,y_n))$.

NOTATION

To say the random pair (X,Y) has the value (x,y) [i.e. (X,Y) = (x,y)] is the same as saying that X = x and Y = y. For this reason an alternate way of expressing the probability P((X,Y) = (x,y)) is P(X = x and Y = y). Often we replace the "and" with a comma ",". Thus, P((X,Y) = (x,y)), P(X = x and Y = y), P(X = x,Y = y) all denote the same probability. The probability function of (X,Y) is called the often called the joint probability function of X and Y.

We frequently want to compute the probability of some specified relationship between X and Y. For example, $P(X < Y)$. Any relationship between X and Y can always be expressed as $(X,Y) \in A$, where A is a subset of the set of all pairs. For example $X < Y$ is the same as saying $(X,Y) \in A$, where $A = \{(x,y) : x < y\}$.

The distribution of the random pair (X,Y) is the collection of the probabilities $P((X,Y) \in A)$ where A is now a subset of the set of all pairs (x,y). To compute

P((X,Y) A we proceed just as for a single variable. We sum the probabilities P(X = x_i ,Y = y_j) over all the pairs (x_i,y_j) that are in A. In symbols:

$$P((X,Y) \in A) = \sum_{(x,y)\in A} P(X,Y) = (x,y))$$

EXAMPLE 1.3

Suppose the joint probability function of X and Y is given by the following table:

TABLE 1-4

		y		
		1	2	3
x	1	.05	0.1	0.1
	2	0.2	0.05	0.25
	3	0.05	0.1	0.1

Find (a) P(X = Y) (b) P(X<Y)

SOLUTION

(a) Let A = {(x,y):x = y} = {(1,1), (2,2), (3,3)} Then P(X = Y) = P((X,Y)$\in$ A) = P(X = 1,Y = 1) + P(X=2,Y=2) + P(X=3,Y=3) = .05+.05+0.1 = 0.2

(b) Here A = {(x,y):x < y} ={(1,2),(1,3),(2,3)}. P(X < Y) = P(X = 1,Y = 2) +P(X = 1,Y=3) +P(X =2,Y =3). =0.1+0.1+ 0.25 = .45

If we know the joint probability function of X and Y we can compute the probability function of X and the probability function of Y.

$P(X = x_i) = P(X = x_i, Y = y_1) + \ldots + (X = x_i, Y = y_n)$ (1.2)

$P(Y = y_j) = P(X = x_1, Y = y_j) + \ldots + (X = x_m, Y = y_j)$ (1.3)

When the joint probabilities are given in a table such as Table 1.2 equation (1.2) tells us to compute P(X = x_i) by adding all the probabilities in row i of the table. Similarly, (1.3) tells us to compute P(Y = y_j) by adding all the probabilities in the j^{th} column of the table.

TABLE 1-5

		y				
		y_1	y_2	…	y_n	P(X=x)
	x_1	$P((X,Y) = (x_1,y_1))$.	$P((X,Y) = (x_1,y_2))$.	…	$P((X,Y) = (x_1,y_n))$.	$P(X=x_1)$
	x_2	$P((X,Y) = (x_2,y_1))$.	$P((X,Y) = (x_2,y_2))$.	…		$P(X=x_2)$
x	.					
	.	.	.	…	.	.
	.	.	.	.	.	.
	x_m	$P((X,Y) = (x_m,y_1))$.	$P((X,Y) = (x_m,y_2))$.		$P((X,Y) = (x_m,y_n))$.	$P(X=x_m)$
P(Y = y)		$P(Y=y_1)$	$P(Y=y_2)$		$P(Y=y_n)$	

When the joint probability is given in a table as above we frequently write the probabilities $P(X = X_i)$ and $P(Y = y_j)$ in the margins of the table. For this reason, one often calls the probability function for X the “marginal probability function of X” and the probability function for Y the “marginal probability function of Y” [However, the word marginal is really superfluous].

EXAMPLE 1.4

Let X and Y be as in Example 1.3. Find the marginal probability functions of X and of Y.

SOLUTION

TABLE 1-6

		y			
		1	2	3	P(Y=y)
	1	.05	.1	.1	.25
x	2	.2	.05	.25	.5
	3	.05	.1	.1	.25
P(X=x)		.3	.25	.45	

EXAMPLE 1.5

Consider an ordinary deck of playing cards. A card can be uniquely identified by its face value and its suit. Let. x be the face value of the card where jack is considered value 11, a queen value 12, and a king value 13. Let y = 1 if the card is a spade =2 if the card is a club = 3 if the card is a diamond and = 4 if it is a heart. Then the pair (x,y) uniquely identifies the card. We model randomly selecting a card from the deck by a random pair, (X,Y), where X is the face value and Y the suit of a card selected, having joint probability

P(X = x,Y = y) = 1/52, x = 1, …,13, y = 1,2,…,4

(a) Find the probability function of X. (b) Find the probability function of Y. (c) Find the probability that the card is either a spade or a club and has face value less than 6.

SOLUTION

The random pair (X,Y) has 52 pairs of values (x,y), x = 1,…,13 and y = 1, …, 4. Each pair has probability 1/52, i.e. P(X = x, Y = y) = 1/52

(a) P(X = x) = P(X = x,Y =1) + P(X = x, Y = 4)

= (1/52) +(1/52) + (1/52) +(1/52) =4/52 = 1/13, x = 1, …, 13

(b) P(Y = y) = (1/52) + … + (1/52) = 13/52 = 1/4, y = 1,.., 4.

(c) There are 5x2 = 10 pairs (x,y) that satisfy the condition x < 6 and y =1 or y =2. So P(X < 6 and Y =1 or Y = 2) = 10/52.

3.2 d-TUPLES

The extension of the idea of an ordered pair to d variables is called a <u>d-tuple</u> and written $(x_1,x_2,\dots,x_d)$. A 1-tuple is just a single number. A 2- tuple is an ordered pair (x_1,x_2), a 3-tuple is an ordered triple (x_1,x_2, x_3). The number x_i is called the ith component of the d-tuple $(x_1,x_2,\dots,x_d)$. Two d-tuples $(x_1,x_2,\dots,x_d)$, and $(y_1,y_2,\dots,y_d)$ are equal, in symbols $(x_1,x_2,\dots,x_d) = (y_1,y_2,\dots,y_d)$, if and only if each of their components are equal. That is, $(x_1,x_2,\dots,x_d) = (y_1,y_2,\dots,y_d)$ if and only if $x_1 = y_1, \dots, x_d = y_d$.

If we wanted to consider d random variables $X_1, \dots, X_d$ simultaneously we would consider the random d -tuple $(X_1, \dots, X_d)$. The probability function for the random d-tuple (or the joint probability function of the d random variables) is $P(X_1 = x_1, \dots, X_d = x_d)$.

NOTATION

We use bold face letters like **X** to denote a random d-tuple $(X_1, \dots, X_d)$ and we denote the values of the d-tuple as **x**. In this notation we would denote the probability function for the random d-tuple as $P(\mathbf{X} = \mathbf{x})$.

We illustrate this extension for the case d = 3.

3.3 RANDOM TRIPLES

If we wanted to consider 3 discrete random variables X, Y, and Z simultaneously, we would use the random triple (X,Y,Z). If X has values $x_1,\dots,x_l$, Y has values $y_1, \dots, y_m$, and Z has values $z_1, \dots z_n$ then the joint probability function of X,Y,Z is written $P((X,Y,Z)=(x_l,y_m,z_n))$ or $P(X = x_l, Y = y_m, Z = z_n)$. Having all these lmn probabilities we compute the probability that the random triple is in some specified set A of triples by summing the probabilities for the triples that are in A. In symbols

$$P((X,Y,Z)\in A) = \sum_{(x,y,z)\in A} P(X = x, Y = y, Z = z) \ .$$

From the joint probability function we can obtain the probability functions of X, of Y and of Z as well as the probability functions of the pairs (X,Y), (X,Z), and (Y,Z).

$$\left.\begin{aligned}
P(X = x_j) &= \sum_{(y,z)} P(X = x_j, Y = y, Z = z) \\
P(Y = y_j) &= \sum_{(x,z)} P(X = x, Y = y_j, Z = z) \\
P(Z = z_j) &= \sum_{(x,y)} P(X = x, Y = y, Z_j = z)
\end{aligned}\right\} \quad (1.4)$$

$$\left.\begin{aligned}
P(X = x_i, Y = y_j) &= \sum_{z} P(X = x_i, Y = y_j, Z = z) \\
P(X = x_i, Z = z_j) &= \sum_{y} P(X = x_i, Y = y, Z = z_j) \\
P(Y = y_i, Z = z_j) &= \sum_{x} P(X = x, Y = y_i, Z = z_j)
\end{aligned}\right\} \quad (1.5)$$

4 EVENTS and THEIR PROBABILITIES

4.1 Events

We can think of observing the value of a random variable X as an "experiment" the "outcome" of which is the value that the random variable assumes. Similarly, we can think of the simultaneous observation of two random variables X and Y as an experiment whose outcomes are the pairs of values that the pair (X, Y) can assume. It is useful to abstract the idea of an experiment. An experiment will refer to a conceptual entity that may correspond to some real experiment or observation or a purely hypothetical one. The salient feature of an experiment is that it has outcomes. [In fact, it is the set of outcomes that actually define what the experiment is.]

Let S denote the set of all possible outcomes of the experiment. An event is a specified set of outcomes of the experiment, i.e. a subset of S. For example, for the experiment of observing the value of a random variable X, we can speak of the event that X takes a value in some set B. This event is denoted $(X \in B)$. For another example we can think of selecting a card from an ordinary deck of cards. Then, (selected card is a spade), (selected card is a king or queen) are events.

Since events are subsets of the set of all outcomes, the set operations of union, intersection, and complement have interpretations in terms of events. We think of the event A as the fact that the

(a) The complement of a set A, A^c, is the collection of all points not in A. if A is an event we think of A^c as the event that the event A does not occur. For example in selecting a card from a deck of cards (card is a spade)^c is the event (card is not a spade).

(b) The intersection of two sets A and B, written $A \cap B$, is the set of all points that are in both A and in B. In terms of events $A \cap B$ is the event both events A and B occur. For selecting a card if A = (card is a spade) and B = (card is king or queen) then $A \cap B$ is the event (card is the king or queen of spades).

(c) Two sets A and B are disjoint if they have no points in common. That is, A and B are disjoint if their intersection is empty, i.e. $A \cap B = \varnothing$.We say k sets, $A_1, \ldots, A_k$ are disjoint if any two different ones are disjoint. That is, $A_1, \ldots, A_k$, are disjoint if for $i \neq j$, $A_i \cap A_j = \varnothing$. For selecting a card The events (card is a spade), (card is a club), (card is a diamond), and (card is a heart) are disjoint. On the other hand the events (card is 1 or 2 or 3) and (card is 3 or 4 or 5) are not disjoint since they have the outcome card is a 3 in common.

(d) The union of two sets A and B, written $A \cup B$, is the set of all points that are either in A or in B or in both A and B. The union of two events is the event that either A or B (or both) occurs.

(e) A is a subset of B, written $A \subset B$ if every point in A is also A point in B. In terms of events if event A occurs implies event B occurs then $A \subset B$. For example id drawing a card the event A = (card has value < 4) implies the event B = (card has value < 6) occurs, so $A \subset B$.

4.2 Probabilities of events

We assume that probabilities are assigned to events. The probability of the event A. is denoted by P(A). The assignment of probabilities to events satisfies the following three properties:

$$\left.\begin{array}{l} \text{i } 0 \leq P(A) \leq 1 \\ \text{ii } P(S) = 1 \\ \text{iii if A and B are disjoint,} P(A \cup B) = P(A) + P(B) \end{array}\right\} \qquad (1.6)$$

From (1.6) one can derive additional rules that probabilities obey. These rules are sometimes called “the calculus of probability”.

$P(A^c) = 1-P(A)$ (1.7)

If $A_1, \ldots, A_n$ are mutually disjoint, then

$$P(A_1 \cup \ldots \cup A_n) = P(A_1) + \ldots + P(A_n) \quad (1.8)$$

A partition of S is a collection of subsets $A_1, \ldots, A_n$ that are mutually disjoint with union is S. If $A_1, \ldots, A_n$ is a partition of S, then for any event B

$$P(B) = P(A_1 \cap B) + \ldots + P(A_n \cap B) \quad (1.9)$$

If A and B are any two events (not necessarily disjoint)

$$P(A \cup B) = P(A) + P(B) + P(A \cap B) \quad (1.10)$$

Sometimes we write $A \cup B$ as A or B

If $A \subset B$, $P(A) \leq P(B)$ (1.11)

There is an extension of (1.10) to more than two events. As the number of events increases this general rule gets more and more complicated. It will be discussed in Section 13.6.

Suppose X is a random variable. Observing a value of X can be considered an experiment whose set of outcomes S consists of the values that X can take. For a subset A of S we can think of A as the event that X takes a value in A. Observe that A is being used in two ways, viz. as the event and as a set of values for X. The confusion between the two uses is permissible because they can be identified. From the point of view of experiments we could write P(A) for the probability that X takes a value in A and from the point of view of random variables we would write $P(X \in A)$ for this probability.

EXAMPLE 1.6

Suppose 80% of the people with carcinoma of the tongue are heavy smokers, 40% are heavy drinkers, and 30% are both. (a) Find the probability that a person with carcinoma of the tongue is either a heavy drinker or a heavy smoker. (b) Find the probability that the person is a smoker but not a drinker. (c) Let X =1 if the person is a smoker and let X =0 if not. Let Y = 1 if the person is a drinker and Y = 0 if not. Find the joint probability function of X and Y.

SOLUTION

Let A be the event that the person is a smoker and B be the event that the person is a drinker. (a) We want $P(A \cup B)$. We know P(A) = 0.8, P(B) = 0.4, and $P(A \cap B) = 0.3$. By (1.10), $P(A \cup B) = 0.8 + 0.4 - 0.3 = 0.9$. (b) The required probability is $P((A \cup B)^c) = 1 - 0.9 = 0.1$.

(c) First record what we know in a table as below

TABLE 1-7

	y		
x	1	0	P(X=x)
1	.3		.4
0			
P(Y = y)	.8		

We can fill in the marginal entries getting the following:

TABLE 1-8

	y		
x	1	0	P(X=x)
1	.3		.4
0			.6
P(Y = y)	.8	.2	

From this we can get the remaining entries.

.4 = P(X = 1) =P(X = 1, Y = 1) + P(X =1, Y =0).= .3 + P(X =1, Y =0). So P(X =1, Y =0) = .4 - .3 = .1. Similarly, P(X = 0, Y =1) = .8 - .3 = .5, Since the sum of all of the probabilities for the 4 possible pairs of values must add to 1 we find P(X = 0, Y = 0) = 1-.3-.5-.1=.1. Table 1.8 gives the final results.

TABLE 1-9

	y		
x	1	0	P(X=x)
1	.3	.1	.4
0	.5	.1	.6
P(Y = y)	.8	.2	1

EXAMPLE 1.7

Suppose 70% of women with breast cancer have an abnormal gene A or an abnormal gene B, 50% have an abnormal gene A, and 30% have an abnormal gene B.

(a) Find the probability that a woman with breast cancer has both genes abnormal.

(b) Find the probability that a woman with breast cancer has neither gene abnormal.

SOLUTION

Let X = 1 or 0 according as the woman does or does not have an abnormal gene A. Let Y = 1 or 0 according as the woman does or does not have an abnormal gene B.

(a) We want $P(X = 1, Y = 1)$. By (1.10)

$P(X = 1, Y = 1) = P(X = 1) + P(Y = 1) – P(X = 1 \text{ or } Y = 1) = 0.5 + 0.3 - 0.7 = 0.1.$

(b) We want $P(X = 0, Y = 0) = 1 – P(X = 1 \text{ or } Y = 1) = 1 - 0.7 = 0.3.$

4.3 INTERPRETATIONS OF PROBABILITIES

What is a probability? If I say the probability that a coin falls heads is 1/2 when tossed once what does that really mean? From the mathematical point of view it doesn’t mean anything at all! This may first disturb you but it is not unfamiliar. In geometry we talk about a point. But what is a point? If you think back to your geometry days you will see that it was never defined. Geometry deals with points and lines and how you can connect them up in various ways, but it never defines what they are. In the same way the mathematics of probability theory deals with how to calculate numbers called probabilities from other probabilities. These rules of calculation are valid no matter what interpretation you may decide to give to those numbers.

The interpretation of probabilities is a problem in philosophy not mathematics. Non the less, we will give two common interpretations to what probabilities mean.

Relative Frequency

Suppose we repeat the experiment a large number of times. Then the proportion of times that we find that the event A is occurs is about P(A). From this point of view to say that a coin tossed once has probability 1/2 of landing heads means that if we repeated tossing the coin a large number of times then we would observe that about 50% of the time the coin would land heads. Often this is stated by saying that if we would toss the coin infinitely many times then 50% of the time it would land heads. However if you think about it you will see that the last statement is really meaningless.

Degree of Certainty

In this interpretation truth or certainty is measured on a scale of 0 to 1 and P(A) is interpreted as how certain we are of the fact that A will occur, with 0 being certain it will not occur and 1 being certainty that it will occur.

4.4 Assigning probabilities

From the point of view of the probability calculus it doesn't matter how the probabilities are assigned to events or how they are interpreted as along as the assignment satisfies (1.6). But how do we assign probabilities to events? The answer is that we just make them up to be what we want them to be. This at first sounds strange but it is not. The assignment of probabilities makes a model of the experiment in question. That model may or may not correspond to the real situation we are trying to model. For example, consider tossing a coin that is once. Our model is that it is a random variable having the values 0 and 1 with. $P(X = 1) = p$. If we choose $p = 0.5$ we get one model for the experiment of tossing the coin, if we choose $p = 0.25$ we get another, etc. Which choice is correct? From the mathematical point of view all of them are correct. If the coin is supposed to be balanced most people would take $p = 0.5$. Doing so is our model of the behavior of a balanced coin. To ask if a real coin behaves like $p = 0.5$ is to ask how well the model agrees with reality. That is an entirely separate issue that has nothing to do with probability theory. Such questions are in the domain of statistics.

5 EQUALLY LIKELY OUTCOMES

Probability theory began with the consideration of experiments whose outcomes were "equally likely" and such experiments still play an important role in many applications. For such experiments the set S of outcomes consists of a finite number m of points and the probability assigned to an event A is simply the number of points, #(A), divided by m, i.e.

$$P(A) = \frac{\#(A)}{m} \tag{1.12}$$

Note that the probability of an individual outcome x is $\#(\{x\})/m = 1/m$. We often view an experiment whose outcomes that are equally as "choosing a point at completely at random" from the set S of outcomes.

5.1 COUNTING RULES

If A has only a few points it is easy to just enumerate the points in A to find #(A). However, as soon as A gets even moderately large, this method becomes difficult, and some rules for counting are needed.

Rule 1

Suppose A_1 has r_1 points and A_2 has r_2 points. The number of ordered pairs (x_1,x_2) with $x_1 \in A_1$ and $x_2 \in A_2$ is r_1r_2.

For example, if $A_1 = \{1,2,3\}$ and $A_2 = \{1,2\}$ then there are 3x2 = 6 pairs with $x_1 \in A_1$ and $x_2 \in A_2$, viz. (1,1), (1,2), (2, 1), (2,2), (3,1), (3,2).

More generally, if A_1 has r_1 points, A_2 has r_2 points, …, A_k has r_k points, then there are $r_1r_2\cdots r_k$ k-tuples $(x_1,x_2, \ldots,x_k)$ with $x_1 \in A_1, \ldots, x_k \in A_k$.

Rule 2

Suppose A has r points. Then there are r^k k-tuples $(x_1,x_2, \ldots,x_k)$ with $x_i \in A$, $i = 1,\ldots,k$.

This is just a special case of Rule 1 with each of the sets A_1, …, A_k being the same set A. For example this rule says there are r^2 pairs that can be formed from a set of r points. Note carefully that in this count pairs such as (1,3) and (3,1) are considered as different pairs. For example, if A = {1,0} then there are 8 triples (x_1,x_2,x_3) for $x_i \in$ A These are (1,1,1), (1,1,0), (1,0,1), (0,1,1), (0,0,1), (0,1,0), (1,0,0), (0,0,0).

Often rule 2 is phrased in terms of sampling. Imagine that A is a box with r tickets numbered 1,…,r. Think of drawing a ticket from the box, recording the number on the ticket and returning the ticket to the box. Repeat this procedure k times and let x_i be the number on the i^{th} ticket. This is called sampling with replacement. Rule 2 says there are r^k different samples of size k that can be drawn with replacement from the box of r tickets.

Rule 3

There are $r(r-1)\cdots(r-k+1)$ k-tuples $(x_1,x_2, \ldots,x_k)$ with x_i = 1,2 …,r and all x_i distinct. For example, if A = {1,2,3}, then there are 3x2 = 6 pairs (x_1,x_2) with $x_1 \neq x_2$. These are (1,2), (1,3), (2,3), (2,1), (3,1), (3,2). Notice that if two k-tuples consist of the same k numbers but in a different order they are considered different k-tuples. For example, the two triples (1,2,3) and (1,3,2) are considered different.

It is convenient to have a symbol for quantities like $r(r-1)\cdots(r-k+1)$ since they occur frequently in mathematics. For r and k non-negative integers, let

$$(r)_k = \begin{cases} r(r-1)\cdots(r-k+1), & k > 0 \\ 1, & k = 0 \end{cases} \qquad (1.13)$$

Imagine that A is a box with r tickets numbered 1,…,r. Think of drawing a ticket from the box, recording the number on the ticket and discarding the ticket. Repeat this procedure k times and let x_i be the number on the i^{th} ticket. This is called sampling without replacement. Rule 3 says there are $(r)_k$ different samples of size k that can be drawn without replacement from the box of r tickets.

Rule 4

There are $k(k-1)\cdots(2)(1)$ different permutations (also called rearrangements) of the k distinct numbers.

For example there are $3 \cdot 2 \cdot 1 = 6$ permutations of (1,2,3). These are (1,2,3), (1,3,2), (2,1,3), (2,3,2), (3,1,2),(3,2,1). We use the symbol k! to denote the permutation of the numbers 1,2,…,k. For convenience in certain formulas we define 0! to be 1.

Rule 5

In the count of the number of k tuples having distinct x_i, those k-tuples that consist of the same k numbers permuted are considered different k-tuples. For example, the triples (1,4,2) and (4,2,1) selected from the set {1,2,3,4} are considered different k-tuples. In contrast, the subset consisting of the points 1,2,4 can be written as {1,2,4} or {1,4,2} or any other permutation of 1,2,4. These all denote the same object, namely the subset consisting of the points 1,2,4. From a sampling point of view a subset of size k (i.e. consisting of k points) from a population of r things can be thought of as a sample of size k without replacement from this population ignoring the order in which the k sample elements were drawn.

Let c(r,k) be the number of subsets of size k from a set of r objects. The objects in each of these subsets can be permuted in k! ways. Therefore c(r,k)k! is the total number of (ordered) k-tuples of distinct objects. Thus $c(r,k)k! = (r)_k$ and therefore $c(r,k) = \frac{(r)_k}{k!}$. We denote $\frac{(r)_k}{k!}$ by the binomial coefficient $\binom{r}{k}$. This coefficient is read, “r choose k”. [Because it is the number of ways we can choose k things from a set of r things.]. So

$$\binom{r}{k} = \frac{(r)_k}{k!}.$$

There is an alternate formula for $\binom{r}{k}$, namely

$$\binom{r}{k} = \frac{r!}{k!(r-k)!}.$$

Observe that if $k > r$ then $\binom{r}{k} = 0$.

In this notation, the number of subsets of size k from a set of r points is $\binom{r}{k}$. In sampling language, $\binom{r}{k}$ is the number of samples of size k that can be selected from a population of r things without replacement and ignoring order. For example, in poker 5 cards are dealt. For example, two poker hands are the same if the consist of the same 5 cards. The order in which you hold the cards is irrelevant. Consequently there are $\binom{52}{5}$ different poker hands and not $(52)_5$.

Rule 6

A population consists of r objects of type 1 and b objects of type 2. The number of ways a sample of size $n \leq r + b$ can be selected without replacement and have exactly x objects of type 1 is

$\binom{r}{x}\binom{b}{n-x}$.

EXAMPLE 1.8Suppose there are 10 cards numbered 1,…,10. A card is selected completely at random from these 10. Let X be the number on the card. A second card is selected completely at random from the 9 remaining cards. Let Y be th number on that card. (a) Find the joint probability function for X,Y. (b) Find the probability that X + Y < 5. (c) Find the probability function for X. (d) Find the probability function for Y.

SOLUTION

The possible values of (X,Y) are pairs (x,y) of numbers 1,…,10 with the restriction that $x \neq y$. In all there are (10)(9) = 90 such pairs and each pair has probability 1/90. Thus the probability function is

P(X = x,Y = y) = 1/90, x,y = 1,…,10 and $x \neq y$.

Let A = {(x,y) : x,y = 1,…,10 and $x \neq y$ and x+y <5} Then A = {(1,3), (3,1) (1,2), (2,1)}. P(X + Y <5) = P((X,Y) $\in$ A) =#(A)/90 = 4/90.

There are 9 pairs with x = 2, namely (2,1), (2,3), …,(2,10). In general there are 9 pairs for any specified value of x =a. Therefore P(X = x) = 9/90 = 1/10.

There are 9 pairs for every specified value of y = 1, ,10. Thus P(Y = y) = 1/10.

EXAMPLE 1.9

(a)A man has n keys exactly one of which hit the lock. He tries the keys completely at random discarding the wrong keys until he finds one that fits the lock. Let X be the number of keys tried. Find the probability function for x. SOLUTION

How can X = x, x =1,...,n. The xth key tried must be the correct key and the previous x - 1 the wrong keys. There are $(n\text{-}1)_{x\text{-}1}$ ways of picking the first x -1 wrong keys and only one way of picking the xth key. The number of choices for the x keys is $(n)_x$. Therefore

$$P(X = x) = \frac{(n-1)_{x-1}}{(n)_x} = \frac{(n-1)\cdots(n-x+1)}{(n)\cdots(n-x+1)} = \frac{1}{n}, \ x = 1, \ldots, n$$

EXAMPLE 1.10

A box 12 cards numbered 1, ..., 12. Suppose 5 cards are selected completely at random without replacement. The numbers are recorded and the cards returned to the box. A second sample of size 5 is then randomly selected. What is the probability that exactly 3 of the cards from sample 1 have the same numbers as 3 of the cards from sample 2?

SOLUTION

At first this problem seems quite difficult, but actually it is quite easy if thought about in the correct way. The effect of the first sample 1 is simply to distinguish 5 of the 12 cards (the “chosen” cards). The second sample must now have exactly 3 “chosen” cards. By rule 6 the probability for this to occur is

$$\frac{\binom{5}{3}\binom{7}{2}}{\binom{12}{5}} = 0.265$$

EXAMPLE 1.11

(a) A closet has 10 pairs of shoes scattered completely at random in the closet. Suppose 8 shoes are randomly selected from the closet. Find the probability that there is no complete pair.(b) Suppose there are n pairs of shoes and we select 2r < n of them. What is the probability of no complete pair?

SOLUTION

There are $\binom{20}{8}$ ways of selecting the 8 shoes from the 20 in the closet. Each of these ways are equally probable. (a) If the 8 shoes form no complete pair then each must come from a separate pair. The pairs the 8 shoes come from can be selected in $\binom{10}{8}$ ways from the 10 pairs. From each pair we have 2 choices of which shoe to pick. Thus the probability of no complete pair is

$$\frac{\binom{10}{8}2^8}{\binom{20}{8}} = 0.0916.$$

(b) There are $\binom{2n}{2r}$ equally likely ways of picking the 2r shoes. If there are no complete pairs then these must be picked from different pairs and there are $\binom{n}{2r}$ ways of picking the pairs. From each of the pairs we can choose one of two shoes. So the required probability is

$$\frac{\binom{n}{2r}2^r}{\binom{2n}{2r}}$$

EXAMPLE 1.12

Suppose there are n + 1 cells including cell A which is infected with a virus. One of the other cells is selected completely at random from the n other cells and infected. That cell then chooses a cell completely at random from the n cells other than himself and infects it, etc. (a) Find the probability that after this process is repeated r times the infection does not return to cell A. (b) Find the probability that after r repetitions no cell has been infected more than once. (c) Find these probabilities for 11 cells and 3 repetitions.

SOLUTION

(a) There are n^r choices for the r cells. These are all equally likely. If the infection is not to return to A then the first cell infected can be any one of n and the remaining r - 1 cells can be any one of n - 1 cells other than A and the one infected by A. Thus the total number of choices for the r cells is $n(n-1)^{r-1}$. Therefore, the required probability is

$$\frac{n(n-1)^{r-1}}{n^r} = \left(1 - \frac{1}{n}\right)^{r-1}.$$

(b) Each of the infected cells must be different, so there are $(n)_r$ choices for these cells. The required probability is

$$\frac{(n)_r}{n^r}.$$

(c) For (a) 0.81 and for (b) 0.72.

***6 POKER**

A poker hand consists of 5 cards drawn without replacement from an ordinary deck of cards. The order of the cards does not count. There are $\binom{52}{5}$ = 2,598,960 different poker hands. Any hand of 5 cards has the same probability q = $\frac{1}{\binom{52}{5}}$ = 0.00000039.

TERMINOLOGY

Pair: 2 cards of the same numerical value.

Triple: 3 cards of the same numerical value.

4 of a kind: 4 cards of the same numerical value.

Full house: a triple and a pair.

Straight is 5 cards in sequence, e.g. 3,4,5,6,7. An ace counts high or low, i.e. either as a 1 or a 14.

Flush: 5 cards of the same suit.

Straight flush is a straight that is also a flush.

Royal flush is the straight flush that is 10, jack, queen, king ace.

We will compute the probabilities of various poker hands and leave the computation of others as exercises. Since any 5 cards have the same probability q, to compute the probability of a specific hand is equivalent to counting the number of ways the 5 cards can be chosen to satisfy the specifications of the particular hand. Examples 1.13 – 1.17 all pertain to poker hands.

EXAMPLE 1.13

Find the probability that there are 2 distinct pairs.

SOLUTION

(a) let x and y be the face values for the cards in the 2 pairs and z the face value for the single remaining card. That is the 5 cards in the hand are to have 2 cards with face value x, 2 with face value y, and 1 with face value z, where $x \neq y \neq z$. How many ways can these face values be chosen? It is here that we must be careful. At first we might say there are (13)(12)(11) ways. However, that is not correct. To see why, consider the hand having face values (1,3,7) indicating that a pair of 1's, a pair of 3's and a 7. The above count of (13)(12)(11) would say any permutation of the numbers 1, 3, and 7 would give a different poker hand. Is the poker hand with face values (3,1,7) different from one with (1,3,7)? No. The former says we have a pair of 3', a pair of 1's and a 7. The latter says we have a pair of 1's, a pair of 3's and a 7. But these are the same. Notice that the hand (1,7,3) is a different hand since this is the hand having a pair of 1's, a pair of 7's and a 3. In general, any permutation of the values (x,y,z) that just switch the values of x and y yield the same hand. Thus the correct number of choices is $\frac{(13)(12)(11)}{2}$. For each choice of face value x there are $\binom{4}{2}$ choices for the suits of the 2 cards in the pair with face value x. Similarly, there $\binom{4}{2}$ choices for the suits of the pair with face value y. Finally, there is $\binom{4}{1}$ choices for the suit of the single card with face value z. In total, there are $\frac{(13)(12)(11)}{2}\binom{4}{2}\binom{4}{2}\binom{4}{1}$ =

12,352 ways of choosing 5 cards so that there are exactly 2 dstinct pairs. The required probability is (12,352)q = 0.0475.

EXAMPLE 1.14

Find the probability of exactly one pair.

SOLUTION

The face values are to be (x,y,z,w) with $x \neq y \neq z \neq w$, with x being the face value of the cards in the pair and y, z, w the face values of the 3 single cards. These can be chosen in any order. [Any permutation of the values y, z, w yield the same hand.]. The suits for the 2 cards with face value x can be chosen in can be selected in $\binom{4}{2}$ ways and each of the singles has $\binom{4}{1}$ choices for its suit. Thus the number of ways we can have exactly one pair is $\frac{(13)(12)(11)(10)}{3!}\binom{4}{2}\binom{4}{1}^3 =$ 1,098,240. The probability of exactly one pair is (1,098,240)q = 0.4226

EXAMPLE 1.15

What is the probability of a straight?

SOLUTION

The face values for the 5 cards must go in sequence. Consequently, the face values for the cards in the straight are fixed by the initial value. This could be any value from 1 (ace) to 10, so there are 10 choices for the face values in the flush. For each face value there are $\binom{4}{1}$ choices for its suit. In all, the straight can be chosen in $10\binom{4}{1}^5 =$ 10,240 ways. The probability of a straight is (10,240)q = 0.0039.

EXAMPLE 1.16

Find the probability of a flush.

SOLUTION

There are 4 choices for the suit and $\frac{(13)(12)(11)(10)(9)}{5!}$ choices for the face values. The total number of ways of getting a flush is $4\frac{(13)(12)(11)(10)(9)}{5!}=$ 51480. The required probability is (51480)q = 0.0020.

EXAMPLE 1.17

Find the probability of a triple.

SOLUTION

The 5 cards are to have 3 with face value x and 1 each with face values y and z, $x \neq y \neq z$. These face values can be chosen in $\frac{(13)(12)(11)}{2}$ ways. The suits for the 3 cards in the triple can be chosen in $\binom{4}{3}$ ways and the suits for the singles can be selected in $A \cap B$ ways each. Thus there are $\frac{(13)(12)(11)}{2}\binom{4}{3}\binom{4}{1}\binom{4}{1}=$ 54912 choices. The probability of a triple is (54912)q = 0.0211.

7 CONDITIONAL PROBABILITY

7.1 DEFINITION OF CONDITIONAL PROBABILITY

I choose a card at random from a deck of cards. You would give probability 1/52 that the card is the ace of spades. But suppose I tell you that the card drawn is a spade. Now what would you give for the probability that it is the ace of spades? This new probability is called the conditional probability that the card is the ace of spades given the information the card is a spade. Most of us would say that the conditional probability is 1/13, based on the belief that knowing the card is a spade we now have only one choice for the required card out of 13 equally likely choices. In this case the intuitive argument is correct. But there are cases when there is either no intuitive argument, or several intuitive arguments leading to a variety of different possible answers. What is needed is a precise definition of conditional probability.

DEFINITION

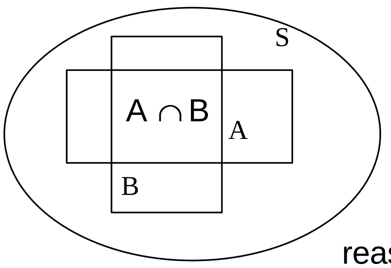
Let A and B be two events. Assume P(B) > 0. Then the conditional probability that the event A occurs given the event B occurs, denoted P(A|B), is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{1.14}$$

If P(B) = 0 then the conditional probability of A given B is undefined. If P(A) > 0 then P(B|A), the conditional probability of B given A, is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{1.15}$$

Figure 1-1



The definition is sensible (See Fig. 1-1). Knowing the event B occurs it now becomes the set of all possible outcomes of the experiment. Therefore, all events, A, are now reassigned probabilities based on the fact that A occurs knowing B has occurred if and only if they both occur, i.e. $A \cap B$ occurs. If we let $Q_B(A)$ be the probability that we would reassign to A knowing B took place then it is reasonable to require that it be proportional to $P(A \cap B)$ and that $Q_B(B) = 1$. With this requirement, $Q_B(A) = cP(A \cap B)$. Since $Q_B(B)$ must be 1 it follows that $c = \frac{1}{P(B)}$. This says we should define the conditional probability of B given A as done in (1.14).

If we multiply both sides of (1.14) by P(B) or we multiply both sides of (1.15) by P(A) we obtain two formulas for $P(A \cap B)$

$$P(A \cap B) = \begin{cases} P(A\,|\,B)P(B) \\ P(B\,|\,A)P(A) \end{cases} \tag{1.16}$$

Formula (1.16) is called the <u>multiplication rule</u>.

In the special case when the event A implies the event B, i.e. $A \subset B$, $A \cap B = A$ and becomes

$$P(A|B) = \frac{P(A)}{P(B)},\ A \subset B. \tag{1.17}$$

EXAMPLE 1.18 (Continuation of Example 1.6)

Consider the information in Example 1.6. (a) What is the probability that a person with carcinoma of the tongue who is known to be a smoker is a drinker? What is the probability that a person with carcinoma of the tongue who is known to be a drinker is not a smoker?

SOLUTION

Let A and B be as in Example 1.6.

We want P(B|A). Using equation (1.15) we see that

$$P(B|A) = \frac{0.3}{0.4} = 0.75$$

We want $P(A^c|B)$. By definition

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

and by (1.9)

$$P(A^c \cap B) = P(B) - P(A \cap B) = 0.4 - 0.3 = 0.1$$

Therefore $P(A^c|B) = \frac{0.1}{0.4} = 0.25$.

EXAMPLE 1.19

Suppose a box has 6 red balls and 4 white balls. Two balls are selected from the box. (a) What is the probability that they are both red? (b) What is the probability that one of them is red and the other is white?

SOLUTION

Think of drawing the balls one at a time. Let A be the event that the first ball is red and let B be the event that the second ball selected is red.

(a)We want $P(A \cap B)$. We know P(A) = 6/10 and P(B|A) = 5/9. By the multiplication rule

$$P(A \cap B) = (6/10)(5/9) = 1/3.$$

(b) We want $P(A \cap B^c) + P(A^c \cap B)$ By the multiplication rule

$$P(A \cap B^c) = P(B^c|A)P(A) = (4/9)(6/10) = 4/15$$

$$P(A^c \cap B) = P(B|A^c)P(A^c) = (6/9)(4/10) = 4/15$$

Therefore the desired probability is 8/15.

$$P(A|B) = \frac{P(A)}{P(B)} \quad (1.18)$$

EXAMPLE 1.20

A balanced coin is tosses twice. What is the probability that both tosses are heads given that at least one of the tosses is a head?

SOLUTION

Let A be the event that toss 1 is a head and B be the event that toss 2 is a head. We want $P(A \cap B | A \cup B)$. Since $A \cap B \subset A \cup B$

Error! Reference source not found. shows

$$P(A \cap B | A \cup B) = \frac{P(A \cap B)}{P(A \cup B} = \frac{1/4}{3/4} = \frac{1}{3}.$$

7.2 LAW of TOTAL PROBABILITY

Suppose $A_1, \ldots, A_n$ is a partition of S. The multiplication rule shows that $P(B \cap A_i) = P(B | A_i)P(A_i)$. From (1.9) we find

$$P(B) = P(B|A_1)P(A_1) + \ldots + P(B|A_n)P(A_n) \quad (1.19)$$

This equation is known as the law of total probability.

Let X be a discrete random variable with possible values $x_1, \ldots, x_n$. The events $(X = x_i)$, $i = 1, \ldots, n$ are a partition. The law of total probability shows that

$$P(B) = \sum_{i=1}^{n} P(B| X = x_i)P(X = x_i) \quad (1.20)$$

This rule is often applied when B is the event Y = y for some discrete random variable Y. In (1.20) we can replace X by a random pair or, more generally by a random d -tuple.

EXAMPLE 1.21

Suppose there are 3 boxes each with 2 coins. Box1 has 2 gold coins. Box 2 has a gold coin and a silver coin, and box 3 has 2 silver coins. A box is selected at random and then a coin is selected at random. (a) Find the probability that it is a gold coin. (b) Find the probability that the other coin in the selected box is also a gold coin. [This classical puzzle frequently appears in various magazines.]

SOLUTION

(a) Let X = i if box i is selected. Let Y = 1 if the first coin selected is a gold coin and let Y = 0 if not. We want P(Y = 1). What do we know? We know P(X = i) = 1/3, i = 1,2,3.

We also know various conditional probabilities. In particular, we know P(Y = 1|X = 1) = 1, P(Y = 1|X = 2) = 1/2, and P(Y = 1|X = 3) = 0. Using the law of total probability this is all we need to compute P(Y = 1) since by that law

P(Y = 1) = P(Y = 1|X = 1)P(X = 1)+P(Y = 1|X = 2)P(X = 2)

+P(Y = 1|X = 3)P(X = 3)= (1/3)1+(1/2)(1/3)+(0)(1/3) = 1/2.

(b Lets think. In order for the 2nd coin to also be a gold coin we must have selected box 1. So the required probability is P(X = 1|Y = 1). To compute this probability we first use the definition of the conditional probability

P(X = 1|Y = 1) = $\frac{P(X=1, Y=1)}{P(Y=1)}$.

We have already found P(Y = 1) = $\frac{1}{2}$.

But P(X = 1, Y = 1) = P(Y =1|X =1) P(X = 1) =(1)(1/3) = 1/3.

Thus

P(X = 1|Y = 1) = (1/3)/(1/2) = 2/3.

7.2.1 SENSITIVE QUESTIONS

Directly asking a person if they engage in some illegal activity such as drug use, cheating on income tax, etc. or in some socially unacceptable behavior such as extramarital affairs often lead to untruthful answers. How then do we gain accurate information about such activities? Here is how it is done.

Instead of just asking the question about sensitive issue you ask two questions, A and B. Question A is the sensitive issue question and question B is a completely innocuous question such as does your social security number end in an even or odd digit. You tell the person to flip a coin. If its heads the person is to answer question A and if its tails to answer question B. You tell the person the not to tell you which of the 2 questions is being answered but just tell you the response yes or no. Under these conditions people are much more likely to answer question A truthfully since the investigator cannot determine from the answer if it is question A or question B being answered.

So how can we use these responses to determine the proportion a of the population engaging in the behavior of interest? The key to this method is to choose question B so that you know the proportion of the population to which the answer to question B is yes is say, b. Here is why it works.

Let p be the probability that the person answers yes. Let X = 1 or 0 as the person answers question A or B. Let Y be 1 or 0 as the person gives you a yes or no answer. Let $a = P(Y = 1|X = 1)$. This is the probability you wish to determine. Note that $b = P(Y = 1|X = 0)$, and $p = P(Y = 1)$. By the law of total probability,

$$p = P(Y = 1) = P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 0)P(X = 0) = a/2 + b/2.$$

Therefore,

$$a = 2p - b.$$

One now can estimate a from an estimate of p. Suppose you sample n people and the proportion of yes answers is $\hat{p}$. Estimating p by $\hat{p}$ in the above gives $2\hat{p} - b$ as an estimate, $\hat{a}$, of a .

As an example, suppose the you sample 1000 people and the number of yes answers is 280. Suppose b = 0.1. Then $\hat{p} = 0.280$ and your estimate $\hat{a}$ is 0.560 - 0.1 = 0.46.

*7.3 DEATH BY MURDER

Amongst the developed nations the United States is the clear leader in death by violent acts. How likely are you to get murdered?

The table below gives the homicide rate per 100,000 in the United States per year based on the 1991-1993 data.

Source: Health United States 1995

NOTE ON TERMINOLOGY

Incidence and death probabilities for various things effecting humans, as a number between 0 and 1, are rarely reported in the literature. This is because they tend to be very small numbers that are difficult for humans to comprehend. Instead, these probabilities are multiplied by some factor such as 1,000, 10,000 or

100,000 to put them on a scale that is more to human liking. If they are multiplied by 100,000, then they are called the incidence or death rate per 100,000. These are always given over a prescribed time period (usually a year) If some years are combined then the rates are per year for each of the years. For example, in the table given below, death rate per 100,000 for children under a year that died by murder in the time period 1991-1993 in the United States was 8.8 per 100,000. This means that for each of the years 1991, 1992, 1993 the probability that a child under 1 year of age would be murdered is 8.8/100,000

=0.000088.

TABLE 1-10

Death Rates for Homicide 1991-1993

	All races and Sexes	White male	Black male	White female	Black female
All ages crude	10.2	9.0	69.8	2.9	13.6
Age (in years)					
Under 1 (age group 1)	8.8	7.0	22.9	1.7	20.3
1-4 (age group 2)	2.8	2.1	8.4	1.5	7.5
5-14 (age group 3)	1.6	1.3	6.0	.8	3.3
15-24 (age group 4)	22.7	17.7	160.1	4.2	21.0
25-34 (age group 5)	17.6	15.0	119.2	4.4	25.8
35-44 (age group 6)	11.3	10.7	73.8	3.5	15.4
45-54 (age group 7)	7.6	8.2	47.6	2.7	8.2
55-64 (age group 8)	5.0	5.7	29.5	1.9	5.7
65-74 (age group 9)	3.8	3.8	27.5	2.0	6.6
75-84 (age group 10)	3.9	3.6	25.2	2.4	10.7
85 and over (age group 11)	4.1	4.5	26.3	2.8	9.8

How to read the table .

Starting from the top the table has 12 rows. Call these rows 0, …, 11 Starting from the left it has 5 columns. Call these columns 1, ,5. Imagine each of the entries in the table divided by 100,000 to form probabilities. For the purpose of this table the entire population consists of whites and blacks. Assume we select a person at random from the population. let

D = 1 if the person murdered and let D = 0 if not.

S =1 if the individual is male and S = 0 if not.

R = 1 if the individual is white and R = 0 if not.

X =i if the individual is in age group i, i =1,…,11.

In terms of these random variables the table is reporting the following probabilities multiplied by 100,000.

row 0, column 1: P(D = 1).

row 0, column 2: P(D = 1|R = 1, S = 1).

row 0, column 3: P(D = 1|R = 0, S = 1).

row 0, column 3: P(D = 1|R = 1, S = 0).

row 0, column 5: P(D = 1|R = 0, S = 0).

For rows i = 1, …,11

row i, , column 1: P(D =1|X = i).

row i, column 2: P(D = 1|R = 1, S = 1, X = i).

row i, column 3: P(D = 1|R = 0, S = 1,X = i).

row i, column 3: P(D = 1|R = 1, S = 0,X = i).

row i, column 5: P(D = 1|R = 0, S = 0,X = i).

Notice how the probability that a person is murdered changes as you are given various additional information about the person.

8 DIAGNOSTIC TESTS

8.1 TEST CHARACTERISTICS

We are all familiar with the fact that medicine uses various types of tests to determine the presence or absence of diseases. Typically these tests are not perfect. On the one hand a person not having the disease may test positive for the disease. That is called a <u>false positive</u>. On the other hand a person having the disease may test negative for the disease. That is called a <u>false negative</u>. Let us consider a specific disease and let D = 1 or 0 according as person has or does not have the disease, and let us also consider a particular test for the disease. Set T = 1 or 0 according as that person tests positive or negative for the disease.

The sensitivity of the test is defined to be the probability a = P(T = 1|D = 1) and the specificity of the test is defined to be the probability b = P(T = 0|D = 0). The false positive rate is 1-b and the false negative rate is 1-a. In the medical literature these are usually reported as percents. The disease incidence p is P(D = 1). In the medical literature this is usually reported as a rate per 100,000 or rate per 10,000. Thus the incidence rate of lung cancer in males in 1992 was 312.3 per 100,000 or p = 312.3/100,000 = 0.00312. Often the disease incidence rate is called the prior probability for the disease [It is called the prior because it is the probability for the disease before the results of a test for the disease are known]. The positive predictive value of the test is P(D = 1|T = 1). The negative predictive value of the test is P(D = 0|T = 0).

Find

the probability that a person tests positive for the disease,

the joint probability function of (T,D)

(c) Find the positive predictive value

(d) Find the that someone who tests positive does not have the disease.

(e) Find the negative predictive value

SOLUTION

(a) We want P(T = 1). Using the law of total probability we find

P(T = 1) = P(T = 1|D = 0)P(D = 0)+P(T = 1|D = 1)P(D = 1) = (1-b)(1-p)+ap.

(b) We give the joint probabilities as a table:

	D	
T	1	0
1	ap	(1-b)(1-p)
0	(1-a)p	b(1-p)

(c) We want the probabilities P(D = 1|T = 1). Now

$P(D = 1|T = 1) = P(D = 1 \text{ and } T = 1)/P(T = 1)$.

Since $P(D = 1 \text{ and } T = 1) = P(T = 1|D = 1)P(D = 1) = ap$ we find from (a) that

$$P(D = 1|T = 1) = \frac{ap}{(1-b)(1-p)+ap}. \qquad (1.21)$$

(d) Since $P(D = 0|T = 1) = 1-P(D = 1|T = 1)$,

$$P(D = 0|T = 1) = \frac{(1-b)(1-p)}{(1-b)(1-p)+ap}.$$

(e) We want $P(D = 0|T = 0)$. Proceeding as in (b) we find

$$P(D = 0|T = 0) = \frac{b(1-p)}{b(1-p)+p(1-a)}. \qquad (1.22)$$

The numerical consequences of these formulas are quite interesting. Equation (1.21) shows that if $b = 1$ then no matter what the values of p and a are, $P(D = 1|T = 1) = 1$ and $P(D = 0|T = 1) = 0$. In this case everyone that tests positive for the disease will have the disease. However, if $b<1$ these facts can drastically change, especially if p is small (as is usually the case). For (1.21)shows that if $b \neq 1$, $P(D = 0|T = 1)$ goes to zero as p goes to 0 . A consequence of this fact is that for essentially all tests and diseases most of those who test positive for the disease will in fact not have the disease!

8.2 HIV INFECTION

Consider HIV infection in the heterosexual non-IV-drug using population of the USA. According to the CDC (center for disease control) the best estimate for HIV in this population is $p = 0.001$. The test most used to detect HIV infection (ELISA test) has an official specificity of 98% but it is probably considerably lower. Given that the test is taken 6 or more months after the exposure to the disease we can take the sensitivity to be 100%. Using this test we find from (1.21) that

$P(D = 1|T = 1) = (.001)/((.02)(.999)+.001) = 0.0477$

$P(D = 0|T = 1) = 0.9523$.

In this case we see that the false positive rate is 95%, i.e. 95% of those who test positive for HIV will in fact not be HIV infected! The odds of not having HIV to having HIV is .9523/.0477 = 19.9644 or 20 to 1. This fact is often used as an argument against HIV screening for the entire US population.

However, there is another side to this story. Before the test, our individual had probability 0.001 of having HIV. However, after that individual tests positive, the probability of the person having the disease increases to 0.0447 so the individual is now about 48 times as likely to have HIV as before.

9 INDEPENDENCE

9.1 DEFINITION OF INDEPENDENCE

Suppose A and B are two events and P(B) > 0. Intuitively we say say A and B are independent if knowing that B occurs does not alter the chance that A occurs, i.e. if P(A|B) = P(A). Using **Error! Reference source not found.** we see that if A and B are independent then

$$P(A \cap B) = P(A)P(B) \tag{1.23}$$

Observe that if P(A) > 0 then P(B|A) = P(B) so knowing that A occurs does not change the chance that B occurs. We use (1.23) as the formal definition of two events being independent.

DEFINITION

Two events A and B are said to be independent if and only if (1.23) holds.

This definition does not require the events to have non-zero probabilities. If P(A) = 0 then $P(A \cap B) = 0$ and P(A)P(B) = 0. Thus an event having probability 0 is independent of any other event.

Let X and Y be two discrete random variables. We say that X and Y are independent if and only if the events (X = x) and (Y = y) are independent for all choices of x and y, i.e. if for all x and y,

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

This extends to n discrete random variables.

DEFINITION

Let $X_1, \ldots, X_n$ be discrete random variables. We say these n random variables are independent if for all values of $x_1, \ldots, x_n$,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1) \ldots P(X_n = x_n). \tag{1.24}$$

In other words, $X_1, \ldots, X_n$ are independent if their joint probability function is the product of the marginal probability functions.

Intuitively, we say these random variables <u>are independent</u> if knowing the values of some them gives no additional information on what values the others will take.

To see that this definition is in accord with the intuitive picture suppose X_1 and X_2 are independent. Then

$$P(X_2 = x_2 | X_1 = x_1) = \frac{P(X_1 = x_1)P(X_2 = x_2)}{P(X_1 = x_1)} = P(X_2 = x_2).$$

and similarly,

$$P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_2 = x_2)P(X_1 = x_1)}{P(X_2 = x_2)} = P(X_1 = x_1).$$

Thus, knowing the value of one of the random variables does not change the probability for the other. Similar facts are true for more than 2 random variables. For example, if X_1, X_2, and X_3 are independent then

$P(X_2 = x_2 | X_1 = x_1, X_3 = x_3) = P(X_2 = x_2)$,

$P(X_1 = x_1, X_3 = x_3 | X_2 = x_2) = P(X_1 = x_1, X_3 = x_3)$,

etc.

The concept of independence extends to d-tuples of various sizes. In general we say that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent if their joint distribution is the product of their marginal distributions.

Suppose **X** and **Y** are independent. Then it is intuitively clear that any function of **X** is independent of any function of **Y**. For, if any knowledge of **X** gives no information about **Y,** then any knowledge of a function, say g(**X**), of **X** will also give no information about **Y** and hence about any function of h(**Y**) of **Y**. For example, suppose X, Y, Z are independent. Then X + Y and Z^2 are also independent. More generally, if $g_1(\mathbf{X}),\ldots,g_k(\mathbf{X})$ are functions of **X** and $h_1(\mathbf{Y}), \ldots, h_i(\mathbf{Y})$ are functions of **Y** then $(g_1(\mathbf{X}), \ldots, g_k(\mathbf{X}))$ and $(h_1(\mathbf{Y}), \ldots, h_i(\mathbf{Y}))$ are also independent. For example, (X + Y, X – Y) are independent of Z^2.

9.2 INDEPENDENCE AND MODELS

Independence of random d-tuples is usually not a property that is established mathematically. Instead, it is usually a property that we impose as part of a model we are trying to build.

One of the most fundamental uses of independence is in modeling successive observations of the same basic quantity under identical conditions. In this case, if we model a single observation by a random variable X, then we model n observations by taking n independent random variables $X_1, \ldots, X_n$ each having the same distribution as X.

9.3 CONDITIONAL INDEPENDENCE

We say n discrete random variables $X_1, \ldots, X_n$ are conditionally independent given the event A if

$$P(X_1 = x_1, \ldots, X_n = x_n|A) = P(X_1 = x_1|A) \ldots P(X_n = x_n|A) \quad (1.25)$$

EXAMPLE 1.22

A box contains 5 coins. One of them is a 2 headed coin and the other 4 are ordinary balanced coins. (a) A coin is randomly selected from the box and that coin is tossed 10 times. Find the probability that there is at least one head.(b) A coin is randomly selected from the box and tossed. That procedure is repeated 5 times. Find the probability of at least on head.

SOLUTION

(a) Let X_i = 1 or 0 according as the i^{th} toss is a head or tail. Let Y = 1 if the 2 headed coin is selected and let Y = 0 if a balanced coin is selected. We want

P(at least one X_i = 1). Here the random variables X_1, , X_5 are not independent, but they are conditionally independent given Y = 0. By the law of total probability

P(at least one X_i = 1). = P(at least one X_i = 1| Y = 1)P(Y = 1)

+.P(at least one X_i = 1| Y = 0)P(Y = 0)

Now P(at least one X_i = 1| Y = 1) = 1

and

P(at least one X_i = 1| Y = 0) = 1- $(.5)^5$ = 0.9688.

Thus

P(at least one X_i = 1). = (1)(.1) + (0.9688)(0.9) = 0.9719.

(b) Let Y_i = 1 if the i^{th} coin selected is a 2 headed coin and Y_i = 0 if not. Let Let X_i = 1 or 0 according as the i^{th} toss is a head or tail. Then, by the law of total probability:

$P(X_i = 1) = P(X_i = 1|Y_i = 1)P(Y_i = 1) + P(X_i = 1|Y_i = 0)P(Y_i = 0)$

$= 0.1 + (0.5)(0.9) = .55$. Here $X_1, \ldots, X_5$ are independent indicators with the same success probability p = 0.55. Therefore,

P(at least one $X_i = 1$). $= 1 - (.45)^5 = 0.9816$.

9.4 COIN TOSSING

We will refer to any success-failure experiment a coin toss. Suppose a coin having probability p for heads is tossed once. A single toss of the coin is modeled by an indicator variable X having success probability p. Suppose the coin is tossed n times. Let $X_i = 1$ if the i-th toss is a head and let it be 0 if not.

Our intuitive ideas about coin tossing suggest that we should model these n coin tosses by taking the n random variables $X_1, \cdots, X_n$ to be independent indicators with the same success probability p.

The probability function of X is

$$P(X = x) = \begin{cases} p, \text{if } x = 1 \\ 1 - p, \text{if } x = 0 \end{cases}.$$

The joint probability function for the n coin tosses is

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1)\ldots P(X_n = x_n) = .\, p^s(1-p)^{n-s}, \quad (1.26)$$

where s is the number of the $x_i = 1$ and n - s id the number of the $x_i = 0$.

COMMENT

Observe that in the successive coin tossing model independence is not a mathematical consequence of the model but is built into the model. The successive tosses of the coin are independent because we decree they are independent. Are real coin tosses independent? What the question is asking is if real tosses behave in the way the independence model says they should. That is not a question for probability theory. Rather it is a question for statistics.

One can set out modeling successive tosses of a <u>balanced coin</u> in a different manner. The outcomes are described as before $(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$ but now the probabilities are assigned by symmetry; all outcomes are equally likely. In all, there are 2^n possible outcomes. We assign each of these probabilities $1/2^n$. Does this give a different model than modeling by independence? No, in both models each n-tuple is assigned the same probability, namely $1/2^n$. In this second model, however, the independence of the random variables is not built in but is a mathematical consequence of the model.

EXAMPLE 1.23

Suppose we roll a balanced die twice. Let X = 1 if the first roll yields a 6 and let X = 0 if not. Similarly, let Y be 1 or 0 according as the second roll does or does not yield a 6. Find the joint probability function of X and Y.

SOLUTION

We consider getting a 6 a success. Using (1.26)we find the joint probability function is given by the following table:

	y		
		1	0
x	1	1/36	5/36
	0	5/36	25/36

EXAMPLE 1.24 (Death by murder continued)

The data for this example is Table 1.9 in the discussion of death by murder together with the following demographic facts from the 1990 US census. The population is 10.7% black. For the purpose of this example take the white population to be the non-black population. The population is 51.8% female. Using the random variables introduced in the death by murder example we can assume that sex S and race R are independent random variables.[This is reasonable because we have no reason to believe that the chance that a black person is male is different from the chance that a white person is male.]

(a) The table below is to give the joint probability function of (S,R). Fill in the table.

		race	
		0	1
sex	0		
	1		

(b) Find the probability that a female is murdered.

(c) Find the probability that a murdered person is a white female.

SOLUTION

(a)By independence, P(S = 0,R = 0) = (0.518)(0.107) = 0.0554. This gives the first table entry. The rest are obtained in a similar manner.

TABLE 1-11

	race	
sex	0	1
0	0.0554	0.4626
1	0.0516	0.4303

(b)We want to compute P(D = 1|S =1).

By definition,

$$P(D = 1|S = 1) = \frac{P(D = 1, S = 0)}{P(S = 0)}$$

We know P(S=0) = 0.518.

As for the numerator,

P(D = 1,S=0) = P(D = 1,S=0,R=1) + P(D = 1,S=0,R=0).

But

P(D = 1,S=0,R=1) = P(D = 1|S=0,R=1)P(S = 0,R = 1) = (2.9/100,000)(0.4626) = 0.00001

and

P(D = 1,S=0,R=0) = P(D = 1|S=0,R=0)P(S = 0,R = 0) =(13.6/100,000)(0.0554) = 0.00001.

Therefore P(D = 1,S=0) 0.00002 and

$$P(D = 1|S = 1) = \frac{0.00002}{0.518} = 0.00004$$

(c) What do we want to compute? We want P(S= 0,R =1|D =1). What do we know? We know from the table P(D =1| S = i, R = j) for i, j = 0, 1 and from the above table we know P(S = i,R = j). Now,

$$P(S = 0, R = 1|D = 1) = \frac{P(S = 0, R = 1, D = 1)}{P(D = 1)}$$

But

P(S= 0,R =1,D =1) = P(D =1|S= 0,R =1)P(S= 0 ,R =1)

= (2.9/100,000)(0.4303),

and $P(D = 1) = 10.2/100{,}000$.

Thus $P(S= 0,R =1|D =1) = [(2.9)(.4303)]/10.2 = 0.1223$.

10 CHANCE OF AT LEAST ONE SUCCESS

Let $A_1, \ldots, A_n$ be n events. There are many situations in which we are interested in determining the probability that there is at least one of these events occur, i.e. the probability $P(A_1 \cup \cdots \cup A_n)$. For example, the probability that at least one person has a specific disease amongst 100 people tested for the disease. Equation (1.10) shows how to calculate this probability for n = 2. A direct extension of that formula will be taken up in Section 14.6.4. Sometimes we can compute the complement of the event $A_1 \cup \cdots \cup A_n$ fairly easily. In that case $P(A_1 \cup \cdots \cup A_n) = 1 - P((A_1 \cup \cdots \cup A_n)^c)$. Now the complement of the event that at least one of the A_i occur is the event that none of them occur, i.e. $A_1^c \cap \cdots \cap A_n^c$. Therefore

$$P(A_1 \cup \cdots \cup A_n) = 1 - P(A_1^c \cap \cdots \cap A_n^c) \qquad (1.27)$$

We can formalize the above in terms of indicator random variables. Let $X_i = 1$ if the event A_i occurs and let $X_i = 0$ if not. The event $A_1 \cup \cdots \cup A_n$ is the event (at least one of the $X_i = 1$) and the event $A_1^c \cap \cdots \cap A_n^c$ is the event ($X_1 = 0, \ldots, X_n = 0$). In terms of these indicators (1.27) becomes

$$P(\text{at least one of the } X_i = 1) = 1 - P(X_1 = 0, \ldots, X_n =0). \qquad (1.28)$$

This equation is useful in those situations when it is simple to compute $P(X_1 = 0, \ldots, X_n =0)$. An alternate method for computing this probability will be given in Section 13.6. The $P(X_1 = 0, \ldots, X_n =0)$ is easy to compute when the indicator random variables are independent. In this case

$P(X_1 = 0, \ldots, X_n =0). = P(X_1 = 0)\ldots P(X_n =0)$.

Substituting the above into the right hand side of (1.28) we find

$P() = 1 - P(X_1 = 0)\ldots P(X_n =0)$...

But $P(X_i = 0) = 1 - P(X_i = 1)$, so

$$P(\text{at least one of the } X_i = 1) = 1 - [1 - P(X_1 = 1)]\ldots[1 - P(X_1 = 1)] \qquad (1.29)$$

Specializing still further, if additionally the independent indicators have the same success probability p, then

$P(X_1 = 0, \ldots, X_n = 0). = (1 - p)\ldots(1 - p) = (1 - p)^n$

Thus

If <u>If $X_1, \ldots, X_n$ are independent indicators each with the same success probability p, then</u>

$$P(X_i = 1 \text{ for some } i) = 1 - (1-p)^n \qquad (1.30)$$

An interesting consequence of (1.30) is that no matter how small a probability p of an event may be may be, if the experiment is repeated often enough the event is certain to occur. If we envision being able to continue the repetitions indefinitely then (1.30) shows that

$P(X_i = 1 \text{ for some } i \le n) = 1 - (1-p)^n$

If $0 < p < 1$ then $0 < 1-p < 1$ and $(1 - p)^n$ converges to 0 as n goes to infinity. Therefore

$$P(X_i = 1 \text{ for some } i) = \lim_{n\to\infty} [1 - (1-p)^n] = 1 \qquad (1.31)$$

EXAMPLE 1.25

Suppose a balanced coin is tossed 5 times. Find (a) the probability that there is at least one head amongst these 5 tosses. (b) Find the probability that all 5 tosses are heads given there is at least one head.

SOLUTION

Let $X_i = 1$ if toss i is a head and let $X_i = 0$ if not, $i = 1,2,\ldots,5$. These are independent indicator random variables each with success probability p = .5. Using (1.24) for d = 5 we find that the probability of at least one head is $1-.5^5$ .9688.

We want

$P(X_1 = 1, \ldots, X_5 = 1 |\text{ at least one } X_i = 1)$.

Since $(X_1 = 1, \ldots, X_5 = 1) \subset (\text{at least one } X_i = 1)$ we see that

$P(X_1 = 1, \ldots, X_5 = 1 |\text{ at least one } X_i = 1)$.

$= P(X_1 = 1, \ldots, X_5 = 1)/P(\text{at least one } X_i = 1) = .5^5/0.9688 = 0.0323$.

EXAMPLE 1.26

In females 60-64 years old the probabilities for four common cancers are given in the table below.

type	1	2	3	4
	breast	cervix	colon	lung
probability	0.0036	0.0002	0.0008	0.0017

(a) Find the probability that a woman in this age group has at least one of these diseases. (b) Find the probability that a woman has breast or lung cancer given that she has at least one of these diseases.

SOLUTION

Let $X_i = 1$ if the woman has cancer of type i and let it be 0 if not.

We want P(at least one of the $X_i = 1$). Here we can assume that these indicator random variables are independent. Using (1.29) we find the desired probability to be 1-(0.99636)(0.99981)(0.99917)(0.99834) = 1-0.99370 = 0.00630

Here we want $P(X_1 = 1 \text{ or } X_4 = 1|$ at least one of the $X_i = 1)$.

Now $(P(X_1 = 1 \text{ or } X_4 = 1) = 1- [1 – P(X_1 = 1)][1 – P(X_4 = 1)]$

= 1-(0.99636)(0.99834)). Therefore

$P(X_1 = 1 \text{ or } X_4 = 1|$ at least one of the $X_i = 1)$.

= (1-(0.99636)(0.99834))/0.00630 = 0.84031.

EXAMPLE 1.27

The purpose of this example is to show that an event having very small probability is virtually certain to occur if tried often enough.

Suppose you toss a balanced coin 10 times. Find the probability that all 10 tosses are a head.

Suppose n persons each toss a balanced coin 10 times. Find the smallest value of n so that the probability that at least on of them gets 10 heads is .999

SOLUTION

$(.5)^{10} = 0.00098$

By (1.30), with p = 0.00098, we see that the probability of at least one success in n trials is $1 – [1 - .00098]^n$. We require n so that this probability is .999. Since 1 - .00098 = .99902 we want n so that

$1 - (0.99902)^n = .999$ or $0.001 = (0.9990)^n$. Taking logs,

$n\ln(0.9990) = \ln(0.0001)$. Therefore

$$n = \frac{\ln(0.001)}{\ln(0.9990)} = \frac{-6.9078}{-0.001} = 6904.3$$

Rounding up, the number of people required is 6905.

This shows that amongst 6905 people it is virtually certain that at least one of them will get 10 heads.

EXAMPLE 1.28 (HIV)

We all know that just being properly exposed to a contagious disease does not mean we will with certainty come down with the disease. (By proper exposure we mean that one is exposed to the disease in a manner in which it can be transmitted.) The transmission rate of a disease is the probability that on a single proper exposure to that disease an individual will contract the disease. HIV is a sexually transmitted disease. No one is very certain about its transmission rate but it apparently is not very large and seems to greatly depend on the method of sexual contact. For unprotected heterosexual vaginal intercourse it seems to be 0.001 from male to female. The transmission rate from female to male is even more uncertain but it seems to be substantially less. Here we will use the transmission rate of 0.001 in both directions. The incidence rate of HIV in the heterosexual population is 0.001.

(a) Suppose Joe has a casual sexual encounter with a woman he picks up. What is the chance that Joe contracts HIV from this encounter? (b) Suppose Joe is very promiscuous and has a casual encounter with a different woman every day for a year. What are his chances of contracting HIV? (c) Jim meets a woman of whom he is fond of and has 365 sexual contacts with her. What are Jim's chances of contracting HIV? Is Joe's or Jim's behavior more risky? By how much? (d) If Joe is now twenty and he kept his promiscuous behavior going for 40 years what would be his chance of becoming HIV infected during this period? (e) How many different casual encounters would Joe have to have to have at least a 50% chance of becoming HIV positive?

SOLUTION

(a) Let X = 1 or 0 according as the person picked up is HIV infected or not and let Y = 1 if the Joe (Jim) contracts HIV on a single encounter. The transmission rate is $P(Y = 1|X = 1) = .001$ Then $P(Y = 1) = P(Y = 1|X = 1)P(X = 1) + P(Y = 1|X = 0)P(X = 0) = (0.001)(0.001) + (0)(0.999) = 0.000001$. So, on a single contact, there is only one chance in a million of contracting HIV.

(b) Formally, we let $Y_i = 1$ or 0 according as Joe contracts HIV from his i-th encounter. These random variables are independent. Part (a) shows their common success probability is q = 0.000001. If Joe has d encounters formula shows that the probability that he contracts HIV from at least one of these encounters is $1-(1-q)^d$. For d = 365 and q = 0.000001 this formula yields the probability 0.000365.

(c) As for Jim, Let $Z_i = 1$ or 0 according as Jim gets infected on his i^{th} contact with his girlfriend. The probability we want is $P(Z_i = 1$ for some i). In this case the indictors are conditionally independent given X and $P(Z_i = 1|X = 1) = .001$ and $P(Z_i = 1|X = 0) = 0$. $P(Z_i = 1$ for some $i| X = 1) = 1-(1-0.001)^{365} = .30953$. Hence $P(Z_i = 1$ for some i) = (0.30953)(0.001) = 0.000306. Since this probability is slightly smaller than Joe's is we see that Joe's behavior is slightly more risky.

(d) We proceed just as in part (b).This yields the probability $1-(0.999999)^{14,600} = 0.01449$.

(e)To answer the question posed we use the formula in part (b). We want to find the smallest d so that $1-(1-q)^d = 1-(.999999)^d = 0.5$. Solving for d we find d = (ln(.5)/ln(.999999) = 693,146.8 so Joe would have to have 693,147 casual contacts to have an even chance of becoming HIV infected!

11 BAYES' RULE

In diagnostic tests we are given the probability function for the disease incidence D and the conditional probabilities of T given D, where T is the test result, and we want the conditional probabilities of D given T. The procedure we used to compute "such turn around conditional probabilities" is an example of a general rule called Bays' rule (or Bays' Theorem). Although we will give this rule below, it definitely is a formula that should not be memorized. Instead, just as we have done in the past, one should just derive it in any special case in which it is needed.

The general problem is as follows. Given a partition A_1, ..., A_n of the set of all outcomes with each A_i having non-zero probability and an event B then

$$P(A_i\,|B) = \frac{P(B|\,A_i)P(A_i)}{P(B|\,A_1)P(A_1)+\cdots+P(B|\,A_n)P(A_n)} \tag{1.32}$$

Equation (1.32) was first written down by the Reverend Thomas Bayes in 1763 and is called Bayes' rule (or Bayes theorem). Trite as it is, Bayes rule forms the basis of a major philosophical system statistical inference. The proponents of this school are known as "Baysians" and the methods the proposed are known as Baysian procedures. In the applied literature, one often sees references to this rule or the procedures based on it, sometimes in almost mystical terms. The basis of system is the interpretation of Bayes' rule as an updating of the probabilities of the A_i given some information about the outcome of the experiment. Since the events A_i partition the set of outcomes one and only one of them can occur. The probability that A_i occurs before the experiment is undertaken is $P(A_i)$. After the experiment is performed we are given the information that the event B occurred. Given that information we now update the probabilities for the A_i occurring by reassigning A_i the probability given in (1.32) . In Bayesian jargon $P(A_i)$ is called the "prior" probability of A_i and $P(A_i\,|B)$ is called the "posterior" probability of A_i given the experiment resulted in B occurring.

If X is a discrete random variable the original distribution of X is called the prior distribution of X and the conditional probabilities $P(X = x|A)$ are called the posterior distribution of X given A. This can be cycled. We can take the posterior distribution given A as a new prior and then find a new posterior distribution given an event B. This is the same as the posterior distribution of X given the event $A \cap B$.

EXAMPLE 1.29

Suppose there are 3 boxes. Box 1 has 5 red an 5 white balls, box 2 has 10 red and 5 white balls, and box 3 has 5 red and 10 white balls. A box is selected at random and then a ball is selected from the box. It is observed to be red. Another ball is selected from the same box. It is also red. Let X be the box selected, Y = 1 or 0 as the first ball selected is red or white, and Z = 1 or 0 as the second ball selected is red or white. (a) What is the prior distribution of X? (b) What is the posterior distribution of X given Y = 1. (c) What is the posterior distribution of X given Z = 1 when we take the prior to the distribution in (b)

SOLUTION

The prior is P(X=1) = P(X = 2) = P(X = 3) = 1/3.

The posterior distribution is given by the following table.

Table 1-12

x	1	2	3
P(X = x)| Y = 1)	3/9	4/9	2/9

The prior distribution is now has probability function given by Table 1-11 rather than the probability function given in part (a). Since we are told Y = 1 the boxes now all have one less red ball. The numerator of (1.32) for x = 1 is (4/9)(3/9) = 0.1481 and the denominator is

(4/9)(3/9) + (9/14)(4/9) + (4/14)(2/9) = 0.4973.

Thus the new posterior probability of X = 1 is 0.1481/ 0.4973 = 0.2978. The new posterior probabilities of X = 2 and X = 3 are calculated similarly. These are given by Table 1-13 below.

Table 1-13.

x	1	2	3
posterior probability	0.2978	0.5787	0.1277

The posterior probabilities in Table 1-12 are just P(X = x|Y =1, Z = 1). We leave the verification as an exercise.

EXAMPLE 1.30

Of normal men under 55, 1% of normal men have a PSA level greater than 2, 35% of men with benign prostate enlargement have a PSA level greater than 2 and 60% of men with carcinoma of the prostate have a PSA greater than 2. The incidence rate of prostate cancer in this age group is 0.0002 and the incidence rate of benign enlargement is 0.006. (a) Jones is 54. He has a PSA test and it is greater than 2. What is the chance that he has each of the 3 possibilities? (b) A week later the test is repeated and again his PSA is greater than 2. Assume the two tests are conditionally independent given X. Now what are the chances of the 3 possibilities?

SOLUTION

Let X = 1 if Jones is normal, = 2 if he has benign enlargement, and = 3 if he has cancer. The prior distribution of X is given in the following table.

Table 1-14

x	1	2	3
P(X = x)	0.9938	0.002	0.0006

Let Y = 1 or 0 according as Jones' PSA is greater than 2 or not on the first test. The conditional probability of Y = 1 given X is given below.

Table 1-15

x	1	2	3
P(Y=1|X =x)	0.01	0.35	0.60

The P(Y = 1) = (.01)(.99380) + (.35)(.006) + (.60)(.0006) = 0.01216. Multiplying the entry for P(Y=1|X =x) in the above table by $\frac{P(X = x)}{P(Y = 1)}$ we obtain the posterior distribution of X given Y = 1.

Table 1-16

x	1	2	3
P(X = x|Y=1)	0.8174	0.1727	0.0099

Let Z = 1 or 0 according as his PSA is greater than 0 or not on the second test. The posterior distribution we want is P(X = x|Y = 1, Z = 1). The assumption that Y and Z are conditionally independent given X = x shows that PY = 1, Z = 1|X = x) = P(Y = 1|X = x)P(Z = 1|X =x). Note that P(Y = 1|X = x) = P(Z = 1|X =x) (because both are just the probability that a test is positive given X = x). Therefore

$P(Y = 1, Z =1) = (.01)^2 P(X = 1) +(.35)^2 P(X = 2) + (.60)^2 P(X =3) = 0.00091.$

Using Bayes' rule we find the requored posterior distribution is given by Table 1-16 below.

Table 1-17

x	1	2	3
P(X = x|Y=1,Z=1)	0.1096	0.8109	0.0794

Observe how the chance that Jones is normal has decreased. Originally he had a 99% chance of being normal. After 1 positive test this dropped to 82% and after 2 positive tests it decreased to just 11%. After 2 positive tests his chance of having cancer increased from 0.02% to 8%.

12 SOME CLASSICAL PROBLEMS

12.1 THE BIRTH OF PROBABILITY THEORY

It is said that probability theory began in order to solve a problem that puzzled a certain Count De Mere, a French 17th century aristocrat. As one knows from the movies, these folks spent their days drinking, womanizing, and gambling. De Mere was found of playing the following game. He would bet on getting at least one six on four rolls of a die. His reasoning led him to believe that his chance of getting the at least one six was 2/3. He thought that since at each role he had probability 1/6 of getting a six his chance of at least one six in four rolls was 1/6 + 1/6 + 1/6 + 1/6.

Show that this reasoning is wrong and compute the correct probability.

Even though De Mere's reasoning was wrong he would win more times than he lost. Somewhat later he switched to betting that he would get at least one 12 in 24 rolls of 2 dice. But now he began to loose more times than he won. This of course puzzled him because he again reasoned incorrectly that his chance of this winning was 2/3 by saying the probability was 24(1/36) = 2/3. Frustrated, he sought the help of the mathematician Pascal who corresponded with Fermat and so probability theory was born!

Calculate the correct probability for De Mere's winning his second game.

SOLUTION

Let $X_i = 1$ if the i^{th} roll of the die is 6 and let $X_i = 0$ if not, i = 1,2,3,4. What's required is P(at lest on $X_i = 1$). As we know this is not the sum of the probabilities $P(X_i = 1)$ but is $1-(5/6)^4 = 0.5177$.

We leave it as an exercise to show the answer is 0.4914.

12.2 BIRTHDAYS

Suppose n tickets are randomly selected with replacement from m tickets. What is the probability all are distinct? There are m^n samples of size n and there are $(m)_n$ ways all the tickets in the sample can all be distinct, so the required probability is

$$P(\text{all are distinct}) = \frac{m(m-1)\cdots(m-n+1)}{m^n}. \tag{1.33}$$

The birthday problem is to find the probability that no two people in a room of n people have the same birthday? This is the special case of the above with m = 365. The table below gives these probabilities for various n. Note that the probability is already less than 1/2 for n = 22.

n	5	10	20	21	22	30	40
p	0.9595	0.8529	0.5563	0.5243	0.4927	0.2696	0.0097

*12.3 WILL THE SUN RISE TOMORROW?

Consider the following problem. We have a coin but we know absolutely nothing about the probability for heads. The coin is tossed r times and all r tosses are heads. What is the probability that yet another toss will be a head?

Before the development of modern statistical theory in the 20th century the notion of equally likely chances was considered as synonymous with no advanced knowledge. From this point of view we would model tossing a coin where we have no knowledge of what the probability of heads is by saying that all values of p are equally likely. But there are infinitely many values that p can have so what does it mean that these values are equally likely?

Laplace modeled this is by assuming we have a large number, say n+1, coins. Coin i has probability i/n for heads. A coin is selected at random and that coin is tossed r times. He then computed the conditional probability that given the r tosses are all heads an (r + 1)st toss will also be heads and let $n \to \infty$. The resulting limiting probability is what we will take for the probability of obtaining an (r +1) first head given we have r heads for a coin whose success probability is "chosen at random".

To put this into mathematics, let X = i if box i is selected and let Y_i = 1 or 0 according as the i^{th} toss is head or tail. These indicator random variables are not independent but they are conditionally independent given X. Then $P(X = i) = \frac{i}{n}$ and

$$P(Y_1 = 1, \ldots, Y_r = 1 | X = i) = \left(\frac{i}{n}\right)^r.$$

By the law of total probability,

$$P(Y_1 = 1, \ldots, Y_r = 1) = \sum_{i=0}^{n}\left(\frac{i}{n}\right)^r P(X = i) = \frac{1}{n+1}\sum_{i=0}^{n}\left(\frac{i}{n}\right)^r. \tag{1.34}$$

We will now approximate last term for large n. Recall how the definite integral of a continuous function f is defined in terms of approximating sums. Suppose we want $\int_0^1 f(x)dx$. We partition the interval [0,1] into n subintervals each of length 1/n. We then choose a point x_i in the i^{th} subinterval, [(i-1)/n, i/n) and form the sums $\frac{1}{n}\sum_{i=1}^{n} f(x_i)$. We can take x_i = i/n to be the right end point of the i^{th} subinterval. Doing so we find that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i) = \int_0^1 f(x)dx.$$

(1.35)

Take f(x) to be x^r in equation (1.35). This gives

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=1}^{n}(i/n)^r = \int_0^1 x^r dx = \frac{1}{r+1}. \tag{1.36}$$

Now

$$\lim_{n\to\infty} \frac{n}{n+1} = 1. \tag{1.37}$$

Writing the left hand side of (1.36) as

$$\frac{n}{n+1}\frac{1}{n}\sum_{i=1}^{n}(i/n)^{r}$$

and using (1.37) and (1.34) we see that

$$\lim_{n\to\infty} P(Y_1 = 1, \ldots, Y_r = 1) = \frac{1}{r+1} \qquad (1.38)$$

Thus the probability of getting r heads on tossing a coin r times whose success probability is chosen at random is $\frac{1}{r+1}$.

We now consider the problem of calculating the probability of getting yet another head if the coin is tossed again given that the previous r tosses were all heads. We want $P(Y_{r+1} = 1|Y_1 = 1, \ldots, Y_r = 1)$. By definition of conditional probability,

$$P(Y_{r+1} = 1|Y_1 = 1, \ldots, Y_r = 1) = \frac{P(Y_1 = 1, \ldots, Y_{r+1} = 1)}{P(Y_1 = 1, \ldots, Y_r = 1)}. \qquad (1.39)$$

Replacing r by r + 1 in (1.38) we find that

$$\lim_{n\to\infty} P(Y_1 = 1, \ldots, Y_{r+1} = 1) = \frac{1}{r+2}. \qquad (1.40)$$

Thus

$$\lim_{n\to\infty} P(Y_{r+1} = 1|Y_1 = 1, \ldots, Y_r = 1) = \frac{r+1}{r+2} \qquad (1.41)$$

The formula $\frac{r+1}{r+2}$ as the probability that an event that has occurred r times will occur yet again is known as Laplace's Law of Succession. Laplace himself (1812) used his law to compute the probability that the sun will rise tomorrow given that it has done so for the last 5000 years (=1,826,250 days). Using his rule we find the probability that it will rise tomorrow is1,826,251/1,826,252 = 0.99999945. The odds for it rising tomorrow are 1,818,182. It is said that Laplace was willing to give these odds for the sun to once again grace the morning skies. However, he must have had some doubts about his theory since he was unwilling to take the opposite bet!

Laplace's law enjoyed quite a following in the 19th Century as a way of predicting some future event given it has occurred r successive times in the past. It was applied to all sorts of things, some sensible, some outright absurd. Whether or not its use is valid depends on the basic assumption in its derivation, namely that we can view whatever we are considering the successive repetitions of a success-failure experiment chosen from a collection of such experiments in which all conceivable possibilities are equally likely.

For example, if we apply the rule we find that a 1 year old has probability 2/3 of living to be 2 years old. However, its 70-year-old grandfather has probability 71/72 of making it one more year. This answer goes contrary to our experience. What's wrong is that the rule depends on the chance that an individual survives another year is equally likely to be any probability between 0 and 1. That certainly is far from true.

12.4 IMPERFECT INFORMATION AND CONDITIONAL PROBABILITIES

To determine any probability one must have a model for the mechanism that gives rise to the probability. Different models of the same situation can yield different probabilities. Sometimes, especially for conditional probabilities, one fails to explicitly formulate the model. This ambiguity can then lead to different persons having different models in mind, and subsequently to different answers. Often, especially in real scientific endeavors, this causes much debate on which answer is "correct". But in such cases there is no single correct answer. Two persons using different models may both come up with different answers, each of which is correct for the model they are (implicitly) using. We will illustrate the role of a model in conditional probability by considering some simple problems that often appear as puzzles in various magazines.

Consider the following problem. John is walking down Westwood blvd. and meets Mary whom he has not seen in 10 years. Since his last contact with Mary John has married and has two children. One of them is with him and he proudly introduces Mary to his son. What is the probability that his other child is also a boy? The most popular answers are 1/3 and 1/2. Which, if any, of these two is the correct answer?

Often this problem (and many others like it) is stated as follows. John has 2 children and one of them is a boy. What is the probability that the other child is a boy? Stated this way it is often forgotten that a model must be in place to determine how the information that one is a boy is obtained.

Problems such as this one that involve conditional probabilities can be tricky. Often what seems like a perfectly straightforward problem is not because the problem as stated does not completely specify an underlying model that is needed to compute all of the required conditional probabilities. In such cases different models can lead to different answers.

As we shall see, as stated, this problem does not really have a single correct answer. The answer depends on what assumptions are made about how and what the information that the child walking with the man is a son conveys. That is, what the correct answer is depends on what you assume the model is for which you are computing the probabilities.

We will analyze this problem step by step so you can see exactly where the model plays a role Suppose $Y_i = 1$ or 0 as child i is a boy or a girl, $i = 1, 2$. Let $X = 1$ or 0 as the child with the man is a boy or a girl. Everyone agrees that the appropriate model is the one that takes the joint probability function of $Y_1$and Y_2 to be the one that assigns the 4 possible values for the pair (Y_1,Y_2) equal probability. Let us take that to be the case. But as we shall see that is not enough and the model is incomplete.

Before computing anything let us look at some of the arguments for the various answers. (i) Knowing the child with John is a boy eliminates the possibility $Y_1 = 0$ and $Y_2 = 0$. This leaves three equally likely possibilities of which one is $Y_1 = 1$ and $Y_2 = 1$, so the answer is 1/3. (ii) By "independence" the probability that the other child is a boy is 1/2.

Now let us attempt to compute the answer. We want to compute

$$P(Y_1 = 1,Y_2 = 1 \mid X = 1)$$

Since $Y_1 = 1$ and $Y_2 = 1$ implies $X = 1$,

$$P(Y_1 = 1,Y_2 = 1 \mid X = 1) \; \frac{P(Y_1 = 1, Y_2 = 1)}{P(X = 1)}$$

Clearly,

$P(Y_1 = 1, Y_2 = 1) = .25.$

The law of total probability shows that

$P(X = 1) = P(X = 1|Y_1 = 1, Y_2 = 1)\ P(Y_1 = 1, Y_2 = 1)$

$+ \quad P(X = 1|Y_1 = 0, Y_2 = 1)\ P(Y_1 = 0,\ Y_2 = 1)$

$+ \quad P(X = 1|Y_1 = 1, Y_2 = 0)\ P(Y_1 = 1, Y_2 = 0)$

$+ \quad P(X = 1|Y_1 = 0, Y_2 = 0)\ P(Y_1 = 0, Y_2 = 0).$

Let's see what we know. We know;

(a)$P(Y_1 = i, Y_2 = j) = 1/4$ for $i,j = 1,0$.

(b) $P(X = 1|Y_1 = 1, Y_2 = 1) = 1$.

(c) $P(X = 1|Y_1 = 0, Y_2 = 0) = 0$.

What about the probabilities $P(X = 1|Y_1 = 0, Y_2 = 1)$ and $P(X = 1|Y_1 = 1, Y_2 = 0)$?

That is, what is the conditional probability that the child with John is a boy if John has a choice to take a boy or a girl with him? Often, almost without thinking, one takes these to be 1/2. But if we look closely at the statement of the problem we see that in fact <u>these probabilities are not specified</u>. Thus the model is not complete. Setting $p = P(X = 1|Y_1 = 0, Y_2 = 1)$ and $q = P(X = 1|Y_1 = 1, Y_2 = 0$ all we can write is

$$P(Y_1 = 1, Y_2 = 1 \mid X = 1) = \frac{.25}{.25(1+p+q+0)} = \frac{1}{1+p+q}.$$

Since p and q can have any value between 0 and 1, $P(Y_1 = 1, Y_2 = 1 \mid X = 1)$ can be as small as 1/3 ($p = q = 1$) or as large as 1 ($p = q = 0$). How do we determine the probabilities p and q? It is here that we need additions to the model. The model must specify how John chooses a child to accompany him on his walk given he has a choice. If we he picks equally likely between the two possibilities, the $p = q =1/2$ and the answer to our problem is

$P(Y_1 = 1, Y_2 = 1 \mid X = 1) = 1/2.$

On the other hand, if he always chooses a boy (if he can), then $p = q = 1$ and $P(Y_1 = 1, Y_2 = 1 \mid X = 1) = 1/3$. Finally, if he always choose a girl provided he can then $p = q = 0$ and $P(Y_1 = 1, Y_2 = 1 \mid X = 1) = 1$.

Here is another example of this kind of problem.

The following problem gives three variants of a classical problem known as the "two ace problem". It has appeared in Parade magazine.

There are 4 cards Card 1 is the ace of spades, card 2, the ace of hearts, card 3, the king of spades, and card 4, the king of hearts. You choose two of the four cards. We will now discuss several variants of "the two ace problem" Variant (ii) is the classical version of the problem. Throughout it is assumed that you do not lie.

I ask you to tell me what one of the two cards is that you have chosen. You oblige, and tell me (a)"I have an ace". What is the probability that I would calculate for the other card being an ace?

I ask you, "is one of the cards an ace". You say "yes I have an ace. What is the probability that I would calculate for the other card being an ace?

The two cards you select are placed face down on the table. You choose one of them at random and look at it. You then inform me that it was an ace. The problem is to calculate the probability that the other card is also an ace?

To many, there is no difference between these three variants. Others think there is a difference between variant (iii) and variants (i) and (ii) but see no difference between (i) and (ii). Actually all three are different. Variants (ii) and (iii) have unique (but different) answers while variant (i) does not. We will consider (i) and (ii).

Let T = 1 if you say that you have an ace and let T = 0 if not. Let X be 1 or 0 according as the first card is or is not an ace and let Y = 1 or 0 according as the second card is or is not an ace. The probability function for (X,Y) is given in the table below:

	Y	
X	1	0
1	1/6	2/6
0	2/6	1/6

We want $P(X = 1, Y = 1 | T = 1)$. Now

$$P(X = 1, Y = 1 | T = 1) = (P(T = 1 | X = 1, Y = 1)\, P(X = 1, Y = 1))/P(T = 1).$$

By the law of total probability,

P(T = 1) = P(T = 1| X = 1, Y = 1)P(X = 1, Y = 1)

+ P(T = 1| X = 1, Y = 0) P(X = 1, Y = 0)

+ P(T = 1| X = 0, Y = 1) P(X = 0, Y = 1)

+ P(T = 1| X = 0, Y = 0) P(X = 0, Y = 0).

In both (i) and (ii) the conditional probabilities P(T = 1| X = 1, Y = 1) (=1) and P(T = 1| X = 0, Y = 0) (= 0) are specified by the fact that you do not lie. However, in (i) the conditional probabilities p = P(T = 1| X = 1, Y = 0) and q = P(T = 1| X = 0, Y = 1) are not specified but they are specified in (ii). In (i) you are asked to name a card and if you have both a king and an ace you have a choice on what to say and not lie. However in (ii) you are asked if you have an ace and in the situation when you have both an ace and a king you have no choice but to say you have an ace in order not to lie. So in (ii) p =1 and q = 1.

Thus in (i)

$$P(X = 1, Y = 1| T = 1) = \frac{\frac{1}{6}}{\frac{1}{6}(1)+\frac{1}{6}(0)+\frac{2}{6}(p)+\frac{2}{6}(q)} = \frac{1}{(1+2p+2q)}$$

while in (ii),

P(X = 1, Y = 1| T = 1) = 1/5.

13 CONTINUOUS DISTRIBUTIONS

13.1 ONE VARIABLE

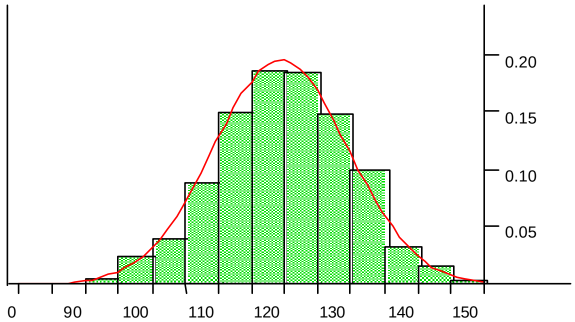
So far we have discussed distributions for random variables that can only take finitely many values. We now turn our attention to random variables such as a blood pressure measurement that measure quantities on a continuous scale. To motivate the method used to determine the distribution of such random variables consider the 1380 measurements of systolic blood pressures that are part of the data given in Table 1-3. Using the original 1380 measurements, we summarize these blood pressures in the frequency table below.

Table 1-18

pressure	Count	Percent	pressure	Count	Percent
85 To 90	3	0.22	120 To 125	258	18.70
90 To 95	9	0.65	125 To 130	206	14.93
95 To 100	36	2.61	130 To 135	139	10.07
100 To 105	58	4.20	135 To 140	47	3.41
105 To 110	126	9.13	140 To 145	24	1.74
110 To 115	208	15.07	145 To 150	6	0.43
115 To 120	260	18.84			

We can graphically represent the information in this frequency table by means of a histogram. In a histogram, we first lay off the cells in the first column as contiguous intervals on the x-axis. Then we represent the proportion of pressures that fall in a particular interval as the area of a rectangle over that interval.

Figure 1-2 r



The proportion of observations that fall between a and b (a < b) is represented as the area of the rectangles (or portions thereof) that are between a and b. Suppose one now approximates the overall pattern of the histogram by a smooth curve passing through the tops of the rectangles. This is illustrated in Fig. 1-2. One could now approximate the proportion of blood pressures that fall between a and b as the area under the curve between a and b. Suppose that the approximating smooth curve is actually the graph of a non-negative function f. Then the area under the curve between a and b is ,so that the proportion of blood pressures between a and b is approximately $\int_a^b f(x)dx$.

In general, to model a random variable X that has a continuum of values we assume we are given a non-negative function f that has integral 1. Such a function is called a density function. We then define the probability that X has a value in the interval A =(a,b] by

$$P(X\in A) = \int_a^b f(x)dx. \quad (1.42)$$

An continuous distribution is one where the probabilities are given by (1.42) .To make a model for a random variable X that has values in some interval we take its distribution to be absolutely continuous. We specify such a distribution by giving its density function f.

COMMENTS

(1) Any non-negative function g on R such that $\int_{-\infty}^{\infty} g(x)dx = a < \infty$ can be made into a density by taking f(x) = g(x)/a.

(2) To say, “X is a random variable with density f” means X is a random variable whose distribution is absolutely continuous with density f.

(3) Unlike the probability function for a discrete random variable the density function f can have values >1. This makes it clear that one cannot interpret f(x) as the probability that X = x (or for that matter as the probability of anything else). In fact, For X having an continuous distribution, P(X = a) for any a is 0 because $\int_a^a f(x)dx = 0$ .Owing to this fact P(X<a) = P(X$\leq$ a) and P(x>a) = P(X $\geq$ a). Consequently, if I is an interval with left end point a and right end point b, then P(X$\in$ I) is the same no matter which of the endpoints are included. That is, P(a<X<b) = P(a$\leq$X<b) = P(a<X$\leq$b) = P(a$\leq$ X $\leq$ b). Often one thinks of f(x)dx as the probability that X takes a vlaue in the “infinitesimal” interval (x, x + dx).

EXAMPLE 1.31

Let

$$f(x)=\begin{cases} cx \text{ if } 0<x<1 \\ \text{elsewhere} \end{cases}.$$

(a) Find c so that f is a density.

(b) If X is a random variable having this density find P(X < .33).

SOLUTION

(a) We need

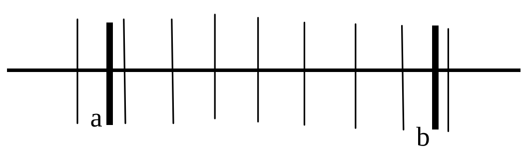
$$1=\int_R f(x)dx=\int_0^1 cxdx=c\frac{x^2}{2}\Bigg|_0^1=\frac{c}{2}.$$

So c = 2.

(b) $P(X < .33) = \int_{(0,.33)} f(x)dx=\int_0^{.33} 2xdx=(.33)^2=0.1089.$

13.2 CONTINUOUS UNIFORM DISTRIBUTION

How do we model selecting a point X at random from the interval (0,1)? To model selecting a point at random from a finite set having m points we use a discrete random variable whose values each carry probability $\frac{1}{m}$. Clearly this



cannot be applied when the set has infinitely many points. Instead, X is here continuous random variable say with density f. The question is, what is f? Our intuition says that if we partition (0,1) into n disjoint subintervals, each of length $\frac{1}{n}$, then the probability that the chosen point is in a specified one of

Figure 1-3

those subintervals is $\frac{1}{n}$. Now consider a subinterval (a, b) of (0,1). A certain number, say #(a,b) of intervals of the partition are contained completely in the interval (a,b). Let B be the union of these subintervals. Then (see Fig.1-3)

$B \subset (a,b) \subset B \cup C_a \cup C_b$

Where C_a is the subinterval of the partition containing the left end point a and C_b is the subinterval of the partition containing the right end point b.

Then

$$P(X \in B) \le P(X \in (a,b)) \le P(X \in B) + P(X \in C_a) + P(X \in C_b). \tag{1.43}$$

But

$$\frac{\#(a,b)}{n} \le b - a \le \frac{\#(a,b)}{n} + \frac{2}{n}$$

Therefore, as $n \to \infty$,

$$\frac{\#(a,b)}{n} \to b - a \tag{1.44}$$

It follows from (1.43) and (1.44) that P(B) = b - a. Since $\int_a^b 1dx = b - a$ it follows that X should have density

$$f(x) = \begin{cases} \text{if } 0 < x < 1 \\ 0, \text{elsewhere} \end{cases}$$

This density is called the <u>uniform density</u> on (0,1) and we say X is <u>uniformly distributed</u> on (0,1). More generally, the uniform distribution on a finite interval (r,s), r < s, is the distribution having density

$$f(x) = \begin{cases} \frac{1}{s-r}, r < x < s \\ 0, \text{elsewhere} \end{cases} \tag{1.45}$$

EXAMPLE 1.32

Let X be uniformly distributed on (1,5). Find (a) P(2 <X < 2), (b) P(X > 3), (c) P(2 <X < 4) or X > 3)

SOLUTION

$$f(x) = \begin{cases} \frac{1}{4}, 1 < x < 5 \\ 0, \text{elsewhere} \end{cases}$$

(a) P(2 <X < 4) = $\int_2^4 \frac{1}{4} d = \frac{2}{4} = 0.5$

(b) P(X > 3) = $\int_3^5 \frac{1}{4} d = \frac{2}{4} = 0.5$

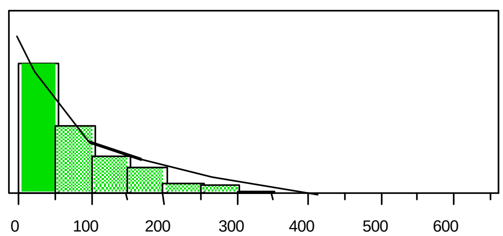
(c) Let A be the event (1 <X < 3) and B be the event (X > 2). Then

P(2 <X < 4) or X > 3) = P(A) + P(B) - P($A \cap B$). But ($A \cap B$) is the interval (3,4) so P($A \cap B$) = 0.25. Therefore

P(2 <X < 4) or X > 3) = 0.5 + 0.5 - 0.25 = 0.75.

EXAMPLE 1.33

The histogram below is for the time X (in minutes) of the first arrival in an emergency room past midnight requiring emergency major surgery for 756



observations.

The shape of the histogram suggests that a smooth curve approximation could be given by an exponential function, i.e. a function of the form

$$f(x) = \begin{cases} ce^{-x/a}, x > 0 \\ 0, \quad x < 0 \end{cases}$$

for positive constants c and a.

(a) Find c so f is a density.

(b) If X has this density with a = 93 find

P(X >100), (ii) P(85 < X < 95), and (iii) P(X < 95|X > 85).

SOLUTION

(a) We need c so that $1 = \int_0^\infty f)x)dx = \int_0^\infty ce^{-x/a}dx = (-ac)e^{-x/a}\Big|_0^\infty = ac$.

so c = 1/a.

(b) (i)P(X > 100) = $= \int_{100}^\infty f(x)dx = \int_{100}^\infty \frac{1}{93} e^{-\frac{100}{93}} = e^{-100/93} = 0.34$

(ii) P(85<X<95) =

$$\int_{85}^{95} f(x)dx = \int_{85}^{95} \frac{1}{93} e^{-\frac{x}{93}} dx = e^{-85/93} - e^{-95/93} = 0.04$$

(iii) P(X < 95|X > 85) = $\dfrac{P(X < 95 \text{ and } X > 85)}{P(X > 85)} = \dfrac{P(85 < X < 95)}{P(X > 85)}$.

From (ii), the numerator is .04. The denominator is

$$\int_{85}^{100} \frac{1}{93} e^{-x/93} dx = 0.40.$$

Thus P(X< 95|X > 85) = .04/.4 = 0.10.

EXAMPLE 1.34

Suppose X has density

$$f(x) = \begin{cases} 1/(1+x)^2, & x > 0. \\ 0, & \text{elsewhere.} \end{cases}$$

Find P(X < 4 or X > 6).

SOLUTION

Let $A = (-\infty, 4)$ and let $B = (6, \infty)$. Then

P(X < 4 or X > 6). = P(X$\in$ $A \cup B$) = $P(X \in A) + P(X \in B)$.

Now P(X$\in$ A) = $\int_{-\infty}^{4} f(x)dx \int_{0}^{4} \frac{1}{(1+x)^2} dx = -(1+x)^{-1}\Big|_0^4 = 1-(1/5) = 4/5$ = 0.80,

and

$$P(X \in B) = \int_{6}^{\infty} \frac{1}{(1+x)^2} dx = 1/7 = 0.143.$$

Thus P(X < 4 or X> 6) = 0.80 + 0.143 = 0.943.

13.3 SEVERAL VARIABLES

13.3.1 MULTIPLR INTEGRALS

The concept of the definite integral extends to functions of more than one variable. Such integrals are called multiple integrals. The evaluation of multiple integrals can be quite complicated. In this course we will not spend any time on the explicit computation of these integrals. For our purposes we need only to know that there are such integrals and that they can be used to define a certain family of probability distributions on R^d.

The simplest way to define the definite integral of a function of more than one variable is by repeated integration. Suppose f(x,y) is a non-negative function of (x,y). For each y, f(x,y) is a function of x. The integral of this function over the variable x, $\int_{-\infty}^{\infty} f(x,y)dx$, produces a function of y. This can then be integrated over the variable y to produce a number. That number is called the (multiple) integral of f over R^2 and is denoted by

$$\iint_{R^2} f(x,y)dxdy \text{ or by} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy.$$

In other words the integral of f over R^2 is defined to be

$$\iint_{R^2} f(x,y)dxdy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x,y)dx\right]dy.$$

In words, the multiple integral of f is produced by first integrating on the variable x and the integrating on the variable y. By advanced mathematical methods it can be proved that we can obtain the same number by first integrating on y and then integrating on x. In other words

$$\iint_{R^2} f(x,y)dxdy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x,y)dy\right]dx..$$

The same procedure can be used to define the integral of a non-negative function of any number of variables. Let $f(x_1, \dots, x_k)$ be a non-negative function of $(x_1, \dots, x_k)$. We proceed inductively on the number of variables. Suppose we have already defined the integral for a function of k-1 variables. Then the integral of f over R^k is defined by

$$\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} f(x_1,\dots x_k)dx_1\dots dx_k = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty} f(x_1,\dots x_{k-1},x_k)dx_1\dots dx_{k-1}\right]dx_k .$$

By using this equation over and over we see that the integral is produced by first integrating on the variable x_1, to produce a function of k-1 variables. Then we integrate that function on x_2 to produce a function of k-2 variables. We continue until we have integrated over all the variables. Again it can be shown that we can perform these k integrations in any order.

Let **x** = $(x_1, \ldots, x_k)$. We will abbreviate the multiple integral by $\int_{R^k} f(\mathbf{x})d\mathbf{x}$.

Let B be a subset of R^d and let

$$1_B(\mathbf{x}) = \begin{cases} 1 \text{ if } \mathbf{x} \in B \\ 0 \text{ if } \mathbf{x} \notin B \end{cases}$$

The integral of f over the set B is defined to be the integral of the function $f1_B$. We write the integral over B as $\int_B f(\mathbf{x})d\mathbf{x}$.

Let **X** be a random k-tuple on R^k. We say **X** has an absolutely continuous distribution if there is a non-negative function f on R^k having integral 1 such that

$P(\mathbf{X} \in B) = \int_B f(\mathbf{x})d\mathbf{x}$. The function f is called the density of **X** (or the joint density of $X_1, \ldots, X_k$).

13.3 INDEPENDENCE

Let **X** = $(X_1, \ldots, X_k)$ be a random k-tuple. We have previously defined independence of k discrete random variables. It is not difficult to show that k discrete independent random variables satisfy

$$P(X_1 \in B_1, \ldots, X_k \in B_k) = P(X_1 \in B_1) \ldots P(X_1 \in B_k). \quad (1.46)$$

for all subsets $B_1, \ldots, B_k$ of R. We use this property to define independence of random variables in general.

In general k random variables are independent if (1.46) holds for all subsets

By using advanced mathematical methods it can be shown that the following holds.

FACTORIZATION CONDITION

Let $X_1, \ldots, X_k$ be a k random variables each having an absolutely continuous distribution. Then $X_1, \ldots, X_k$ are independent if and only if

$$f(x_1, \ldots, x_k) = f_1(x_1) \ldots f_k(x_k), \quad (1.47)$$

where f_i is the density of X_i.

14 EXPECTATION

14.1 EXPECTED VALUE

Suppose X is a random variable. The concept of an “average value” is an attempt to give by means of a single number a “typical value” that X takes. There are several averages that are used. The most widely used is that of the mean or the expected value.

DEFINITION

Let X be a discrete random variable with $x_1, \ldots, x_n$ as its possible values. The expected value of X ,denoted EX ,is the quantity

$$EX = x_1P(X = x_1) + \ldots + x_nP(X = x_n) \quad (1.48)$$

Let X be a random variable having an absolutely continuous distribution with density f. We say the expected value of X exists if $\int_R |x| f(x)\,dx < \infty$. In that case we define the expected value of X to be the quantity

$$EX = \int_{-\infty}^{\infty} xf(x)dx \quad (1.49).$$

The expected value is also called the mean and it is also denoted by μ. The term mean applies to both the distributions and random variables. Thus we can speak of a random variable with mean μ or a distribution with mean μ.

EXAMPLE 1.35

Let X be an indictor random variable with success probability p. Find EX.

SOLUTION

EX = 0f(0) + 1f(1) = 0 + p = p.

EXAMPLE 1.36.

Let X be the discrete random variable with probability function give in the table below. Find EX.

x	-1	1	2	4
P(X = x)	0.2	0.3	.0.4	.0.1

SOLUTION

EX = (-1)(.2) + (1)(.3) + (2)(.4) = (4)(.1) = 1.3

EXAMPLE 1.37

Let X have the discrete uniform distribution on {1, …, n}. Find EX.

SOLUTION

EX = 1(1/n) +2(1/n) + … + n(1/n) = (1/n)(1 +2 = … + n).

The formula for the sum of the first n integers shows1 +2 = … +n = $\frac{n(n+1)}{2}$.

Thus

EX = (n+1)/2.

EXAMPLE 1.38

$$f(x) = \begin{cases} 2x, & \text{if } 0 < x < 1. \\ 0, & \text{elsewhere.} \end{cases}$$

Find EX.

SOLUTION

$$EX = \int_{-\infty}^{\infty} xf(x)dx = 2\int_0^1 x^2 dx = \left.\frac{2x^3}{3}\right|_0^1 = \frac{2}{3}.$$

One of the reasons that the expected value is so widely used is that it has some very desirable mathematical properties. These properties can sometimes be used to find the expected value for a random variable X from the expected values of other random variables when the distribution of X is unknown.

14,2 RULES OF EXPECTATION

a. If c is a constant then Ec = c.
b. If X is a random variable then E(cX) = cEX.
c. If X and Y are random variables then E(X + Y) = EX + EY.
d. If X $\geq$ Y then EX $\geq$ EY. In particular, if X $\geq$ 0, then EX $\geq$ 0.
e. If X and Y are independent then E(XY) = (EX)(EY).
f. If $c_1, \ldots, c_n$ are constants and $X_1, \ldots, X_n$ are random variables, then $E(c_1X_1 + \ldots + c_nX_n) = c_1EX_1 + \ldots + c_nEX_n$.
g. If $X_1, \ldots, X_n$ are independent, then $E(X_{1\ldots}X_n) = (EX_1)\ldots(EX_n)$.

EXAMPLE 1.39

Suppose X is a random variable with mean 3 and Y is a random variable with mean –2 find the expectation of 2X + 4Y.

SOLUTION

Using rule 6 we see that $E(2X + 4Y) = 2EX + 4EY = 2(3) + 4(-2) = -2$.

EXAMPLE 1.40

Suppose X and Y are independent with means 2 and –4 respectively Find E(XY).

SOLUTION

By rule 5, $E(XY) = (EX)(EY) = (2)(-4) = -8$.

COMMENT

If X and Y are not independent then it is usually not the case that $E(XY) = (EX)(EY)$. For example, let X be an indicator random variable with success probability p and let Y = X. Then $E(XY) = EX^2$. Now X^2 is a random variable having only 2 possible values 0,1 and $P(X^2 = 1) = P(X = 1) = p$. Thus $EX^2 = 0P(X^2 = 0) + 1P(X^2 = 1) = p$. But $(EX)(EY) = (EX)(EX) = p^2$.

EXAMPLE 1.41

Suppose $X_1, \ldots, X_n$ are random variables each with the same mean μ. Find

(a) $E(X_1 + \ldots + X_n)$ and (b) $E[\frac{1}{n}(X_1 + \ldots + X_n)]$.

SOLUTION

By rule 6, $E(X_1 + \ldots + X_n) = \mu + \ldots + \mu = n\mu$.

Using this result and rule 2 we find $E[\frac{1}{n}(X_1 + \ldots + X_n)]. \frac{1}{n} n\mu = \mu$.

EXAMPLE 1.42

Let $X_1, \ldots, X_n$ be indicator random variables each with the same success probability p and Let $N = X_1 + \ldots + X_n$. Find EN.

SOLUTION

From Example 1.36, $EX_i = p$. Using rule 6

$EN = EX_1 + \ldots + EX_n = p + \ldots + p = np$.

14.3 EXPECTATION of A FUNCTION OF X

Suppose g(**x**) is a function of k-tuples **x** and **X** is a random k-tuple. Let Y = g(**X**). Then Y is a random variable. An extremely useful fact is that we can compute EY without explicitly knowing the distribution of Y. All we need is the distribution of **X**. Here is the rule:

If **X** is discrete

$$Eg(\mathbf{X}) = \sum_{x} g(\mathbf{x})P(\mathbf{X} = \mathbf{x}). \quad (1.50)$$

If **X** has an absolutely continuous distribution with density f then

$$Eg(\mathbf{X}) = \int g(\mathbf{x})f(\mathbf{x})d\mathbf{x}. \quad (1.51)$$

EXAMPLE 1.43

Let X have the density in Example1.39 and let m>0. Find EX^m.

SOLUTION

We take $g(x) = x^m$ and use (1.51) .

$$EX^m = \int_{-\infty}^{\infty} g(x)f(x)dx = \int_0^1 2x^{m+1}dx = \left.\frac{2x^{m+2}}{m+2}\right|_0^1 = \frac{2}{m+2}.$$

EXAMPLE 1.44

Let X be the random variable in Example 1.37. Find E|X|.

SOLUTION

Using (1.50) we find

E|X| = (1)(.2) + (1)(.3) + (2)(.4) = (4)(.1) = 1.7

EXAMPLE 1.45

A box has r red and b black balls. Suppose 2 balls are randomly selected from the box. Let X = 1 or 0 according as the first ball is red or black. Let Y = 1 or 0 according as the second ball is red or black. (a) Find E(XY) if the balls are selected with replacement. (b) Find E(XY) if the balls are selected without replacement.

SOLUTION

In this case X and Y are independent. By rule 5 and Example 1.36, E(XY) = (EX)(EY) $=(r/r+b)^2$

Note that XY = 0 except for X = Y = 1. Using (1.50) E(XY) = (1)P(X =1,Y = 1) = $\frac{r(r-1)}{(r+b)(r+b-1)}$.

14.4 CONDITIONAL EXPECTATION

Since the conditional distribution of the discrete random variable Y given the value of another random variable X is a distribution one can speak about its mean. This is called the conditional expectation of Y given X = x and denoted as E[Y|X=x]. More precisely,

$$E[Y|X = x\} = \sum_y yP(Y = y \mid X = x) \qquad (1.52)$$

Analogous to the law of total probability there is a law of total expectation. This says

$$EY = \sum_x E[Y \mid X = x]P(X = x) \qquad (1.53)$$

Equation (1.53) follows easily from the definition of expectation, (1.52), and the law of total probability. The details are left as an exercise.

EXAMPLE 1.46

A box has 10 coins. Eight of the coins are ordinary balanced coins and two of the coins are two-headed. A coin is randomly selected from the box and that coin is tossed n times. What is the expected number of heads.

SOLUTION

Let X = 1 if an ordinary coin is selected and let X = 0 if a two-headed coin is selected. Let Y_i = 1 if the ith toss is a head and let Y_i = 0 if not. Let N = $Y_1 + \ldots + Y_n$. From Example 1.43, E[N|X = 1] n/2 = 5 and E[N|X = 0] = n. The law of total expectation now shows

EN = (4/5)(n/2) + (1/5)(n) = (3/5)n.

EXAMPLE 1.47

Suppose N is a discrete random variable having values 1,2, m. and that the conditional of S given N = n is nμ Find ES.

SOLUTION

The law of total expectation shows

$$ES = \sum_n n\mu P(N = n) = \mu EN \qquad (1.54)$$

***14.5 GAMBLING SYSTEMS**

Many popular gambling casino games such as red-black on roulette are just coin flips. Let p be the probability of heads and q = 1- p be the probability of tails. The game is called favorable, fair, or unfavorable (to the player) according as p > 1/2, = 1/2, or < 1/2. All casino games are unfavorable. Suppose you bet $1 on each trial. If the coin falls heads you win a $1 and if it falls tails you loose $1. Let X_i be 1 or -1 according as you win or loose the i^{th} play. Then your total winning (loss if negative) is

$S_n = X_1 + \dots + X_n$.

It is left as an exercise to show that

$ES_n = n(p - q)$.

Notice that if q = p your expected winning is 0 and if q > p it is negative. Many gamblers are staunch believers that they can altar this expectation and transform either a 0 or negative expectation to a positive one by a betting system. A betting system is simply a procedure for betting variable amounts of money at each play instead of a constant amount, as above. Since no gambler is clairvoyant, the wagers cannot depend on the results of future plays. However, the amount wagered at the i^{th} play typically depends on the results of the past i - 1 plays. In fact, it is a choice of how the wagers depend on the past history that constitute the betting system. The number of betting systems is legion, and the gambling casinos love the people that use them. Let's see why.

Let V_i be the amount bet on the i^{th} trial. The following assumptions are made on the bets V_i:

The V_i depends only on the random variables $X_1, \dots, X_{i-1}$.

$V_i \geq 0$

The first assumption (a) is that the bets depend only on the past outcomes. The 2nd assumption is that the least amount you can bet is nothing at all.[A 0 bet means that you skip betting on that play]. The 3rd assumption (c) says your bets are bounded.

The amount won or lost on this trial is V_iX_i. The amount won after n plays of the game is

$$S_n = V_1X_1 + V_2X_2 + \ldots + V_nX_n. \quad (1.55)$$

Since the random variables $X_1, X_2, \ldots$ are independent the random variables and V_i depends only on the random variables $X_1, \ldots, X_{i-1}$, V_i and X_i are independent. Thus $E(V_iX_i) = EV_iEX_i$. Using (1.55) and the fact that $EX_i = p - q$ we see that

$$ES_n = E(V_1X_1 + V_2X_2 + \ldots + V_nX_n) = (EV_1 + \ldots + EV_n)(p - q) \quad (1.56)$$

Notice that if $p = q$, $ES_n = 0$ and if $q > p$, $ES_n < 0$. Thus the gamblers are quite wrong. <u>If the number of plays is fixed in advanced, no betting system can change an expected gain that is ≤ 0 to one a positive expected gain</u>.

14.5.1 MARTINGALE SYSTEM

Most real gambling systems that work on the basis that you can continue play indefinitely and not that you stop playing after a fixed number of plays. One of the most popular gambling systems of this type is the martingale. In this system your initial bet is 1. Thereafter, you double the amount previously wagered if you get a tail and you bet 1 at each head. In other words, $V_1 = 1$ and for $n > 1$,

$$V_n = \begin{cases} 2V_{n-1} & \text{if } X_{n-1} = -1 \\ 1 & \text{if } X_{n-1} = 1 \end{cases}$$

The appeal of the martingale is that each time you toss a head you eradicate your past losses and are \$1 ahead. To see why this is true we need the formula for the sum of a geometric progression.

A geometric progression with n terms is $1 + a + \ldots + a^n$. These add up to

$$1 + a + \ldots + a^n = \begin{cases} \dfrac{1-a^{n+1}}{1-a} & \text{if } a \neq 1 \\ n & \text{if } a = 1 \end{cases}. \quad (1.57)$$

For $a = 2$ we find

$$1 + 2 + \ldots + 2^n = 2^{n+1} - 1.$$

Consequently, if you loose the first n bets you will have lost $\$(2^{n+1} – 1)$. But your next bet is $\$2^{n+1}$. Therefore, if you win you will have a gain of $\$(2^{n+1} (2^{n+1} – 1)) = \$1$.

Since you must ultimately toss a head it seems that by playing long enough you are sure to win the $1 and by repeating this procedure over and over you could win any amount of money. This is true! The catch is that you need an infinite amount of money and no limit on the size of the bets in order to make it work. This is due to the fact that the bets can become very large. [If you get k tails in a row your next bet would be $\$2^{k+1}$ which could either exceed your funds or the house limit]. All casinos place a limit on the size of the bet you can make in any one play (the house limit).

Let's see how the martingale really works in practice. Assume the house limit is $M and you play until either you win $1 or are forced to terminate because your next bet would exceed M. If the first j - 1 tosses are all tails your next bet would be $\$2^j$ The maximum number of times you can double is k where k is the largest integer such that $2^k \leq M$. Taking logs we see that k is the greatest integer $\leq \frac{\ln(M)}{\ln(2)}$. For example, for M = 1000, ln(1000)/ln(2) = 9.9657, so k = 9.

Since we will play until either we get a head or are forced to stop the series of games must end in at most k plays. Let S* be your gain at the termination of the series. Then S* can have one of two values, viz. 1 and $-(2^k – 1)$. [Either you get a head before k tosses in which case S* = 1 or you get k tails in a row and are forced to stop. In the second case you have lost $\$(2^k – 1)$.

The expected gain at termination is

$$ES^* = (1 – q^k) – q^k(2^k – 1) = 1 – (2q)^k \qquad (1.58)$$

Note that for q > 1/2 the expectation is negative and for q = 1/2 it is 0 as it should be. Thus the presence of the house limit destroys the possibility of the martingale system working even if you have unlimited resources.

A roulette wheel has 38 slots of which 18 are red, 18 are black, and 2 are green. If you bet $1 on red you win or loose $1 according as the ball does or does not fall in a red slot. You win with probability p =18/38 = 0.4737 and loose with probability q = 0.5263. Suppose the house limit \$1000 and play the martingale. Then you have chance $1 - (.5263)^9 = 0.9969$ of winning \$1 and chance 0.0031 of loosing $\$2^9 - 1$ = \$511. Thus the gambler has a large chance of winning a small sum and a small chance of loosing a large sum.

14.6 STRONG LAW OF LARGE NUMBERS

14.6.1 THE LAW OF AVERAGES

Let $\overline{X}_n = \frac{1}{n}(X_1 + \ldots + X_n)$

be the arithmetic average of the n random variables $X_1, \ldots, X_n$. Suppose the random variables are independent and that they each have the same distribution. Let μ be the mean of their common distribution. A theorem in advanced probability theory known as the <u>strong law of large numbers</u> asserts that as n gets large the averages $\overline{X}_n$ converge to μ. In symbols,

$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1. \qquad (1.59)$$

The strong law of large numbers is often referred to as the <u>law of averages</u>. If the random variables X_i are indicator variables with success probability p, then $N_n = X_1 + \ldots + X_n$ is just the number of successes in the n trials and $\overline{X}_n = N_n/n$ is the proportion of successes in the n trials. This is typically denoted by $\hat{p}_n$ (i.e. $\hat{p}_n = N_n/n$). Since μ = p in this case, (1.59) asserts that

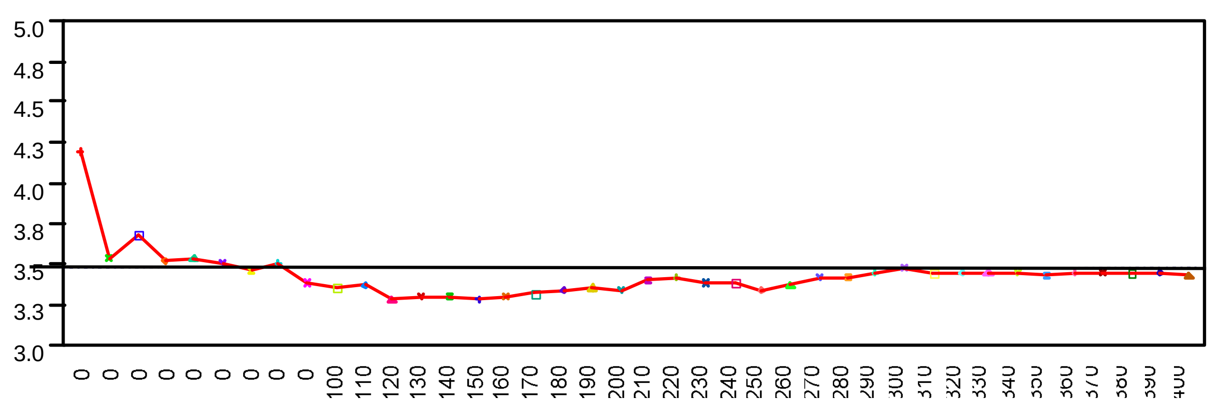
$$P\left(\lim_{n\to\infty} \hat{p}_n = p\right) = 1. \qquad (1.60)$$

Let's return to De Mere's gambling schemes in section 12.1. Why was he winning money with his first gambling scheme and loosing money with his second? Because, in the first scheme his probability of winning was a bit better than 1/2 and in his second it was less than 1/2. The law of averages (1.60) tells us that after a large number of plays the results that De Mere observed must happen.

The law of averages is often misunderstood, especially by gamblers. Suppose you are tossing a balanced coin repeatedly. Lets suppose the first 4 tosses are all heads. A common argument is then to believe that the next toss is much more likely to be a tail because by the law of averages in 5 tosses there should be about half heads and half tails. You now know that that is not true; the successive tosses are independent so that the next toss, like all the others, has probability 1/2 of falling heads.

The law of averages tell us that any event that can happen with non-zero probability is certain to happen infinitely often with probability one in infinitely many independent repetitions. To see this, let $X_i = 1$ if in the i^{th} repetition the event occurs and let $X_i = 0$ if not. Let p be the probability of an occurrence. If $p > 0$ then we know from (1.60) that $\lim_{n\to\infty}\frac{N_n}{n} = p$. So for large n, $N_n \approx np$, But if $p > 0$, $np \to \infty$ as $n \to \infty$. Therefore, $N_n \to \infty$. Thus the event must occur infinitely many times.

Although $\overline{X}_n$ converges to μ the convergence is quite slow. Consider rolling a balanced die. Let X be the number on the face showing. Then X is a random variable taking values 1, …, 6 with probability 1/6. and E(X)=3.5. Consider rolling the die 400 times. We then have 400 independent random variables, $X_1, \ldots, X_{400}$, each distributed like X. Let $\overline{X}_n$ be the arithmetic average of the first n of these random variables. In the chart below we plot the values of $\overline{X}_n$ for n=10,20, …,400



Observe that even for 400 rolls the total sample average is still a little away from the theoretical mean μ =.3.5. This illustrates that the law of large numbers does not take hold at a rapid rate.

14.6.2 EXTENSIONS OF THE STRONG LAW OF LARGE NUMBERS

We recall the following fact from calculus. Suppose g(x) is a continuous function on an interval containing a. If the sequence $\{x_n)$ converges to a, then the sequence $\{g(x_n)\}$ converges to g(a). This fact extends to the random sequence $\{\bar{X}_n\}$ as follows. Let g be continuous on an interval containing μ. Then

$$P(\lim_n g(\bar{X}_n) = g(\mu)) = 1. \tag{1.61}$$

We express (1.61)in words by saying $g(\bar{X}_n)$ converges to $g(\mu)$ with probability 1. Notice that the special case of g(x) = x is the same as(1.59).

EXAMPLE 1.48

Suppose X_1, X_2, … are independent indicator random variables with success probability p. What does $\sqrt{\hat{p}_n(1-\hat{p}_n)}$ tend to as n $\to \infty$?.

SOLUTION

Take g(x) = $\sqrt{x(1-x)}$ in (1.61). This shows that $\sqrt{\hat{p}_n(1-\hat{p}_n}$ converges to $\sqrt{p(1-p)}$.

***14.6.3.MONTE CARLO**

Suppose we wish to compute the integral of some function g of k variables. For functions of one variable, i.e. k =1, there are very efficient numerical integration methods. The situation is quite different for functions of more than one variable. As the number of variables increase the standard numerical integration methods become increasingly inefficient. Interestingly, one can use the strong law of large numbers to approximate the definite integral of a function g. The procedure for doing this is known as Monte Carlo Simulation. We will illustrate the basic idea involved for functions of one variable.

Let g be a function on the finite interval (a,b). Suppose we want I = $\int_a^b f(x)dx$.

Let X have the uniform distribution on the interval (a,b). That is, let X have density

$$f(x) = \begin{cases} \dfrac{1}{b-a} \text{ for } a < x < b. \\ 0, \text{ elsewhere.} \end{cases}$$

Then

$$(b-a)Eg(X) = \int_a^b g(x)f(x)dx = \frac{b-a}{b-a}\int_a^b g(x)dx = \int_a^b g(x)dx. \tag{1.62}$$

Now let $X_1, \ldots, X_n$ be independent random variables each with the uniform distribution on (a,b). By the strong law of large numbers,

$$\lim_{n\to\infty} \frac{g(X_1)+\ldots+g(X_n)}{n} = Eg(X).$$

Therefore

$$Eg(X) \approx \frac{g(X_1)+\ldots+g(X_n)}{n},$$

so

$$I \approx (b-a)\frac{g(X_1)+\ldots+g(X_n)}{n}. \tag{1.63}$$

Now one can use random number generators to obtain the $X_1, \ldots, X_n$, and therefore approximate I by the right hand side of (1.63).

Let's first try this procedure an example for which we know the answer. Suppose $g(x) = x^2$ and a = 0, b=2. Then exactly I = 8/3 = 2.667. Taking a sample of 2000 the right hand side of (1.63)came out to be 2.636.

Now lets try it on a function that cannot be integrated by ordinary calculus but can only be evaluated by numerical approximations. Let g(x) = e^{-x^2} and let a = 0, b = 1. Let I be the integral of g over the interval (0,1). Taking a sample of size 2000 the right hand side of (1.63) produced 0.745. Using a very accurate numerical integration technique produced a value of 0.747 for the integral.

***14.6.4 LAW OF INCLUSION-EXCLUSION**

Let $X_1,\ldots, X_n$ be indicator random variables. In (1.29) we gave a formula for P(at least one X_i = 1). We will derive an alternate method for computing this probability. Let

$$T_n = 1-[(1-X_1)\cdots(1-X_n)]$$

The random variable T_n can only have the values 0 or 1. If all of the X_i = 0 then T_n = 0. On the other hand, if at least one of the X_i = 1, then T_n = 1 Thus P(T_n = 1) = P(at least one X_i = 1). Since ET_n = P(T_n = 1) we see that

$$P(\text{at least one } X_i = 1) = ET_n = E[1-(1-X_1)\cdots(1-X_n)] \tag{1.64}$$

We now expand the product $(1-X_1)\cdots(1-X_n)$

$$(1-X_1)\cdots(1-X_n)=1-\sum_i X_i+\sum_{i<j}X_iX_j+\sum_{i<j<k}X_iX_jX_k+\ldots+(-1)^nX_1\cdots X_n.$$

Therefore

$$E[1-(1-X_1)\cdots(1-X_n)]$$

$$=\sum_i EX_i-\sum_{i<j}E(X_iX_j)+\sum_{i<j<k}E(X_iX_jX_k)+\ldots+(-1)^{n-1}E(X_1\cdots X_n). \qquad (1.65)$$

Now $EX_i = P(X_i = 1)$. The random variable X_iX_j is 1 if both $X_i = 1$ and $X_j = 1$. Otherwise it is 0. Therefore $E(X_iX_j) = P(X_i = 1, X_j = 1)$. Similarly, $E(X_iX_jX_k) = P(X_i = 1, X_j + 1, X_k = 1)$, etc. Using (1.64) and (1.65) we get the following extension of (1.10).

LAW OF INCLUSION - EXCLUSION

P(at least one $X_i = 1$) =

$$\sum_i P(X_i=1)-\sum_{i<j}P(X_i=1,X_j=1)$$
$$+\sum_{i<j<k}P(X_i=1,X_j=1,X_k=1)+\ldots+(-1)^{n-1}P(X_1=1,\ldots,X_n=1). \qquad (1.66)$$

For example, for n = 3 this equation says

$P(X_1 = 1 \text{ or } X_2 = 1 \text{ or } X_3 = 1) = P(X_1 =1) + P(X_2 =1) + P(X_3 =1)$

$- P(X_1 = 1, X_2 = 1) - P(X_1 = 1, X_3 = 1) - P(X_2 = 1, X_3 = 1)$

$+ P(X_1 = 1, X_2 = 1, X_3 = 1)$.

The k^{th} term in (1.66) is the sum of $\binom{n}{k}$ terms. In many applications of the inclusion-exclusion law the probability that any particular set of k of the indicators equal 1 is the same as for any other set of k. This produces a considerable simplification to (1.66). In this case, the k^{th} term is the sum of $\binom{n}{k}$ probabilities all of which are the same as $P(X_1 = 1, \ldots, X_k =1)$. So becomes (1.66)

$$\text{P(at least one of the } X_i = 1) = \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}P(X_1=1,\ldots,X_k=1). \qquad (1.67)$$

***14.6.5 MATCHING**

Suppose there are n boxes labeled 1,…, n and n balls also labeled 1, …, n. Balls are placed at random into the boxes with only one ball per box. A match occurs at i if ball i is in box i. Let $X_i = 1$ if there is a match at i and let $X_i = 0$ if not. The number of matches is $N = X_1 + \ldots + X_n$. We want to find $P(N > 0)$ which is the same as P(at least one of the $X_i = 1$).

Consider how we can have a match at k specified positions, say 1, …,k. To do so box 1 must have ball numbered 1, box 2 ball numbered 2, …, box k must have ball number k. The remaining n - k balls can go into any of the remaining n - k boxes. There are n! ways of distributing the balls into the boxes and there are (n - k)! ways of distributing them into the boxes subject to the condition that k specified boxes have k specified balls. Thus the probability for any such specification is (n - k)!/n!. The k^{th} term in (1.67) becomes $\binom{n}{k}\frac{(n-k)!}{n!} = \frac{1}{k!}$ and using this we get from (1.67) that

$$P(N > 0) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}. \qquad (1.68)$$

From the above equation we find the probability p_n of no matches is

$$p_n = P(N = 0) = 1 - \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}. = 1 - 1 + \frac{1}{2!} + \cdots + (-1)^n \frac{1}{n!}. \qquad (1.69)$$

Later, in Chapter 2, we will show that for n >10 the sum on the right is very closely approximated by 1/e ≈ 0.3679. Therefore the chance of at least one match is about 0.6321. This is quite interesting since the number is independent of n. For example, for n = 100 the chance of a match at any specified position is 1/100. For n = 1,000,000 the chance of a match at any specified position is 1/1,000,000. Nonetheless, the chance of at least one match is the same, namely 0.6321 for either case.

It is now not difficult to compute P(N = k) for any k. Let m_n be the number of ways of having no match with n boxes. Then $p_n = \frac{m_n}{n!}$ so $m_n = n!p_n$. Suppose k boxes are specified, say boxes 1, ... ,k . To have a match occurring in each of these k boxes and have no other matches we must have a match in these boxes and the remaining n –k balls must go into the remaining n –k boxes with no match. Therefore the number of ways of having exactly k matches in k specified boxes is $\frac{m_{n-k}}{n!} = \frac{(n-k)!p_{n-k}}{n!}$. Without specifying which k boxes are to have the k matches we have $\binom{n}{k}$ choices for the k boxes that are to have the match.

Therefore

$$P(N = k) = \binom{n}{k}\frac{(n-k)!p_{n-k}}{n!} = \frac{p_{n-k}}{k!} \approx e^{-1}\frac{1}{k!} \tag{1.70}$$

EXAMPLE 1.49

20 people go to a picnic. Each brings a boxed lunch. These are gathered together and then randomly assigned to the 20 people. (a) What is the approximate probability that no on receives the lunch they brought? (b) What is the approximate probability that exactly 2 people receive their own lunch?

SOLUTION

This is the matching problem with n = 20 (a) $P(N_{20} = 0) = 0.3679$. (b) $PN_{20} = 2) = 0.1839$.

15 VARIANCE

15.1 VARIANCE AND STANDARD DEVIATION

The mean attempts to summarize a distribution by a single number that is supposed to be a typical value for a quantity having that distribution. By itself, this gives little idea of what the distribution is really like. Consider the systolic blood pressure measurements. If I tell you the mean of the distribution of systolic blood pressure is 120 what does this tell you? Roughly speaking it tells you that you would anticipate a blood pressure to be near that number with perhaps some exceptions. But how many exceptions and by how much will these differ from 120? As can be seen from Table 1-1 most will be near 120 but we get some as large as 140 and some as small as 90. What we need is a measure of how much spread there is in the distribution about its mean. That is supplied by the variance of a distribution and the square root of the variance called the standard deviation.

Let X be a random variable with mean μ. The variance of X denoted Var(X) or σ^2 is the quantity

$\mathrm{Var}(X) = E(X - \mu)^2$. (1.71)

Since μ = EX, we often write the above as

$$\mathrm{Var}(X) = E(X - EX)^2.$$

The standard deviation of X is $\sqrt{\mathrm{Var}(X)} = \sigma$.

As with the mean, we refer to the variance and standard deviation as both a property of a random variable and of a distribution. Thus for example we speak of a distribution having standard deviation σ.

By using the rules of expectation and expanding the right hand side (1.71) of we obtain an alternate formula for Var (X) that is more useful for computation.

$\mathrm{Var}(X) = EX^2 - \mu^2$. (1.72)

Here is the derivation of (1.72).

By algebra,

$(X-\mu)^2 = X^2 - 2\mu X + (EX)^2$. Using the rules for expectation we find $E(X-\mu)^2 =$ $E[X^2 - 2\mu X + (EX)^2] = EX^2 + E[-2\mu X] + E^2\mu^2$

$= EX^2 - 2\mu EX + \mu^2 = EX^2 - 2\mu^2 + \mu^2 .EX^2 - \mu^2$.

EXAMPLE 1.50

Let X be an indicator random variable with success probability p. Find the variance and standard deviation of X.

SOLUTION

We use (1.72). We already know from Example 1.36 that EX = p. Now $EX^2 =$ $(1)^2p + (0)^2(1-p) = p$. So by (1.72) Var (X) = $p - p^2 = p(1-p)$. The standard deviation is $\sqrt{p(1-p)}$.

EXAMPLE 1.51

Let X have the density in Example 1.39. Find the variance of X.

SOLUTION

We have already found EX = 2/3. We need EX^2. Now using the result in Example 1.44 with m = 2 we find $EX^2 = 2/4$. By (1.72) Var (X) = $(2/4) - (2/3)^2 =$ 0.06.

15.2 RULES FOR VARIANCE

Below, we list three of the four basic properties of variance.

Let X be a random variable and c be a constant.

a. Var (c) = c

b. Var (cX) = c^2 Var(X)

c. Var (X + c) - Var (X)

These rules follow easily from the definition of variance and the properties of expectation. For example,

Var(cX) = $E(cX - E(cX))^2 = E(cX - cEX)^2 = E(c(X\text{-}EX))^2 = c^2E(X\text{-}EX)^2 = c^2$Var(X).

EXAMPLE 1.52

Suppose X has variance 2. Find (a) Var (X –3), (b) Var (3X).

SOLUTION

(a) Var (X –3) = Var (X) = 2. (b) Var (3X) = 3^2Var(X) = (9)(2) = 18.

EXAMPLE 1.53 (change of scale)

Suppose X is a random variable with mean μ and standard deviation σ. Let a $\neq$ 0 and b be constants. Let Y = aX + b. [The random variable Y often arises as a scale change on the random variable X. For example, if X is a temperature in degrees centigrade then this temperature in degrees Fahrenheit is Y = (9/5) X +32. Find (i) the mean of Y, (ii) the variance of Y, and (iii) the standard deviation of Y

SOLUTION

(i) EY = E(aX +b) = a EX +b = aμ + b.

(ii) Var (Y) = Var(aX + b) = Var(aX) = a^2Var (X) = $a^2\sigma^2$.

(iii) The standard deviation of Y is $\sqrt{a^2\sigma^2} = |a|\sigma$.

15.3 VARIANCE OF A SUM

Let X and Y be random variables. What is the Var(X + Y)? By definition

Var(X + Y) = $E(X + Y - E(X + Y))^2$.

Now

$(X + Y - E(X + Y))^2 = (X + Y - EX - EY)^2 = ((X\text{-}EX) + (Y - EY))^2$.

Expanding, we see that

$((X\text{-}EX) + (Y - EY))^2 = (X - EX)^2 + (Y - EY)^2 + 2((X - EX)(Y - EY))$.

Thus,

$$Var(X + Y) = E(X - EX)^2 + E(Y - EY)^2 + 2E((X - EX)(Y - EY))$$

$$= Var(X) + Var(Y) + 2E((X - EX)(Y - EY)). \qquad (1.73)$$

In general $E((X - EX)(Y - EY)). \neq 0$. Thus, in general, the variance of the sum is not the sum of the variances.

For an extreme example Take Y to be a multiple of X, say $Y = cX$, where $c \neq 0$. Then

$X + Y = X + cX = (1 + c)X$.

So

$Var(X + Y) = Var((1 + c)X) = (1 + c)^2 Var(X)$.

Now

$Var(X) + Var(Y) = Var(X) + Var(cX) = (1 + c^2)Var(X)$.

But $(1 + c^2) = (1 + c)^2$ if and only if $c = 0$.

15.3.1 INDEPENDENT RANDOM VARIABLES

In the special case when X and Y are independent $E((X - EX)(Y - EY)) = 0$. To see that this is the case note that if X and Y are independent, then so are $X - EX$ and $Y - EY$. Thus by Rule 7 on expectations, $E((X - EX)(Y - EY)) = (E(X - EX)E(Y - EY)$. But $E(X - EX) = EX - EX = 0$ so $E((X - EX)(Y - EY)) = 0$. Thus for independent random variables

$Var(X + Y) = Var(X) + Var(Y)$.

This last formula extends by induction from 2 to n independent random variables and leads to the following rule:

Let $X_1, \dots, X_n$ be independent random variables. Then

$$Var(X_1 + \dots + X_n) = \sum_{i=1}^{n} Var(X_i). \qquad (1.74)$$

EXAMPLE 1.54

Suppose X and Y are independent. If $Var(X) = 2$ and $Var(Y) = 3$ find (i) $Var(X + Y)$ and (ii) $Var\ X - Y$).

SOLUTION

(i) Var (X + Y) = Var (X) + Var (Y) = 2+3 = 5

(ii) Var (X – Y) = Var (X) + Var (-Y) = Var (X) + $(-1)^2$ Var (Y) = 2 +3 = 5.

COMMENT

A common error is to think for X and Y independent, Var (X – Y) = Var (X) – Var (Y). This is not true. In fact with X and Y as in this last example this wrong formula would yield as the variance of X-Y the answer 2-3 = -1, a clearly wrong answer.

EXAMPLE 1.55 (Independent indicators)

Suppose n repetitions are made of the same success- failure experiment having success probability p. Let X_i = 1 if the ith repetition is a success and let X_i = 0 if not. The number of successes in these n repetitions is N = $X_1 + \ldots + X_n$. Find the mean and variance of N.

SOLUTION

We already know from Example 1.43 that

EN = $E(X_1 + \ldots + X_n) = EX_1 + \ldots + EX_n$ = np. Example 1.50 shows $\mathrm{Var}(X_i)$ = p(1 – p). Using (1.74), we see that

$\mathrm{Var}(N) = \mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n) = np(1-p)$. Thus

$$\mathrm{Var}(N) = np(1 - p) \qquad (1.75)$$

***15.3.2 GENERAL CASE**

Equation (1.73) gives the general formula for the variance of the sum of two random variables. The quantity E((X – EX)(Y – EY)) is called the <u>covariance</u> between X and Y and is denoted as cov(X,Y). That is

$$\mathrm{cov}(X,Y) = E((X - EX)(Y - EY)). \qquad (1.76)$$

From the formula above it is easy to verify that id a and b are constants the

$$\mathrm{cov}\,(aX,bY) = ab\mathrm{Cov}\,(X,Y). \qquad (1.77)$$

By using the rules of expectation we can find a formula for the variance of the sum of n random variables $X_1, X_2, \ldots, X_n$.

$$\mathrm{var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathrm{var}(X_i) + 2\sum_{i,j,i<j}^{n} \mathrm{cov}(X_i, X_j). \qquad (1.78)$$

The proof of formula (1.78) is given in the appendix.

For 3 random variables the above formula is

$$Var(X_1 + X_2 + X_3) = Var(X_1) + Var(X_2) + Var(X_3) + 2[cov(X_1, X_2) + cov(X_1, X_3) + cov(X_2, X_3)].$$

EXAMPLE 1.56

Suppose X and Y are random variables such that Var (X) = Var (Y) = 1 and Cov (X,Y) = -1/2. Find Var (X – Y)

SOLUTION

Using (1.78)

Var X – Y) = Var (X) + Var (-Y) +2Cov (X, –Y) = Var (X) + $(-1)^2$Var (Y)

+ 2(-1)Cov (X,Y) = 1 + 1 + 2(-1/2) = 1.

***16 BLOOD TESTS**

A large number n of people are to be given a test for a certain disease having incidence rate p. For example, suppose we want to test everyone in the USA for HIV. For simplicity, assume both the sensitivity and specificity of the test is 1.[This unrealistic assumption will not materially effect the results given here.] Testing all individuals one by one requires n tests. If p is small most of the time the test will be negative. Is there a more efficient method that requires fewer tests? Exactly this problem confronted military authorities during Word War II when they needed to test for venereal disease. Here is the solution they came up with.

Divide the population of n individuals into groups of m individuals each (except perhaps for a last group which could have fewer than m individuals). Samples of the blood from these m individuals are mixed together to form a pooled sample and this pooled sample is tested. If the pooled sample is negative we know that none of the m individuals in that group have the disease. In this case only one test was required for the m individuals. On the other hand, if the pooled sample is positive, then each of the individuals in that group must be separately tested. This requires m additional tests to the one used to test the group or m+1 tests total to screen the m individuals in that group. With this procedure the number of tests needed is a random variable having only two possible values 1 and m+1.

For simplicity, we will ignore the last group and assume that n is an exact multiple of n, say n = km. Then there are k groups each with m individuals. Let X_i be the number of tests required to screen the individuals in group i. Let Z_i = 1 if the pooled sample from group i is positive and let $Z_i = 0$ if not. Then $X_i = 1+mZ_i$. The number of tests needed are $N = X_1 + \ldots + X_k$. Notice in using this procedure we are gambling on the individuals in the group to all be negative. In that case we win by only having to use one test. However, if someone in that group does have the disease we loose over having tested everyone because we require m + 1 test to test these m individuals. With the second procedure the number of tests required is random so we can't say for certain how many tests will be needed. We can however determine the expected number of tests required.

Now,

$EX_i = E(1+mZ_i) = 1 + m\,P(Z_i = 1)$.

Equation (1.30) shows $P(Z_i = 1) = 1-(1-p)^m$, so

$EX_i = 1 + m(1-(1-p)^m)$.

Consequently,

$$EN = \sum_{i=1}^{k} EX_i = \underbrace{[1+(1-(1-p)^m)]+\cdots+[1+(1-(1-p)^m)]}_{k\text{ terms}} = k(1+m(1-(1-p)^m)).$$

Since k = n/m,

$$EN = \frac{n}{m}(1+m(1-(1-p)^m)) = n\left(\frac{1}{m}-(1-p)^m+1\right) = n + n\left(\frac{1}{m}-(1-p)^m\right).$$

Thus

$$EN = n + n\left(\frac{1}{m}-(1-p)^m\right). \qquad (1.79)$$

We will now find the variance of N.

$$\mathrm{Var}(N) = \mathrm{Var}(X_1 + \ldots + X_m) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_m) \qquad (1.80)$$

and

$\mathrm{Var}(X_i) = \mathrm{Var}(1 + mZ_i) = m^2\mathrm{Var}(Z_i)$.

Since Z_i is an indicator random variable with success probability $1-(1-p)^m$, it follows from Example 1.52 that

$\mathrm{Var}(Z_i) = (1-(1-p)^m)(1-p)^m$.

Thus by (1.80) and the fact that k = n/m,

$$\mathrm{Var}(N) = nm(1-(1-p)^m)(1-p)^m. \qquad (1.81)$$

On average, the number of tests required when using the pooled method is given by (1.79). Looking at the right hand side of this expression we see that it is possible for EN to be larger than n (which happens if the expression in the brackets is positive) or it could be less than n (which happens if the expression in the brackets is negative). If EN >n then the pooling is ineffective and we should test everyone. On the other hand, if EN<n, the pooling is effective and we should use it. These possibilities raise several questions since we are at liberty to choose the group size m. Is there a choice of m which makes the expression in brackets negative? If so, what choice of m will make it as negative as possible? That choice of m would be the optimal group size since it will make EN as small as possible.

To answer these questions only requires calculus. Looking at (1.79) we see that what is at issue is if for a given p the function f(m) = $\frac{1}{m}-(1-p)^m$ can ever be negative and if so what value of m makes it as negative as possible.

To get started, we use a time-honored procedure in mathematics, namely that of looking at the extreme cases, for a clue to what is going on. If p = 0, (1/m) –1 is negative for all m and it takes its minimum at the largest value that m can have, namely, n. This tells us that if p = 0 all the blood samples should be pooled into one group. At the other extreme, if p = 1, the function is (1/m) and no choice of m will make it negative. These extremes suggest that for small p the pooling should be effective while for large p it should not. Consequently, we would anticipate that there should be a cutoff value of p, say $p_{0,}$ so that for p > $p_{0,}$ $\frac{1}{m}-(1-p)^m > 0$ for all m and for p < p_0 there are m for which this expression can be negative.

To find this p_0 we argue as follows. Suppose p is such that for all m $\frac{1}{m}-(1-p)^m \geq 0$. Then (1/m) $\geq$ $(1-p)^m$. This is equivalent to

$$(1/m)^{1/m} \geq 1-p \text{ for all m.} \tag{1.82}$$

We will now find where $(1/m)^{1/m}$ has a minimum. For this purpose consider the function $h(t) = (1/t)^{1/t} = e^{(1/t)\ln(1/t)}$ for $t>0$. Taking the derivative we find $h'(t) = h(t)((-1 + \ln(t))/t^2)$. Setting this equal to 0 and solving for t we find t = e, so $h'(e) = 0$. The 2nd derivative test shows that e is a minimum of the function h. Hence $(1/m)^{1/m} \geq (1/e)^{1/e}$. Thus for (1.82) to hold for all m it must be that $(1/e)^{1/e} \geq 1-p$ or $p \geq 1-(1/e)^{1/e}$. Consequently, the smallest value that p can have for this to be true is $p = p_0 = 1 - (1/e)^{1/e} = 0.3078$.

Thus if the disease incidence is greater than 0.3078 pooling will not be effective. On average we will use more tests than testing all n people. If the disease incidence is smaller than 0.3078 then there are choices of m that will make $\frac{1}{m} - (1-p)^m$ negative. Fig1-2 shows this function for p = .2.

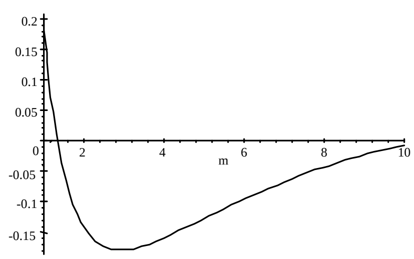


Figure 1-4

To find the value of m that makes this as negative we must find the minimum of this function. One way is to start with m =1 and compute successive values of $\frac{1}{m} - (1-p)^m$ until we find the minimum. The table below shows this method for p = 0.1

m	1	2	3	4	5
$\frac{1}{m} - (1-p)^m$	0.1	0.31	0.39567	0.4061	0.3905

As can be seen from the table the minimum occurs for m = 4. This method is fine unless p is very small (in which case the minimum occurs at a large value of m)

A more efficient method is to use calculus to find the minimum. But this leads to an equation that cannot be solved explicitly.

Fortunately, for small p, there is an excellent approximation to $(1-p)^m$ that can be used. Using L'Hopital's rule it is easy to verify that

$$\lim_{p\to 0}\frac{1-(1-p)^m}{mp}=1$$

Consequently, for small p, $(1-p)^m \approx 1-mp$. [We use the symbol $\approx$ as in $a\approx b$ to mean a is approximately equal to b] Replacing $(1-p)^m$ with $1-mp$ leads to the expression $\frac{1}{m}-1+mp$. Setting the derivative of this function to 0 and solving we find it has a unique minimum at $m = 1/\sqrt{p}$. Thus for small p, we take $m = 1/\sqrt{p}$ as the optimal m. Substituting this value for m in the expression $\frac{1}{m}-1+mp$ yields the value $2\sqrt{p}-1$. So replacing $\frac{1}{m}-(1-p)^m$ with $2\sqrt{p}-1$ in (1.79) we obtain the following expression for EN:

$$EN \approx 2n\sqrt{p}. \qquad (1.83)$$

In the table below we give the exact optimal m, the value $1/\sqrt{p}$, the exact EN, and the approximate EN for various values of p when we take $m = 1/\sqrt{p}$.

Table 1-19

p	m optimal	EN (exact)	$1/\sqrt{p}$	$2n\sqrt{p}$
0.3	3	0.9903n	1.8257	1.0954n
0.1	4	0.5939	3.1623	0.6325n
0.05	5	0.4263n	4.4721	0.4472n
0.01	10	0.1956n	10	0.2000n
0.001	32	0.0629n	31.6228	0.0632n

The above table makes it clear that the approximation is very good for $p \le 0.01$. It also makes it clear that for small p (which is usually the case) there seem to be substantial savings using the pooled method.

For example, we have seen that for HIV infection the incidence rate in the heterosexual population is 0.001. If that population was to be screened for HIV, then the expected number of tests needed using the pooled method with the optimal group size of 32 would on average require only 6,290 tests per 100,000 persons screened. This is about 6% of the number of tests needed to test each individual separately!

Using the approximations $m = 1/\sqrt{p}$, $1-(1-p)^m \approx mp$, and $(1-p)^m \approx 1-mp$ we find from (1.81) that

$Var(N) \approx n(1-\sqrt{p})$.

Since p is small $(1-\sqrt{p}). \approx 1$.

Thus

$$Var(N) \approx n \quad (1.84)$$

Now what have we accomplished? Have we shown that pooling is really better than testing everyone? No. All we have shown is that the average number of tests is much smaller. The actual number of tests N will surely be different than this average. How much different will it be? We can't answer that for sure because N is random. Lets measure things in units of standard deviation. Equation (1.84) shows that the standard deviation is $\sqrt{n}$. What we would like to know is the probability

$$P(N > EN + a\sqrt{n}) \quad (1.85)$$

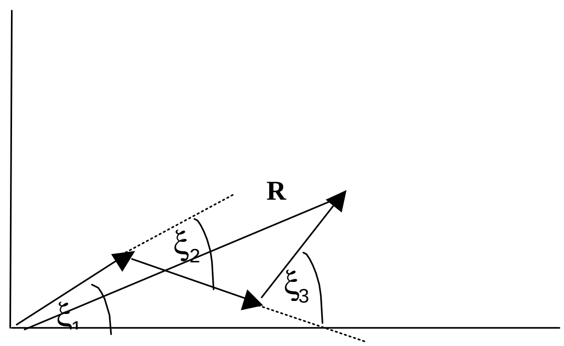
for various a like 1, 2 , etc. Suppose we new that for a = 4 that that probability was less than 0.0001. Suppose that p = 0.001. Using Table 1-18 we then could say with essential certainty that we would not need more than $0.0632n + 4\sqrt{n}$ tests. For n = 1,000,000 this would say that not more than 67,200 tests would be needed. Thus being able to compute the probabilities in (1.85) would certainly provide the information needed to assert that the pooling is really effective.

For the moment we cannot compute these probabilities. That will be rectified in Chapter II by using one of the most remarkable results in probability theory known as the Central Limit Theorem.

***17 POLYMER CHAINS**

The following is a simple model for a polymer chain in the plane. Suppose each link has a fixed length 1. The first link makes angle ξ_1 with the x axis and the link i +1 makes angle ξ_i with the x axis and link i, i > 1 makes angle ξ_i with link i - 1. (See Fig.1-3). We assume the $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables having a symmetric distribution. Let R^2 be the square of the distance from the origin to the end of the chain. We will compute ER^2.

Figure 1-5



Let $S_i = \xi_1 + \ldots + \xi_i$. Then the x- component of **R** is $\cos(S_1) + \ldots + \cos(S_n)$ and the y component is $\sin(S_1) + \ldots + \sin(S_n)$.

Therefore

$$R^2 = \left[\cos(\xi_1) + \ldots + \cos(\xi_1 + \ldots + \xi_n)\right]^2 + \left[\sin(\xi_1) + \ldots + \sin(\xi_1 + \ldots + \xi_n)\right]^2 .$$

Expanding and using the identities cos(a-b) = cos (a)cos(b) + sin(a)sin(b) and $\cos^2(a) + \sin^2(a) = 1$, we find

$$R^2 = n + 2 \sum_{1 \le i < j \le n} \cos(S_j - S_i). . \quad (1.86)$$

Let $E\cos(\xi) = a$ and $E\sin(\xi) = b$ where ξ is a random variable having the same distribution as the ξ_i. Since the distribution is symmetric and sin is an odd function b = 0.Now

$\cos(S_j - S_i) = \cos(\xi_{i+1} + \xi_{i+2} + \ldots + \xi_j)$

Using the identity cos(a-b) = cos (a)cos(b) + sin(a)sin(b) we can write

$\cos(S_j - S_i) = \cos(\xi_{i+1})\cos(\xi_{i+2} + ... + \xi_j) + \sin(\xi_{i+1})\sin(\xi_{i+2} + ... + \xi_j)$. Since the ξ_i are independent and $E\sin(\xi_{i+1}) = 0$,

$E\cos(S_j - S_i) = E\cos(\xi_{i+1})E\cos(\xi_{i+2} + ... + \xi_j) + E\sin(\xi_{i+1})E\sin(\xi_{i+2} + ... + \xi_j)$.

$= E\cos(\xi_{i+1})E\cos(\xi_{i+2} + ... + \xi_j) = aE\cos(\xi_{i+2} + ... + \xi_j)$. Repeating this computation again and again we find

$$E\cos(\xi_{i+1} + ... + \xi_j) = a^{j-i}. \tag{1.87}$$

Substituting (1.87) into (1.86) we find

$$ER^2 = n + 2\sum_{1\le i<j\le n} E\cos(S_j - S_i) = n + 2\sum_{1\le i<j\le n} a^{j-i} = n + 2\sum_{j=1}^{n}(a + ...a^{j-1}).$$

So

$$ER^2 = n + 2\sum_{j=1}^{n}(a + ...a^{j-1}). \tag{1.88}$$

At this point we can make use (1.57) of the formula for the sum of a geometric progression. This states that for $a \neq 1$,

$$1+a+\ldots+a^{k-1} = \frac{1-a^k}{1-a}.$$

Using this formula we find

$$a+\ldots+a^{j-1} = a(1+\ldots+a^{j-2}) = \frac{a(1-a^{j-1})}{1-a}.$$

Thus the 2nd term on the right in (1.88) is

$$2\sum_{j=1}^{n}\frac{a(1-a^{j-1})}{1-a} = 2n\frac{a}{1-a} - 2\frac{a(1-a^n)}{(1-a)^2}.$$

Substituting this back into (1.88) and combining terms we find

$$ER^2 = n\frac{1+a}{1-a} - 2a\frac{1-a^n}{1-a}. \tag{1.89}$$

CHAPTER 1 PROBLEMS

SECTIONS 1-3

1. Let X have probability function

x	-3	-2	-1	0	1	2
P(X=x)	4/20	2/20	3/20	1/20	4/20	6/20

Find (a) $P(X \leq 0)$, (b) $P(X < 0)$, (c) $P(X \geq -1)$, (d) $P(X < -1 \text{ or } X > 1)$.

2. Let P(X = x) = ax for x = 1,2,3 = 0 and P(X = x) = 0 elsewhere. Find a.

3. Suppose X has probability function

x	0	1	2	3	4
P(X=x)	.1	.4.	.1		.1

Find P(X=x)

4. Let X have probability function

x	-2	-1	0	2	3	4
P(X=x)	1/15	2/15	4/15	3/15	2/15	3/15

Find (a) $P(X < 0)$ (b) $P(X \geq -1)$ (c) $P(-2 \leq X < 3)$

5. A box has 5 red, 10 white, and 5 green balls. A ball is selected from the box. Let Y = 1 if it is red, Y = 2 if it is white, and Y = 3 if it is green. Find the probability function of Y.

6. Let X and Y have joint probability function given by the following table:

	y		
x	-1	0	1
-1	1/20	4/20	4/20
0	3/20	1/20	2/20
1	2/20	0	3/20

Find (a) P(X = Y) (b) P(X < Y) (c) The probability function of X (d) The probability function of Y.(e) $P(X= 0 \text{ and } Y \in \{-1,1\})$

7. Let X and Y have joint probability function given by the following table:

	y			
x	1	2	3	4
1	3/20	2/20	1/20	1/20
2	1/20	3/20	0	1/20
3	2/20	1/20	5/20	0

a) Find the marginal probability function for X and the marginal probability function for Y

(b) Find P(X $\leq$ 2, Y > 2)

(c) Find P(X $\leq$ 2 or Y > 2)

8 Suppose a card is selected at random from an ordinary deck of playing cards. Let X and Y be as in Example 1.6, i. e. let X = i, i =1, 2, …, 10 if the card has face value i and let X = 11 if the card is a jack, X =12 if the card is a queen, and X = 13 if the card is a king. Let Y = 1 if the card is a club, Y =2 if the card is a spade, Y = 3 if the card is a heart, and Y = 4 if the card is a diamond. (a)Find P(X <6 and Y >1) (b) Find P(X <5 and Y =2 or 4).

9 Suppose we roll two balanced dice. Let X = i if the first die shows face i and let Y = j if the second die shows face j, i, j = 1,2 …, 6. Take the joint probability function of X and Y to be P((X.Y) = (x,y)) = 1/36 for x = 1,…, 6 and y = 1,…,6. Find (a) P(X = 2 or Y = 3), (b) P(X $\leq$ 2 or Y > 4).

10 Let S be the sum of the face values obtained in throwing two balanced dice. Verify that the entries in the following table are valid.

Table 1-20

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

11 A person is either rh positive or rh negative. There are four blood types A, B , AB and O. Suppose X = 1 if the person is rh positive and X = 0 if not. Let Y = 1 if the person is type A, Y = 2 if the person is type B, Y = 3 if the person is type

AB, and Y =4 if the person is type O. Suppose we know the following information about the joint probability of X and Y:

x	1	2	3	4	P(X =x)
1	.35		.05		.9
0				0	
P(Y = y	.4	.3	.1		

Fill in the remaining entries of the table.

SECTIONS 4-5

12. Suppose 80% of the population have antibodies to Barr-Epstein virus, 50% have antibodies to Clamydia, and 40% have both antibodies. (a) What is the probability a person has either of the two antibodies? What is the probability that a person has antibodies to Barr-Epstein but does not hav the Clamydia antibody?

13. Three cards are randomly selected from an ordinary deck of cards. Find the probability of a triple, i.e. exactly 3 cards of the same face value.

14. 4 cards are randomly selected from an ordinary deck of cards. (a) Find the probability that all 4 cards have different face values. (b) Find the probability that all 4 cards have the same face value.

15. Parking overnight is not allowed on a certain street. (a) If you receive 6 parking tickets and they are all given on a Saturday or Sunday are you justified in concluding that the tickets are not equally likely to be given on any of the 7 days? [Compute the probability that you get 10 tickets on the specified days under the assumption that they are equally likely to be given on any day and see how small it is.] (b) If none are given on Sunday is it reasonable to conclude that no tickets are issued on Sunday? (c) If you get 12 tickets and none are issued on Sunday can you conclude that tickets are not issued on Sunday? .

16. Suppose 10 people are randomly arranged in a row. (a) Find the prob ability that Harry and Sally are next to each other. (b) If there are n people what is the probability that Harry and Sally are next to each other?

17. There are twelve parking spaces all in a row. Cars are supposed to be parked completely at random in these 12 spaces. Suppose eight cars are parked. What is the probability that the four empty spaces are all next to each other.

18. Chromosomes occur in pairs. Suppose n pairs of chromosomes are separated to form 2n chromosomes. Chromosomes are then randomly picked in pairs. Let $X_i = 1$ if the ith pair of chromosomes selected are from the same original pair and let $X_i = 0$ if not. (a) Find $P(X_1 = 1)$. (b) Find $P(X_2 = 1|X_1 =1)$ and $P(X_1 = 1, X_2 = 1)$. (c) Find $P(X_3 = 1|\ X_1 = 1.\ X_2 = 1)$ and $P(X_1 = 1, X_2 = 1, X_3 = 1)$. (d) Find $P(X_k = 1|X_1 = 1, ..., X_{k-1} = 1)$ where $1 < k \leq n$. and use this to find $P(X_1 = 1, ..., X_k = 1)$. (e) Find the probability that all n pairs picked have their two components from the original pairs.

19. Suppose n cards, numbered 1,2, …,n, are randomly shuffled. (a) Find the probability that the i^{th} card from the top is card number i.(b) Find the probability that two distinct cards i and j from the top are cards number i and j respectively.

20. Suppose n balls are randomly distributed into r boxes. (a) Find the probability that box i is empty. (b) find the probability that two different boxes i and j are empty.

SECTION 6

These problems all are for poker.

21. Find the probability of 4 of a kind.
22. Find the probability of a full house.
23. Find the probability of a straight flush.
24. Find the probability of a royal flush
25. The computation of the probability of a straight includes that of a straight flush. Find the probability of a straight that is not a straight flush.
26. The computation of the probability of a flush includes that of a straight flush. Find the probability of a flush that is not a straight flush.

27. Find the probability of a straight flush that is not a royal flush.
28. (a) Find the probability that the 5 cards have different face values. (b) Find the probability that the 5 cards have different face values and neither a straight nor a flush.

SECTION 7

29. (a) Let X and Y be as in problem 6. Find the conditional probability function of Y given X and the conditional probability function of X given Y. (b) Find $P(Y \leq 0 \mid X = 0)$ (c) Find $P(Y \leq 0 \mid X \leq 0)$.
30. The following problem caused quite a stir in the popular press in 1990-1991 because of its appearance in the "Ask Merilyn" column in Parade magazine. It related to a television show called "Lets Make A Deal" and the problem is named after the host of that Host of that show Monty Hall and is called the "Monty Hall Problem". The problem is as follows. There are three doors. Behind door 1 is a prize and there is nothing behind the other two doors. A contestant chooses a door. Clearly the chance of winning the prize is 1/3. Monty chooses another door and opens it to show the contestant there is nothing behind his door. He then offers the contestant the opportunity to switch from the door first picked to the remaining unopened door. Monty knows the door that hides the prize and he never opens that door. The problem is what is the contestant's chance of winning if the contestant switches. The columnist Merilyn vos Savant claimed the chance of winning remained the same, namely 1/3. A large number of people claimed she was wrong. Some claimed the chances were now 1/2.Others claimed it was now 2/3. Show the correct probability is 2/3. [Hint let Y=1 if you switch and win the prize and let Y = 0 if you switch and do not win the prize. Let X = i if the contestant first chooses door i. We want P(Y = 1). We can easily determine P(Y=1|X=i)]. [You can play this game for yourself on the WEB. Go to http://www.stat.sc.edu/~west/javahtml/LetsMakeaDeal.html Try it for yourself. Play it 20 times where you switch and 20 times where you don't switch to see

that the correct answer is 2/3. [If your reading this on the web just press the above link to play the game.]

31. A multiple-choice exam has 5 choices. A student knows the correct answer to 40% of the questions. On a given question, if the student knows the answer he marks the correct answer. If not, he chooses an answer at random. (a) Find the probability that the student gets the correct answer to a given question. (b) Find the probability that the student guessed the answer to a question that he got correct.

32. In Seattle it rains 3 days out of 4. If rain is forecast Jones takes his umbrella. When rain is not forecast Jones takes his umbrella 1/3 of the time. The forecaster is correct 3/4 of the time. (a) What is the probability Jones is caught in the rain with no umbrella? (b) What is the probability that Jones carries his umbrella when it does not rain?

SECTION 8 - 9.

33. The odds for success for an indicator random variable are p/1-p, where p is the success probability. In diagnostic testing, the prior odds O for the disease is p/1-p, where p is the disease incidence. The posterior odds O* for having the disease is $\frac{P(D=1|T=1)}{1-P(D=1|T=1)}$ (a) Show that O* = (a/1 - b)O. (b) Let r = a/1 - b. Give conditions on r so that O* > O, O* = O, and O* < O.

34. Find a formula relating the posterior odds for not having the disease $\frac{P(D=0|T=0)}{1-P(D=0|T=0)}$ to the prior odds against the disease (1 - p)/p.

35. Let X and Y be two indicator random variables. (a) Show that P(X = 1, Y = 1| X = 1 or Y = 1) = $\frac{P(X=1,Y=1)}{1-P(X=0,Y=0)}$ (b) If X and Y are independent each with success probability p what is P(X = 1, Y = 1| X = 1 or Y = 1)?

36. Verify the entries in Table 1.10.

37. Box 1 has 6 red and 6 green balls. Box 2 has 4 red and 6 green balls. A ball is selected from each box and placed in the opposite box. What is the probability that the box compositions remain the same? (Hint. Let X = 1 or 0 as a red ball

is selected from box 1 and let Y = 1 or 0 as a red ball is selected from box 2. Then the box compositions remain the same i and only if X = Y.)

38. A man rolls a red die and a white die. What is the probability that the number showing on the red die is smaller than the number showing on the white die?

39. (a) In throwing 2 dice is it more likely to get a 9 than a 10? (b)In throwing 3 dice is it more likely to get a 9 than a 10?

40. Genes occur in pairs. A particular gene has two variants (alleles) Suppose that the probability that the two genes in the pair are same allele is p and that the probability that they are different is 1 - p. One of the two genes is chosen at random and copied. The process is the repeated. What is the probability that the two copies are the same allele?

41. Consider a disease and a risk factor for the disease. The joint probabilities of R and D, the conditional probabilities of D given R and the conditional probabilities of R given D are given in the table below.

	Joint		D given R		R given D	
	D		D		D	
R	1	0	1	0	1	0
1	a	b	a*	b*	a**	b**
0	c	d	c*	d*	c**	d**

Show $.\frac{ad}{bc} = \frac{a*d*}{b*c*} = \frac{a**d**}{b**c**}$

This shows that the odds ratio is the same is the same no matter how a study is carried out.

42. Show the odds ratio is 1 if R and D are independent and conversely if the odds ratio is 1 then R and D are independent.

43. In death by murder compute the relative risk of being murdered for a black male to a white male.

44. Jack and Jim went partying and missed the final exam in their statistics course. They decided to tell the professor that they missed the exam because they got two flat tires. The professor accepted their excuse and agreed to a make up

exam. Relieved, the boys came the next day for the exam. They were shocked to find that the exam had only one question. That question was, which two tires were flat. What is the probability that they both give the same answer?

SECTION 10

45. (a) A balanced coin is tossed 15 times. (a) What is the probability of getting all heads? (b) Suppose 250,000 people each toss a balanced coin 15 times. What is the probability that at least one of these people get 15 heads? The point of this problem is that something having very small probability must happen to someone in a large enough population.

46. What is the smallest number of people needed to toss a balanced coin 15 times so that the probability that at least one of them gets 15 heads is .8?

47. In our discussion of we saw that probability of becoming HIV infected in a single heterosexual contact was .000001. AIDS alarmists often make much of cases such as that of Janice who had a single sexual encounter and subsequently became HIV infected. In a large metropolitan area there must be many such encounters. Find the probability that in 3,000,000 such encounters someone becomes HIV infected.

48. The story of HIV infection is entirely different in certain high risk groups such as male homosexuals in the so-called “bath house culture” than in the non IV drug heterosexual population. In this population the conservative estimate of the incidence rate of HIV is 20% and owing to the mode of sexual contact used the transmission rate is 1%. Also the number of different partners per year can be very high, as many as 1000 or more. (a) Find the probability that a male homosexual in this population becomes HIV infected in a single contact. (b) Find the probability that such an individual becomes infected if he has 1000 such contacts each with a different partner (c) How many contacts with different partners are needed so that the probability is .5 that he becomes infected?

49. D'Alembert was one of the greatest French 18^{th} Century scientists. He claimed that the chance of getting at least on head in two tosses of a balanced coin was 2/3 because each of the 3 outcomes HH, HT, TT have the same

probability, namely 2/3. What is wrong with his argument and what is the correct answer?

SECTION 11

50. A box 3 coins. Coin 1 has probability 1/3, coin 2 probability1/2, and coin 3 probability 2/3 of falling heads. A coin is selected at random and tossed.(a) What is the posterior probability that coin i was selected for i = 1,2,3. (b) It is tossed again and falls heads. Now what is the posterior probability coin i was selected.

51. A box has 3 tickets. The probability Q that all three tickets have the same number is p, the probability that two of the three have the same number is q and the probability that all three have different numbers is 1 - p - q. Two tickets are selected with replacement. (a) Find the probability that they have the same number. (b) Find the probability that the three original tickets have the same number if the selected two tickets have the same number. (c) Find the probability that the three original tickets have different numbers if the selected two tickets have the same number.

52. A man has 4 children. He tells you they are not all of the same sex. What is the probability that he has 2 boys and 2 girls?

SECTION 12

53. Verify that the probability of at least one 12 in 24 rolls of 2 dice is 0.4914.

54. Consider rolling 2 dice. What is the smallest number of rolls n so that the probability of at least one 12 in n rolls is > .5?

55. A famous puzzle is the “Prisoner’s Dilemma. There are 3 prisoners Manny, Moe, and Joe. One will be chosen at random and set free and the other two will be executed. Moe asks the guard to tell him the name of one of the other two who will be executed claiming that this information will be of no additional value since he already knows one of them at least will be executed. The jailer agrees and says Many will be executed. “Ah” says Moe ”it is now only between Joe and me, so my chance of going free is 1/2 rather than 1/3”. The problem is to decide if Moe is correct, i.e. to determine the probability that Moe goes free given the information about Manny. Assume that the Jailer does not lie and

that he never tells Moe that he will die if he is one of the two to be executed. Let X = 1 if Many goes free, X = 2 if Moe goes free, and X = 3 if Joe goes free. Let Y = 1 if the jailer says Manny is to be executed and let Y=0 if he says that Joe is to be executed. Let p = P(Y=1|X=2). (a) Is p specified? (b) The probability that Moe goes free given the Jailer says Many will die is P(X = 2|Y=1). Show this probability is p/(1+p). (c) What strategy does the Jailer need to follow for this probability to be 1/2? (d) What strategy does the Jailer need to follow for this probability to be 1/3?(d) If the Jailer follows the strategy of never saying Manny will die unless he must then what is the probability p/(1+p)?

56. Sam is one of two children. What is the probability that Sam has a sister?

57. In the problem of 4 cards show the desired probability in variant (iii) is 1/3 and corresponds to p = q = 1/2 in variant (i).

58. A building has 12 floors. 10 people get on at the ground floor. What is the probability they all get off at different floors?

59. Suppose a bus makes 10 stops. Initially there are 5 people on the bus. What is the probability that they all get off at different stops? Hint. Proceed as in the birthday problem

SECTION 13

60. Let X have density $f(x) = \begin{cases} cx, & 0 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$

Find c, (b) Find P(.75 < X < 1.5), (c) Find P(X < 1| X < 1.5).

61. Let a> 0 and let X have density $f(x) = \begin{cases} cx^a, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$

(a) Find c, (b) Find P(.2 < X < .4).

62. Let X have the density $f(x) = \begin{cases} \frac{1}{a} e^{-\frac{x}{a}}, & x > 0 \\ 0, & x < 0 \end{cases}$.

For t and s positive numbers (a) Find P(X > t) (b) Find P(X > t+s | X > t).

SECTION 14

63. Find the mean of the distribution having the density given in Problem 60.

64. Find the mean of the distribution having the density given in Problem 61

65. If $EX = 2$, $EY = 3$, and $EZ = -1$ what is $E(3X + 2Y + Z)$?

66. Suppose X is a random variable such that $P(X = -1) = 1 - p$ and $P(X = 1) = p$, $0 < p < 1$. If $EX = 0$ what is p?

67. Suppose X has the density given in problem 47. Find EX^m for $m > 0$.

68. Suppose X has probability function given in the following table:

x	0	1	2
P(X=x)	.4	.1	.6

Find EX and EX^2

69. Jack takes a multiple-choice exam that has 20 questions. Each question has 5 alternatives of which exactly one is the correct answer. If jack knows the correct answer he marks it. Otherwise, he guesses from amongst the 5 choices. Let p be the proportion of questions for which Jack knows the answer. (a) Find the probability that on a specific question Jack gets the correct answer. (b) If each question counts 5 points, then what is Jack's expected score on the exam? (c) What must p be in order for Jack to expect to get an 80?

70. A coin having probability p for heads is tossed n times. At each play, if the coin falls heads you win $1 and if it falls tails you loose $1. What are your expected winnings? [Hint let $X_i = 1$ if the coin falls heads and let $X_i = -1$ if it falls tails. Your winnings after n tosses are $N = X_1 + \ldots + X_n$

71. A random variable is said to have a symmetric distribution if X and -X have the same distribution. A function g is called odd if $g(-x) = -g(x)$. For example sin(x) is an odd function. Show that if X ihas a symmetric distribution and g is an odd function then $Eg(X) = 0$. [Hint consider Eg(X) and Eg(-X)]

72. Suppose X takes values a and –a each with probability 1/2. Show $E\cos(X) = \cos(a)$.

73. Cancer is the result of several mutations in the genes that regulate cell growth. Mutations are very rare. Suppose that in a given day there is chance 10^{-11} that a regulatory gene in a particular cell undergoes a mutation. The human body has about 3×10^{12} cells. (a) What is the expected number of cells that

undergo such a mutation in a given day? (b) What is the probability that there is at least one such mutation in a given day?

74. Suppose n cards numbered 1 to n are shuffled. Say there is a match at i if the j^{th} card from the top is card j. Let X_j = 1 or 0 according as there is a match at j or not. (a) Show $P(X_j = 1) = 1/n$. (b) The number of matches is $N = X_1 + \ldots + X_n$. Show EN = 1.

75. Suppose there are r boxes and n balls. The balls are distributed completely at random into the boxes. Let $X_i = 1$ if box i is empty and = 0 if not. The number of empty boxes is $N = X_1 + \ldots + X_n$. (a) Find $P(X_i = 1)$. (b) Find EN

76. There are r boxes. The probability of putting a ball in box i is p_i. Let $X_i = 1$ if box i is empty and = 0 if not. Suppose n balls are distributed into these boxes. The number of empty boxes is $N = X_1 + \ldots + X_n$.(a) Show that the probability that box i is empty (i.e. $P(X_i = 1) = (1 - p_i)^n$ (b) Find EN. (c) What choice of probabilities p_i make the EN a maximum. (d) What choice of probabilities p_i make the EN a minimum.

SECTION 15

77. Find the variance of the distribution with density given in Problem 61

78. Suppose X and Y are independent, Var (X) = 2, and Var(Y) = 1. (a) Find the Var(X + Y). (b) Find the standard deviation of X + Y.

79. If X has variance 2 what is the variance of $-3X + 2$?

80. Suppose you are asked to predict a value c for a random variable X. We measure the error g(c) of the prediction by the so called “mean square error” given by $g(c) = E(X - c)^2$. Find the value of c that makes this error a minimum and find the value of the error at the minimum. [Hint- expand $E(X - c)^2$ using the rules of expectation to find a polynomial of degree 2 in the variable c and use calculus methods to find its minimum.]

81. If Var(X) = Var(Y) = 1 and Cov(X,Y) =.25 what is Var(X + Y)?

82. Let $X_1, X_2, \ldots$ be independent random variables such that $P(X_i = 1) = p$ and $P(X_i = -1) = q = 1 - p$. (a) Find $Var(X_i)$ (b) Find $Var(X_1 + \ldots + X_n)$.

83. Suppose n cards numbered 1 to n are shuffled. Say there is a match at i if the j^{th} card from the top is card j. Let X_j = 1 or 0 according as there is a match at j

or not. The number of matches is $N = X_1 + \ldots + X_n$.(a) Show that for $j \neq k$, $P(X_j = 1, X_k = 1) = \frac{1}{n(n-1)}$. (b) use (a) to show that for $n > 0$, $Var(N) = 1$.

CHAPTER 2

SOME SPECIAL DISTRIBUTIONS

In this chapter we will describe and use some special distributions that find wide applications. To ease the burden of computing with these special distributions you may use a probability calculator that may be part of a calculator or program that you already own. For any computer, you may use the online probability calculators on the statistics web cite. For PC's you can download and use the free NCSS probability calculator offline.

1 BINOMIAL DISTRIBUTION

1.1 DEFINITION OF THE BINOMIAL DISTRIBUTION

Consider the observation of n success-failure experiments. We will assume that:

(a) Each success-failure experiment has the same success probability p.

(b) The n experiments are independent of each other.

The most basic example of the above situation is tossing the same coin n times, or equivalently, tossing n identical coins once. We will be interested in the distribution of the random variable N_n of the number of successes in these experiments. Let $X_i = 1$ if the i^{th} experiment is a success and let $X_i = 0$ if not. Assumptions (a) and (b) are equivalent to the assumption that the random variables $X_1, \dots, X_n$ are independent indicators each with the same success probability p.

We have already found the mean and variance of N_n [See Chapter 1, Example 1.55]. Now we will derive the probability function of $N_n = X_1 + \dots + X_n$.

The joint probability function of $X_1,\dots,X_n$ was given in Chapter1. We repeat that formula here. Let x_i =1 0r 0 and k be the number of $x_i = 1$. Observe that $k = x_1 + \dots + x_n$. Then

$$P(X_1 = x_1, X_2 = x_2, \dots ,X_n = x_n) = p^k(1-p)^{n-k} \qquad (2.1)$$

Equation (2.1) shows that every n tuple having the same number of 1's has the same probability. Now how can $N_n = 0$? That can happen if and only if all of variables $X_1, \dots ,X_n$ are 0. So

$P(N_n = 0) = P(X_1 = 0, \dots ,X_n = 0) = (1 - p)^n$.

How can $N_n = 1$? That can happen if and only if exactly one of the random variables is 1 (and therefore the rest are 0). So

$P(N_n = 1) = P(\text{exactly 1 of the } X_1,\dots ,X_n \text{ is 1})$. There are n different choices for which of these variables will be 1. Since every n-tuple with exactly one 1 carries the probability, namely $p(1 - p)^{n-1}$, we see that

$P(N_n = 1) = P(\text{exactly 1 of the } X_1,\dots ,X_n \text{ is 1}) = np(1 - p)^{n-1}$

In the same manner,

$P(N_n = k) = P(\text{exactly k of the } X_1,\dots ,X_n \text{ is 1}) = C(n,k)p^k(1-p)^{n-k}$

where C(n,k) is the number of choices of which of the n variables will be 1.To finish the computation of the probability function of N_n we need to determine C(n,k). This is a counting problem that can be solved by the rules in Section 5.1.of Chapter 1 Think of n the $X_1,\dots ,X_n$ as n tickets. We are to select k of them without replacement and ignoring order(these are the ones that will get the 1's) C(n,k) is the number of ways this can be done. However we know from Rule 5 of Section 3.3.2 of Chapter 1 that this is given by the binomial coefficient $\binom{n}{k}$. Thus

$$P(N_n = k) = \binom{n}{k}p^k(1-p)^{n-k},\ k = 0, 1, \dots,n \qquad (2.2)$$

The distribution with probability function given by (2.2) is known as the binomial distribution with parameters n and p. This is often abbreviated as b(n,p). Example 1.50 of Chapter 1 shows

$$EN_n = np \qquad (2.3)$$

$$Var(N_n) = np(1-p). \qquad (2.4)$$

EXAMPLE 2.1

Assume the probability of a child being male and female is the same, namely 1/2. A family has 4 children Find (a) the probability that there are 2 boys and 2 girls. (b) Find the probability that there are 3 boys. (c). Is it more likely that they have 3 of one sex or two boys and two girls?

SOLUTION

Let N be the number of male children. Then N is b(4,.5) distributed.

P(N = 2) = $\binom{4}{2}(.5)^2(.5)^2 = 6(.5)^4 = 0.3750$.

P(N = 3) = $\binom{4}{3}(.5)^3(.5) = 4(.5)^4 = 0.2500$.

Since p = .5 the probability of 3 girls is the same as the probability of 3 boys so the probability of 3 of one sex is twice the probability in (b). This is .5. Thus, contrary to popular belief, it is more likely that they have 3 of one sex than 2 boys and 2 girls.

EXAMPLE 2.2

The standard chemistry blood panel measures the blood serum level of 18 important substances such as glucose, total protein, etc. The normal range for each substance is determined so that 5% of healthy people will not fall in the normal range. (a) Find the probability that a healthy person will have all tests in the normal range. (b) Find) the probability that a healthy person will have 2 or more tests outside the normal range.

SOLUTION

Assuming the tests are independent the number N that fall outside the normal range is b(18, .05) distributed.

(a) $P(N = 0) = (.95)^{18} = 0.3972$

(b) $P(N \geq 2) = 1 - [P(N = 0) + P(N = 1)] = 1 - [.3972 + 18(.05)(.95)^{17}]$

$= 1-[.3972 + .3763] = 0.265.$

*1.2 TOURNAMENTS

The following is a classic problem known as “Probleme de Partie”.Teams A and B play a series of games. For each game team A wins with probability p and team B wins with probability q = 1 – p. The successive plays are independent. Team A wins the tournament if it wins a games before Team B wins b games. The World Series is an example of such a tournament with a = b = 4. The tournament cannot end before the smaller of a and b games have been played and it cannot continue for more than a + b – 1 games.

Find the probability that Team A wins the tournament and does so on game n.

Find the probability the Team A wins the tournament.

Find the probability that Team B wins the tournament on game n.

Find the probability that Team B wins the tournament.

Find the probability that the tournament requires n games.

SOLUTION

Let T be the number of games played in the tournament and let X be 1 or 0 according as Team A wins or looses the tournament.

(a) Two things must occur for Team A to win the tournament on the n^{th} game. (i) It must win game n. (ii) There must be a – 1 wins and fewer than b losses in the previous n – 1 games. Consequently, n can only have the values a, a + 1, , a + b –1. For each such n

$$P(X = 1, T = n) = p\binom{n-1}{a-1}p^{a-1}q^{n-a} = \binom{n-1}{a-1}p^a q^{n-a}, \; n = a, ..., a + b - 1 \qquad (2.5)$$

(b) $$P(X = 1) = \sum_{n=a}^{a+b-1}\binom{n-1}{a-1}p^a q^{n-a}.$$

(2.6)

(c) If we reverse the roles of A and B we find

$$P(X = 0, T = n) = \binom{n-1}{b-1}q^b p^{n-b} \quad n = b, \ldots, a + b - 1. \qquad (2.7)$$

(d) $$P(X = 0) = \sum_{n=b}^{a+b-1}\binom{n-1}{b-1}q^b p^{n-b}$$

(2.8)

Summing (2.5) and (2.7) we find

(e) $$P(T = n) = \binom{n-1}{a-1}p^a q^{n-a} + \binom{n-1}{b-1}q^b q^{n-b}, \; n = \min(a, b), \quad , a + b - 1. \qquad (2.9)$$

EXAMPLE 2.3

Find the probabilities in parts (a)-(e) for two equally matched teams in the World Series.

SOLUTION

Since the teams are equally matched p = q = 1/2.

For p = .5 the joint probability function and the probability functions of N and Y are given in the table 2-1 below.

Table 2-1

N	Y: 1	Y: 0	P(N =)
4	.0625	.0625	.125
5	.1250	.1250	.25
6	.15625	.15625	.3125
7	.15625	.15625	.3125
P(Y=)	.5	.5	***

The coefficients $\binom{n}{k}$ are called the binomial coefficients. They first arose in connection with the binomial theorem that states that for any 2 numbers a and b

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}. \tag{2.10}$$

Observe that if we take a = p and b = 1 - p in (2.10) we obtain

$$1 = \sum_{k=0}^{n} p^k (1-p)^{n-k}. \tag{2.11}$$

This of course is just a verification that the binomial probability function sums to 1. Here is a more interesting use of (2.10).

EXAMPLE 2.4

Find the probability that in tossing a coin having probability p for heads that you get an even number of heads.

SOLUTION

Let N be the number of heads obtained. We want P(N is even).

But this is

$$\sum_{k\,even} \binom{n}{k} p^k (1-p)^{n-k}$$

To find this sum take a = -p and b = (1 - p) in (2.10). The left hand side becomes $(1 - 2p)^{n.}$ The right hand side becomes

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} p^k (1-p)^{n-k}. \qquad (2.12).$$

Now add (2.11) and (2.12). The left hand side of the sum becomes $1 + (1 - 2p)^n$. In adding the right hand sides the odd terms subtract out so you get that the right hand sides summed are

$$2\sum_{k\,even} \binom{n}{k} p^k (1-p)^{n-k}.\text{Thus}$$

$$P(\text{N is even}) = \frac{1+(1-2p)^n}{2} \qquad (2.13)$$

1.2 ADDITIVE PROPERTY OF THE BINOMIAL DISTRIBUTION

Suppose N is b(n,p) distributed and N* is b(m, p) distributed. Then N + N* is b(n + m, p) distributed. This property is known as the additive property of the binomial distribution .

To see why the additive property is true go back to the basic fact that the b(n,p) distribution arises as the distribution of the sum of n independent indicator random variables each having the same success probability p. Let $X_1,\ldots,X_n$, $Y_1, \ldots, Y_m$ be n + m independent indicators all with success probability p. Let $N = X_1 + \ldots + X_n$ and let $N^* = Y_1 + \ldots + Y_m$. Then N is b(n,p) distributed, N* is b(m,p) and N and N* are independent. Thus $N + N^* = X_1 + \ldots + X_n + Y_1 + \ldots + Y_m$. is b(n + m, p) distributed.

EXAMPLE 2.5

John tosses a balanced quarter 3 times and Joe tosses a balanced dime 4 times. What is the probability that between them they get exactly 3 heads.

SOLUTION

Let N be the number of heads that John gets and let N* be the number of heads that Joe gets. We want P(N + N* = 3). Since N is b(3, .5) distributed and N* is b(4, .5) distributed, N + N* is b(7, .5) distributed. Therefore

$$P(N + N^* = 3) = \binom{7}{3}(.5)^7 = .2734.$$

EXAMPLE 2.6

Suppose a balanced coin is tossed 5 times. What is the probability that there are 3 heads in the 5 tosses if in the last two tosses there is 1 head?

SOLUTION

Let N be the number of heads in the first 3 tosses and N* the number of heads in the last two tosses. Then N and N* are independent, N is b(3, .5) distributed and N* is b(2, .5) distributed. We want $P(N + N^* = 3|N^* =1)$.

Now,

$$P(N + N^* = 3|N^* =1) = \frac{P(N + N^* = 3,\ N^* = 1)}{P(N^* = 1)}.$$

To say N + N* = 3 and N* = 1 is the same as saying N = 2 and N* = 1. So we can write the numerator above as P(N = 2, N* = 1). But N and N* are independent, so P(N = 2 and N* = 1) = P(N = 2)P(N* = 1). Therefore,

$$P(N + N^* = 3|N^* =1) = P(N =2) =(\ 3)(.5)^3 = 0.3750.$$

2 SAMPLING

A sample of size n from a population of m elements is a subset of n elements from the population. There are two basic methods by which a sample can be drawn. A sample is drawn with replacement if an element is selected and then replaced. A sample is drawn without replacement if an element is drawn and not replaced. The first obvious difference between the two sampling schemes is that in sampling with replacement the same element may be chosen again and again while in sampling without replacement that is impossible. But there are other important differences as well.

The main difference is independence. In sampling with replacement the successive elements that enter the sample are independent of each other. In sampling without replacement they are not. Because of the independence, sampling with replacement is mathematically simpler than sampling without replacement.

2.1 HYPERGEOMETRIC DISTRIBUTION

Suppose the population consists m objects of which r are of a specific kind (say red) and m – r are of a different kind (say black). We select a sample of size n from this population. Let N_n be the number of red objects in the sample. Let X_i = 1 or 0 according as the i^{th} element in the sample is red or not. Then N = X_1 + …, X_n.

If the sample is drawn with replacement then the successive draws are independent of each other, i.e the random variables X_1, …,X_n are independent. In this case N_n has the b(n, r/r + b) distribution. If the sampling is done without replacement then the random variables X_1, …,X_n are not independent In this case N does not have a binomial distribution. Instead, it has a distribution known as the <u>hypergeometric distribution</u>.

There are $\binom{m}{n}$ ways of selecting n tickets from a box of m tickets without replacement and ignoring order. If exactly k of these n are to be red they can be chosen from the r red tickets in $\binom{r}{k}$ ways. The remaining n – k tickets amongst the n selected are black. These can be selected from the m – k black tickets in $\binom{m-r}{n-k}$ ways. Thus the probability of getting exactly k red tickets in the sample of n is

$$P(N_n = k) = \frac{\binom{r}{k}\binom{m-r}{n-k}}{\binom{m}{n}} \qquad (2.14)$$

2.2 MEAN AND VARIANCE OF THE HYPERGEOMETRIC DISTRIBUTION

Even though (2.14) provides an explicit formula for the probability function of the hypergeometric distribution it is much simpler to compute the mean and variance by using the fact that N_n = X_1 + …+ X_n and the rules of expectation and variance.

Now

$EN_n = E(X_1 + \dots, X_n) = EX_1 + \dots + EX_n = P(X_1 = 1) + \dots + P(X_n = 1)$ (2.15)

To find $P(X_i = 1)$ note that there are $(m)_n$ samples of size n that can be drawn without replacement. The i^{th} element must be red. There are r choices for which red element it can be. There are no restrictions on the other n – 1 elements in the sample so there are $(m-1)_{n-1}$ choices for these n – 1 sample elements. Thus

$$P(X_i = 1) = \frac{r(m-1)_{n-1}}{(m)_n} = \frac{r}{m}.$$

Using (2.15) we see that

$$EN_n = n\frac{r}{m} \qquad (2.16)$$

To find $Var(N_n)$ we need to use the general formula **(1.78)** for the variance of a sum. We will defer the computations to the appendix. We end up with

$$Var(N) = n\frac{r}{m}(1-\frac{r}{m})(1-\frac{n-1}{m-1}). \qquad (2.17)$$

It is interesting to compare the mean and variance with that for a sample drawn with replacement. If the sampling was done with replacement, then N_n would have the binomial distribution with success probability r/m. Consequently it would have mean $n\frac{r}{m}$, and variance $n\frac{r}{m}(1-\frac{r}{m})$. Thus whether or not the sampling is done with replacement the expected number of red elements in the sample is the same. However, equation (2.17)shows that the variance in sampling without replacement is smaller by the factor $(1-\frac{n-1}{m-1})$. This quantity is called the finite population correction. Observe that if n is very small compared to m then this correction factor is essentially 1. In this case there is little difference between the two variances. Notice however that as the sample size starts to approach the population size the variance in sampling without replacement gets very small. In fact, when the sample size is the population size, i. e. n = m, then the variance is 0. That the variance should be 0 in this case is obvious. If we select everyone in the population, then the number of red elements N_n is not random It must be that $N_n = r$, so N_n must have variance 0.

2.3 BINOMIAL APPROXIMATION TO THE HYPERGEOMETRIC

Sampling with replacement can often be applied to give excellent approximations for sampling without replacement. What this means is that even though the sampling is done without replacement we can, under the appropriate conditions, just pretend that the sampling was done is done with replacement and use the results for sampling with replacement.

In most applications the sample size is very small compared to the population size. Typically one has samples in the thousands from populations in the millions or samples in the tens of thousands from populations in the hundreds of millions. In such situations one can accurately approximate the hypergeometric distribution with the binomial b(n,r/m) distribution. What we do is simply calculate the desired probabilities using the b(n,r/m) distribution instead of the hypergeometric distribution. However when the sample size is an appreciable fraction of the population size this approximation is not very good and should not be used. As a rule of thumb we will consider it O.K. to use the binomial approximation when sample size is less than a tenth of the population size.

EXAMPLE 2.7

In 1995, 738,781 of 2,312,203 deaths in the USA were due to coronary heart disease. Suppose a sample of 1000 names is selected at random from the death certificates filled in 1995. Let N be the number that died of coronary heart disease. Strictly speaking, N does not have a binomial distribution. But here the population size is > 10n = 10,000 so we ignore this fact and just go a head and use b(1000, 738,781/2,312,203) as the distribution of N.

EXAMPLE 2.8

In the following see if it is permissible to use a binomial distribution for N. If so give the parameters of the distribution. If it is not permissible to use a binomial distribution give a reason why it is not permissible.

(a) Suppose 72% of the adults in Los Angeles favor stronger gun control legislation. N is the number of people in a sample of sample of 500 adults that favor such legislation.

(b) A small company has 400 employees of which 320 are women. A sample of 100 employees is drawn and N is the number of females in the sample.

SOLUTION

Binomial is O.K. (The population of LA is certainly > (10)(500)). Here N is b(500, .72) distributed.

Binomial cannot be used. The population is too small compared to the sample size.

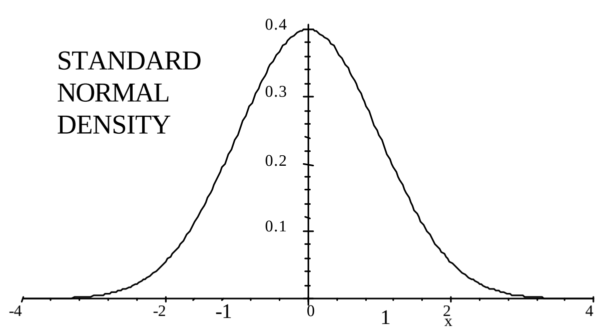
3 NORMAL DISTRIBUTIONS

We will discuss a two-parameter family of distributions that are by far the most important distributions in probability and statistics. They are variously called <u>Gaussian distributions</u> (after the German mathematician Carl Fredrich Gauss who did not invent them) or ,more commonly, <u>normal distributions</u> (for no really good reason).

3.1 STANDARD NORMAL DISTRIBUTION

Let $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Figure 2-1



It can be shown that $\int_{-\infty}^{\infty} \phi(x)dx = 1$, so $\phi(x)$ is a density. This density is called <u>the standard normal density</u> and the distribution with this density is called the <u>standard normal distribution</u>

Analysis of the function $\phi(x)$ shows:

(a) that it is symmetric about 0,

(b) has points of inflection at $\pm$ 1,

(c) has a unique maximum at 0.

Let Z have the standard normal distribution. It is easily calculated that EZ = 0. A more involved calculation that involves integrating by parts shows Var(Z) =1.

Suppose Z has a standard normal distribution. For any real number x let

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt. \qquad (2.18)$$

Note that $\Phi(x)$ is the area under the standard normal density curve from $-\infty$ to x.

Then for any $-\infty \le a < b \le \infty$

$$P(a< Z < b) = \int_{a}^{b} \phi(x)dx = \int_{-\infty}^{b} \phi(x)dx - \int_{-\infty}^{a} \phi(x)dx,$$

with the interpretation that $\Phi(-\infty) = 0$ and $\Phi(\infty) = 1$.

In other words, for any $-\infty \le a < b \le \infty$,

$$P(a< Z < b) = \Phi(b) - \Phi(a). \qquad (2.19)$$

The function $\Phi(x)$ is called the (cumulative) distribution function of the standard normal distribution. A table of $\Phi(x)$ is provided in these notes. Using this table one can compute P(a< Z < b) via (2.19). Alternately, one can use the normal probability calculator to compute these probabilities.

EXAMPLE 2.9

Let Z have the standard normal distribution. Find (a) P(Z < -1) (b) P(Z > 1.23) (c) P(|Z| > 1.23) (d) P(.2 < Z <1.6).

SOLUTION

(a) Using the table we find P(Z < -1) = 0.1587. (b) From the table we find P(Z < 1.23) = 0.8907. Therefore P(Z > 1.23) = 1- P(Z < 1.23) = 1 - .8907 = 0.1093. (c) P(|Z| > 1.23) = P(Z < -1.23) + P(Z > 1.23). From the table and (b) we find P(Z < -1.23) + P(Z > 1.23)= .1093 + .1093 = 0.2186.(d) From the table P(Z < 1.6) = 0.9452 and P(Z < .2) =0 .5793. Therefore P(.2 < Z <1.6) = 0.9452 - 0.5793 = 0.3659.

EXAMPLE 2.10

Show that if Z has the standard normal distribution then for any a > 0,

(i) $P(Z < -a) = P(Z > a)$.

(ii) $P(|Z| > a) = 2P(Z > a) = 2[1 - \Phi(a)]$.

(iii) $P(|Z| < a) = 2\Phi(a) - 1$.

SOLUTION

(i) $P(Z < -a) = \int_{-\infty}^{-a} \phi(x)dx$.

Make the change of variable x = -y in this integral. This gives

$$\int_{-\infty}^{-a} \phi(x)dx = -\int_{\infty}^{a} \phi(y)dy = \int_{a}^{\infty} \phi(y)dy = P(Z > a) \qquad (2.20)$$

(ii)) $P(|Z| > a) = P(Z > a) + P(Z < -a)$. By (i) $P(Z <- a) = P(Z > a)$. So $P(|Z| > a) = 2P(Z > a) = 2[1 - \Phi(a)]$.

(iii) $P(|Z| < a) = 1 - P(|Z| > a) = 1 - 2[1 - \Phi(a)] = 2\Phi(a) - 1$.

3.2 NORMAL DISTRIBUTIONS

We will now define a 2 parameter family of densities by scaling Z. Let $a \neq 0$ and b be numbers and let X =aZ + b Then it turns out that X has density

$$f(x) = \frac{1}{\sqrt{2\pi}|a|} e^{-\frac{(x-b)^2}{2a^2}} \qquad (2.21)$$

Now using the rules of expectation and variance

EX = E(aZ + b) = aEZ + b = b,

and

$\mathrm{Var}(X) = a^2 \mathrm{Var}(Z) = a^2$.

Thus the parameter $b = \mu$ the mean of the distribution with density given by (2.9) and the parameter $|a| = \sigma$ the standard deviation of this distribution. Owing to this fact we parameterize the distribution by these two quantities and write the density in(2.21) as

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \qquad (2.22)$$

The normal distribution with mean μ and standard deviation σ is the distribution having density given by (2.22). Often we denote this distribution by $N(\mu, \sigma)$.

The basic properties of a normal distribution are given below.

1. A normal distribution is specified by its two parameters that are the mean μ and the standard deviation σ.

2. If X is N(μ, σ).distributed then X has the distribution of $\sigma Z + \mu$.

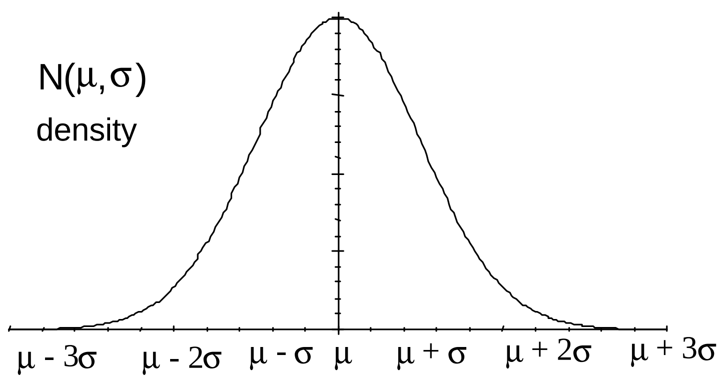
3. If X is N(μ, σ).distributed then $(X - \mu)/\sigma$ has the standard normal distribution.

4. (a) P(|X - | μ < σ) = 0.6827

   (b) P(|X - |μ < 2σ) = 0.9545

   (c) P(|X - | μ < 3σ) =0.9973

Figure 2-2



The graph is given N(μ, σ).in Fig.2-2. How does the density curve change with the parameters? As indicated in Fig.2-3, for fixed σ, changes in μ just translate the center of the distribution but do not change its shape. For fixed μ changes σ change the shape of the density curves. As illustrated in Fig. 2-4, the smaller σ is the more peaked the curve becomes about. μ

Figures 2-3 show the normal curves for mean 1 and mean 2 with standard deviations 1. Figure 2-4 shows normal curves with mean 0 and standard deviations 1, 0.5, and 0.1. These curves and Property 4 in the box show that for a normal distribution the mean and standard deviation are excellent measures of location and spread.

Figure 2-3

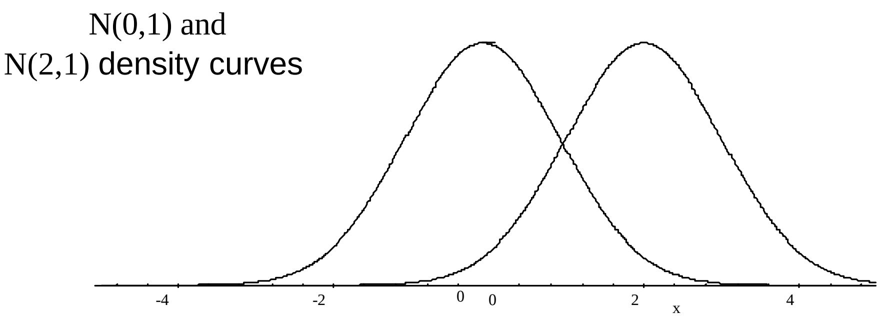
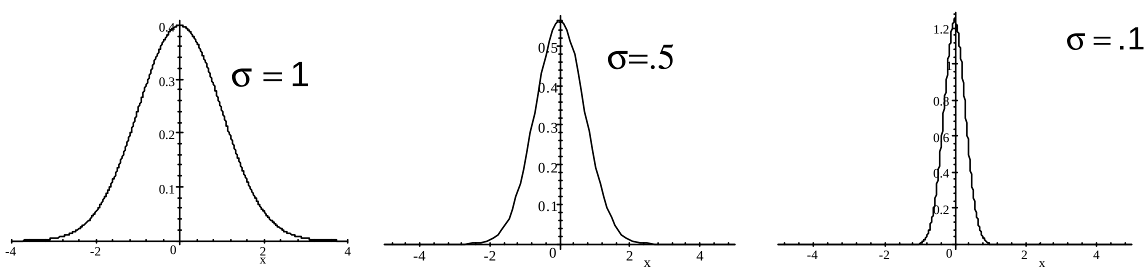


Figure 2-4



3.3 CALCULATIONS WITH NORMAL DISTRIBUTIONS

Suppose X is $N(\mu,\sigma)$ distributed. Then Z = $\dfrac{X-\mu}{\sigma}$ has the standard normal distribution. We use this fact to perform calculations involving X. For $-\infty \le a < b \le \infty$,

$P((a - \mu)/\sigma < (X- \mu)/\sigma < (b - \mu)/\sigma)$

$= P((a - \mu)/\sigma < Z < (b - \mu)/\sigma) = \Phi((b-\mu)/\sigma) - \Phi((a-\mu)/\sigma)$.

Thus

$$P(a < X < b) = \Phi((b-\mu)/\sigma) - \Phi((a-\mu)/\sigma) \qquad (2.23)$$

EXAMPLE 2.11

The mean diastolic blood pressure is 80 with a standard deviation of 7. Assume blood pressures follow a normal distribution. Find the probability that a person has a diastolic blood pressure greater than 95.

SOLUTION

Let X be the diastolic blood pressure. We want P(X > 95). Using (2.23),

$$P(X > 95) = 1 - \Phi\left(\frac{95-80}{7}\right) = 1 - \Phi(2.14) = 0.016$$

EXAMPLE 2.12

(a) In Example 2.11 find the probability that a diastolic blood pressure is between 80 and 90. (b) Find the probability that in 5 such measurements exactly 2 are between 80 and 90.

SOLUTION

(a) Let X be the diastolic blood pressure. We want P(80 < X < 90).By (2.23),

$$P(80 < X < 90).= \Phi\left(\frac{90-80}{7}\right) - \Phi\left(\frac{80-80}{7}\right) = \Phi(1.4286) - \Phi(0) = 0.9234 - 0.5$$

= 0.4234.

(b) Let N be the number of pressures between 80 and 90. Then N is b(5,0.4236) distributed. We want P(N = 2). This is

$$P(N = 2) = \binom{5}{2}(.4236)^2(.5764)^3\ 0.3436.$$

3.4 INVERSE NORMAL CALCULATIONS

.The distribution function

$$\Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

is a strictly increasing and continuous function of x with values between 0 and 1. Therefore, for any 0<p<1 there is a unique $x = x_p$ such that

$$\Phi\left(\frac{x_p-\mu}{\sigma}\right) = p. \qquad (2.24)$$

The number x_p is called the p^{th} <u>quantile</u> of the distribution and 100 x_p is the p^{th} <u>percentile</u> of the distribution. The simplest way to find x_p is to use a probability calculator. In the absence of such a calculator you can use the standard normal table backwards. To do so first find the value z'_p such that $\Phi(z'_p) = p$ by scanning the table entries for p and taking the corresponding z value. In general, you will not find p. Instead, you will find two values, $p_1 < p < p_2$, that straddle p. Corresponding to these will be the values z'_{p_1} and z'_{p_2}. We then take $z_p = (z'_{p_1} + z'_{p_2})/2$ as the value for p. For example, for p = 0.99 we find the straddling table entries are z_{p_1} = 0.9898, which corresponds to z_1 = 2.32, and p_2 = 0.9901, which corresponds to z_{p_2} = 2.33. So we take $z_{.99}$ = 2.325. Having found z'_p, we can now find x_p by using the fact that $z'_p = (x_p - \mu)/\sigma$, so

$$x_p = \sigma z'_p + \mu. \qquad (2.25)$$

EXAMPLE 2.13

Suppose systolic blood pressures follow a normal distribution with mean 120 and standard deviation 12. At what point x will 99% of the population have a systolic blood pressure $\le$ x?

SOLUTION

The value x is the value $x_{.99}$. That is, it satisfies equation (2.24) for p = 0.99 with μ = 120 and σ = 12. In the discussion above, we found $z'_{.99}$ = 2.325. Therefore BY (2.25)

$$x_{.99} = (2.325)(12) + 120 = 147.9.$$

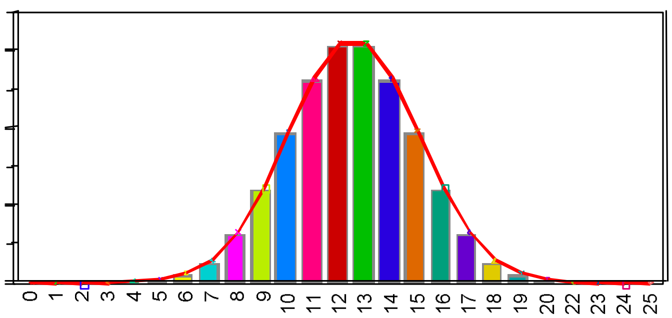
4 NORMAL APPROXIMATION TO THE BINOMIAL

Suppose N is b(n,p) distributed. Unless n is very small (or unless one has a probability calculator that will compute binomial probabilities) it is tedious to compute probabilities such as $P(N \le k)$. We are going to show that there is a simple way to approximate these probabilities accurately by using a normal distribution. This approximation, known as the normal approximation to the binomial, is exceedingly useful.

You might ask, "Why should we bother with the normal approximation to the binomial at all. After all, these days we can exactly compute the binomial probabilities with the “push of a button”. The reason is twofold. The first is a practical one. You just might want to compute some binomial probability when all you have available is a normal table. Also, there are some things we can do with the normal approximation that are difficult to do with the exact distribution. The second is theoretical. The normal approximation to the binomial is just the tip of the iceberg. It is the first example in a long series of approximations of various distributions (most of which have no probability calculator) by the normal distribution. These approximations go under the subject of the Central Limit Theorem that we will look at a little later.

A histogram is made from the probability function of the b(n,p) distribution by centering a rectangle of width 1 about each j = 0, 1, …, n and taking its height to be $\binom{n}{j}p^j(1-p)^{n-j}$. Below we give such a histogram for p =0.5 and n =25.

Figure 2-5



Consider the histogram. A smooth curve drawn through the tops of the rectangles rather resembles that of a normal density curve having mean 25(.5) =12.5 and standard deviation $\sqrt{25(.5)(.5)}$ = 2.5. This is no accident. It is in fact the case that a normal distribution can be used to accurately approximate a binomial distribution unless n is quite small. Actually, it's the relation between n and p that determines the accuracy of the approximation. In order to use the normal approximation what is needed is for np ≥ 5 and n(1 - p) ≥ 5. This approximation is extremely useful.

<u>The normal distribution is used to approximate a b(n, p) distributed random variable N by just pretending that it is normally distributed with mean np and</u>

standard deviation $\sqrt{np(1-p)}$. For x =0, 1, … ,n we approximate P(N≤x) by

$\Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$

For example, suppose n=25 an p = 0.5 and we wish to compute P(N ≤ 15). We would approximate this probability by pretending that N was normally distributed with mean 25(.5) = 12.5 and standard deviation $\sqrt{25(.5)(.5)}$ = 2.5. Thus P(N≤ 15) ≈ $\Phi\left(\frac{15-12.5}{2.5}\right)$= 0.8413.

4.1 HALF INTEGER CORRECTION

The accuracy of the normal approximation depends on how large n is. The larger n the more the accurate it is. The rule of thumb is that to apply the normal approximation one should have np ≥ 5 and n (1 - p) ≥ 5. However, it can be substantially improved for moderate values of n by using what is known as the half integer correction. This is how the correction is done. If N is b(n, p) distributed we approximate P(N ≤ x) for x an integer, 0, …, n by pretending N is normal with mean np and standard deviation $\sqrt{np(1-p)}$ and replacing x by x + .5. Thus

$$P(N \le x) = \Phi\left(\frac{x-np+.5}{\sqrt{np(1-p)}}\right) \qquad (2.26)$$

We can now approximate P(N < x), P(N > x), and P(N ≥ x) by using (2.26) and the fact that N is integer valued. For example, to compute P(N < x) we first note that P(N < x) is the same as P(N ≤ x-1) and use (2.26) with x replaced by x-1. Since P(N > x) = 1 – P(N ≤ x), and P(N ≥ x) = 1 - P(N < x) = 1 - P(N ≤ x-1) Equation (2.26) suffices to calculate these probabilities.

Suppose we want to approximate P(N = x) for x an integer between 0 and n. We cannot do this by saying P(N = x) $\approx P(Z = \frac{x-np}{\sqrt{np(1-p)}})$ since

$$P(Z = \frac{x-np}{\sqrt{np(1-p)}}) = 0$$

Here is how to approximate P(N = x). Observe that for the exact binomial distribution P(N = x) = P(x - .5 < N < x + .5). Now use the normal approximation to approximate the probability P(x - .5 < X < x + .5). This yields the formula

$$P(N = x) \approx P(x -.5 < X < x + .5).= \Phi(\frac{x-np+.5}{\sqrt{np(1-p)}}) - \Phi(\frac{x-np-.5}{\sqrt{np(1-p)}}) . \qquad (2.27)$$

COMMENT

The half integer corrections provide greater accuracy for moderate values of n but become more and more irrelevant as n grows large [The correction factor only changes the argument of Φ by a factor of $\frac{.5}{\sqrt{np(1-p)}} \to 0$ as n $\to \infty$.] .

EXAMPLE 2.14 Let N be b(25, .5) distributed. Use the normal approximation with half integer correction to find (a) P(N $\leq$ 15), (b) (P(N < 15), (c) P(N > 15), (d) P(N $\geq$ 15).

SOLUTION

(a) P(N $\leq$ 15) $\approx P(Z \leq \frac{15-12.5+.5}{\sqrt{25(.5)(.5)}}) = \Phi(1.2).$ = 0.8849

(b) P(N < 15) $\approx P(Z \leq \frac{15-12.5-.5}{\sqrt{25(.5)(.5)}}) = \Phi(.8).$ = 0.7881

(c) P(N > 15) = 1 - P(N $\leq$ 15) = 1 - .8849 = 01151

(d) P(N $\geq$ 15).= 1 - P(N < 15) = 1 - .7881 =0.2119

Lets see how well the normal approximation works. The table below gives the exact and the approximate values both with and without the half integer correction for binomials with p = .5 for various values of n for an x that is close to 1 standard deviation above the mean.

Table 2-2

n	x	P(N ≤ x) exact	Approximate with half integer correction	Approximate without half integer correction
10	7	.9543	.9432	.8972
25	15	.8852	.8850	.8413
50	29	.8987	.8982	.8707
100	55	.8644	.8643	.8413
500	260	.8262	.8262	.8145

EXAMPLE 2.15

Suppose that 5% of people reserving seats on an airline fail to show for the flight. If a plane has 250 seats and sells 260 tickets what is the probability that everyone who shows for the flight will get a seat?

SOLUTION

Let N be the number of people that show for the flight. Then N is b(260, .95) distributed. This has mean (260)(.95) = 247 and standard deviation $\sqrt{(260)(.95)(.05)}$ = 3.51 Everyone who shows for the flight gets a seat if N $\leq$ 250. So the probability we want is P(N $\leq$ 250). By the normal approximation,

$$P(N \leq 250) \approx \Phi(\frac{250-247+.5}{3.51}) = \Phi(.9972) = 0.8407 .$$

EXAMPLE 2.16

Suppose the proportion of people that survive 2 years after diagnosis with a certain type of leukemia is 60%. A new chemotherapy is discovered and tried on 15 subjects. Of these, 11 are alive after 2 years. (a) Compute the probability that 11 or more of 15 subjects given no treatment survive 2 years. (b) Suppose the treatment is tried on 150 subjects and 110 are alive after 2 years. Compute the probability that 110 or more of 150 subjects given no treatment survive 2 years

SOLUTION

(a) The number N that survive 2 years is b(15,.6) distributed. We want P(N $\geq$ 11) = 1 - P(N $\leq$ 10) = .2173.

(b) Here the number N that survive 2 years without treatment is b(150, .6). We compute P(N $\geq$ 110). Using the normal approximation and ignoring the half integer correction P(N $\geq$ 110) = 1 - $\Phi(\frac{109-90}{6})$ = 1 - $\Phi(3.17)$ = .0008.

Often its more convenient to work with the <u>sample proportions</u>, $\hat{p}_n$ = N_n/n, in n trials than with the number N_n of such successes. Now for 0<x<1,

$P(\hat{p}_n \leq x)$ = P(N_n/n $\leq$ x) = P(N_n $\leq$ nx).

Using the normal approximation to the binomial we find

$$P(\hat{p}_n \leq x) \approx \Phi\left(\frac{x-p}{\sqrt{\frac{p(1-p)}{n}}}\right) \quad (2.28)$$

In words, this equation tells us that $\hat{p}_n$ is approximately normally distributed with mean p and standard deviation $\sqrt{\frac{p(1-p)}{n}}$.

EXAMPLE 2.17

One of the main sources of cancer data is the SEER (standing for Surveillance, Epidemiology, and End Results) reports published by the National Cancer Institute. The 5-year survival rate for a particular cancer is the percent of people that survive 5 years after the diagnosis of the cancer. This is the most common overall index on the severity of the cancer and the effectiveness of therapy. [This index is sometimes a bit cloudy because often one does not know of those that did not survive if they died of the cancer or some other cause.] Overall, across all stages, (A stage is a classification scheme on how far the cancer has progressed when it is diagnosed. The stages used in SEER are localized, regional spread, distant spread, and unknown.) and all ages the 5 year survival rates (1986-1991 SEER rates) for breast cancer in women is 84.4% for white women while only 69% of black women survive 5 years. The combined 5 year survival rate is 83.2%. [Can you explain why the combined survival rate is almost the same as the white survival rate? (This is a mathematical question)].

Can you supply some scenarios that might explain the huge discrepancy between the black and white rate? (This is not a mathematical question).

Find the probability that of 342 black women just diagnosed with breast cancer 71% or more will be alive 5 years from now.

SOLUTION

Let $\hat{p}$ be the proportion of the 342 that survive 5 years. We want $P(\hat{p} \geq .71)$. Using the normal approximation

$$P(\hat{p} \geq .71) = 1 - \Phi\left(\frac{.71-.69}{\sqrt{\frac{(.69)(.31)}{342}}}\right) = 1 - \Phi(.7997) = 0.2119 .$$

5 DISCRETE DISTRIBUTIONS EXTENDED

Consider tossing a coin repeatedly. Let X be the first toss for which we get a head. What values can X take? No matter how large a positive integer n we select we cannot guarantee that the coin will fall heads at least once by toss n. Indeed, the probability that the first n tosses all yield tails is $(1 - p)^n$ which is > 0 for every n. The random variable X is an example of one that has every positive integer as a possible value. In order to deal with such random variables we need to extend the concept of a discrete random variable to allow it to have an infinite sequence of possible values. Such an extension is easily accomplished. In order to carry out this extension we need the concept of the sum of an infinite series.

5.1 INFINITE SERIES

Let $\{a_n\}$, n = 1, 2, … be a sequence of non-negative numbers. We define their “sum” a as

$$a = \lim_{n\to\infty} \sum_{i=1}^{n} a_i . \quad (2.29)$$

In technical terms, the expression on the right hand side of (2.28) is called an <u>infinite series</u> and the value of the limit a is called the <u>sum of the infinite series</u>. The numbers a_n are the <u>terms of the series</u> and a_n is the <u>n-th term</u>.

There are various ways we denote an infinite series. Often we simply write $a_1 + a_2 \ldots$ Other notation used is $\sum_n a_n$ or $\sum_{n=1}^{\infty} a_n$. Sometimes we label the terms so they start at a value other than 1. Suppose they start at k. That is, the terms are $a_k, a_{k+1}, \ldots$ Then we would write their sum as either $a_k + a_{k+1} + \ldots$ or as $\sum_{n=k}^{\infty} a_n$.

All of the rules for finite sums are also valid for infinite series of non-negative terms. For example:

for a constant b,

$$b\sum_n a_n = \sum_n ba_n .$$

If $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ are two non-negative sequences, then

$$\sum_n a_n + \sum_n b_n = \sum_n (a_n + b_n).$$

It is possible that the limiting value in (2.29) is ∞. For example, if $a_n = 1$ for all n, then $a_1 + a_2 \ldots + a_n = n$ and $\lim_{n\to\infty}\sum_{i=1}^{n} a_i = \lim_{n\to\infty} n = \infty$. In general, except for a very few special cases, special tricks are required to find the sum of an infinite series.

5.2 EXTENDED DISCRETE DISTRIBUTIONS

We extend the concept of a discrete distribution by allowing it to have an infinite sequence of possible values, $x_1, x_2, \ldots$ For example, it could have values 1,2, …

Let A be a subset of R. Then $\sum_{x_i \in A} P(X = x_i)$ is interpreted as an ordinary finite sum if A contains only finitely many terms of the infinite sequence $\{x_n\}$ or as an infinite series if it has infinitely many terms of the infinite sequence $\{x_n\}$. As in the finite case we will abbreviate $\sum_{x_i \in A} P(X = x_i)$ as $\sum_{x \in A} P(X = x)$.

We will now consider a discrete random variable as one that either has a finite number of possible values or an infinite sequence of possible values. We will only be concerned with discrete random variables that can have an infinite sequence of values when those values are non-negative integers. For such random variables we define P(X $\in$ A) as

$$P(X \in A) = \sum_{x \in A} P(X = x) \tag{2.30}$$

All the rules developed in the finite case also hold in the infinite sequence case. We will use these freely in the sequel. In particular the mean is

$$EX = \sum_{x} xP(X = x) \tag{2.31}$$

where we allow ∞ as a value. If $EX^2 = \sum_{x} x^2 P(X = x) < \infty$ we then the mean is also finite and we define the variance as just as in the case when X has only finitely many values, var(X) = $E[X - EX]^2$ and, as in the previous case

$E[X - EX]^2 = EX^2 - (EX)^2$.

5.3 GEOMETRIC DISTRIBUTIONS

5.3.1 GEOMETRIC SERIES

An expression of the form

$$\sum_{i=1}^{n} a^i = 1 + a + \ldots + a^n \tag{2.32}$$

is called a geometric progression. It is easy to prove by induction that

$$\sum_{i=1}^{n} a^i = \begin{cases} \dfrac{1-a^{n+1}}{1-a}, \text{if } a \neq 1 \\ n+1, \quad \text{if } a = 1 \end{cases} \tag{2.33}$$

If |a| < 1, $\lim_{n \to \infty} a^n = 0$. Consequently (2.33) shows that

$$\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}, |a| < 1 \tag{2.34}$$

The infinite series in (2.34) called the geometric series.

5.3.2 GEOMETRIC DISTRIBUTION

Consider performing independent successive repetitions of a success-failure experiment having probability p for success. Let X be the number of trials required to first get a success. Then X can have any positive integer as a possible value. To say X > n is the same as saying that the first n trials are all failures. Thus P(X

> n) = $(1 - p)^n > 0$. Thus, no matter how large an n we choose, there is always a non-zero chance that no success will occur in the first n trials. It fact it is conceivable that there is a non-zero chance that we will perform the trials forever and not get a success. As we will see momentarily that is not the case and we must ultimately get a success.

In order that the first success to occur at the n^{th} trial we must get a success at trial n and the first n – 1 trials must all be failures. Thus

$$P(X = n) = p(1 - p)^{n-1},\ n = 1, 2, \ldots \qquad (2.35)$$

The distribution with probability function given by the right hand side of (2.35) is called the geometric distribution with parameter p. Using (2.34) with a = 1- p shows that

$$P(X < \infty) = \sum_{n=1}^{\infty} P(X = n) = p\sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1. \qquad (2.36)$$

Equation (2.36) shows that with probability 1 we must ultimately get a success.

To find the mean and variance of the geometric distribution we need to find the sum of the two infinite series $\sum_n np(1-p)^{n-1}$ and $\sum_n n^2 p(1-p)^{n-1}$. This can be accomplished by more advanced mathematical techniques. Using the values for these sums we find the mean to be

$$EX = \frac{1}{p},\ Var(X) = \frac{1-p}{p^2}. \qquad (2.37)$$

EXAMPLE 2.18

Suppose two balanced dice are rolled until the sum of the face values is 12. (a)Find the probability that 12 rolls are needed. (b) Find the expected number of rolls that are required.

SOLUTION

Let X be the number of rolls required. Then X has the geometric distribution with parameter p = 1/36. (a) $P(X = 12) = (1/36)(35/36)^{11} = .0204$. (b) EX = 1/p = 36.

EXAMPLE 2.19

Jack and Jane each toss a balanced coin repeatedly. (a) Find the probability that they both first toss a head at the n-th flip. (b) Find the probability that they both first toss a head at the same flip.

SOLUTION

Let X be the flip at which Jack first gets a head and let Y be the flip at which Jane first gets a head. We can assume X and Y are independent. They each have the geometric distribution with parameter 1/2 (a) want P(X = n,Y = n). Since X and Y are independent, P(X = n, Y = n) = P(X = n)P(Y = n) = $(1/2)^n(1/2)^n = (1/4)^n$ (b) We want P(X = Y). Now,

$$P(X = Y) = \sum_{n=1}^{\infty} P(X = n, Y = n) = \sum_{n=1}^{\infty} (1/4)^n.$$

To evaluate the infinite series note that

$$\sum_{n=1}^{\infty} (1/4)^n = (1/4)\sum_{n=1}^{\infty} (1/4)^{n-1} = (1/4)\sum_{k=0}^{\infty} (1/4)^k$$

Using (2.34)with a = 1/4 we find

$$P(X = Y) = \frac{\frac{1}{4}}{1-\frac{1}{4}} = \frac{1}{3}.$$

EXAMPLE 2.20

Suppose we toss m balanced coins repeatedly. Say a success occurs if all m coins are heads and a failure if not. What is the expected number of times we need to toss these coins to get a success?

SOLUTION

Here $p = 2^{-m}$. By (2.37), the expected number of tosses is $1/2^{-m} = 2^m$.

*5.3.3 CRAPS

One of the most popular casino games is Craps. The game supposedly originated from the game of “Hazard” invented by 12th Century Crusaders. Formal rules were laid down and the game became very popular in early 19th Century England and France. In British clubs the casts of 2, 3, or 12 were referred to as “Crabs”. French mispronunciation led to the word “Craps”. The game immigrated to the French colony of New Orleans and was picked up by Afro-Americans. It spread up the Mississippi and outward across the USA. It developed an unsavory reputation in the late 19th Century because its manipulation by crooked gamblers traveling on steamboats and train sleeping cars. It regained respect because of its immense popularity with the soldiers in World War I. Since then it has been enshrined with social acceptability by its adoption in the Casinos of the world where it is no longer played in alleys but on fancy tables decorated with cabalistic symbols.

The game of Craps is played as follows. The shooter starts by rolling 2 dice. (this is called the come-out). If they come up 7 or 11 (called a natural) the player wins. If they come up (2,3,or 12) (called craps) the player looses. Otherwise, the number x that comes up is called the point. The point must be one of the numbers 4,5,6,8,9 or 10. The player then continues until he either throws his point again or throws a 7. He wins if he throws his point and looses if he throws the 7.

Let X be the value on the come-out. Let Y = 1 if the shooter wins and Y = 0 if not. We consider P(Y = 1|X = x) where x is one of the points. Given X =x let T be the number of additional rolls needed for the shooter to either win or loose. Then Y =1 and T = n if and only if the shooter gets his point on roll n after the come-out. For this to happen role n must be the point value and the other n – 1 rolls must be any value other than the point or 7. To calculate these probabilities let p_x be the probability that we get x on the role of 2 dice. These probabilities are given in Table 1-19. The probability of getting a 7 is 1/6. Therefore the probability that on the role of 2 dice we neither get x nor 7 is $1 - p_x - 1/6$. Hence

$$P(Y = 1, T = n \mid X = x) = p_x(1 - p_x - 1/6)^{n-1} \quad (2.38)$$

By (2.33) and the above

$$P(Y = 1|X = x) = p_x \sum_{n=1}^{\infty}\left(1 - p_x - \frac{1}{6}\right)^{n-1} = \frac{p_x}{p_x + \frac{1}{6}}.$$

(2.39)

The numerical values are given in the following table.

Table 2-3

x	4	5	6	8	9	10
P(Y=1|X=x	1/3	2/5	5/11	5/11	2/5	1/3

For x = 7 or 11, P(Y = 1|X = x) = 1. For X = 1,3,12 P(Y = 1| X = x) = 0. Using the law of total probability we now can find the probability that the shooter wins the game.

$$P(Y = 1) = \left(\frac{6}{36} + \frac{2}{36}\right) + \frac{3}{36}\frac{1}{3} + \frac{4}{36}\frac{2}{5} + \frac{5}{36}\frac{5}{11} + \frac{5}{36}\frac{5}{11} + \frac{4}{36}\frac{2}{5} + \frac{3}{36}\frac{1}{3} = 0.4930.$$

This is the closest one can get to a fair game in the casinos. Observers can place a bet on the shooter to win (called pass). Clearly the casinos will not let you bet directly that the shooter looses because this has probability greater than 0.5, but you can almost place such a bet called "no pass bar 12". In this bet if the first roll is a 12 it is considered a tie for the person betting on "no pass" so the probability of winning is 1 - P(pass) - $\frac{1}{2}$P(12) = 1 - 0.4930 - 0.0139 = 0.4931. The bets "come" and "no - come" are the same as pass and no pass respectively.

6 THE CENTRAL LIMIT THEOREM

6.1 MORE ON THE NORMAL DISTRIBUTION

Here are three additional properties of normally distributed random variables:

Suppose X is $N(\mu,\sigma)$ distributed. Then for any a $\neq$ 0 and b, aX + b is normally distributed with mean $a\mu$ + b and standard deviation $\sqrt{a^2\sigma^2} = |a|\sigma$.

If X is $N(\mu_1,\sigma_1)$ distributed and Y is $N(\mu_2,\sigma_2)$ distributed and X and Y are independent, then

X + Y is $N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ distributed,

and

X - Y is $N(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$ distributed

If X_i, i = 1, …, n is $N(\mu_i, \sigma_i)$ distributed and if the X_i are independent, then X_1 +…+ X_n is $N(\mu, \sigma)$ distributed with $\mu = \mu_1 + ... + \mu_n$ and $\sigma = \sqrt{\sigma_1^2 + ... + \sigma_n^2}$.

.EXAMPLE 2.21

Suppose systolic blood pressure is N(120, 10) distributed. What is the probability that Jon's blood pressure is at least 5 points lower than Jacks?

SOLUTION

Let X be Jon's blood pressure and let Y be Jacks. Then we can take X and Y to be independent and identically normally N(120, 10) distributed. We want P(X < Y -5). But this is the same as P(X - Y < - 5). Now X - Y has mean 0 and standard deviation $\sqrt{100 + 100}$ = 14.14. By additional property (ii) X - Y is N(0, 14.44) distributed. Therefore P(X - Y < - 5) = $\Phi(\frac{-5}{14.14})$ = 0.3642.

6.2 CENTRAL LIMIT THEOREM

Random variables X_1, …, X_n are called <u>independent and identically distributed</u> if they are independent and they all have the same distribution. Suppose X_1, …, X_n are independent and identically distributed with mean μ.and standard deviation σ. The quantity $\overline{X}_n = \frac{X_1 + ... + X_n}{n}$ is called the <u>sample average</u> (or sample mean) of the observations. By the rules for expectation and variance,

$$E(\overline{X}_n) = \mu, \qquad (2.40)$$

$$Var(\overline{X}_n) = \frac{\sigma^2}{n}, \qquad (2.41)$$

and

$$\text{standard deviation of } \overline{X}_n = \frac{\sigma}{\sqrt{n}}. \qquad (2.42)$$

Observe that equation (2.40) shows that the expected value of any particular X_i and that of the sample average $\overline{X}_n$ are the same. However, (2.42)shows their standard deviations are quite different. That for an individual X_i is σ while that for their average $\overline{X}_n$ is smaller, namely $\frac{\sigma}{\sqrt{n}}$. In fact by taking a large enough sample the standard deviation of $\overline{X}$ can be made as close to 0 as we wish.

Suppose $X_1, \ldots, X_n$ are independent and identically $N(\mu,\sigma)$ distributed. From additional property (iii) of the normal distribution, we see that $S_n = X_1 + \ldots + X_n$ is $N(n\mu, \sqrt{n}\sigma)$ distributed, and by additional property (i), $\overline{X}_n = S_n/n$ is $N(\mu, \frac{\sigma}{\sqrt{n}})$ distributed. One of the most remarkable and important results in all mathematics is the fact that the distributions of S_n and $\overline{X}_n$ are approximately normally distributed no matter what the common distribution of the X_i is!. This fact is known as the Central Limit Theorem.

Let us first point out that the normal approximation to the binomial distribution

CENTRAL LIMIT THEOREM

Suppose $X_1, \ldots, X_n$ are independent and identically distributed with mean μ and standard deviation σ. Then no matter what the distribution of X is, for n sufficiently large,

(a) S_n is approximately $N(n\mu, \sqrt{n}\sigma)$ distributed,
and

b. $\overline{X}_n$ is approximately $N(\mu, \sqrt{\frac{\sigma}{n}})$ distributed.

(without the half integer correction) is just a special case of the Central Limit Theorem. Just take X to be an indicator random variable with success probability p. Then $\mu = p$, $\sigma = \sqrt{p(1-p)}$, and by (2.1) S_n is b(n,p) distributed. The normal approximation to the binomial is exactly statement (a) in the box and the normal approximation to $\hat{p}$ is statement (b) in the box.

The really amazing part of the Central Limit Theorem is that for large n, other than for its mean and standard deviation, the approximate distributions of S_n and $\overline{X}_n$ do not at all depend on common distribution of the X_i. The practical importance of this is that to calculate things about S_n and $\overline{X}_n$ we do not need to know this distribution. of X. That's good, because in reality we almost never know this distribution.

COMMENT

The statement in the box is how we use the CLT. Just as for the normal approximation to the binomial, we simply pretend that S_n is $N(n\mu, \sqrt{n}\sigma)$ distributed. Mathematically speaking, the CLT is a limit result. Precisely, the CLT states that for any x,

$$\lim_{n\to\infty}\left[P(S_n \le x) - \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)\right] = 0\,, \qquad (2.43)$$

or, by a change of variable,

$$\lim_{n\to\infty}\left[P(S_n \le x\sigma\sqrt{n} + n\mu) - \Phi(x)\right] = 0\,. \qquad (2.44)$$

This limit result tells us that for large enough n, $P(S_n \le x)$ will differ from $\Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right)$ by as small an amount as we wish. The question of how large an n is needed so that the approximation has a specified accuracy is not easy to answer. In practice, we usually just go ahead and use it whenever $n \ge 20$.

EXAMPLE 2.22

Suppose flights from LA to San Francisco have a mean of 45 minutes and a standard deviation of 10 minutes.

(a) Find the probability that a flight lasts longer than 50 minutes.

(b) Find the probability that the average of 30 flights is greater than 50 minutes.

(c) Find the probability that the total time of 20 flights is less 950 minutes.

SOLUTION

(a) This cannot be found from the information given. We would need the distribution of X to perform the calculation and this is not given.

(b) By the CLT, the average $\overline{X}_{30}$ of the 30 flights can be taken to be approximately $N(45, \frac{10}{\sqrt{30}})$ distributed. Thus

$$P(\overline{X}_{30} > 50) \approx 1 - \Phi\left(\frac{50-45}{1.8257}\right) = 0.0031.$$

(c) The total time of 20 flights is the sum S_{20} of the 20 flight times. By the CLT

$$P(S_{20} < 950) \approx \Phi\left(\frac{950-(20)(45)}{10\sqrt{(20)}}\right) = \Phi(1.12) = 0.8686.$$

EXAMPLE 2.23

The mean serum calcium in healthy individuals is 9.47 with a standard deviation of 0.3. Find the approximate probability that the average of 20 measurements of serum calcium are between 9.3 and 9.5.

SOLUTION

We want $P(9.3 < \overline{X}_{20} < 9.5)$. By the CLT,

$$P(9.3 < \overline{X}_{20} < 9.5) \approx \Phi\left(\frac{9.5-9.47}{\frac{.3}{\sqrt{20}}}\right) - \Phi\left(\frac{9.3-9.47}{\frac{.3}{\sqrt{20}}}\right) = \Phi(.45) - \Phi(-2.53) =$$

0.6679.

EXAMPLE 2.24 (Blood Tests)

Recall the discussion of blood tests from Chapter 1. There we showed that if the disease incidence rate $p < 0.3078$ then pooling would be effective. What we actually showed was that the expected number of tests, EN, using the optimal group size $m \approx \frac{1}{\sqrt{p}}$ would about $2n\sqrt{p}$. For p small this is substantially less than n, the number needed to test everyone. However, what we really want to know is the actual number N of tests that we will need using the pooling. Since N is a random variable we cannot know with certainty what it will be. But we are now in a position to calculate $P(N \leq x)$ for any x. These probabilities will tell us considerably more than just knowing EN.

From Chapter1, the number of tests used is $N = X_1 + \ldots + X_k$, where the X_i are independent and identically distributed random variables. Using the optimal group size of m $\approx \frac{1}{\sqrt{p}}$ and the approximation $(1 - p)^m \approx 1 - mp = 1 - \sqrt{p}$ we find from the results in Chapter 1 that

$$\mu = EX_i = 1 + m[1 - (1 - p)^m)] \approx 2.$$

$$\sigma^2 = Var(X_i) = m^2(1 - p)^m[1 - (1 - p)^m] \approx \frac{1-\sqrt{p}}{\sqrt{p}}.$$

Applying the CLT we see that

$$P(N \le x) \approx \Phi\left(\frac{x - k\mu}{\sqrt{k}\sigma}\right).$$

Now $k = \frac{n}{m} \approx n\sqrt{p}$ so $\sigma\sqrt{k} \approx \sqrt{n(1-\sqrt{p}} \approx \sqrt{n}$.

Thus

$$P(N \le x) \approx \Phi\left(\frac{x - 2n\sqrt{p}}{\sqrt{n}}\right). \qquad (2.45)$$

What does (2.45) tell us? Consider for example that the 99.99th percentile of the standard normal distribution is 3.719. Using (2.45) and taking $x = 2n\sqrt{p} + 3.719\sqrt{n}$ we find that

$$P\left(N - 2n\sqrt{p} \le 3.719\sqrt{n}\right) \approx \Phi(3.719) = 0.9999 \qquad (2.46)$$

In words (2.46) tells us that with 99.99% certainty, the number of tests required will not exceed the mean of $EN = 2n\sqrt{p}$ by more than $3.719\sqrt{n}$. For example, if 100,000 people were to be tested, then with probability at least 0.9999 the number of tests would not exceed the mean by more than $(3.719)(\sqrt{100{,}000}$ = 1176.05 if p = .01 then EN = .2n = .2(100,000) = 20,000. Thus with at least 99.99% certainty we will not need more than 21,177 tests to test the group of 100,000 people. Observe how much more we now know than just knowing the mean number of tests.

EXAMPLE 2.25 (Martingale System)

In Chapter 1 (section 14.5.1) we discussed the martingale gambling system. Recall that the player initially bets $1 and doubles his bet at each tail and terminates play when either he gets a head or he is forced to quit by the house limit. Call that round 1.Now suppose upon termination he starts anew with a $1 bet and continues as before forming round 2. When round 2 finishes he starts on round 3, etc. Let Y_i be his gain on round i. Then Y_i can have only 2 values: 1 with probability 1 - q^k and -(2^k - 1) with probability q^k. The random variables $Y_1,\ldots,Y_m$ are independent. In Chapter 1 we found $EY_i = 1 - (2q)^k$ and in the exercises that the variance is $(2^k - 1)q^k(1 - q^k)$. Using the CLT we can now approximate how much money our gambler wins. His total gain is $G = Y_1 + \ldots + Y_m$.

$$P\left(G \le m(1-(2q)^k + \sqrt{m(2^k-1)q^k(1-q^k)}b\right) = \Phi(b) \qquad (2.47)$$

For a numerical illustration consider playing red on roulette. The probability q is 0.5263. Suppose M = 1000. Then k = 9 and (2.47) shows that for b =1.645

$$P(G \le -0.5862m + 1.5782\sqrt{m}) = 0.95. \qquad (2.48)$$

We know from Chapter 1 that there is probability 0.998 of winning a $1. Since the mean is negative, are winnings after m rounds must ultimately become negative. We can use (2.48) to estimate the smallest m for which we have probability 0.95 of having a negative gain. What is required is the smallest m so that $-0.5862m + 1.5782\sqrt{m}) = 0$. This is $\sqrt{m}$ = 1.5782/0.5862 = 2.69 . So m = 7.248. Thus our gambler has 95% chance of a loss 8 rounds.

7 POISSON PROCESSES

7.1 THE POISSON DISTRIBUTION

For any real number t the infinite series $\sum_{n=0}^{\infty}\frac{t^n}{n!}$ has sum e^t

i.e.,

$$\sum_{n=0}^{\infty}\frac{t^n}{n!} = e^t. \qquad (2.49)$$

Consequently, $e^{-t}\sum_{n=0}^{\infty}\frac{t^n}{n!} = 1$.

We say X has a <u>Poisson distribution with parameter a</u>, a > 0 if

$$P(X = x) = e^{-a}\frac{a^x}{x!}, \ x = 0, 1, \dots \qquad (2.50)$$

It is easy to compute the mean and variance of X.

$$EX = \sum_{x=0}^{\infty} x e^{-a}\frac{a^x}{x!} = \sum_{x=1}^{\infty} e^{-a}\frac{a^x}{(x-1)!} = a\sum_{x=0}^{\infty} e^{-a}\frac{a^x}{x!} = a.$$

A similar computation shows $EX^2 = a + a^2$, so $Var(X) = a + a^2 - a^2 = a$.

Thus the parameter of a Poisson distribution is both its mean and its variance. Henceforth, we shall refer to a Poisson distribution with mean a (rather than with parameter a).

EXAMPLE 2.26

Let X be Poisson distributed with mean 2. Find (a) P(X = 2), (b) P(X > 1).

SOLUTION

$P(X = 2) = e^{-2}(2)^2/2!$ =0 2707 (b) $P(X > 1) = 1 - P(X \le 1) = 1 - (e^{-2} + 2e^{-2}) =$ 0.5940.

7.2 POISSON APPROXIMATION TO THE BINOMIAL

The Poisson distribution was named after one of the great 19th Century French mathematicians, Simeon Denis Poisson. (1781-1840) Originally, the Poisson distribution arose as an approximation to the binomial for those cases in which p is small and n is large. This approximation, which is credited to Poisson but may actually be due to the French mathematician De Moivre, appeared in 1837. The Poisson approximation simply pretends that a random variable having the binomial b(n,p) distribution has a Poisson distribution with mean np. How accurate is the Poisson approximation? That depends on the value of n and p. It is of use only in those cases where n is large np^2 is small. (Note that these conditions imply that p is small)

We now have two methods to approximate the binomial distribution, viz. the normal and the Poisson. The normal is quite accurate when $np \ge 5$ and should be used when this is the case. The Poisson approximation is usually used in those situations in which p is very small and n is not large enough to make $np \ge 5$. In these situations np^2 will usually be quite small.

EXAMPLE 2.27

Suppose the incidence rate of a certain genetic defect is 0.001. (a) Find the probability that amongst 2000 randomly chosen people not more than 2 have the defect.(b) Approximate this probability by the Poisson approximation. (c) Approximate this probability using the normal approximation.(d) compare the results in parts (a)-(c).

SOLUTION

As usual we assume that the binomial model is valid. Therefore the number of people N having the defect amongst the 2000 is b(2000, .001) distributed. (a) P(N $\leq$ 2) = $(.999)^{2000} + \binom{2000}{1}(.001)(.999)^{1999} + \binom{2000}{2}(.001)^2(.999)^{1998}$ = 0.6767

We pretend N is Poisson distributed with mean 2000(.001) = 2. So

P(N $\leq$ 2) $\approx e^{-2}(1 + 2 + 2^2/2)$ 0.6767.

Here np = 2 and $\sqrt{np(1-p)}$ = 1.4135. The normal approximation gives

P(N $\leq$ 2) $\approx \Phi(\frac{2.5-2}{14135})$ = 0.6382

As we can seen, the Poisson approximation is exceptionally accurate. Observe that $np^2 = 2000(.001)^2$ = 0.002 is very small. However, the normal approximation, even with the half integer correction, is not nearly as good (though it isn't all that bad). Without the half integer correction however it gives the value 0.5, which is a quite poor approximation to the true value.

7.3 SOME POISSON MODELS

Computationally, the Poisson approximation is not anywhere near as useful as the normal approximation already discussed. True, if computations are done by hand, it can replace a boring computation involving binomials with a more pleasant one, but if that's all the Poisson distribution was good for no one today would care about either it or the approximation. Indeed, even when it first appeared it, was merely viewed as a useful method to avoid a tedious calculation but otherwise caused no real excitement. In fact, for about 50 years after its introduction the Poisson distribution itself found no real application and was all but forgotten. Then in 1898 the work of a certain German professor named Ladislaus von Bortkiewicz investigating of all things the fatalities of soldiers in the Prussian cavalry due to horse kicks suddenly propelled the Poisson distribution to prominence. His work was the first of many to show that many interesting phenomena are intimately connected with the Poisson distribution.

7.3.1HORSE KICKS

Bortkiewicz analyzed information gathered by the Prussian cavalry about the number of fatalities amongst soldiers due to horse kicks. He was interested in the distribution of the number of fatal horse kicks per corps-year. That is, in the number of fatal horse kicks per corps per year. The information was gathered on 10 cavalry corps over a 20-year period. These in total gave 200 corps-years of observations.

Here is the data.

Table 2-4

fatalities	0	1	2	3	4
Observed number of corps-years	109	65	22	3	1

As a model for the distribution of fatal horse kicks Bortkiewicz now assumed that the number, X, of fatal horse kicks in a corps year followed a Poisson distribution. Having decided that X should be Poisson distributed in order to "fit" the data to the Poisson Bortkiewicz now needed to establish the value for its parameter a. (In the jargon of statistics to estimate a.) If we suppose that the distribution of fatal horse kicks is the same in each of the 200 corps-years and that these fatalities are independent from corps-year to corps-year, then these 200 observation are a random sample of size 200 from the unknown Poisson distribution. To determine the parameter a he noted that a is the expectation of the distribution. Since a is the mean of the Poisson distribution he estimated a by taking it to be the sample average of the number of fatal kicks over his 200 corps-years of observations. In all, there were (0)(109) + (1)(65) + (2)(22) + (3)(3) +(4)(1) = 122 fatalities over the 200 corps-years period. Thus the average rate per corps-year was 122/200 = 0.61. He then took X to have the Poisson distribution with mean 0.61 and computed the probabilities P(X = x) for x = 0,…,4. These probabilities are the <u>predicted proportions</u> using the Poisson model. The table below gives theses probabilities together with the observed proportions.

Table 2-5

fatalities	0	1	2	3	4
predicted proportion	.5435	.3315	.1010	.0205	.0030
Observed proportion	.5450	.3250	.1100	.0150	.0050

As can be seen, the agreement between the observed and predicted is excellent. This shows that the model that takes the number of fatal horse kicks to follow a Poisson distribution (with mean .61) is quite a good. Notice that here the Poisson distribution is being used as a basic distribution and not as an approximation to a binomial distribution.

There is another way that we can present this data. Under the model in a given corps-year there is probability P(X = x) that there are exactly x fatal kicks. Let $Z_i = 1$ if the i^{th} corps-year has exactly x fatal kicks and let $Z_i = 0$ if not. Then $Z_1 + \ldots + Z_{200}$ is the number of corps-years with exactly x horse kicks. Since $P(Z_i = 1) = P(X = x)$ for all i, the expected number of corps-years with exactly x kicks is 200P(X = x). We can then compare these expected numbers with the actual observed numbers as in the table below.

Table 2-6

fatalities	0	1	2	3	4
Expected number of corps-years	108.7	66.3	20.2	4.1	.6
Observed number of corps-years	109	65	22	3	1

This second way of presenting the data (which amounts to just multiplying the entries in the previous table by the number of cases (=200)) is used more frequently than the first. We will use it in the next two illustrations.

7.3.2 RADIOACTIVE DISINTEGRATION'S

Another early use of the Poisson distribution was connected with radioactive disintigrations. In the famous early 20th century experiment of Rutherford, Chadwick, and Ellis they observed the number of α-particles reaching a counter in a 7.5 second interval for 2608 such time intervals. In all, they observed 10,094 particles in these 2608 intervals giving the average of 10,094/2608 = 3.870

particles per 7.5 second time interval. We now proceed as with the horse kicks. Let X be the number of particles that arrive in a 7.5 second time interval and assume X has a Poisson distribution. Estimate its parameter to be the observed average number of particles per time interval, namely 3.870 and then compute P(X =x). If we multiply these probabilities by the number of time intervals (= 2608) we find the expected number of intervals that have x particles. The table below gives the observed and expected number of intervals with x.

Table 2-7

k	observed	expected	k	observed	expected
0	57	54.399	5	408	393.515
1	203	210.523	6	273	253.817
2	383	407.361	7	139	140.325
3	525	525.496	8	45	67.882
4	532	508.418	9	16	17.075

As can be seen the model that says the number of particles in 7.5-second interval follows a Poisson distribution fits this data quite well.

7.3.3 FLYING BOMB HITS

During World War II London was subject to attack by unguided missiles called flying bombs launched from Germany. We consider the number of hits by such bombs in the south of London. The entire region was divided into 576 sub-regions each having area equal to 1/4 sq. kilometer. The entire region had 537 hits. The average number of hits was 537/576 = 0.9323. The table below records the observed and expected number of sub-regions with k hits. The expected number with k hits is 576P(X = k) where X is Poisson distributed with mean 0.9323.

k	0	1	2	3	4	≥5
observed	229	211	93	35	7	1
expected	226.74	211.39	98.54	30.62	7.14	1.57

Again, the model that says the number of hits in a 1/4 sq. kilometer area is Poisson distributed fits the data quite well.

7.4 POISSON PROCESSES

Why does the Poisson distribution arise as both an approximation to the binomial and at the same time fit the phenomena discussed above so well? What properties do these seemingly unrelated things have in common?

In all cases we are investigating either "events" that occur randomly in time or "things" that are distributed randomly on a subset of either the line, the plane, or of 3-space. We can think of events occurring in time as the distribution of things (i.e. the events) on the positive half line. Therefore we can model all of these phenomena as the random distribution of hypothetical entities which we will call particles that are distributed on some specified subset S of either the lne, the plane, or of 3-space.

One way we can imagine describing the distribution of the particles on S is by giving the number of particles, N(I), in a subset I of S for every subset I of S. In all, this enormous collection of random variables describes our distribution of particles. These random variables are linked to each other. Observe that if A_1, A_2, …, A_n are disjoint subsets of S with union A then we must have

$$N(A) = N(A_1) + \dots + N(A_n). \qquad (2.51)$$

The model we will make for the distribution of the particles on S is that of particles that are distributed in a "completely random", "homogeneous", and "sparse" manner. We must of course define precisely each of the three terms in quotes. This we will do by imposing conditions I-III below.

Intuitively, if particles are distributed completely at random then knowing how many are in one subset should give no information about how many are in another disjoint subset. Thus we formalize the concept of a completely random distribution of particles by the following condition.

Let A_1, …, A_n be disjoint subsets of S. Then the random variables $N(A_1),\ldots,N(A_n)$ are independent.

Let |A| denote the length of A if S is a subset of the line. If S is a subset of the plane let |A| denote its area, and let |A| be the volume of A if S is a subset of 3-space. Homogeneity is formalized by demanding that for a subset I of S the distribution of N(I) depends only on the size of I as measured by |I| and not on its shape. This means that if A and B are two subsets of S such that |A| = |B| then P(N(A) = k) = P(N(B) = k) for all k = 0,1, … We therefore formalize the concept of homogeneity by the following condition.

The probability function of N(I) depends on I only as a function of |I|.

The third concept of the distribution of particles being sparse is the most difficult to make precise. In rough terms the distribution is sparse provided that for a small subset I, i.e. a subset with |I| small, the probability of having any particle at all is proportional to |I| and the probability of having more than one particle is essentially 0. To make this precise we note that by (II) the probabilities P(N(I) > 0) and the probabilities P(N(I) > 1) are functions of |I|. We impose the following condition.

There is a constant $\lambda \geq 0$ such that

(IIIa) $$\lim_{|I| \to 0} \frac{P(N(I) > 0)}{|I|} = \lambda .$$

(IIIb) $$\lim_{|I| \to 0} \frac{P(N(I) > 1)}{|I|} = 0.$$

What IIIa tells us is that for a small set P(N(I) > 0) is approximately proportional to |I|.

In general, one might think that there would be many different ways of distributing particles onto S to satisfy these three conditions. That is, that the random variables N(A) could have many possible distributions. But remarkably, that is not the case. These 3 conditions actually completely fix the distributions of the random variables N(A) for all subsets A of S!

Under conditions I-III if |A| is finite then N(A) is Poisson distributed with mean λ |A|.

A Poisson process on S with intensity λ is a collection of random variables {N(A)} for A running over the subsets of S that satisfy the following 2 conditions (i) Let A_1, …, A_n be disjoint subsets of S. Then the random variables $N(A_1),\ldots,N(A_n)$ are independent.(ii) For A a subset of S, N(A) has the Poisson distribution with parameter $\lambda|A|$.

The implications of our previous discussion is that any distribution of particles on S that satisfies conditions I-III must be a Poisson process. It is this fact that makes the Poisson distribution so important.

To work problems involving the Poisson process we need to know the intensity λ. This may either be directly specified (See Example 2.26) or preliminary information can be given that enables one to compute λ.(See Examples 2.27 – 2.30).

EXAMPLE 2.28

Assume that cosmic rays arrive in time as a Poisson process with intensity of 2 per minute. Find the probability that no cosmic ray arrives in a 1/2 minute interval of time.

SOLUTION

Let us measure time in minutes. The number of cosmic rays N in a 1/2 minute interval is Poisson distributed with parameter (2)(.5) = 1. The required probability is $P(N = 0) = e^{-1} = 0.3679$.

EXAMPLE 2.29

Suppose the expected number of bacteria in a 2 cm^2 portion of a dish is 3. Find the probability that a 3 cm^2 portion of the dish has at most 2 bacteria. (Assume the Poisson process holds)

SOLUTION

We will measure in units cm^2. Let N be the number of bacteria in a 2 cm^2 portion. Then N has the Poisson distribution with parameter $\lambda(2)$. We are told EN = 3. Thus $(\lambda 2)$.= 3 so $\lambda = 1.5$. The number of bacteria M in a 3 cm^2 portion is Poisson distributed with parameter $3\lambda = 4.5$. The required probability is

$$P(M \le 2) = P(M = 0) + P(M = 1) + P(M = 2) = e^{-4.5}(1 + 4.5 + \frac{(4.5)^2}{2}) = 0.1736.$$

EXAMPLE 2.30

Suppose that flaws in a particular type of sheet metal occur at the average rate of 1.5 per $10ft^2$. Find the probability that a circular piece of metal of radius 3ft. has more than one flaw.

SOLUTION

Let us work in units of ft^2. Then we are told that 10λ = 1.5 so = λ = . Let N be the number of flaws in the circular piece. Then N is Poisson distributed with mean $(.15)(\pi(3)^2)$ = 4.2412. So P(N > 1) = 1- (P(N = 0) + P(N = 1)) =1 – $e^{-4.2412}$(1 + 4.2412) = .9246.

EXAMPLE 2.31

Suppose Frank has angina. The probability that he has at least one angina attack in a given day is .1. What is the probability that he has at most 2 attacks in a 3-day period?

SOLUTION

Let X be the number of attacks in a day. Then X is Poisson distributed with mean λ. We are told P(X > 0) = .1. Therefore 1 – $e^{-\lambda}$ = .1 so –ln(.9) = λ. Let Y be the number of attacks in a 3-day period. Then Y is Poisson distributed with mean 3λ = - 3ln(.9).= .3161. Then P(Y $\le$ 2) = $e^{-.3161}(1 + .3161 + (.3161)^2/2)$ = 9958.

EXAMPLE 2.32

Two genes on the same chromosome are called linked. During meiosis chromosomes break and recombine by a process called crossing over. Consider two non-sex chromosomes. The paternal and maternal chromosomes duplicate and the 4 chromosomes line up in a bundle. Think of these as 4 straight lines. Breaks occur along this bundle as a Poisson process. At each break one of each of the two maternal and paternal chromosomes exchange their upper parts. This is a crossover. If the maternal and paternal genes are different a crossover results in a different genotype than either of the parents. This is called a recombination. There can be several crossovers between the two genes. Only an odd number of crossovers give rise to a recombination. The "distance" d between two linked genes is defined to be the expected number of crossovers between them. The recombination fraction p is the probability of a recombination between the two genes. Find d in terms of p.

SOLUTION

Let N be the number of crossovers between the 2 genes. Then p = P(N is odd) and d = EN. Now

$$P(N \text{ is odd}) = e^{-d} \sum_{n \text{ odd}} \frac{d^n}{n!}.$$

To evaluate the sum note that from (2.49)

$$e^d = 1 + d + \frac{d^2}{2!} + \ldots$$

$$e^{-d} = 1 - d + \frac{d^2}{2!} + \ldots$$

so

$$\frac{e^d - e^{-d}}{2} = d + \frac{d^3}{3!} + \ldots = \sum_{n \text{ odd}} \frac{d^n}{n!}$$

Thus

$$p = e^{-d} \frac{e^d - e^{-d}}{2} = \frac{1 - e^{-2d}}{2}. \qquad (2.52)$$

Solving for d we find

$$d = \frac{-\ln(1 - 2p)}{2} \qquad (2.53)$$

In genetics, formula (2.52), are called Haldane's mapping function.

7.5 MATCHING

In Section 13.6.1 of Chapter 1 we considered matching. If N_n is the number of matches then we found that

$$P(N_n > 0) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}.$$

Therefore,

$$P(N_n = 0) = 1 - \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}. = 1 - 1 + \ldots + \frac{(-1)^n}{n!}.$$

But the expression on the right is just the sum of the first n terms of the infinite series whose sum is e^{-1}. Thus

$$\lim_{n\to\infty} P(N_n = 0) = e^{-1}. \quad (2.54)$$

The convergence of the infinite series is quite rapid. Consequently, for n > 10 we can well approximate $P(N_n = 0)$ by $e^{-1} = .3679$.

It is now quite simple to find the formula for $P(N_n = k)$ for any k. To this end note that $N_n = k$ if and only if there are matches at exactly k of the n positions and no matches at the other n - k remaining positions. The k positions for the matches can be selected from the n positions in $\binom{n}{k}$ ways. Let b_n be the number of ways in matching n things that there can be no match. Then $P(N_n = 0) = B_n/n!$. Turning this around,

$B_n = (n!)P(N_n = 0)$.

So the probability of no match at the n - k remaining positions is $B_{n-k} = (n-k)!P(N_{n-k} = 0)$. Thus

$$P(N_n = k) = \frac{\binom{n}{k}B_{n-k}}{n!} = \frac{1}{k!}P(N_{n-k} = 0). \quad (2.55)$$

For fixed k, as n gets large we see that

$$\lim_{n\to\infty} P(N_n = k) = e^{-1}\frac{1}{k!}$$

Thus N_n is approximately Poisson distributed with mean 1

8 DIFFERENCE EQUATIONS

Let $c_1, \ldots, c_r$ be specified numbers such that $c_r \neq 0$. We desire to find sequences $a_0, a_1, \ldots$ that satisfy the relationship

$$a_n = c_1a_{n-1} + c_2a_{n-2} + \ldots + c_ra_{n-r},\ n \geq r \quad (2.56)$$

A homogeneous difference (or recurrence) equation of order r is an expression of the form (2.56). The constants $c_1, \ldots, c_r$ are called the coefficients. Difference equations arise often in probability theory and their use is a powerful tool for solving many probability problems.

8.1 HOMOGENEOUS DIFFERENCE EQUATIONS OF ORDER 1

The homogeneous difference equation of order 1 is simply

$a_n = c_1 a_{n-1}$ (2.57)

To illustrate the type of reasoning that leads to difference equations consider the number of tosses of a coin needed to first get a head. Let a_{n-1} be the probability that this first occurs with toss n. For n > 1, in order for the first head to occur at toss n it must be that the first toss is tails (Otherwise the first head occurs at toss 1). Looking what happens after the first toss the first head must happen at toss n – 1. For example, for the first toss to happen at toss 5 the first toss must be tails and looking from tosses 2, 3 ,.. the first toss must be at the 4^{th} one of those tosses. Since all the tosses are independent we see that

$a_n = qa_{n-1}$, n > 1. (2.58)

To illustrate the general procedure of solving homogeneous difference equation we try to see if taking $a_n = bt^{n-1}$will satisfy the equation for suitable choice of b and t. Substituting this expression into (2.57) yields

$bt^n = c_1 bt^{n-1}$

Dividing both sides of the above equation by bt^{n-1} we see that we obtain a solution for any value of b provided $t = c_1$. Thus $a_n = b(c_1)^n$ is a solution for any b. If now we specify the value of the first term a_0 then the value of b is fixed. For taking n = 0 we see that $a_0 = b(c_1)^0 = b$.

8 2 HOMOGENEOUS RECURRENCE EQUATIONS OF ORDER 2

The homogeneous difference equation of order 2 is

$$a_n = c_1 a_{n-1} + c_2 a_{n-2}, \quad (2.59)$$

There are 2 types of problems for these difference equations called initial value problems and boundary value problems. In the former we seek a solution that has specified values for the first 2 terms, a_0 and a_1. That is why it is called the initial value problem. In the latter problem we require that (2.59) hold only for n between two values l < n < u and for given values d_1 and d_2 (called the boundary conditions), $a_l = d_1$ and $a_u = d_2$.

To solve (2.59) we first ignore the initial conditions or the boundary conditions and look for solutions to (2.59). Now there are in general many solutions. We point out an important property that solutions of (2.59) posses.

Suppose a_n and b_n are solutions. Then for constants k and d, $ka_n + d\, b_n$ is also a solution because

$ka_n + db_n - c_1(ka_{n-1} + db_{n-1}) - c_2(ka_{n-2} + db_{n-2}) =$

$k[a_n - c_1a_{n-1} - c_2a_{n-2}] + d[b_n - c_1b_{n-1} - c_2b_{n-2}]$.

Since the terms in the brackets are 0 we see that $ka_n + d\, b_n$ is a solution.

To solve (2.59) we see if there is some t so that $t^{n.}$ is a solution. To do so we take $a_n = t^n$ in .(2.59) This yields the equation

$t^n = c_1 t^{n-1} + c_2 t^{n-2}$

Dividing both sides by t^{n-2} and subtracting the left hand side from the right hand side yields the quadratic equation

$$t^2 - c_1 t - c_2 = 0. \qquad (2.60)$$

This shows that t^n is a solution of (2.59) if and only if t satisfies the quadratic equation (2.60). This equation is called the characteristic equation.

8.2.1 INITIAL VALUE PROBLEM

We must now consider the case when there are two different roots when there is one multiple root.

CASE 1 Two Distinct Roots.

Let these be r_1 and r_2, where we label the roots so that $|r_1| > |r_2|$. Then for constants b_1 and b_2,

$$a_n = b_1 r_1^n + b_2 r_2^n \qquad (2.61)$$

is also a solution.

Now we see if we can determine the constants b_1 and b_2 so that this solution satisfies the initial conditions.

Taking n = 0 and n = 1 in (2.61) we obtain two equations for the unknown coefficients in terms of the two roots and the initial values.

$$\begin{aligned} a_1 &= b_1 + b_2 \\ a_2 &= b_1 r_1 + b_2 r_2 \end{aligned} \qquad (2.62)$$

Solving these equations for b_1 and b_2 we find

$$b_1 = \frac{r_2 a_0 - a_1}{(r_2 - r_1)}$$
$$b_2 = \frac{-r_1 a_0 + a_1}{(r_2 - r_1)} \qquad (2.63)$$

Thus (2.61) with the coefficients given by (2.63) is the unique solution to (2.59) satisfying the initial conditions when there are two different roots to the characteristic equation.

It is of interest to determine how this solution behaves for large n. Since $\frac{|r_2|}{|r_1|} < 1$,

$$\lim_{n\to\infty}\left(\frac{|r_2|}{|r_1|}\right)^n = 0.$$

Thus

$$\lim_{n\to\infty}\left(\frac{r_2}{r_1}\right)^n = 0.$$

(2.64)

Now (2.61) shows

$$\frac{a_n}{r_1^n} = b_1 + b_2\left|\frac{r_2}{r_1}\right|^n$$

Using (2.64) we see that

$$\lim_{n\to\infty}\frac{a_n}{r_1^n} = b_1.$$

This is usually written as

$$a_n \approx b_1 r_1^n \qquad (2.65)$$

It says that for large n, a_n is about $b_1 r_1^n$.

CASE 2 Both Roots Equal

Call the common root r. Then $(t - r)^2 = 0$. Expanding we see that $t^2 - 2tr + r^2 = 0$. Equating the coefficients of t in this equation with those in (2.60) we find that

$$2r = c_1 \text{ and } r^2 = -c_2. \qquad (2.66)$$

We will use these facts to show that nr^n is also a solution of (2.59). Writing that equation as $a_n - c_1a_{n-1} - c_2a_{n-2} = 0$ and substituting $a_n = nr^n$ in the left hand side of the equation we get $nr^n - c_1(n-1)r^{n-1} - c_2(n-2)r^{n-2}$. We need to show this is 0. To do so we write it as

$$nr^{n-2}(r^2 + c_1r + c_2) + r^{n-2}(-c_1r + 2c_2). \qquad (2.67)$$

The first term in (2.67) is 0 because r is a root of the characteristic equation. By (2.66) second term is $-2r^2 + 2r^2 = 0$. Thus nr^n is a solution. Consequently, for any two constants b_1 and b_2,

$$a_n = b_1r^n + b_2nr^n \qquad (2.68)$$

is a solution. Taking n = 0 and n = 1 we find

$$b_1 = a_0$$

$$b_2 = \frac{a_1}{r} - a_0$$

(2.69)

Thus (2.68) with the coefficients given by (2.69) is the unique solution of (2.59) satisfying the initial conditions in the case when the characteristic equation has a single root.

To determine how solution (2.68) behaves for large n we proceed just as in the previous case. This shows that

$$\lim_{n\to\infty} \frac{a_n}{nr^n} = b_1 \qquad (2.70)$$

so

$$a_n \approx b_1 nr^n. \qquad (2.71)$$

EXAMPLE 2.33 (Success Runs Of Length 2)

Consider tossing a coin having probability p for heads. A success run of length 2 first occurs at toss n if tosses n and n - 1 are both heads and amongst the first n - 1 tosses there are not 2 heads in succession. Let a_n be the probability that a success run of length 2 first occurs at toss n. Let $X_i = 1$ if toss i is a head and $X_i = 0$ if not. Let T = n if the first success run of length 2 is at toss n and let $a_{n-1} = P(T = n)$. [The reason for this shift to n -1 is that we number the sequence $\{a_n\}$ starting from n = 0 while T has values 1, 2, …] Consider what must happen at the first two tosses. If the first toss is a tail, then for a success run of length 2 to first occur at toss n it must be that the first success run of length 2 starting from toss 2 occurs at toss n, i.e at the (n -1)st toss starting from toss 2. Thus $P(T = n|X_1 = 0) = a_{n-1}$. If toss 1 is a head then toss 2 must be a tail. (Otherwise the first success run will occur at the 2nd toss.) Thereafter, we must get the first success run of length 2 in n - 2 additional tosses. Thus $P(T = n|X_1 = 1) = qpa_{n-2}$. Using the law of total probability we obtain the following difference equation:

$$a_n = qa_{n-1} + pqa_{n-2}, \; n \geq 2. \qquad (2.72)$$

The initial conditions are $a_0 = 0$, $a_1 = p^2$.

The characteristic equation is $r^2 = qr + pq$. It has the 2 distinct roots

$$\frac{q \pm \sqrt{q^2 + 4pq}}{2}.$$

Using (2.63) we find

$$a_n = \frac{p^2}{\sqrt{q^2+4pq}}\left[\left(\frac{q+\sqrt{q^2+4pq}}{2}\right)^n - \left(\frac{q-\sqrt{q^2+4pq}}{2}\right)^n\right], \; n \geq 0.$$

(2.73)

In particular, foe p = q = 1/2,

$$a_n = \frac{1}{2\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{4}\right)^n - \left(\frac{1-\sqrt{5}}{4}\right)^n\right].$$

For large n (2.65) shows

$$a_n \approx \frac{1}{2\sqrt{5}}\left(\frac{1+\sqrt{5}}{4}\right)^n.$$

EXAMPLE 2.34

Consider a particular allele of a gene on the X-chromosome. Assume generation n + 1 is formed by randomly selecting a male and female and mating them. A male has a single X-chromosome that it receives from its mother. A female has 2 X-chromosomes. It receives from each parent. Let f_n be the probability that a female in generation n has the allele. Let g_n be the probability that male in generation n has the allele. For a male in generation n, the conditional probability that he has the allele is 1 given his mother has the allele and 0 given his mother does not. The probability his mother has the allele is f_{n-1}. Therefore, by the law of total probability,

$$g_n = f_{n-1} \quad n > 0. \qquad (2.74)$$

Consider a female in generation n. She has the allele with probability 1 given both her parents have the allele. She has the allele with probability 1/2 if her mother does and her father doesn't or vice-versa. She has the allele with probability 0 if neither parent has the allele. Using the law of total probability we find

$$f_n = (1)f_{n-1}g_{n-1} + (1/2)[f_{n-1}(1 - g_{n-1}) + g_{n-1}(1 - f_{n-1})] = (f_{n-1} + g_{n-1})/2, \; n > 0 \qquad (2.75)$$

Taking n = 1 we see that $f_1 = (1/2)(f_0 + g_0)$, where f_0 and g_0 are the allele frequencies for the initial generation. Using (2.74) we see that $g_{n-1} = f_{n-2}$. Therefore, for $n \geq 2$ we can write (2.75) as the difference equation

$$f_n = (f_{n-1} + f_{n-2})/2 \qquad (2.76)$$

We leave the solution as an exercise.

EXAMPLE 2.35

Fibonacci's rabbits

Each pair of mature rabbits produces exactly one new pair of immature rabbits. Immature rabbits take exactly one time period to become mature rabbits.

Let f_n be the number of pairs of mature rabbits at time n and let g_n be the number of pairs of immature rabbits at time n. At time 0, $f_0 = 1$ and $g_0 = 0$. Find f_n and g_n.

SOLUTION

$$g_n = f_{n-1}, \; , n > 0 \qquad (2.77)$$

$$f_n = g_{n-1} + f_{n-1}, \; n > 0 \qquad (2.78)$$

From (2.77) we see that $g_{n-1} = f_{n-2}$. Substituting this into (2.78) we find that

$$f_n = f_{n-1} + f_{n-2}, \; n > 1 \qquad (2.79)$$

The initial conditions are $f_0 = f_1 = 1$.

The solution to (2.79) yields one of the most famous sequences of numbers called the Fibonacci numbers. As can be seen from (2.79) each successive Fibonacci number is the sum of the previous 2. These numbers have a way of appearing in all sorts of strange places. For example, the number of petals on a sunflower are always one of the Fibonacci numbers as are the number of cones on a pine cone. We leave the solution of (2.79) as an execise.

8.2.2 BOUNDARY VALUE PROBLEM

Equation (2.61) gives the general solution when there are two distinct roots to the characteristic equation. We now determine the coefficients b_1 and b_2 by using the boundary conditions. Taking n = l and n = u yields two equations

$$\left.\begin{array}{l} b_1 r_1^l + b_2 r_2^l = d_1 \\ b_1 r_1^u + b_2 r_2^u = d_2 \end{array}\right\} \qquad (2.80)$$

The solution is

$$\left.\begin{array}{l} b_1 = \dfrac{r_2^u d_1 - r_2^l d_2}{r_2^u r_1^l - r_1^u r_2^l} \\ b_2 = \dfrac{-r_1^u d_1 + r_1^l d_2}{r_2^u r_1^l - r_1^u r_2^l} \end{array}\right\} \qquad (2.81)$$

If there is a single root r the general solution is given by (2.68). Taking n = l and n = u yields the equations

$$\left.\begin{array}{l} d_1 = b_1 r^l + b_2 l r^l \\ d_2 = b_1 r^u + b_2 u r^u \end{array}\right\} \qquad (2.82)$$

The solution of these equations is

$$\left.\begin{array}{l} b_1 = \dfrac{u r^u d_1 - l r^l d_2}{(u-l) r^{u+l}} \\ b_2 = \dfrac{-r^u d_1 + r^l d_2}{(u-l) r^{u+l}} \end{array}\right\} \qquad (2.83)$$

EXAMPLE 2.36 (Gambler's Ruin)

The following is one of the most famous classical problems in probability theory. Suppose you play a game such as red or black in roulette. You at each play you can either win a dollar with probability p or loose a dollar with probability q = 1-p. The successive plays are independent of each other. You start with x dollars and play until either you are broke or you have $d \geq x$ dollars. It is a classic problem known as the gambler's ruin to calculate the probability that you go broke before your fortune increases to d.

Let a_x be the probability that starting with x dollars you go broke before your fortune increases to d. Clearly $a_0 = 1$ and $a_d = 0$. Suppose $0 < x < d$. After the first play of the game you have x + 1 dollars with probability p or x - 1 dollars with probability q. Thereafter, things proceed just as before with the new amount of money you have. Thus the b_x the recursion

$$a_x = pa_{x+1} + qa_{x-1},\ 0 < x < d \tag{2.84}$$

Here $l = 0$, $u = d$, $d_1 = 1$, $d_2 = 0$.

The characteristic equation is

$r = pr^2 + q.$

The solutions of this quadratic equation are

$$\frac{1 \pm \sqrt{1-4pq}}{2p}$$

A little algebra and use of the fact that $q = 1-p$ shows that $1 - 4pq = (q - p)^2$. Thus the roots are $\frac{1 \pm |q-p|}{2p}$. If $p \neq q$ then these roots are distinct. However if p = q, i.e. if p = 1/2 then both the roots are 1.

Let us first consider the case when q > p. In that case the two roots are $r_1 = q/p$ and $r_2 = 1$. The general solution is then

$$a_x = b_1 + b_2(q/p)^x. \tag{2.85}$$

Using (2.81) we obtain

$$\left.\begin{aligned} b_1 &= \frac{1}{1-(q/p)^d} \\ b_2 &= \frac{-(q/p)^d}{1-(q/p)^d} \end{aligned}\right\}$$

substituting into (2.85) we find

$$a_x = \frac{(q/p)^x - (q/p)^d}{1-(q/p)^d}, 0 \le x \le d. \tag{2.86}$$

If $q < p$ the characteristic equation has the same roots except now $r_1 = 1$ and $r_2 = q/p$. Consequently, the probabilities a_x are again given by (2.86). Finally, for $p = q = 1/2$ (2.83) shows

$$a_x = 1 - \frac{x}{d}, \; 0 \le x \le d \qquad (2.87)$$

We leave the details as an exercise.

Is it possible that one could play and never arrive at 0 or d? The answer is no. If h_x is the probability that starting from x you never get to either 0 or d then h_x satisfies (2.84) (with b_x replaced by h_x) with boundary condition $h_0 = h_d = 0$. It follows from this fact that $h_x = 0$ for all x, $0 \le x \le d$. The probability that you get to d before getting to 0 is $1 - b_x$.

Consider the game of roulette. In playing red and black you have probability p = 18/38 0.4737 of winning and probability q = 20/38 = 0.5263 of loosing. This is an unfavorable game (to the player). A fair game is one where p = 1/2. The table below gives the probability of getting to d before going broke starting with x = $100 for roulette and for a fair game.

d	101	110	120	150	200
p =.4737	.900	.3487	.1216	.0052	.00003
p = .5000	.9901	.9091	.8333	.6667	.5000

Notice that in a fair game the chance that you double your money is .5 but is essentially 0 in roulette. In roulette there is a 90% chance that you will win a dollar before you loose your 100 but there is therefore a 10% chance that you will loose the entire 100 before winning a $1.

PROBLEMS CHAPTER 2

SECTIONS 1 and 2

1. Suppose 5 balanced dice are rolled. What is the probability of (a) exactly 2 5's, (b) exactly 1 5, (c) either 1 or 2 5's.
2. An exam has 10 multiple choice questions each with 5 alternatives. Joe takes the exam and knows the answer to only 2 of the questions. For the other 8 questions he selects an answer at random from the 5 choices. (a) Find the probability that he gets exactly 5 answers correct. (b) Find the probability he gets at least 5 answers correct.
3. You toss a balanced coin 6 times. Are you more likely to get exactly 3 heads or 4 of one and 2 of the other kind?
4. A balanced die is rolled 6 times. Find the probability that there are exactly 2 sixes in the first 4 rolls if in the 6 rolls there are exactly 3 sixes total.
5. Suppose a balanced coin is tossed 3 times. Find the probability that all of the tosses are heads given that at least two of the three tosses is heads.
6. A test gives a false positive 30% of the time. Suppose this test is given to 10 different people chosen from a large population. Find the probability that the number of false positives is (a) > 0, (b) < 2 (c) $= 3$ (d) at most 2 (e) at least 2.
7. Substance x causes a toxic reaction in 20% of the people to whom it is given. What is the probability that it causes a toxic reaction in exactly 2 of 5 people to whom it is given. Assume the people are selected from a large population
8. In the following if N has a binomial distribution give its parameters. If the binomial does not apply give a reason why not.
   (a) A box has 30 red balls and 20 black balls. A sample of12 balls are selected with replacement from the box and N is the number of red balls in the sample.
   (b) A certain town has 4500 registered voters. 60% of them favor merging the town with the adjacent town.
   (c) A sample of 10 voters is taken and N is the number of them that favor the merger.

(d) A certain town has 4500 registered voters. 60% of them favor merging the town with the adjacent town. A sample of 1000 voters is taken and N is the number of them that favor the merger.

(e)) In the age group 60-64 58% have resting systolic blood pressure above 120. In a sample of 125 such people N is the number that have systolic blood pressure above 120.

(f) The incidence rate of breast cancer in females 50-54 is 119.4 per 100,000 and in females 55-59 it is 145.2 per 100,000. N is the number of cases of breast cancer in female 50-60 found in the 1550 women in this age group who visited a large clinic.

9. The population of a certain small town having 10,000 adults is 58% black and 42% white. Twelve people are supposedly picked at random from these adults to serve as jurors. If the number of blacks selected is ≤ 4 would you believe the selection process was random? [Hint. Compute the probability that if 12 people are picked at random then the number of blacks is ≤ 4]

10. According to the 1994 United States Statistical Abstracts 25% of all pregnancies end in abortion. Suppose 5 pregnant women are randomly selected. What is the probability that at least 3 of these women will have an abortion?

11. Suppose 5 cards are selected with replacement from an ordinary deck of cards. (a) What is the expected number of aces? (b) What is the variance of the number of aces?

12. Prevailing medical opinion is that no type of cancer is contagious. Childhood leukemia is a rare form of cancer. Because of the existence of so called “leukemia clusters” in which a large number of cases appear in a small local over a short period of time there is debate on whether this form of cancer is in fact contagious. The most cited of these clusters occurred in a 5.5 year period (late 1956 to mid 1961) in the town of Niles Illinois, a suburb of Chicago. There were 7076 children under the age of 15 during that time period and there were 8 cases leukemia amongst them. The incidence rate for all of Cook county excluding Niles over this 5.5 year period was .000248. (a) How likely is it that

the 8 cases could have happened by chance? Use a binomial probability calculator to compute the probability that that there 8 or more cases amongst 7076 children assuming that the chance is the incidence rate for Cook county. (Ans. .000475). This would tend to argue that it didn't happen by chance. But there is another side of the story. The number of towns like Niles is large. Suppose in the entire United States there are 4500 such towns. (b) Find the probability that at least one of them has a cluster such as Niles assuming the incidence rate for Cook county is the same as that for the nation as a whole. (Ans. .8821) (c) Putting the two facts together what would you conclude? (d) Suppose that in the next 5.5 years there were 7 cases of Leukemia amongst 8024 children under the age of 15 in Niles. Now what would you say?

13. A certain gene can be one of two alleles. There is a population of N genes of which m are allele 1 and N- m are allele 2. A new generation of N genes is formed by randomly selecting N genes with replacement from the initial population and copying them. (a) Find the probability that this new generation has x alleles 1. (b)Find the expected number of alleles 1 in the new generation.

SECTION 3

14. Let Z have the standard normal distribution. Use the normal table to find $P(-1.2 < Z < 2)$.
15. Let X be N(10, 5) distributed. Use the normal table to find $P(X > 18)$.
16. Suppose fasting blood glucose levels have a mean of 90 and a standard deviation of 12. Assume blood glucose levels follow a normal distribution. (a) Find the probability that a person has a fasting blood glucose level > 110. (b) Find the probability that in 10 randomly selected people at least one has a fasting blood glucose >110.
17. The length of life of a light bulb is N(2500, 200) distributed. The manufacture states that the bulb will last longer than t hours. How should t be chosen so that the manufacture is correct 99% of the time?
18. Employees of a certain company are given a test. Those in the top 5% will considered for top management positions. Suppose the scores are normally

distributed with a mean of 100 and standard deviation of 20. What is the minimum score needed for advancement?

SECTION 4

19. (Continuation of Problem 16). Use the normal approximation to the binomial to find the probability that in 350 randomly selected people at least 12 have a fasting blood glucose > 110

20. Dr Wiz noticed that of 14 patients that he saw in the last 5 years who had carcinoma of the tongue 10 of them drank 3 or more cups of coffee per day. He concluded that coffee drinking is a risk factor for this cancer. (a)If 65% of the population drink three or more cups per day of coffee is his conclusion warranted? (b) If 30% of the population drink 3 or more cups of coffee per day is his conclusion warranted.[Use the normal approximation with half integer correction or the binomial calculator to compute the probability that 10 or more of 14 people drink 3 or more cups per day.]

21. According to the 1994 United States Statistical Abstracts 25% of all pregnancies end in abortion. Abortion is officially forbidden for Catholics but many Catholics get abortions anyway. Suppose that in a random sample of 120 pregnant Catholic women 22% ended in abortion. Does it seem that Catholics are having fewer abortions than the nation as a whole? [Hint assume not and compute the probability that in a sample of 120 the observed proportion of abortions $\leq$ .22.]

22. (a) Use the probability calculator to exactly compute the probability that in 30 rolls of a balanced die the number of 1's is exactly 7 (b) . Compute the probability in (a) using the normal approximation with half integer correction.

23. A balanced coin is tossed 100 times. Use the normal approximation to find the probability that there are exactly 50 heads. Use the probability calculator to exactly compute this probability.

24. Suppose 25% of the female population has venereal warts. Find the probability that in a sample of 120 women 24 % or less have venereal warts?

25. Suppose blood clotting time follows a normal distribution with mean 3.6 min. and standard deviation 1 min. (a) Find the probability that a clotting time is

greater than 4.2 min. (b) Find the probability that amongst 100 people not more than 30 have a clotting time greater than 4.2 min.

SECTION 5

26. Suppose two balanced dice are rolled repeatedly. (a) Find the probability that the first 12 occurs at the 10^{th} roll. (b) What is the expected number of rolls needed?
27. Suppose during flu season there is a 10% chance of contracting the flu on contact with any individual. Find the probability that you contract the flu on the 7^{th} individual you come in contact with.
28. Suppose 3 people play the game of “odd man out”. In this game people toss a coin having probability p for heads. If exactly one of the 3 has an outcome that is different than that of the other 2 that player is considered the “odd man out” (a) Find the probability r that on a given trial there is an odd man out. What is this probability for p = 1/2? (b) Find the probability that there are n trials until there is an odd man out. What is this probability for p = 1/2. (c) Find the expected number of trials for an odd man out to occur. What is the expected number for p = 1/2.
29. (continuation) Work parts (a) - (c) of the previous problem if m people play “odd man out”.
30. A coin having probability p for heads is tossed repeatedly. Let T_r be the number of tosses needed to get the r^{th} head. (a) Find the probability that $T_r = r$ (b). Find the probability that $T_r = x$ for x = r, r +1, ... ≥

SECTION 6

31. Assume that total serum protein is N(6.74, .375) distributed. (a) Find the probability that a total serum protein measurement is < 6.5. (b) Find the probability that of 86 determinations of total serum protein from 86 individuals 28 or more were < 6.5. (Ignore the half integer correction.)
32. The diameter X of a cylinder is normally distributed with mean 31.5 cm. and standard deviation .22 cm. The diameter Y of a piston is normally distributed with mean 31.4 cm. and standard deviation .31 cm. If a piston has a larger diameter than a cylinder it must be reworked to fit. Find the probability that a piston must be reworked.

33. The survival time from diagnosis of lung cancer has mean 2.4 years with standard deviation of 1.8 years. In a group of 312 people just diagnosed with lung cancer what is the probability that the average survival time of this group is between 2.3 and 2.5 years?
34. Suppose the resting systolic blood pressure of a healthy person has a mean of 120 and a standard deviation of 10. A diagnosis of hypertension is made if this blood pressure is consistently above 140. (a) Assuming that blood pressures follow a normal distribution what is the probability that a healthy person has a blood pressure > 140 on a single reading? (b) What is the probability that a healthy person would have the average of 10 readings of systolic blood pressure > 140?
35. You toss a coin having probability p for heads. If it falls heads you win $1 and if it falls tails you lose $1. (a) Let X_i be the amount you win or loose at toss i. Find the mean and standard deviation of X_i. (b) The amount of winnings (or losses) after n tosses is $S_n = X_1 + ...+ X_n$. For n = 100 and p = .49 find the approximate probability that you have a gain.
36. (Continuation) How many tosses are needed to have the probability of a gain = .01?

SECTION 7

37. Suppose that 5% of people booking a flight fail to show for the flight. Use the Poisson approximation to evaluate the probability of 110 people booking the flight not more than 5 fail to show for the flight.

    In problems 38 – 42 assume the Poisson process is operative.
38. Suppose the average number of auto accidents on the corner of Wilshire and Westwood is .85 per week. (a) Find the probability that a 4 day period is accident free. (b) Find the probability of exactly one accident in a given week.
39. Raisons are stirred into cake batter so that on average there is 2 raisons per 5cm^3 of batter. A cupcake has 20 cm^3 of batter. What is the expected number of raisons in the cupcake?
40. The length of a chromosome is measured in units called base pairs. Mutations occur along a chromosome at the rate of 2 per 10,000 base pairs. Find the probability that a 35,000 base-pair segment has exactly 4 mutations.

41. The probability that Sam has an angina attack in any given day is .1. Find the probability that in a 12 hour period period Sam has at least one attack.

42. Suppose the average number of bacteria in a $10cm^3$ sample of urine is 30. Five $1cm^3$ vessels are filled with from this sample and incubated with nutrients. If one or more bacteria are present in a $1cm^3$ vessel after incubation a colony will be visible. (a) Find the probability that a $1cm^3$ vessel has a visible colony. (b) Find the probability that at least one of the 5 vessels has a visible colony

SECTION 8

43. Let $f_n = f_{n-1} + f_{n-2}$, n > 1. The initial conditions are $f_0 = f_1 = 1$. The sequence satisfying this difference equation is called the Fibonacci sequence and the terms are the Fibonacci numbers. (a) Find the formula for the n^{th} Fibonacci number. (b) Find $\lim_{n\to\infty}\frac{f_{n+1}}{f_n}$. This is called the golden ratio.

44. (a) Solve the difference equation (2.76) with the initial conditions $f_1 =(1/2)(f_0 + g_0)$.
(b) Find $\lim_{n} f_n$

CHAPTER 3
SIGNIFICANCE TESTS

1 INTRODUCTION

Consider cardiovascular disease, the nations "numero uno" disease. (One of every two deaths is from cardiovascular disease.) What causes cardiovascular disease? If by cause you mean one or two agents that can lead to the disease and in whose absence the disease cannot occur, then the answer is no one knows. For today, even after 50 years of research, no such agents have been found. Instead, what we have is a myriad of variables (called factors) that are associated with the disease. Such things as cigarette smoking, blood pressure, cholesterol, body weight, age, etc. are such variables. These are called risk factors. How do we know that cigarette smoking is a risk factor for cardiovascular disease? Will cessation of smoking reduce or eliminate the risk? How do we know that elevated cholesterol is a risk factor for cardiovascular disease? How do we determine if drug intervention to lower cholesterol protects against cardiovascular disease? Given several different types of drugs that lower cholesterol, which are the most effective? Is there any gain (or loss) to lowering cholesterol by diet over drug therapy?

One of the most challenging problems in medicine is to answer questions of the above type. To do so requires substantial use of probability and statistics. In general, these kinds of problems can be very complicated. In this final chapter we will discuss some very special simple cases.

1.1 SOME EXAMPLES

We will start with some artificial examples to illustrate some of the basic ideas.

EXAMPLE 3.1 (Wiz and Liz)

Consider the following dialog between Dr. Wiz, an oncologist, and Liz, a research assistant somewhat knowledgeable in statistics.

Dr. Wiz: I have developed a new treatment for acute myeloid leukemia that looks quite good. As you know, this disease has a 36% 2 year survival rate. I tried my new treatment on 25 newly diagnosed patients with this disease and 12 of them survived 2 years. Now 12 is 48% of 25 so it looks as if my new treatment is doing some good.

Liz: Not so fast Wiz. It could just be chance fluctuation. There may in fact be no real benefit from your treatment.

Dr Wiz: Chance fluctuation, no benefit, I don't understand what you mean. Anybody can see that 48% is bigger than 36%. The treatment must be doing something.

Liz: Well Wiz, let me try to explain it to you. Suppose we look at 25 newly diagnosed patients with this disease who are not given your treatment. The proportion of these, $\hat{p}$, that survive 2 years is a random variable whose <u>expected value</u> is .36. If we examined many such groups of size 25 then the $\hat{p}$ for each of these groups would not always be .36. Some would be bigger and some would be smaller. This bouncing around in the values is what we call chance fluctuation (around the mean value .36). With or without your treatment, the proportion $\hat{p}$ that survive 2 years is a random quantity. Without your treatment $\hat{p}$ has mean .36. With your treatment $\hat{p}$ has mean p that is unknown to us. (We can think of this p as the proportion of patients with the disease that survives two years under your treatment from some imaginary population of all possible such patients.) Now there are two possibilities for what your treatment may be doing. One, it may have no effect. This is the same as saying $p = .36$. Two, it may have some real effect. This is the same as saying that $p > .36$. What I'm saying is that your observed 49% could either be due to the fact that your treatment has benefit, i. e. $p > .36$ or due to chance fluctuation when in fact $p = .36$.

Dr. Wiz Well LIz, I sort of see what you mean. You're saying that for my sample of 25 my 48% could either be due to the fact that my treatment has some benefit so that the true proportion p that survive with my treatment is > .36 or that in fact it was of no real benefit so that the true proportion was still .36 but that I was just lucky and got that 48% survived. Sort of like getting 8 heads in tossing a balanced coin 10 times.

Liz That's right.

Dr Wiz So, how can I tell which is which? How could I show that $p > .36$?

Liz We cannot proceed directly to show $p > .36$. Instead, we proceed to see if we can say if your result could have happened by chance. We do this by assuming that $p = .36$. We will then see if that hypothesis is at all reasonable. If not, then the only other possibility is that the hypothesis that $p = .36$ is wrong. In that case it must be that your 48% is due to something else than chance. If we can then find no other explanation for the 48% other than your treatment, then we could conclude that your treatment has some benefit. Does that make sense to you?

Dr Wiz It sure does. If I understand you correctly then the situation is analogous to the diagnosis of certain ailments such as IBS in medical practice. We cannot directly show that a patient exhibiting a list of symptoms suggestive of IBS has IBS (irritable bowel syndrome). Instead, we perform tests to rule out diseases that exhibit similar symptoms such as various FB (functional bowel) diseases and draw the conclusion that the patient has IBS from the fact that the patient has none of the other possibilities. O. K, How do we go about "testing" to see if $p > .36$.

Liz: Here is the procedure. We assume that $p = .36$. Let us call this the null-hypothesis.

(1) <u>First we put the measurements on a scale of deviations from the mean in units of standard deviation</u>. Under the null hypothesis $\hat{p}$ has mean .36 and standard deviation $\sqrt{\frac{(.36)(1-.36)}{25}} = .096$. We re-scale $\hat{p}$ by subtracting the mean and dividing by the standard deviation. This yields what is called the z-score

$$Z = \frac{\hat{p} - .36}{\sqrt{\frac{(.36)(1-.36)}{25}}} = \frac{\hat{p} - .36}{.096}. \qquad (3.1)$$

If we now substitute the observed value for $\hat{p}$ (.48 in your case) we obtain the observed value, $z_{(obs)}$, of the Z-score. For your sample of 25

$$z_{(obs)} = \frac{.48 - .36}{.096} = 1.25 .$$

This tells us that your observed z-score is 1.25 units of standard deviation above the mean of .36.

(2) Next, we determine how likely it is to get a value as extreme as $z_{(obs)}$ by pure chance. We do this by computing the so called p-value of the score. The p-value is computed as follows. Under the null hypothesis the z-score is approximately standard normally distributed. The observed magnitude of the z-score is $|z_{(obs)}|$ = 1.25. We now compute the probability under the null hypothesis that the z-score Z has a magnitude at least this large.

P(|Z| > 1.25) = 2(1 - Φ(1.25)) = .2112.

(3) Finally we try to draw a conclusion. In this case the p-value is .1770. What this tells us is that 21% of the time with a sample of size 25 we would see a deviation from the mean of .36 as large or larger than our observed 1.3542 even if the treatment had no effect. In other words, 21% of the time we would get a result as least as good as you got by pure chance when p = .36. Although it may seem to some that 21% is not a very large chance it is certainly not a chance that is extremely small. All scientific investigators would consider it too big to conclude that your result did not happen by chance. They would not consider your 48% as sufficient evidence to conclude that your treatment is effective, i.e., that p > .36. Most investigators would probably take the following point of view. Your results are suggestive that your treatment has some benefit. However, your sample size is probably just too small to reach a definite conclusion that your treatment has benefit.

Dr. Wiz. That is a little disappointing. But what's this about my sample of 25 being too small? Do you mean that if I got that 48% survived 2 years but my sample was larger than 25 I might have definitely been able to conclude that my treatment had benefit?

Liz Yes, that might have been the case. Suppose that instead of doing your trials on 25 subjects you did them on 100 subjects. Let's say that you get the same percentage survive 2 years as before, namely 48%. Now under the null hypothesis the observed proportion $\hat{p}$ of those that survived 2 years would again be a random variable whose mean is .36 but the standard deviation would now be $\sqrt{\frac{(.36)(1-.36)}{100}} = .048$. The z-score would now become $Z = \frac{\hat{p} - .36}{.048}$ and the observed z-score, $z_{(obs)} = 2.0833$.

Once again using the normal approximation we find that the p-value of the score is now P(|Z| > 2.0833) = .0372. This is a fairly small probability. In this case, unlike the case for 25 subjects, there is only a 3% chance that you could have gotten that 48% or more survived 2 years if in fact your treatment had no effect. If now you had used 1000 subjects and got the same 48% result, then your observed z-score would be 6.5881. The p-value of this score is 0. So, this rules out that it could have happened by chance. What's left then is only the conclusion that $p \neq .36$.

EXAMPLE 3.2

This example too is artificial but it is based upon how the relationship between aspirin and cardiovascular disease first came to light.

In the summer of 1982 Dr. Observant, a pain specialist at a large California HMO, was looking over the records of the 7246 patients, age 65^+, with longstanding chronic pain that were in the plan in 1981. Some of them of course had died during that year and Observant was curious as to their cause of death. He discovered that 75 of them died of cardiovascular disease. This struck him as curious. The number seemed low. He looked up the death rate for cardiovascular disease in 1981 for people 65 or older and found that it was 1842.5 per 100,000 (in proportions .0148). From his data the proportion in the 65^+ people with chronic pain is 75/7246 = .0104. Dr. observant then concluded that for some reason persons with chronic pain had less Cardiovascular disease. Is Dr. Observant correct in his assessment?

To answer the question we proceed along the same lines as Liz did in the last example. Let p_0 = 1842.5/100,000 = .0184 be the proportion of 65^+ people that die of cardiovascular disease and let p be the proportion of people with chronic pain that die of cardiovascular disease. The value of p is unknown. Dr Observant's contention is that $p < p_0$. Observant's evidence for this is the fact that his observed proportion of deaths due to cardiovascular disease is smaller than p_0 . What we must decide is if Observant's small value could have happened by chance under the null hypothesis that there is no difference between the cardiovascular disease rate in the chronic pain group and the population as a whole, i.e. if $p = p_0$. To do so we let $\hat{p}$ be the proportion of deaths due to cardiovascular disease and form the Z-score

$$Z = \frac{\hat{p} - .0148}{\sqrt{\frac{(.0148)(1 - .0148)}{7246}}} \qquad (3.2)$$

Under the null hypothesis Z has an approximate standard normal distribution. The observed value of $\hat{p}$ is .0104. Substituting that value into (3.2)gives the observed value of the z-score, $z_{(obs)}$ = -3.1018 The magnitude of this observed score is |-3.1018| = 3.1018. The p-value for this score is the probability under the null hypothesis that we get a value at least as extreme as the one we observed, i. e. $P(|Z| > 3.1018)$.= .0019. This is indeed a very small probability. We conclude that it did not happen by chance. What's left is only the conclusion that $p \neq p_0$. Thus it seems that Dr. Observant is on to something.

2 SIGNIFICANCE TESTS FOR A SINGLE PROPORTION

The last two examples are examples of what is known in statistics as significance tests. More specifically, they are examples of significance tests for a single proportion. We will now discuss the general procedure for doing significance tests for a single proportion. This consists of 5 parts, viz. structure, problem, data, test,conclusion

1. STRUCTURE:

The proportion of elements in a certain population, which we will call the control population, having a specific property, is p_0.

The proportion of elements in another population, which we will call the treatment population, having the specific property is p.

We know p_0 but we do not know p.

Before proceeding further let us see how Example 3.2 fits into the general scheme given above. For that example, the control population is the entire US population 65^+ and $p_0 = .0148$. The treatment population is an imaginary population of people 65^+ with chronic pain in the US.

2 PROBLEM

To decide if there sufficient evidence to conclude that the treatment and control populations are different. The null hypothesis that there is no difference. That is, $p = p_0$. We emphasize the following point. The purpose of our investigation is not to show that the null hypothesis is true but rather that it is false. The null hypothesis is put up as a paper tiger to be knocked down. We hope that the data will show that it is not reasonable to conclude that the null hypothesis is true. Suppose that the data does not show this is the case. In that case we cannot claim that we have "proved" the null hypothesis holds. All we can claim is that the data does not allow us to conclude that it does not.

3.DATA

The proportion of elements $\hat{p}_n$ in a random sample of size n from the treatment population that have the specific property. $\hat{p}_n$ is a random variable. For a particular sample of size n, $\hat{p}_n(\text{obs})$ is the actual value that the random variable $\hat{p}_n$ has for that sample.

4 TEST

The test is based on the test statistic $\hat{p}_n$. We form the score

$$Z = \frac{\hat{p}_n - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}. \qquad (3.3)$$

Under the null hypothesis Z is approximately standard normally distributed. The quantity in (3.3)is called the z-score. The observed value of the z-score is z(obs) where

$$z(\text{obs}) = \frac{\hat{p}_n(\text{obs}) - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

The p-value of the observed score is the probability of observing a score at least as extreme as z(obs). Using the normal approximation

$$\text{p-value of } z(\text{obs}) \approx 2(1 - \Phi(|z(\text{obs})|)\,. \qquad (3.4)$$

5 CONCLUSIONS

Can we explain our result as being due to pure chance under the assumption that p is the same as p_0? To answer this question 2 we consider the p-value of the observed z-score which is the probability of getting a result as least as extreme as the one we observed z-score assuming $p = p_0$. The smaller this p-value is the less likely it is that our observed result could have happened by chance.

There are then two courses of action open to us.

We decide that the p-value is sufficiently large so we are willing to say that our result could have happened by chance. In this case we say we <u>accept the null hypothesis</u>. As pointed out above, this does not mean that we actually believe the null hypothesis to be true nor does it mean there is no real difference between p and p_0. It only means that we do not have sufficient evidence to say they are not the same. Often we refer to this situation as saying there is no (statistically) <u>significant difference</u> between p and p_0.

We decide that the p-value is so small that we are unwilling to say our result could have happened by chance. In this case we say we <u>reject the null hypothesis</u>. Well then, if it could not have happened by chance under the assumption that $p = p_0$ how then did it happen? There is only one possibility, $p < p_0$ or $p < p_0$. Often we refer to this situation by saying there is a (statistically) significant difference between p and p_0.

So, whether a result is statistically significant or not depends on the p-value. This of course immediately raises the question of how small must be the p-value be to say the result is statistically significant. There is no universal answer to that question. It depends on how certain you want to be that it could not have happened by chance. The more certain you want to be the smaller will be the p-value you will require. The cutoff value is called the significance level (or size). The most widely used are of .05 and .01. If the significance level is taken to be .05 then a p-value $\leq$ .05 would be taken as sufficient evidence for the result to be statistically significant while a p-value > .05 would be considered as insufficient evidence. On the other hand, if the significance level is taken to be .01, then a p-value of say .0226 would be considered as insufficient evidence for rejection, but this same p-value would suffice for rejection if the significance level was taken to be .05.

In practice, the significance levels are often set by the editorial policy of the journal in which a particular article is to be published. Most of them will not publish a result as significant if the p-value is above .05 and some even insist on a significance level of .01. In almost all articles either the actual p-value is reported or it is reported as "p-value < α", where α is some specific value such as .05, .01.

EXAMPLE 3.3

This is an artificial example. Past experience shows that 69% of people with moderate arthritic pain claim to experience substantial relief with aspirin. A new pain remedy is developed and it is tried on 122 subjects with such pain. Of these 92 claim to have substantial relief. Can you conclude that the new remedy is superior to aspirin?

SOLUTION

The null hypothesis here is that the proportion p to experience relief with the new remedy is the same as the old, i.e. p = .69. The observed proportion in the sample of 122 is $\hat{p}(\text{obs}) = 92/122 = .75$. The observed z-score is z(obs) = $\frac{.75 - .69}{\sqrt{\frac{(.69)(.31)}{122}}} = 1.43$. The p-value of this score is $2(1 - \Phi(1.43)) = .1528$. The result is not significant. There is not enough evidence to conclude that the new remedy is better than aspirin.

EXAMPLE 3.4

Although this example is based on data from the first study of breast cancer screenings (The HIP trials) and real breast cancer incidence data from SEER it is not how the study was done. (We will examine the real study a little later).

Our purpose in this example is twofold. First, to provide another illustration of how a test of a single proportion is done. Second, to introduce for the first time some of the pitfalls of significance tests.

Breast cancer is the second leading cause of cancer death in women. If this cancer is caught when it is still localized in the breast it is much more curable than if not. In an attempt to detect the cancer early, starting in 1963, a large New York (HMO) type medical provider offered about half of its female members age 40-64 the option of annual screening for breast cancer. Each woman who agreed to the screening was examined both by mammography and manual breast examination. Of those offered the screening 20,200 agreed to the screenings. After 5 years into the study the plan administrator wanted to know if the additional expense of these screenings was worthwhile. That is, he wanted to know if there is reduced mortality from breast cancer due to screening? The administrator noted that the death rate for all women under 65 in the US was essentially the same for each of the 5 years. Using the age specific death rates in the table below and the age distribution of population of women in the US he computes that the death rate for women 40-64 was 49.6 per 100,000 per year.

age	40-44	45-49	50-54	55-59	60-64
Death rate per 100,000	23	37	53.8	70.4	87

He therefore took the death rate for the 5 year period to be 5 times this death rate. Thus he took p_0 = 247.8/100,000 = .0025. During this 5 year period 23 of the women in the study died of breast cancer. Does the screening work?

SOLUTION

Let p be the proportion of deaths due to breast cancer over the 5 year period for a hypothetical population of all women given screening. The null hypothesis is that p = .0025. The observed proportion of deaths in the sample is $\hat{p}(obs)$ =

23/20,200 = .0011. The observed z-score is z(obs) = $\dfrac{.0011 - .0025}{\sqrt{\dfrac{(.0025)(.9975}{20,200}}}$ = -3.98.

The p-value of this score is .00003.

There is no way this could happen by chance. We can definitely conclude that the result is highly statistically significant. For sure, $p \neq p_0$. But can we conclude that the screening works? That is quite another matter. All things being equal the answer is yes. But all things can be far from equal. As we will see that is the case here. Read on.

Discussion of Example 3.4

As we will see a little later, this study, as a way of deciding if screening reduces mortality from the cancer, is very poor. There are much better ways in which such a study should be done. The central problem in comparing the treatment group to the control group (in any study, not just this one) is homogeneity. The women in the two groups must not differ in any in any factor other than the treatment factor(screening or no screening) that could effect whether or not they die from breast cancer. That is what we mean by all things being equal. Why is homogeneity so important? Let see what happens if it fails. In that case there are one or more factors that we are unaware of that influence the outcome (here death from breast cancer) that are different in the treatment and control groups Such factors are called confounding factors. The effects of these confounding factors gets all mixed up with the effect of our factor screening. In comparing the two groups we cannot tell if the result we got is due to our factor screening or to the confounding factors. So, unless we are sure that confounding factors are not present, we cannot conclude that the difference we saw is due to our factor screening.

Unfortunately, the way this study was done there are likely to be many confounding factors. Let me point out just one and show you what kind of influence it might have. The confounding factor is age. Breast cancer death rates increase rapidly with age. If the age distribution of the women in the treatment group differ substantially from that of the nation as a whole as a whole that formed the control group then we have a serious problem.

Is there reason to believe that the women in the treatment group have a different age distribution than the nation as a whole. Yes! In 1963 most people did not belong to HMO's. Those that did tended either to be working women or wives of working men. As such, they would probably tend to be younger than women in the nation as a whole. Let us see what effect this might have. Suppose that the women in the study had an age distribution as follows:

age	40-44	45-49	50-54	55-59	60-64
percent	60	35	3	1.5	.5

Then the administrator should not have used the age distribution for the nation as a whole to compute p_0. He should have used the above age distribution. Based on the national age specific rates and the above age distribution the probability that a women would die of breast cancer in a single year is

(23/100,000)(.6) + (37/100,000)(.35) + (53.8/100,000)(.03) + (70.4/100,000)(.015) + (87/100,000)(.005) = .000299.

Over the 5 year period the probability is (5)(.000299) = .0015. The administrator should have taken this for p_0.Let us now redo the significance test using this p_0. The observed z-score is now

$$z(obs) = \frac{.0011 - .0015}{\sqrt{\frac{(.0015)(.9985)}{20,200}}} = -147 .$$

The p-value of this score is .1416 which is not significant.

Examining the results of the two calculations we see the confounding factor of age drastically altered the results. The first calculation did not take age into account. The second, in the parlance of statistics, "corrected" for it. The first calculation showed the result was highly significant. That was true, The result didn't happen by chance. But, as the second calculation shows, it didn't happen because the screening saved lives. It seems to have happened because the women in the study were substantially younger than the women in the US as a whole.

3 COMPARATIVE STUDIES

We will discuss some of the principles involved in the following basic problem:

To determine if a treatment for a specific disease is beneficial or if a factor is a risk factor for a specific disease.

To begin the discussion let us suppose we want to know if a certain treatment is beneficial. Available to us are the results of trying the treatment on n subjects. These subjects constitute the treatment group. We pretend that the treatment group is a sample of significance level n from a hypothetical population of perhaps infinitely many individuals that are given the treatment. In general, the treatment does not work 100% of the time. So, in order to see if the treatment is beneficial or not we must compare the results we get from the treatment group with something else.

We have already discussed one method of doing this comparison. We compare our result with the corresponding data from the population as a whole. In general, this is a very poor way to decide if some specific treatment is effective or if some factor is a risk factor for a disease. With this procedure there are usually just too many possible confounding factors to reach a sensible conclusion.

Why are confounding factors so bad? Because they produce bias. What is bias? Bias is a systematic error in the study that tends to favor one outcome over the other. For example, in the breast cancer screening, the age factor biased (i.e. prejudiced) the study toward showing the breast cancer screening was effective.

A much better method than comparing against a known standard is to take another sample of subjects who are not given the treatment or who do not have the risk factor of interest. Hopefully, these can be more evenly matched with those in the treatment group. We call this group the control group. Analogous to the treatment group we consider the control group to be a sample from a hypothetical population of all individuals who are not given the treatment or who do not posses the risk factor. In the ideal case, the subjects in both groups would be equivalent in all respects except for the treatment either being given or not given or the risk factor either being present or absent. In reality that can never be achieved but the closer we can come to that situation the better.

3.1 STUDY DESIGNS

There are two basically different types of studies. These are known as observational studies and controlled experiments. The crucial difference between the two is who decides which subjects will be in the treatment group and in the control group.

In an observational study the subjects themselves decide which group they will be in. The investigator merely observes what happens. In contrast, in a controlled experiment the investigator assigns a given subject to one of the two groups. Although at first glance the difference between the two studies may seem minor that is not at all the case. Despite how carefully an observational study may be carried out the possibility of bias can not be eliminated. Because of this, results of observational studies are usually taken as suggestive and not as "proof" that a treatment is beneficial or that a factor is a risk factor. On the other hand, a controlled experiment, if correctly executed, can minimize the possibility of bias, and the conclusions drawn from them can be considered as "proof".

From the above discussion one might think that observational studies are unimportant. This is not at all true. An observational study is usually where one begins the investigation of a treatment or a risk factor. If these show that we are on the right track then one considers running a controlled experiment (if such an experiment is possible) to prove the issue. Not all things can be made into a controlled experiment. Suppose we want to know if smoking a pack or more a day of cigarettes for 10 years or more is a major cause of lung cancer. To study the issue as a controlled experiment we would need a large group of people who never smoked who would agree, as decreed by the investigator, to smoke or not smoke the required pack a day for 10 years. That is of course impossible.

4 SIGNIFICANT TESTS FOR TWO PROPORTIONS

4.1 GENERAL SETTING

No matter which way a study is carried out we are eventually going to compare the results from the treatment group with those from the control group.

In general terms, there are two hypothetical populations that we can call the treatment and control populations respectively. The proportion of elements in the treatment population having some specific property is p^t and the proportion of elements in the control population having this property is p^c. Both of these proportions are unknown. Our data consists of n observations from the treatment population and m observations from the control population. These observations constitute the treatment group and the control group respectively. The proportion of elements in the treatment group that have the specific property is $\hat{p}_n^t$ and the proportion of elements from the control group having the property is $\hat{p}_m^c$. We consider these quantities as estimates of the quantities p^t and p^c respectively. We recall from Chapter II that $E\hat{p}_n^t = p^t$ and the standard deviation $\hat{p}_n^t$ is $\sqrt{\frac{p^t(1-p^t)}{n}}$. The estimate of this standard deviation, which is obtained by replacing p^t by $\hat{p}_n^t$ in the formula for the standard deviation is called the <u>standard error</u> of $\hat{p}_n^t$. This is denoted $SE(\hat{p}_n^t)$ That is

$$SE(\hat{p}_n^t) = \sqrt{\frac{\hat{p}_n^t(1-\hat{p}_n^t)}{n}}.$$

Similarly,

$$SE(\hat{p}_m^c) = \sqrt{\frac{\hat{p}_m^c(1-\hat{p}_m^c)}{m}}$$

Often we present our results as a table.

	sample size	estimate	SE
treatment	n	$\hat{p}_n^t$	$SE(\hat{p}_n^t)$
control	m	$\hat{p}_m^c$	$SE(\hat{p}_m^c)$

The general question to be answered is if an observed difference between $\hat{p}_n^t$ and $\hat{p}_m^c$ is sufficiently large to allow us to conclude that $p^t \neq p^c$ or is the difference just due to chance. As in the case of a single proportion, we play devil's advocate and assume there is no real difference. The <u>null hypothesis is that the two proportions are the same, i. e. $p^t - p^c = 0$.</u> As in the previous case, we hope our evidence will allow us to refute this hypothesis.

Before proceeding further let us consider an example.

EXAMPLE 3.5

Does lowering cholesterol prevent heart attacks? The Helsinki heart study examined this question. Middle aged men were randomly assigned to either receive the drug gemfibrozol or an inert substance (such an inert substance is called a placebo). In all 2051 received the drug and 2030 received the placebo. After 5 years there were 56 men in the drug group that had heart attacks and there were 84 men in the placebo group that had heart attacks.

(a) Is this an observational study or a controlled experiment?

(b) What are the treatment and control groups.

(c) What are n ,m, $\hat{p}^t_n$, and $\hat{p}^c_m$?

SOLUTION

controlled experiment

Those getting the gemfibrizol form the treatment group. The others form the control group.

n = 2051, m = 2030, $\hat{p}^t_n$ = 56/2051 = .0273, $\hat{p}^c_m$ = 84/2030 = .0414

4.2 REAL POPULATIONS

Sometimes we want to decide if there is a difference between two real (rather than hypothetical) populations. We do this by simply thinking of the treatment population as one of the populations and the control population as the other population. For example, suppose we wanted to know if there was a difference in the incidence of prostate cancer between whites and blacks. We would say take the treatment population to be the white population and the control population to be the black population. Then p^c would be the proportion of new cases of prostate cancer in blacks in a given year and p^t would be the proportion of new cases of prostate cancer in whites in a given year.

4.3 MATHEMATICAL MODEL

As in the case of a single proportion, to proceed further we need a probability model. The model will enable us to determine a test statistic, a score, and the distribution of the score under the null hypothesis. We can then use the distribution of the score to compute the p-value for an observed value of this score to see if we can attribute our observed result to chance or not.

The model we are going to build crucially depends on the following:

FUNDAMENTAL ASSUMPTION:
The observations in each group are independent of each other.

There are two ways to compare p^t with p^c. We can consider their difference p^t - p^c or the ratio p^t/p^c. Both are widely used. We will first consider the difference.

The test statistic we are going to use is the estimate of the difference $\hat{p}^t_n - \hat{p}^c_m$. We will figure out what the distribution of this difference should be when the null hypothesis is true and we will use it as a basis for a test.

The normal approximation for proportions tells us that $\hat{p}^t_n$ is approximately $N\left(p^t, \sqrt{\frac{p^t(1-p^t)}{n}}\right)$ distributed and $\hat{p}^c_m$ is approximately $N\left(p^c, \sqrt{\frac{p^c(1-p^c)}{m}}\right)$ distributed.

Under the fundamental assumption the random variables $\hat{p}^t_n$ and $\hat{p}^c_m$ are independent of each other. Consequently, by properties of the normal distribution, it should be that $((\hat{p}^t_n - \hat{p}^c_m) - (p^t - p^c))$ is approximately normally distributed with mean 0 and standard deviation SD given by

$$SD = \sqrt{\frac{p^t(1-p^t)}{n} + \frac{p^c(1-p^c)}{m}}\ . \qquad (3.5)$$

The null hypothesis says that $p^t = p^c$. Call the common (but unknown) value of these two proportions p. Then the SD given by (3.5) becomes

$$SD = \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)p(1-p)}\ . \qquad (3.6)$$

Under the null hypothesis this quantity $(\hat{p}^t_n - \hat{p}^c_m)/SD$ is approximately standard normally distributed. However, this ratio cannot be used as the basis of a test because p is unknown. To estimate p we use the following estimator, known as the pooled estimator. The null hypothesis asserts that the success probabilities in the 2 populations are the same. Thus to estimate the common p we should use

$$\hat{p} = \frac{\#(\text{successes treatment group}) + \#(\text{successes control group})}{n+m}$$

Now #(successes treatment group) = $n\hat{p}_n^t$ and #(successes control group) = $m\hat{p}_m^c$. Thus the pooled estimate is

$$\hat{p} = \frac{n\hat{p}_n^t + m\hat{p}_m^c}{n+m} \qquad (3.7)$$

The standard error (SE) of the estimator of the difference $\hat{p}_n^t$ - $\hat{p}_m^c$ is the estimate of the SD obtained by replacing p by $\hat{p}$ in (3.6). It then can be demonstrated that the z-score

$$Z = \frac{\hat{p}_n^t - \hat{p}_m^c}{SE_h} \qquad (3.8)$$

is approximately standard normally distributed.

EXAMPLE 3.6

What is z(obs) for the cholesterol study in Example 3.5 and its p-value? Would we accept or reject the null hypothesis with .05 as the significance level? Would we accept or reject with .01 as the significance level?

SOLUTION

We summarize the data in Example 3.5 as follows.

Sample size	estimate	SE
treatment	2051	0.0273
control	2030	0.0414

The pooled estimator $\hat{p}$ = (2051)(0.0273) + (2030)(0.0414) = 0.0343. The SE = $\sqrt{\left(\frac{1}{2051}+\frac{1}{2030}\right)(0.0343)(0.9657)}$ = 0.0057.

The z-score is $\frac{0.0273 - 0.0414}{0.0057}$ = -2.47

The p -value of this score is $2[1 - \Phi(2.47)] = 0.014$. Thus we would reject (but barely) at the 0.01 level.

EXAMPLE 3.7

In 1976 Bucher et al examined the racial differences in the occurrence of ABO hemolytic disease by examining the records of infants born at North Carolina Memorial Hospital. Here are the results of that study.

	Sample size	estimate	SE
Black babies	3584	.0112	.001758
White babies	3831	.0044	.001069

(a) Find $\hat{p}$

(b) Find SE

Find the observed z-score

Conclusion

SOLUTION

(a) $\hat{p}$ = [(3584)(.0112) + (3831)(.0044)]/3584 + 3831 = .0081.

(b) $SE_h = \sqrt{\left(\frac{1}{3584} + \frac{1}{3831}\right)(.0081)(1 - .0081)} = .0021$

(c) $z_h(\text{obs}) = .0081/.0021 = 3.89$

(d) The p-value is 0.00010. There is no chance the observed difference could happen by chance. There seems to be a real difference between the races with respect to this disease with it being much more prevalent amongst blacks.

5 CONTROLLED EXPERIMENTS

Problem: To determine if a treatment for a specific disease is effective.

In a controlled experiment the investigator assigns the subjects to either receive or not receive the treatment.

How should the subjects be assigned to the two groups? The best way to do this is by random choice. This is done as follows. The investigator decides that n subjects will be in the treatment group and m will be in the control group. The subjects are numbered 1 to n + m and n + m tickets numbered 1 to n + m are placed in a box. Tickets are selected at random from the box. The subjects corresponding to the first n tickets form the treatment group and the remaining m subjects form the control group. For example, suppose n = 300 and m = 200. If the first ticket drawn from the box has number 423 and Jones is subject number 423 then Jones is assigned to the treatment group.

Why is random choice so good such a good way to proceed? Because lady luck is blind. If the sample sizes n and m are large enough randomization will equal out the differences for those factors that could be confounding factors for the study. For example, smoking is a factor for cardiovascular disease. Suppose we didn't know this and the study is to evaluate a drug that is to prevent heart attacks. The subjects in the study are both smokers and non-smokers. By using randomization for large enough sample sizes the proportion of smokers in the two groups will be about the same. A controlled experiment in which the subjects are assigned at random to the two groups is called a randomized controlled experiment.

Randomization will tend to minimize bias due to confounding factors but it will not eliminate bias due to other sources. One major other source are investigator or patient prejudices. There is a well-known effect in medicine called the placebo effect. This is the phenomenon that if one gives a treatment for disease, which can have no effect on the disease, some people will gain a benefit from that treatment. For example, the administration of a pill made of a non-digestible starch will give pain relief to some people with pain if they are told the pill will relieve pain. To minimize the placebo effect the control group is given a placebo (a substance that resembles the treatment substance but has no physiological effect.) whenever this is possible. To further minimize bias, the experiment is usually carried out so that neither the subject nor the attending physician knows if the subject is getting the real treatment substance or the placebo. This procedure is called double blinding. An experiment carried out so that neither the patient nor physician knows what's being given is said to be double blinded. If the experiment is also randomized then it is called a randomized double blind experiment. This kind of controlled experiment, if carried out on a large enough number of subjects, is considered to be the definitive method to establish if a treatment is effective or not.

EXAMPLE 3.8 (Salk Vaccine Trials, Method 1)

Although Polio was not a rare disease, it was in fact never a common disease. However, because it often had very debilitating effects on young children (and perhaps because President Franklin Roosevelt had the disease) a major effort was undertaken in the late 40's to 50's to find its cause and cure. The cause was found to be a virus. No cure has ever been found, but vaccines to protect against ever getting the disease have been developed. The first such vaccine, Salk Polio vaccine, was to be put to the test in 1954. No vaccine will give immunity 100% of the time. Thus we would expect some cases of polio to occur amongst those that are given the vaccine. Several different trials were proposed to test the vaccine. Here is one of them.

A vaccine (or for that matter any medical treatment) cannot be given to a child unless the parents consent to it being given. A large number of families were contacted to be in the trials. They were told that their child would either receive the vaccine or a placebo. Some of the parents consented to their child being in the trials and others refused. Of those that consented, 400,000 children were randomly selected to be in the trials and these were randomly divided into two groups of 200,000 each. One group got the vaccine and the other the placebo. The procedure was carried out in a double blind manner. Here are the results. Also reported are the results for 350,000 children in the refused group.

	Sample size	Rate per 100,000
vaccinated	200,000	28
placebo	200,000	71
refused	350,000	46

(a) Find $\hat{p}$

(b) Find SE

(c) Find the observed z-score

(d) Conclusion

SOLUTION

(a) ((200,000)(28/100,000) + (200,000)(71/100,000))/400,000 = .000495.

(b) SE = $\sqrt{\left(\frac{2}{200,000}\right)(.000495)(1-.000495)}$ = .000070.

(c) z(obs) = [(28/100,000) - (71/100,000)]/.000070 = - 6.1428.

(d) The p-value is 8×10^{-10} is astronomically small. Since the result is coming from a very good controlled experiment we feel confident in saying that the difference observed between the two groups is due to the vaccine. In short, the vaccine works!

EXAMPLE 3.9.3.10 (Salk Vaccine Trials{ XE "Salk Vaccine Trials" }, Method 2)

As mentioned above the randomized double blind trial discussed in the last example was not the only test of the vaccine. The NFIP (National Foundation for Infantile Paralysis) proceeded to test the vaccine as follows. They too recruited a large number of families for the trials but limited their tests to children most vulnerable for the disease, namely those in grades 1 to 3. They offered the vaccine to all those families having a child in grade 2 and used the children in grades 1 and 3 as the control group. Those offered the vaccine divided themselves into two groups, namely those who accepted the vaccine for their child and those who refused to allow their child to be vaccinated.

Here are the results of the NFIP trials.

	Sample size	Rate per 100,000
Grade 2 vaccinated	225,000	25
Grade 1 & 3 control	725,000	54
Grade 2 refused	125,000	44

Let us now listen in to a dialogue between a medical doctor Jon and a research assistant Don versed in statistics.

Jon Why were two large scale trials of the Salk Vaccine carried out at the same time?

Don As prestigious as the NFIP was at the time, many knowledgeable people recognized that the method that the NFIP wanted to use had serious flaws. Since the NFIP could not be dissuaded the randomized double blind trials discussed in our previous example were also undertaken.

Jon What's wrong with the NFIP method?

Don There are many things wrong. Here is one point. Picking the treatment and control groups from different populations is a poor idea.

Jon Well, why is that? They were all children and not so different in age.

Don Polio is a contagious disease. It's prevalence and rate of spread amongst 1st and 3rd graders could be quite different than amongst 2nd graders. If it were more prevalent and /or the spread rate is higher amongst the 2nd graders then the incidence rate amongst them would be higher. That would tend to bias the results to showing the vaccine was not effective. The revere bias would hold if these factors were less amongst the 2nd graders. Thus unless these factors are the same in both populations the study will be biased.

Jon How should the results of the NFIP trials be analyzed? I suppose we should compare the rate in the vaccinated group with the rate in the control group?

Don At first glance that's what most people not trained in statistics would say. However, if we do that then we cease to have a controlled experiment.

Jon We do? Why is that?

Don The investigators determined that the treatment group was to be the 2nd graders and the control group was to be the 1st and 3rd grades. But within the treatment group it was the <u>parents not the investigators</u> that determined who got vaccinated. So, if we compare the vaccinated group with the control group we have an observational study and not a controlled experiment.

Jon Hmm, o.k., I see that now. But so what, these are only names. Is there anything really wrong using this comparison?

Don Yes.

Jon what's the problem?

Don Suppose the 1st and 3rd graders were offered the vaccine. The parents of the children of these grades would again divide into those that would accept the vaccine and those that would refuse it. Thus the control population is a mixed population of those children whose parents who would allow the vaccine to be used and those that would not. The treatment population is composed solely of children whose parents allow the vaccine to be used. It is reasonable to believe that the children in the refused population might differ in many other factors other than the vaccine being given. These factors could be confounding factors and therefor produce bias. In short the control and the vaccinated populations are not equivalent.

Jon So should we compare the vaccinated with the refused group?

Don That too would be an observational study. It will suffer the same defects as your first suggestion.

Jon Well, what should we do?

Don We should compare the results from the entire treatment group, i.e. both the vaccinated and the refused, with the control group. That way we have a controlled experiment although it is neither randomized nor blinded.

Jon It seems strange. that we put in the refused group results.

Don Remember, the control group has both types of children. Putting in the refused helps balance out that fact.

Jon I get it. So, how do we get the result for the entire treatment group from the data given? What does the NFIP trial show?

Don In total the treatment group has 350,000 subjects. The proportion of cases of polio in this entire group is (225,000)(25/100,000) + (125,000)(44/100,000) = .000318.

	Sample size	proportion
Grade 2 treatment	350,000	.00032
Grades 1 & 3 control	725,000	.00054

The observed z-score is -4.99. This has p-value .000001. Needless to say we can conclude this did not happen by chance. Despite the possible sources of bias in the NFIP study we probably can also conclude that the difference was due to the vaccine.

Jon So our apprehensions about the NFIP trials were unwarranted.

Don Not really. The sources of bias just weren't sufficiently great to reverse our conclusions as in the breast cancer screening example. However, if we compare the results from the randomized trials with those from the NFIP study we see that the NFIP study was biased against the vaccine. To see this, note that the ratio of polio control group to those in the treatment group for the NFIP study was 1.69 while this ratio for the randomized trials was 2.54.

EXAMPLE 3.11 (KAPS)

Cholesterol has been considered a risk factor for cardiovascular disease since the mid 50's. Methods for reducing cholesterol include diet and drugs. There are various classes of drugs. The newest (and most effective) are a class of drugs called Statin drugs. The question is if intervention with statin drugs is effective for cardiovascular disease. There are several endpoint measures that one could use to judge this. One could consider, (i) incidence of the disease, (ii) mortality from the disease, (iii) total mortality. Many randomized double blind trials have been carried out (and are still being undertaken) to answer this question. At issue with any drug therapy is does the drug increase cancer risk So, often cancer data is gathered with the other variables in the trial.

One such study concluded in 1995 was Kaps (Kuopio atherosclerosis prevention study). This study was intended as a primary prevention study although 8% of the participants had a previous history of myocardial infarction (i.e. heart attack). The study consisted 447 men who were selected because they had elevated cholesterol (>240).The mean age of the men was 57 and the mean follow-up time was for 3 years. At entry, the average cholesterol level of the participants was 258. In the treated group the average percent change was -20%. In the control group the average percent change was -1%. Here are the results.

	treatment			control		
	size	Estimate	SE	size	estimate	SE
CVD death	224	.0089	.0063	223	.0135	.0077
All death	224	.0134	.0077	223	.0179	.0089
CHD incidence	224	.0223	.0099	223	.0359	.0125
Cancer incidence	224	.0134	.0077	223	.0224	.0099

CVD = cardiovascular disease, CHD = coronary heart disease

(a) Use the non- pooled method to fill in the missing table entries.

	Estimate difference	SE(difference)	score	p-value
CVD death	-.0046	.00995	-.46	.6456
All death				
CHD incidence	-.0136	.01594	-.86	.3990
Cancer incidence	-.0090	.01254	-.72	.4716

(b) What are your conclusions about this drug?

SOLUTION

Estimate difference = .0134 - 0179 = -.0045, SE(diff) = $\sqrt{(.0077)^2 + (.0089)^2}$ = .0118, score = -.0045/.0118 = .38, p- value = .7040

(b) None of the p-values are anywhere near small enough to eliminate chance as the explanation for the difference. As far as this study shows there is no difference between using the drug or not. In short the drug seems to do nothing. However this is very small study done over a short period of time. There were not many incidences of heart disease or deaths in the entire study.

EXAMPLE 3.12 (CARE)

The CARE (Cholesterol and Recurrent Events) study was another test of cholesterol intervention. This study, which ended in 1996, also used provastatin. Unlike KAPS, CARE was intended as a secondary prevention trial. It consisted 100% of people who had already had a heart attack. Elevated cholesterol was not a selection criteria. In this study there were 4159 subjects of which 14% were women. Their average age at entry was 59 and the median follow up time was 5 years. Initially the average cholesterol of the participants was 209 The percent change in the treated group compared to the control group was -20%. Below is a summary of machine output using the pooled method.

	treatment		control		pooled		
	size	estimate	size	estimate	estimate	Z score	p-value
All death	2081	.0865	2078	.0943	.0904	-.88	.3789
CVD death	2081	.0538	2078	.0626	.0582	-1.20	.2226
CHD incidence	2081	.1019	2078	.1319	.1169	-3.01	.0026
Cancer incidence	2081	.0827	2078	.0775	.0801	.61	.5387

For all deaths, what is the estimate of the difference?

For all deaths, what is the SE of this estimate under the null hypothesis?

Conclusions

SOLUTION

Estimate of difference = .0865 - .0943 = -.0078.

We want SE. Since z = difference/ SE we have -.88 = -.0078/SE. Therefore SE = -.0078/-.88 = .0089.

There is no significant difference in either overall death or CVD death between the treatment and the control group. This says that the drug does not seem extend life. There is a highly significant difference with respect to heart attack incidence. The drug does seem to prevent heart attacks. These results are puzzling and a bit paradoxical. Of importance is the fact that this group was not at risk due to elevated cholesterol. Nevertheless, the drug seems to have played a significant role in preventing heart attacks. This suggests that perhaps it is the drug itself that is doing the prevention and not the fact that it reduced cholesterol.

PROBLEMS CHAPTER 3

In problems 1 - 3 state the null hypothesis, compute the z-score and its p-value and tell if you would reject at significance level .05, at significance level .01.

1 A coin is tossed 25 times and falls heads 16 times. Is the coin unbalanced?

2 Dr. Hocum claims that 80% of his patients never have significant arthritic pain after completing a 6 month course of treatment at his clinic. In a sample of 75 of his former patients, 45 said they are now in pain. What would you say of Hocum's claim?

3 In September 1994 USA today reported that in a telephone survey of 1022 randomly selected adults 470 of them said that they favored the presidents economic policy. A presidential aid said that the majority of adults are satisfied with the way in which the president is handling the economy. Is the aid correct?

In problems 5-6 state if the study is an observational study or a controlled experiment.

4. Does hormone replacement increase the risk of breast cancer in post menopausal women? To find out, an investigator solicited 5000 women who were in menopause. She assigned 3200 of them hormone replacement and the remainder a placebo. At the end of 5 years she compared the number of incidences of breast cancer in the two groups.

5. Does hormone replacement increase the risk of breast cancer in post menopausal women? To find out, an investigator solicited 5000 women who were in menopause for 5 years. In this group 3200 were on hormone replacement and the remainder were not. She noted the incidence of breast cancer in the two groups of women.

6. In a study of the relationship between breast cancer and women on hormone replacement women in the age group 50-64 from Kings County in Washington were investigated. 557 were randomly selected from those diagnosed breast cancer between Jan. 1, 1988 and June 30, 1990. Additionally, 492 women in the same age group were randomly selected from those without breast cancer.

In the cancer group, 57.8% were on hormone replacement while in the non-cancer group 51.2% were on hormone replacement. Does this study show any relationship between hormone replacement and breast cancer? (Use significance level .05)

7. The current screening procedure used for breast cancer fails to detect the cancer in 15% of the women who actually have the disease. A new screening procedure is developed and tried on 70 women known to have breast cancer. It failed to detect the cancer in 6 of them. Does this support the claim that the new method is better than the old method?
8. It is a theory that people often postpone death until after some important event such as their birthday. If this theory is correct we would expect a dip in the death rate in the month preceding a persons birthday. If no such dip is present then the probability that a person dies in any month should be 1/12. Investigating the deaths in a sample of 348 famous people showed that 16 of them died in the month preceding their birthday. Can we conclude the theory has merit from this data?
9. The following is a real study whose results were announced in March 1997. Heart attacks are classified as Q-wave and non-Q-wave. The non-Q-wave are less severe than the Q-wave and account for about half of the heart attacks. Since 1987, both the American College of Cardiology and the American Heart Association have urged aggressive treatment of non-Q-wave heart attacks by methods such as angioplasty and by-pass surgery rather than the less drastic methods of drug therapy. They were certain that was the best approach although the issue was never put to a test. Because if their influence the aggressive approach was consider to be only really acceptable treatment. Starting in 1993 a group of investigators at the VA decided to finally put the issue to the test. Between 1993 and 1996 920 patients at 15 VA centers with non-Q-wave heart attack were randomly divided into 2 groups of 460 each. Those in one group got the conventional aggressive treatment and those in the other group got the milder drug therapy. After 2.5 years the groups were compared. Here are the results.

	Death all causes		2nd heart attack		Death in 1st 9 days	
	Estimate	SE	Estimate	SE	Estimate	SE
aggressive	.1739	.0177	.1522	.0168	.0457	.0097
drugs	.1283	.0156	.1717	.0176	.0130	.0053

When needed use significance level .05 in the following questions.

a) From the point of view of 2nd heart attack is there any difference between the two treatments?

b) From the point of view of total death rate is there a difference between the two treatments?

c) Form the point of view of the death in the first 9 days of treatment is there any difference between the two treatments?

d) Is the position of the American College of Cardiology and the American Heart Association justified by these results?

e) What treatment would you recommend if you had a patient who just had a non-Q-wave heart attack based on these results?

NORMAL TABLE 1

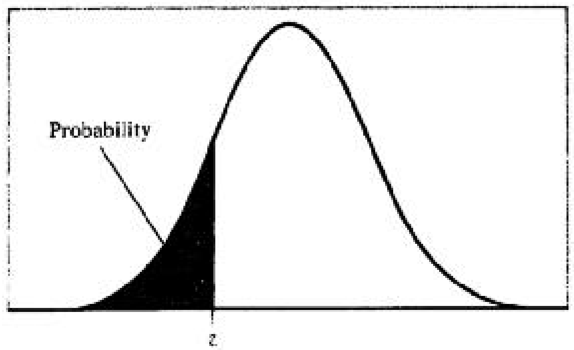


Table entry for z is the probability lying below z.

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

NORMAL TABLE 2

Table entry for z is the probability lying below z.

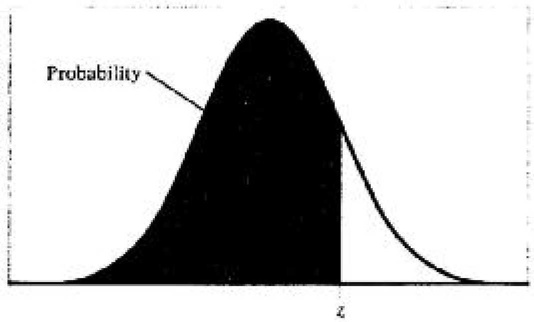


Table A (Continued)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

ANSWERS TO PROBLEMS

PROBLEMS CHAPTER 1

SECTIONS 1-3

1. 10/20, (b) 9/20, (c) 14/20, (d) 12/20
2. a = 1/6
3. f(3) = .3
4. (a) 3/15, (b) 14/15, (c) 10/15
5. P(Y = 1) = 5/20, P(Y = 2) = 10/20, P(Y = 3) = 5/20
6. (a) 5/20, (b) 10/20, (c) $f_X(-1) = 9/20$, $f_X(0) = 6/20$, $f_X(1) = 5/20$, (d) $f_Y(-1) = 6/20$, $f_Y(0) = 5/20$, $f_Y(1) = 9/20$, (e) 5/20
7. (a)

x	1	2	3
P(X=x)	7/20	5/20	8/20

y	1	2	3	4
P(Y=y)	6/20	6/20	6/20	2/20

(b) 3/20 (c) 17/20

8. (a) 15/52, (b) 8/52
9. (a) 11/36 (b) 20/36
10.

k	2	3	4	5	6	7	8	9	10	11	12
P(X+Y=k)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

11.

x	1	2	3	4	P(X =x)
1	.35	.3	.05	.2	.9
0	.05	0	.05	0	.1
P(Y = y	.4	.3	.1	.2	

SECTIONS 4-6

12. Let X = 1 or 0 as a person has Barr-Epstein antibody or not and let Y = 1 or 0 as the person has Clamydia. From the information given we can form the joint probability function of X and Y.

	y		
x	1	0	P(X=x)
1	.4	.4	.8
0	.1	.1	.2
P(Y= y)	.5	.5	

(a) .9 (b) .4

13. $\frac{(13)(12)(4)(3)(2)4^2}{(52)(51)(50)(49)}$

14. (a) $\frac{(13)(12)(11)(10)4^4}{(52)(51)(50)(49)}$

(b) $\frac{13}{(52)(51)(50)(49)}$

15. $\frac{\binom{4}{2}(4)(3)(2)(13)(12)13^2}{(52)(51)(50)(49)}$

16. $\frac{2(9)(8!)}{10!}$

17. Prob(exactly 1 pair) + prob(triple) + prob(4 of kind)

18. (a) 1/n (b) 1/(n)(n-1)

19. (a) $(r - 1)^n/r^n$ (b) $(r -2)^n/r^n$

SECTION 6

20. $(13)(12)\binom{4}{4}\binom{4}{1}q = 0.00024$

21. $(13)(12)\binom{4}{3}\binom{4}{1}q = 0.00097$

22. (4)(10)q = 0.000016

23. 4q = 0.0000016

24. 0.0039 - 0.000016 = 0.00388

25. 0.0020 – 0.000016 = 0.00198

26. 0.000016 – 0.0000016 = 0.000014

27. (a) $\binom{13}{5}\binom{4}{1}^5 q = 0.51398$. (b) 0.51398) –[0.0039 + 0.0020 – 0.000016] = 0.0581

SECTION 7

28. (a)

y	-1	0	1
P(Y=y|X=-1)	1/9	4/9	4/9
P(Y=y|X=0)	3/6	1/6	2/6
P(Y=y|X=1)	2/5	0	3/5

y	-1	0	1
P(X=x|Y=-1)	1/6	3/6	2/6
P(X=x|Y=0)	4/5	1/5	0
P(X=x|Y=1)	4/9	2/9	3/9

(b) 4/6 (c) 9/15

31. (a) 13/25 (b) 3/13

32. (a) (2/3)(1/4) (b) (1/4) + (3/4)(1/3)

SECTION 8

33. (b) O* > O if a > 1 - b,O* < O if a < 1 - b and O* = O if a > 1 - b

34. $$\frac{PV_{-}}{1-PV_{-}} = \frac{b}{1-a}\frac{1-p}{p}$$

SECTION 9

35. (b) $\frac{p^2}{1-(1-p)^2}$

37. 1/2

38. 15/36

39. prob for 9 is 4/36, prob for 10 is 3/36

40. prob for 9 is $\frac{70}{36^3}$ prob for 10 is $\frac{75}{36^3}$

43. $\frac{69.8}{9} = 7.76$

44. $\frac{1}{\binom{4}{2}} = 1/6.$

SECTION 10

45. (a) $(1/2)^{15} = .000031$ (b) .999569

46. ln(.2)/ln(.999569) = 52737.26 so 52738

47. .95021

48. (a) .002 (b) .8649 (c) 347

49. He is ignoring the order of the tosses. The outcomes he is considering are not equally probable. In fact HH and TT have probability 1/4 each but HT has probability 1/2.

SECTION 11

50. (a)

i	1	2	3
Posterior for i	2/9	3/9	4/9

(b)

i	1	2	2
Posterior for i	4/29	9/29	16/29

51. (a) $\frac{p}{p+.01(1-p)}$ (b) 0.99 (c) No, it assumes Jim is guilty with probability 1/2 with no evidence.

52. 0.7145

SECTION 12

54. 25

55. (a) No (c) Always say Many if possible. (d) Choose many with probability 1/2 if there is a choice.

56. 2/3

58. $\frac{(11)_{10}}{11^{10}}$

59. $\frac{(10)_5}{10^5}$

SECTION 13

60. (a) c = 1/2 (b) .4219 (c) .4444(a + 1)/(a + 2)

61. (a) c = a+1 (b) $(.4)^{a+1} - (.2)^{a+1}$

62. (a) $e^{-t/a}$ (b) $e^{-s/a}$

SECTION 14

63. (4/3

64. (a+1)/(a+2)

65. 11

66. 1/2

67. (a+1)/(a+m+1)

68. EX = 1.3, EX^2= 2.5

69. (a) p+(1-p)/5
    (b) 100[p+(1-p)/5]
    (c) 3/4

70. n(2p-1)

74. 300

75. (a) $P(X_i = 1) = \left(1-\frac{1}{r}\right)^n$ (b) $r\left(1-\frac{1}{r}\right)^n$

76. (b) $\sum_{i=1}^{r}(1-p_i)^n$ (c) There are r choices. $p_i = 1$, $p_j = 0$ for $j \neq i$ (d) $p_i = 1/r$ for all i.

SECTION 14

77. $\frac{a+1}{2+a+1}+\left(\frac{a+1}{a+2}\right)^2$

78. (a) 3 (b) $\sqrt{3}$

79. 20

80. 158

81. c = EX, minimum value = var(X)

82. 2.5

PROBLEMS CHAPTER 2

SECTIONS 1 and 2

1 N = number of 5's

(a) $P(N = 2) = \binom{5}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^3 = .1652$

(b) $P(N = 1) = \binom{5}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^4 = .4034$

(c) $P(1 \le N \le 2) = P(N = 1) + P(N = 2) = .5686$

2 N = number of correct answers that he guesses.

(a) $P(N = 3) = \binom{8}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^5 = .1468$

(b) $P(N \ge 3) = .2031$

3 Let N = number of heads

$P(N = 3) = \binom{6}{3}\left(\frac{1}{2}\right)^6 = .3125$

$P(N = 4) + P(N = 2) = 2P(N = 4) = .4688$

More likely to get 4 of one kind

4 Let N = total number of 6's, N_4 = number of 6's in the first 4 rolls, N' = number of 6's in last 2 rolls. N_4 and N' are independent and $N = N_4 + N'$.

$$P(N_4 = 2 | N = 3) = \frac{P(N_4 = 2)P(N' = 1)}{P(N = 3)} = \frac{\binom{4}{2}\binom{2}{1}}{\binom{6}{3}} = .6$$

5 N = number of heads

$$P(N = 3 | N \ge 2) = \frac{P(N = 3)}{P(N \ge 2)} = \frac{(.5)^3}{\binom{3}{2}(.5)^3 + (.5)^3} = .25$$

6 N = number of false positives

(a) $P(N > 0) = 1 - P(N = 0) = 1 - (.7)^{10} = .9718$

(b) $P(N < 2) = P(N = 0) + P(N = 1) = .1493$

(c) P(N = 3) = .2668

(d) P(N $\leq$ 2) = .3828

(e) P(N $\geq$ 2) = 1 - P(N $\leq$ 1) = .8507

7 N = number that get reaction

P(N = 2) = .2048

8 (a) b(12, 3/5)

(b) does not apply because population size is too small compared to the sample size

(c) b(10,.6)

(a) does not apply because population size is too small compared to the sample size

(b) b(125,.58)

(c) does not apply. Here N = N' + N'' but the success probabilities for N' and N'' are not the same

9. P(N $\leq$ 4) = .076. Looks suspicious.

10. P(N $\geq$ 3) = .1035

11. (a)5(4/52)=.3846 (b)5(4/52)(1-(4/52))=.3550

13. (a) $\binom{N}{x}\left(\frac{m}{N}\right)^{x}\left(1-\frac{m}{N}\right)^{N-x}$ (b) N(m/N) = m

SECTION 3

14. $\Phi(2)-\Phi(-12)$ = .9772 - .1151 = .8621

15. $1-\Phi\left(\frac{18-10}{5}\right) = .0548$

16. (a) P(X > 110) = .0478 (b) 1 - (1 - .0478)^{10} = .3873

17. Need t so that P(X > t) = .99. t =200($z'_{.01}$) + 2500 = 200(-2.325) + 2500 = 2035

18. 132.9

SECTION 4

19. p = .0478, np = (350)(.0478) = 16.730, $\sqrt{np(1-p)}$ = 3.9923

P(N $\geq$ 12) = $1-\Phi\left(\frac{12-16.730-.5}{3.9923}\right)$ = 1- Φ (-1.31) = .9049

20. (a) N is b(14,.65), $P(N \geq 10) = .4227$, no (b) N is b(14,.3), $P(N \geq 10) = .002$, yes
21. $P(\hat{p} \leq .22) = 2236$. they don't have fewer
22. exact .1098, approximate .1203
23. exact .0796, approximate .0791
24. $P(\hat{p} \leq .24) = .4413$
25. (a) .2743 (b) .2823 [using normal approximation without correction]

SECTION 5

26. (a) $\frac{1}{36}\left(\frac{35}{36}\right)^9$ (b) 36
27. $(.1)(.9)^6 = .0531$
28. (a) $r = 3[qp^2 + pq^2]$ where $q = 1 - p$, 3/4 (b) $r(1 - r)^n$, $(3/4)(1/4)^n$ (c) ,1/r 4/3
29. $r = m[pq^{m-1} + qp^{m-1}]$, 1/r
30. $p\binom{x-1}{r-1}p^{r-1}(1-p)^{x-r}$

SECTION 6

31. (a) .2611 (b) N = number < 6.5. N is b(86,.2611) $P(N \geq 28) = .1314$
32. $P(Y > X) = P(Y - X > 0)$. Y - X is $N\left(-.1, \sqrt{(.22)^2 + (.31)^2}\right)$ so $P(Y - X > 0) = .3962$
33. .6730
34. (a) $P(X > 140) = .0228$ (b) $P(\overline{X} > 140) = 0$
35. (a) $EX_i = p - q$, $SD = \sqrt{pq}$ (b) via the CLT $P(S_{100} > 0) = .3446$
36. Need n so that $\frac{.02\sqrt{n}}{\sqrt{(.49)(.51)}} = 2.33$, n = 3391.7

SECTION 7

37. (a) (b) $e^{-5.5}(1 + \ldots + (5.5)^5/5!) = .5289$
38. $\lambda = .85/7 = .1214$, so probability is $e^{-4(.1214)} = .6153$
39. 8
40. The number is Poisson with mean (2/10,000)(35,000) = 7. Probability is $e^{-7}(7^4/4!) = .0912$

41. The number per day is Poisson with mean $\lambda = -\ln(.9)$. The number in a 12 hour period is Poisson with mean $-\ln(.9)/2$. Probability of at least one attack is $1 - e^{\ln(.9)/2} = .0513$

42. (a) $1 - e^{-(3)} = .9502$ (b) $1 - (1 - .9502)^5 = .9999$

SECTION 8

43. (a) $f_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right]$ (b) $f_n \sim \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}$ so $\lim_{n\to\infty}\frac{f_{n+1}}{f_n} = \frac{1+\sqrt{5}}{2}$

44. (a) $f_n = \frac{2f_0 + g_0}{3} + \frac{f_0 - g_0}{3}\left(-\frac{1}{2}\right)^n$ (b) $\frac{2f_0 + g_0}{3}$

12. PROBLEMS CHAPTER 3

1 null hypothesis p= .5, z(obs) = 1.4, p-value = .1616, accept at both levels

2 null hypothesis p = .8 z(obs) = -8.66 , p-value .0, reject at .01 and .05 Hocum is wrong

3 null hypothesis p = .5 z(obs) = -2.57, p-value = .0102, Reject at both .01 and .05. Aid is wrong

4 controlled experiment

5 observational study

6 using non pooled method z(obs) = 2.1473 p-value = .0318, significant difference

7 null hypothesis p = .15, z(obs) = -1.51, p-value .1310 no significant difference

8 null hypothesis p = .0833, z(obs) = -2.518, p-value = .0118, reject at .05 but not at .01

9 (a)z(obs) = -.8015 p-value = .4238, no difference

(a) z(obs) = 1.93, p-value = .0536

(b) z(obs) =2.96, p-value = .003

(c) Seem that they are quite wrong

(f) drugs

INDEX

F

G

H

I