THE BINOMIAL DISTRIBUTION

ITS DERIVATION AND APPLICATIONS

by

CHIAM TAH WEN

A Graduation Exercise submitted as
part fulfilment towards the
Degree of Bachelor of Arts
in Economics (Statistics)

Division of Statistics,
Faculty of Economics and Administration,
University of Malaya.

August 1966

TABLE OF CONTENTS

SYNOPSIS

This exercise gives an account of the derivation of the Binomial Distribution from first principles and the applications of the distribution to various fields of human activities.

As the Binomial Distribution is based on the theory of probability, which is the corner-stone of many physical, biological and social sciences, we first discuss some fundamental concepts and the basic calculus of probability for experiments with a number of possible outcomes. A probability measure is first introduced over the events of a sample space; independence of events and trials repeated under identical conditions are then discussed.

Next we introduce the analytic theory of probability in the finite case. Random variables are defined as functions on sample spaces, and probability distributions, means, variance and standard deviation are dealt with.

In the chapter on the Binomial Distribution, which is the most important probability function defined on a finite sample space, we derive the basic properties of a Bernoulli process and a binomially distributed random variable, and discuss some of the important properties of the distribution. We include in this chapter a discussion on the law of large numbers, which serves as a basis for the intuitive notion of probability as a measure of relative frequencies. Without this law, the whole probability theory would lose its intuitive foundation.

In many practical problems, the values of n and k in the Binomial Distribution formula $b(k;n,p)$ are very large; a direct use of the formula becomes almost impossible as the binomial coefficients are difficult to evaluate. Two approximate methods of calculation, based on the Poisson distribution and the Normal distribution, are discussed. We also illustrate, with examples, the computation of the Binomial Distribution for very large n, $(n \geqslant 100)$ by means of a short method.

The Binomial Distribution is applied, to a large extent, in economics, engineering, medicine and genetics. Included in this exercise are the applications of this distribution in industrial quality control, decision-making, testing of a statistical hypothesis, testing of significance for differences in samples, in power supply, in sera or vaccine testing, in random walk, in parking problem and in the Mendelian hereditary theory.

Appendices in this exercise explain the mathematical derivation of some formulae and approximations.

ACKNOWLEDGEMENTS

The writer wishes to express his most sincere gratitude to Dr. Chan Kai Leng, Lecturer, Division of Statistics, Faculty of Economics and Administration, University of Malaya, Kuala Lumpur, for the valuable help and the unflagging interest he has taken in the production of this Graduation Exercise.

The writer also wishes to thank Dr. Saw Swee Hock, Head of Division of Statistics, Faculty of Economics and Administration of the University for his advice which is greatly appreciated.

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Probability concepts are now playing an increasingly important role in various fields. Statistics, the discipline connected with the collection, classification, analyzing and interpretation of data, is based on the theory of probability. Making explicit reference to the nature and effects of chance phenomena, probability theory is the corner-stone of many physical and biological sciences. For instance, telephone engineers use the ideas of probability theory to calculate the density of telephone traffic; physicists employ probabilistic notions to study thermal noise in electric circuits and the Brownian motion of particles immersed in a liquid or gas, while the geneticists attempt to predict, through the use of probability theory, the relative frequency with which various characteristics occur in groups of individuals. Apart from their application in the physical and biological world, probability concepts are finding increased use in the social sciences and business as well: economists use the techniques of game theory and other aspects of operation research analysis to discuss competition and to arrive at optimum investment in order to maximize returns (which generally are achieved at the output where marginal revenue = marginal cost) and the business executives who have to make decisions in the face of uncertainty, invoke the theory of probability as an aid in planning inventory and establishing quality control.

One may ask: what is the peculiar feature in the probability theory that enables it to have such diverse applications? What is the property that is possessed in common by such phenomena as the number of telephone calls made in a town in a day, the number of individuals possessing a genetical composition or the standard of quality of particles manufactured by a certain process?

E. Parzen[1] attributes the wide application of

---

1. E. Parzen: <u>Modern Probability Theory and its Applications</u>, John Wiley and Sons, Inc, 1960.

probability theory to the "randomness" of the phenomena. Each of the above-mentioned may be considered a random phenomenon in the sense that

"A random or chance phenomenon is an empirical phenomenon characterized by the property that its observation under a given set of circumstances does not always lead to the same observed outcome (so that there is no deterministic regularity) but rather to different outcomes in such a way that there is statistical regularity."[2] He means by this that numbers exist between 0 & 1 and that the numbers represent the relative frequency with which the different possible outcomes may be observed in a series of observations of independent occurrences of the phenomenon.

Closely connected with the concept of a random phenomenon are the concepts of a random event and of the probability of a random event. A random event refers to one whose relative frequency of occurrence in a very large number of observations of randomly selected situations in which the event may occur approaches a limit value; and as the number of observation is increased to infinity, the limit value of the relative frequency becomes the probability of the random event.

The modern theory of probability is conceived in terms of axioms. According to W. Feller[3], three aspects of the theory, namely, the formal logical content, the intuitive background and the applications must be distinguished. From the view-point of formal logical content, probability theory, like geometry or analytical mechanics, begins with undefined concepts or axioms from which various logical propositions are deduced. In the matter of intuition, we notice that the axioms of geometry and mechanics refer to an existing intuitive background. Probability, too, derives its notions and terminology from intuition; these notions being just as indefinable and as intuitive as are the notions of a point, line or mass.

---

2. E. Parzen: op. cit., page 3.

3. W. Feller: An Introduction to Probability Theory and its Applications. Vol.1. 2nd edition, John Wiley & Sons, Inc, 1957. Page 1.

With regard to applications, the abstract mathematical models which the probability theory constructs serve as tools of analysis, and different models can describe the same empirical situation. These abstract models are mostly of a qualitative nature; only experience can tell us whether or not these models reasonably describe laws of nature of life.

CHAPTER TWO

SAMPLE SPACES

In probability theory, we are interested in Statistical probability, which is related to the possible outcomes of an experiment, whether real or conceptual; and not with modes of inductive reasoning such as "Ahmad will probably come" or "Fermat's conjecture is probably false". We may conduct or conceive experiments, for instance, tossing a coin, arranging a desk of cards, observing the lifespan of a person, noting the frequency of accidents, or even sampling penguins on the Mars and note their possible outcomes.

Frequently the outcomes are idealized. Take the case of tossing a coin for instance. We ordinarily agree to regard "head" and "tail" as the only possible outcomes, though the coin does not necessarily fall "head" or "tail", as it can roll away or stand on its edge. We call the results of experiments or observations events. We speak of the event, for example, of seven coins tossed, more than four fell heads. The composition of a sample (e.g., two people blind in a sample of 120) and the result of a measurement (e.gs.: height 5' 6"; 9 trunk lines busy) are each an event.

When we throw two dice, we have 36 possible combinations:

1,1	2,1	3,1	4,1	5,1	6,1
1,2	2,2	3,2	4,2	5,2	6,2
1,3	2,3	3,3	4,3	5,3	6,3
1,4	2,4	3,4	4,4	5,4	6,4
1,5	2,5	3,5	4,5	5,5	6,5
1,6	2,6	3,6	4,6	5,6	6,6

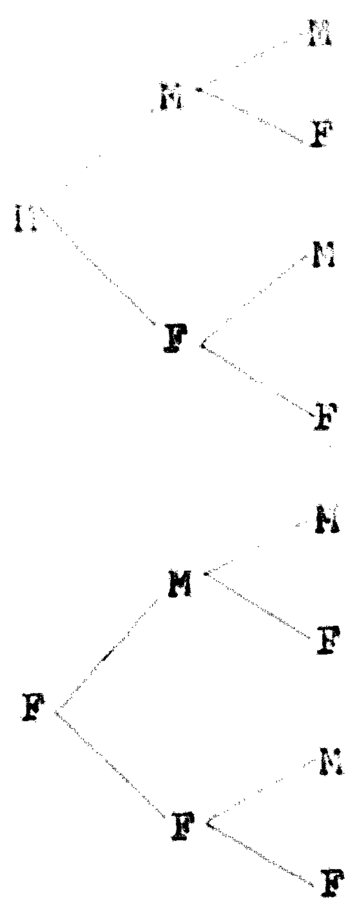
An outcome such as "Sum 4" is a compound event which can be further decomposed: Sum 4 occurs if the outcome is (1,3), (2,2) or (3,1). Thus we distinguish between simple (indivisible) and compound events or outcomes.

The events "two odd faces" can be decomposed into (1,1), (1,3) ...... (5,5), a total of nine simple events.

Each simple outcome or event is called a sample point; their aggregate forms the sample space. Thus an experiment whether real or conceptual is defined by the sample space.

In tossing a coin, we have a sample space consisting of the set $\Omega = \{H, T\}$, the outcomes having been idealized. We note that this experiment is completely defined by the sample space. Similarly, in the experiment of throwing two dice, the 36 combinations stated above from our sample space.

If a survey of families with three children is made and the sexes of children (in order of age, oldest child first) are recorded, we will have the follwing sample space:-



$\therefore \Omega = (\text{MMM, MMF, MFM, MFF, FMM, FMF, FFM, FFF})$

where M and F represent male and female children respectively.

Let us now consider the sample space obtained from the experiment of distributing "three balls in three cells".[1] The twenty-seven outcomes or sample points are listed in Table 1.

TABLE 1

1.	abc	-	-	10.	a	bc	-	19.	-	a	bc
2.	-	abc	-	11.	b	ac	-	20.	-	b	ac
3.	-	-	abc	12.	c	ab	-	21.	-	c	ab
4.	ab	c	-	13.	a	-	bc	22.	a	b	c
5.	ac	b	-	14.	b	-	ac	23.	a	c	b
6.	bc	a	-	15.	c	-	ab	24.	b	a	c
7.	ab	-	c	16.	-	ab	c	25.	b	c	a
8.	ac	-	b	17.	-	ac	b	26.	c	a	b
9.	bc	-	a	18.	-	bc	a	27.	c	b	a

Each of these arrangements represents a simple event, i.e., a sample point. Instead of just three balls in three cells, we can extend our study to the more general case of r balls in n cells. The following situations, whose intuitive background though vary, are abstractly equivalent to the scheme of placing r balls into n cells, in the sense that the outcomes differ only in their verbal descriptions:-

(i) r accidents in 7 days or in n days.

(ii) an elevator or lift starting with r passengers and stops at n floors. The different arrangements of discharging the passengers are replicas of the different distribution of r balls in n cells.

---

1. This example is taken from W. Feller, op. cit., page 9.

(iii) In firing at n targets, r hits among n targets.

(iv) the distribution of r persons; here we have $n=2$ cells and r balls.

(v) the possible distributions of r misprints in the pages of a book ($r <$ the number of letters per page).

We have been discussing the case of placing three distinguishable balls into three cells. If we assume that the balls are indistinguishable, then we have now a sample space of only ten sample points; which are listed in Table 2.

TABLE 2

1. {000 | – | –}	4. {00 | 0 | –}	8. {– | 00 | 0}
2. {– | 000 | –}	5. {00 | – | 0}	9. {– | 0 | 00}
3. {– | – | 000}	6. {0 | 00 | –}	10. {0 | 0 | 0}
	7. {0 | – | 00}	

If we assume further that even the cells are indistinguishable, then we have only three possible different arrangements:

1. {000 | – | –} 2. {00 | 0 | –} 3. {0 | 0 | 0}

A sample space need not necessarily be a finite set -- it can be an infinite set. For example, toss a coin until it falls heads for the first time. It is conceivable that we get an unending sequence of tails and that a head is never obtained. Let us denote this outcome w. If a head is obtained, we specify the outcome by recording the number of the toss that produced the first head. The sample space becomes

$$\Omega = \{w, 1, 2, 3, \ldots\ldots\}, \quad \text{an infinite set.}$$

As our discussion indicates, one way of precisely formulating the notion of an experiment is to write down an associated sample space. Before attributing probabilities to sample points, let us now turn our attention to some important ideas in set theory, which are used to specify relations among events.

Given an event A, we may consider the case that A does not occur. This, the complement of A (denoted by A'), consists of those sample points that do not belong to A. The symbol $C = A \cup B$ indicates that, given two events A and B, the event C is the union of A and B, if either A or B or both occur. $A \cup A'$ is thus the whole sample space which represents certainty. $A \cap B$, which represents the intersection of A and B, consists of points common to A and B. If there are no such common points, as in the case of A and A', A and B cannot occur simultaneously and become "mutually exclusive". Symbolically, it is written $A \cap B = \emptyset$.

Applying these ideas to Table 1, we see that the event A "one cell is multiply occupied" is realized in the arrangements numbered 1 - 21. The event A is thus the aggregate of the sample points 1 - 21. The event B "first cell not empty" is the aggregate of the sample points 1, 4 - 15 and 22 - 27. The event C, "both A and B occur," i.e. $A \cap B$ is the aggregate of the 13 sample points 1, 4 - 15.

2. Probabilities in sample spaces:

Given a (finite) sample space Ω with the sample points $E_1, E_2, \ldots\ldots E_n$, we shall assume that with each point E_j there is associated a number, called the probability of E_j and denoted by $P\{E_j\}$. According to the axiomatic theory of probability, this number must be non-negative and such that

$$P\{E_1\} + P\{E_2\} + \ldots\ldots\ldots + P\{E_n\} = 1.$$

The probability $P\{A\}$ of an event is the sum of the probabilities of all sample points in it. Thus $P\{\Omega\} = 1$. It follows that for any event A, $0 \leq P\{A\} \leq 1$.

To find $P\{A \cup B\}$, we consider all points belonging to either A or B, but those belonging to both A and B are counted only once. Therefore, $P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}$. For mutually exclusive events, i.e. for the events $A \cap B = \emptyset$, we have the important addition principle:

$$P\{A \cup B\} = P\{A\} + P\{B\} \qquad (2.2.1)$$

We can arbitrarily assign probabilities to sample points E_1, E_2, ... E_n as long as

(i) $P\{E_j\}$ is non-negative;

and (ii) $P\{E_1\} + P\{E_2\} + ... P\{E_n\} = 1.$

Thus, for the sample space H, T obtained when we toss a coin, each of the following assignments of probabilities is acceptable:-

(a) $P\{H\} = P\{T\} = \frac{1}{2}$

(b) $P\{H\} = \frac{2}{3}$, $P\{T\} = \frac{1}{3}$

(c) $P\{H\} = 0$ and $P\{T\} = 1.$

Consequently, if p is any real number between 0 and 1 inclusive, there are infinitely many possible acceptable assignments, one for each choice of the number p.

An excellent example showing that different assignments of probabilities are compatible with the same sample space may be seen in Maxwell-Boltzmann statistics and Bose-Einstein statistics. With reference to Table 1, it seems natural to assume that all sample points are equally probable, i.e. that each sample point has probability $\frac{1}{27}$. Here the Maxwell-Boltzmann model applies. For most applications, e.g. birthdays, firing at targets, accidents, sampling, an elevator carrying passengers to different floors, the argument in Maxwell-Boltzmann statistics appears sound. Modern theory has shown, however, that this statistics does not apply to any known particles; in no case are all n^r arrangements approximately equally probable. Bose and Einstein showed that certain particles, like photons, nuclei, and atoms containing an even number of elementary particles, are subject to the Bose-Einstein statistics, for which we consider only distinguishable arrangements, and assign probability $\binom{n+r-1}{r}^{-1}$ to each arrangement. With reference to Table 2, it may be argued that the actual physical experiment is unaffected by our failure to distinguish between the balls; physically there remain 27 different possibilities, even though only 10 different forms are distinguishable. The Bose-Einstein model, which assigns probability to only distinguishable arrangements, attributes probability $\frac{1}{10}$ to each of the sample points.

3. Conditional Probability; Independence:

Conditional Probability, an important tool of probability theory, may be defined as follows:

Given that $P\{B\} \neq 0$, the conditional probability of event A, relative to event B, denoted by $P\{A/B\}$ is $\frac{P\{A \cap B\}}{P\{B\}}$.

If $P\{A/B\} = P\{A\}$, event A is said to be independent of event B.

Two events A and B are said to be stochastically independent if

$$P\{A \cap B\} = P\{A\} \cdot P\{B\} \qquad (2.3.1)$$

This multiplication principle will be applied in the derivation and applications of the Binomial Distribution.

4. Trials repeated under Identical Conditions

Having discussed the notion of stochastic independence, we can now apply it to formulate the intuitive concept of experiments "repeated under identical conditions".

Consider an experiment described by a sample space Ω, assuming that Ω consists of finitely many sample points E_1, E_2, ... E_n. When the same experiment is performed twice successively, the conceivable outcomes are the N^2 pairs of sample points (E_1, E_1), (E_1, E_2), ... (E_n, E_n) and these now constitute the new sample space. We thus have the combinational product of Ω by itself: $\Omega \times \Omega$: with reference to analytical geometry, one speaks of the first and second co-ordinate of the point, (E_i, E_j). This idea is extended to $\Omega \times \Omega \times \Omega$ ....(to n factors).

We can assign probabilities to outcomes in many ways for the new sample space. However, when experiments are performed repeatedly under identical conditions, we imply independence: the first outcome should have no influence on the second.

$$\therefore P\{E_i, E_j\} = P\{E_i\} P\{E_j\} = P_i P_j$$

assuming the probabilities of E_i and E_j are P_i and P_j.

CHAPTER THREE

RANDOM VARIABLES

In Mathematics, we often come across the idea of a function. The quantity y is called a function of the real number x if to every x there corresponds a value y. The idea of a function can be applied to cases where the independent variable is not a real number; e.g. the distance is a function of a pair of points; the binomial coefficient $\binom{x}{k}$ is a function defined for pairs of numbers (x, k) of which k is a non-negative integer.

A real-valued function defined on a sample space is called a random variable. Some examples are the gamblers' gain, the number of multiple birthdays in a company of n people and the energy and temperature of a physical system.

In a finite sample space, we can tabulate any random variable X by enumerating in some order all points of the space and associating with each the corresponding value of X. If we let x_1, x_2, ... be the values which the random variable X assumes, then in a discrete sample space, x_j s being integers, the aggregate of all sample points in which X assumes the value x_j forms the event $X=x_j$. Denoting the probability of this event by $P\{X=x_j\}$, we have the function $P\{X=x_j\} = f(x_j)$, $j=1, 2, \ldots\ldots$ which becomes the probability distribution of the random variable X. Obviously $f(x_j) \geq 0$ and $\sum f(x_j) = 1$.

Example:

Three fair coins are tossed. How many fall heads?

The answer is a number determined by the outcome of the experiment. The number may be 0, 1, 2 or 3.

When 3 fair coins are tossed, we have the following sample space and the probability for each sample point.

Sample Point	No. of Heads	Probability
HHH	3	$\frac{1}{8}$
HHT	2	$\frac{1}{8}$
HTH	2	$\frac{1}{8}$
HTT	1	$\frac{1}{8}$
THH	2	$\frac{1}{8}$
THT	1	$\frac{1}{8}$
TTH	1	$\frac{1}{8}$
TTT	0	$\frac{1}{8}$

Tabulating the probability function for the number of heads, we have

Probability	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
Number of Heads	0	1	2	3

Let the variable X represent the number of heads, X is a random variable since the value of X is a number determined by the outcome of the experiment. Associating $X=x_j$ with a probability, we have

$$P\left\{X=x_j\right\} = f(x_j) \qquad (j=1, 2 \ldots\ldots\ldots)$$

In this 3-coin experiment, we note that

$$P\left\{X=0\right\} = f(0) = \frac{1}{8}$$

$$P\left\{X=1\right\} = f(1) = \frac{3}{8}$$

$$P\left\{X=2\right\} = f(2) = \frac{3}{8}$$

$$P\left\{X=3\right\} = f(3) = \frac{1}{8}$$

Let us see how the idea of Probability Distribution of a random variable is applied in the sample space of Table 1, with probability $\frac{1}{27}$ for each sample point. The number N of occupied cells is a random variable. At the three points 1 - 3, the random variable assumes the value 1; at the eighteen points 4 - 21, the value 2 and at the six points 22 - 27, the value 3. Thus the probability distribution of N is given by

$$P\left\{N=1\right\} = f(1) = \frac{1}{9}$$

$$P\left\{N=2\right\} = f(2) = \frac{2}{3}$$

$$P\left\{N=3\right\} = f(3) = \frac{2}{9}$$

Another random variable is the number X of balls in the first cell. Tabulation I shows that its probability distribution is given by

$$P\left\{X=0\right\} = f(0) = \frac{8}{27}$$

$$P\left\{X=1\right\} = f(1) = \frac{12}{27}$$

$$P\left\{X=2\right\} = f(2) = \frac{6}{27}$$

$$P\left\{X=3\right\} = f(3) = \frac{1}{27}$$

Consider now two random variables X and Y defined on the same sample space, and denote the values they assume by x_1, x_2, ... and y_1, y_2, ... respectively. Let the corresponding probability distributions be $\left\{f(x_j)\right\}$ and $\left\{g(y_k)\right\}$, The aggregate of points in which the two conditions $X=x_j$ and $Y=y_k$ are satisfied forms an event whose probability is denoted by

$$P\left\{X=x_j,\ Y=y_k\right\}$$

We then have the joint probability function:-

$$P\{X=x_j,\ Y=y_k\} = p(x_j,\ y_k) \qquad j,\ k = 1,\ 2,\ \ldots.$$

Clearly $p(x_j,\ y_k) \geq 0$ and $\sum_{j,\ k} p(x_j,\ y_k) = 1.$

For every fixed j,

$$p(x_j,\ y_1)+(x_j,\ y_2)+\ldots\ldots = P\{X=x_j\} = f(x_j)$$

and for every fixed k,

$$p(x_1,\ y_k)+p(x_2,\ y_k)+\ldots\ldots = P\{Y=y_k\} = g(y_k)$$

Thus by adding the probabilities in individual rows and columns, we obtain the probability distribution of X and Y. These probability distributions are sometimes called marginal distributions.

With reference to Table 1, we note that the combination N=1, X=0 occurs at two points where $P\{N=1,\ X=0\} = \frac{2}{27}$. The probabilities of all pairs are given by the Joint Probability Distribution of N and X shown below

N \ X	0	1	2	3	Distribution of N
1	$^2/_{27}$	0	0	$^1/_{27}$	$\frac{3}{27}$
2	$^6/_{27}$	$^6/_{27}$	$^6/_{27}$	0	$\frac{18}{17}$
3	0	$^6/_{27}$	0	0	$\frac{6}{27}$
Distribution of X	$\frac{8}{27}$	$\frac{12}{27}$	$\frac{6}{27}$	$\frac{1}{27}$	.

We notice that adding the entries in the rows and columns gives the probability distribution of N and X respectively.

2. The Mean of a Random Variable or Expectation

The mean of a random value is a measure of location; it roughly indicates a "middle" or "average" value of the random variable.

Let X be a random variable assuming the values x_1, x_2, ... with corresponding probabilities $f(x_1)$, $f(x_2)$ .... The mean or expected value of X is defined by

$$E\ (X) = \sum x_k\ f(x_k)$$

on the assumption that the series converges absolutly.

If X_1, X_2, ... X_n are random variables with expectations, then the expectation of their sum exists and is the sum of their expectations.

$$E(X_1 + X_2 + \dots . X_n) = E(X_1) + E(X_2) + \dots .. E(X_n)$$

If X and Y are mutually independent random variable, with expectations, then their product is a random variable with expectation and

$$E\ (XY) = E(X)\ E(Y)$$

3. The Variance of a Random Variable

The mean of a random variable X, being an average value, does not tell us the variability of the values of X. To study a distribution more accurately, we also require a measure of the variability, the "spread" or "dispersion" of the values of the random variable, as random variables with different probability functions can also have equal means. This is illustrated in the following charts:

x_1

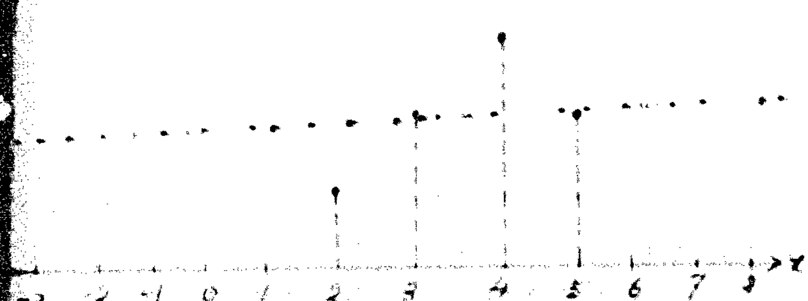
x	2	3	4	5
$f_1(x)$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{2}{8}$

$E(x_1) = \frac{30}{8} = 3.75$

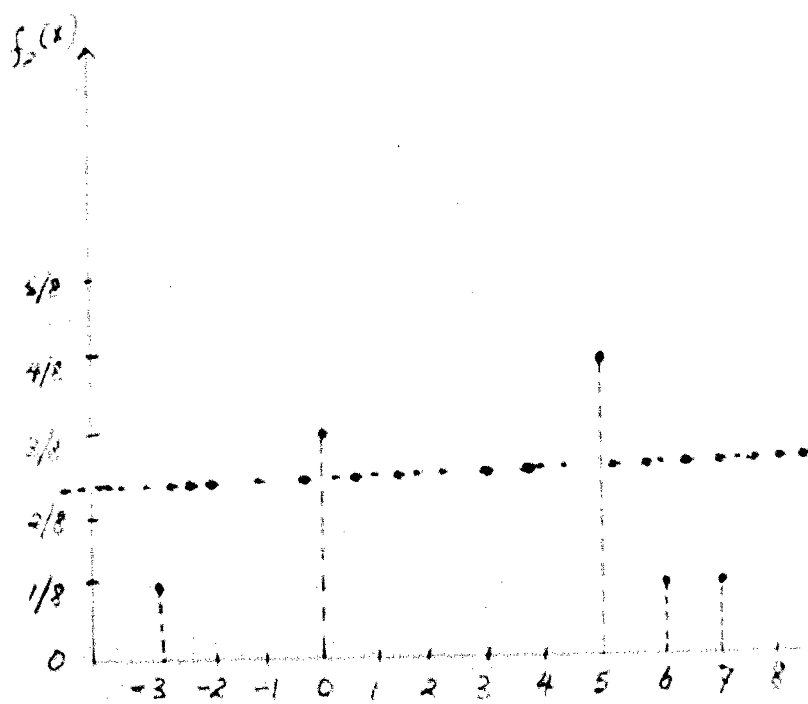
x_2

x	-3	0	5	6	7
$f_2(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

$E(x_2) = \frac{30}{8} = 3.75$



(a) Probability function for x_1



(b) Probability function for x_2

We notice that though their means are the same, the probability distribution of X_1 is less spread out than that of X_2.

If we let X be a random variable with second moment $E(X^2)$ and let $u = E(X)$ be its mean, the variance of X is defined as

$$E\left((X - u)^2 \right)$$

which is equal to

$$E\,(X^2) - u^2$$ [1]

For independent random variables, $Var(Sn) = Var(X_1 + X_2 + \ldots\ldots X_n) = Var(X_1) + Var(X_2) + \ldots\ldots Var(X_n)$.

The standard deviation of X is the positive square root (or zero) of the variance of X.

Let X be any random variable with mean u_x and standard deviation $\sigma_x > 0$. Let the random variable X^* be defined as

$$X^* = \frac{X - u_x}{\sigma_x}$$

(X^* is called the standardized random variable corresponding to X)

Then $E\,(X^*) = 0; \qquad Var\,(X^*) = 1$

In other words, the standardized random variable has mean 0 and standard deviation 1.

---

1. For the mathematical derivation of variance, see <u>Appendix 1</u>.

CHAPTER FOUR

THE BINOMIAL DISTRIBUTION

1. Bernoulli trials[1] and the Binomial Distribution

In probability theory, we are often asked to solve problems involving experiments made up of a number, say n, of individual trials. Each trial is itself really an arbitrary experiment, and is therefore defined in the mathematical theory by some sample space and assignment of probabilities to its simple events. The trials can be independent or dependent, and the simple events of the sample space for the n-trial experiment are assigned probabilities accordingly.

Although each trial may have many possible outcomes, we are often interested only in whether a certain result occurs or not. For example, a machine turns out parts which are classified defective or good; a person is blind or not blind; two dice are rolled and the sum of the numbers showing is five or is different from five. In other words, we simply describe the result of an outcome as A or non-A. To standardize our terminology, we call one of the two possible results of a trial a success, the other a failure; and which result is to be called a success is of course completely arbitrary. As the results of an experiment are just a success or a failure, the sample space for the outcome of a trial will contain only two elements. Generally we denote the two probabilities by p and q and refer to the outcome with probability p as success, S and to the other as failure, F.

To satisfy the axiomatic theory of probability, evidently p and q must be non-negative and $p+q=1$.

---

1. James Bernoulli (1654-1705). His main work, Ars Conjectandi, was published posthumously in 1713.

In repeated independent trials, if there are only two possible outcomes for each trial and their probability remain the same throughout the trials, then we have what the probabilists call Bernoulli trials. The sample space for an experiment made up of n Bernoulli trials is the Cartesian product set

$$\{S . F\} \times \{S . F\} \times \ldots\ldots \{S . F\}$$

containing 2^n n-tuples as elements. Every n-tuple represents an outcome of the n-trial experiment and is made up of n symbols, each a S or a F. Since the trials are independent, the probabilities multiply. For instance, for the sequence SSSFSF, we have for its probability pppqpq.

We deduce from the above discussion that the probability of any simple event whose n-tuples contain k S's and hence n-k F's (in any order) is $p^k q^{n-k}$; k=0, 1, 2, ... n. One such n-tuple is determined by selecting the k trials in which S's occur from among all n trials. This can be done in $\binom{n}{k}$ ways[2]. Therefore there are n-tuples containing k S's and n-k F's, the probability of the corresponding simple events being $p^k q^{n-k}$.

We thus come to the following conclusion:

If $b(k;n,p)$ is the probability that n Bernoulli trials with probabilities p for success and $q=1-p$ for failure result in k successes and n-k failures, $(0 \leq k \leq n)$, then

$$f(k) = b(k;n,p) = \binom{n}{k} p^k q^{n-k} \qquad (4.1)$$

This theorem can also be derived in terms of random variables. In an experiment made up of n Bernoulli

---

2. $\binom{n}{k}$, defined as the number of k-subsets (subsets with exactly k elements) of a set of n elements, is equal to $\frac{n!}{k!(n-k)!}$.

trials, we are interested in determining the probability function of the random variable whose value is the total number of successes obtained in the experiment. This random variable, Sn, has possible values 0, 1, 2, ... n. Now Sn=k, where k assumes any one of these possible values, is the event for which exactly k S's and therefore n-k F's occur. This event is the union of the $\binom{n}{k}$ simple events determined by n-tuples with k S's and n-k F's, the probability of each such simple event being $p^k q^{n-k}$. Hence

$$f(k) = b(k;n,p) = P\{Sn=k\} = \binom{n}{k} p^k q^{n-k}$$

$$k=0,\ 1,\ 2,\ \ldots\ldots\ n. \qquad (4.2)$$

For given values of n and p, the parameters, the probability function defined by P(Sn=k) is called the binomial probability function. The random variable Sn is said to be binomially distributed, the attribute "binomial" referring to the fact that this formula represents the k^{th} term of the binomial expansion of $(q+n)^n$ which is equal to $q + \binom{n}{1} q^{n-1}p + \binom{n}{2} q^{n-2}p^2 + \ldots\ldots$ $\ldots + p^n$ where $\binom{n}{1}$, $\binom{n}{2}$ ..... are binomial coefficients.

This statement also shows that

$$\sum_{k=0}^{n} b(k;n,p) = (q+p)^n = 1$$

as is required by the notion of probability.

We note further that b(k;n,p) represents a family of binomial distributions, the value of each term being dependent on the values of the parameters n and p.

We have been considering the Binomial Distribution where an experiment has two outcomes. A generalization of this will be stochastic independent processes with more than two outcomes.

Assume that the outcomes are $\{a_1, a_2, \ldots a_k\}$ occuring with probabilities $p_1, p_2, \ldots\ldots p_k$.

Let $n = r_1 + r_2 + \ldots\ldots r_k$ where each $r \geq 0$.

The probability of getting exactly r_1 occurrences of a_1; r_2 occurrences of a_2 ..... is

$$f(r_1, r_2 \ldots\ldots r_k) = \frac{n!}{r_1! r_2! \ldots r_k!} p_1^{r_1} p_2^{r_2} \ldots\ldots p_k^{r_k}.$$

2. Some properties of the Binomial Distribution

(i) The Central Term[3].

As k goes from 0 to n, the terms $b(k;n,p)$ first increase monotonically, then decrease monotonically reaching their greatest value when $k=m$, except that $b(m-1;n,p) = b(m;n,p)$ when $m=(n+1)p$.

We call $b(m;n,p)$ the central term. Often m is called "the most probable number of successes", but for large values of n, all terms $b(k;n,p)$ are small.

Illustration

Compute the binomial probabilities for $n=5$, $p=0.4$ and for $n=4$ and $p=0.4$

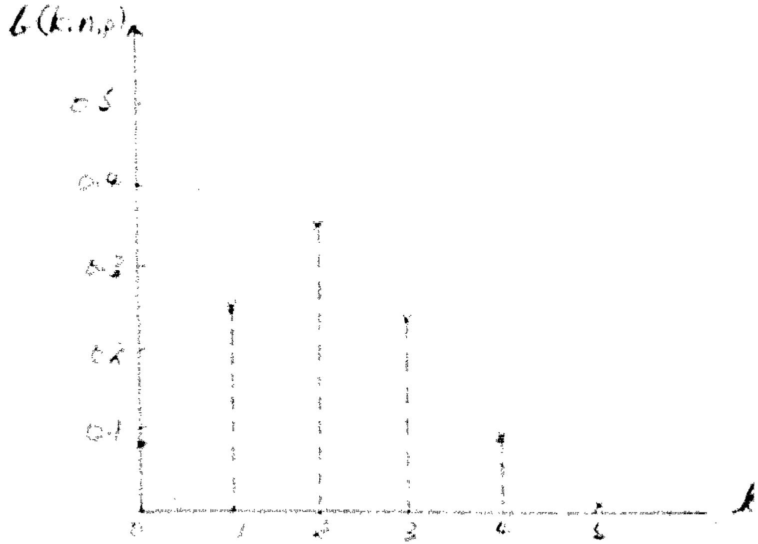
The case of $n=5$, $p=0.4$:

We note that $(n+1)p = (5+1)0.4 = 2.4$ and is not an integer.

For $k=0, 1, 2, \ldots\ldots 5$, we have values of $\binom{n}{k}p^k q^{n-k}$ 0.078, 0.259, 0.346, 0.230, 0.077 0.010. Graphically, we obtain

---

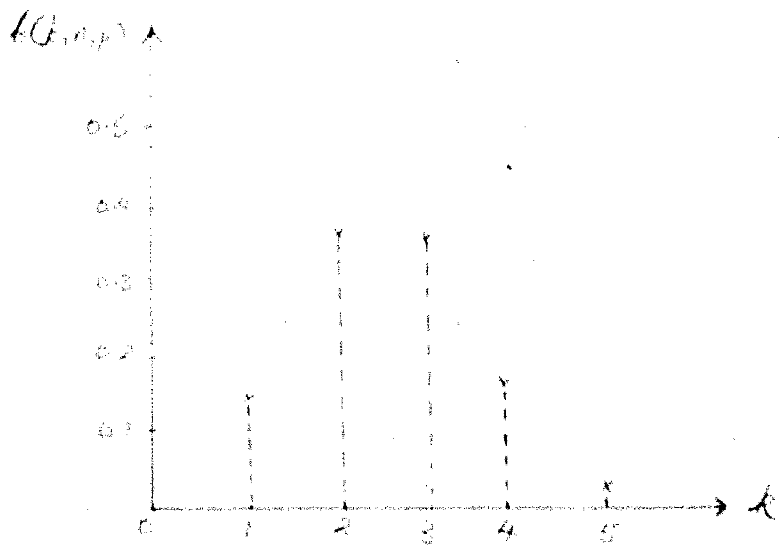
3. For the proof of this theorem, see Appendix 2(A).



These values show that for $(n+1)p$ not an integer, if $k < (n+1)p$, $b(k;n,p)$ is greater than the preceding term and if $k > (n+1)p$, $b(k;n,p)$ is smaller than the preceding term.

<u>The case of $n=4$, $p=0.4$</u>:

We note that $(n+1)p = (4+1)0.4 = 2$ is an integer. For $k=0, 1, 2, \ldots\ldots 5$, the values of $\binom{n}{k} p^k q^{n-k}$ are 0.130, 0.346, 0.346, 0.153 and 0.026. In graph, we show:



This illustrates the fact that for $(n+1)p = m$, an integer, $b(k;n,p)$ increases up to $b(m-1;n,p)$ which is equal to $b(m;n,p)$, and then decreases.

(ii) Theorem on Tails[4].

If $r \geqslant np$, the probability of at least r successes satisfies the inequality

$$\sum_{v=0}^{n-r} b(r+v;n,p) \leq b(r,n,p) \frac{(r+1)q}{r+1-(n+1)p}$$

and if S np, the probability of at most S successes satisfies the inequality

$$\sum_{\phi=0}^{S} b(\phi;n,p) \leqslant b(S;n,p) \frac{(n-S+1)}{(n+1)p-S}$$

(iii) The Mean of the Binomial Distribution.

This is equal to np where n denotes the number of Bernoulli trials and p the probability of success.

Proof: Let Xk be the number of successes scored at the k^{th} trial. This variable assumes the values 0 and 1 with corresponding probabilities q and p.

Hence $E(Xk) = 0.q + 1.p = p$

But $Sn = X_1 + X_2 + \ldots\ldots X_n$ where Sn is the total number of successes in n Bernoulli trials.

4. For the proof of this theorem, see Appendix 2(B).

And since each X_k depends only on the k^{th} trial, $X_1, X_2 \ldots\ldots X_n$ are independent random variables.

$$
\begin{aligned}
E(S_n) &= E(X_1+X_2+\ldots\ldots+X_n) \\
&= E(X_1)+E(X_2)+\ldots\ldots+E(X_n) \\
&= \underbrace{p + p + p \ldots\ldots}_{n \text{ terms}} = np
\end{aligned}
$$

(iv) <u>The Variance of the Binomial Distribution.</u>

In Binomial distribution, the variance is **npq.**

Proof:-

For each of the random variables, X,

$$E(X^2)=0^2 xq+1^2 xp=p \;\&\; u=1xp+0xq=p$$

$$
\begin{aligned}
Var(X) &= E\{(X-u)\}^2 \\
&= E(X^2)-u^2 \quad \text{where } u=E(X) \text{ is the mean.} \\
&= p-p^2 \\
&= p(1-p) \\
&= pq
\end{aligned}
$$

$$
\begin{aligned}
Var(S_n) &= Var(X_1)+Var(X_2)+\ldots Var(X_n) \\
&= pq+pq+\ldots\ldots\ldots\ldots \\
&= npq
\end{aligned}
$$

using the fact that for independent variables, the variance of their sum is the sum of their variances.

3. The Law of Large Numbers

In a very large number of trials, the probability of an event is interpreted as the relative frequency of its occurrence. Let us attempt to justify this intuitive frequency interpretation of probability by means of the Law of Large Numbers.

Our intuitive notion of probability is based on the assumption that if in n identical trials, A occurs x times, and if n is very large, then $\frac{x}{n}$ should be near the probability p of the event A. In terms of Bernoulli trials, with probability p for success, the above notion is equivalent to the concept that if S_n represents the number of successes in n trials, then $\frac{S_n}{n}$, the average number of success, should be near p.

Let us give a theoretical formulation[5] to this.

Consider the probability that $\frac{S_n}{n}$ exceeds $p+E$, where $E > 0$ is an arbitrarily small but fixed number. This probability is the same as $P\{S_n > n(p+E)\}$ and equals $\sum_{v=0}^{n-r} b(r+v;n,p)$ when r is the smallest integer exceeding $n(p+E)$.

Then,

$$\sum_{v=0}^{n-r} b(r+v;n,p) \leq b(r;n,p)\,\frac{(r+1)q}{r+1-(n+1)p}$$

implies

$$P\{S_n > n(p+E)\} \leq b(r;n,p)\,\frac{n(p+E)+q}{nE+q}$$

---

5. See W. Feller, op. cit., page 141.

With increasing n, the fraction on the right remains bounded, whereas $b(r;n,p) \to 0$. Since $b(r;n,p) < b(k;n,p)$ for each k such that $(n+1)p \le k < r$. [because for $k > (n+1)p$, the term $b(k;n,p)$ is smaller than the preceding one, as we have pointed in the discussion on the theorem on Central Term.] and there are about nE such terms $b(k;n,p)$.

It follows that as n increases,

$$P\{Sn > n(p+E)\} \to 0.$$

Using the formula

$$\sum_{p=0}^{S} b(p;n,p) \le b(S;n,p)\,\frac{(n-S+1)p}{(n+1)p-S}$$

we can show that $P\{Sn < n(p-E)\} \to 0$.

We have

$$P\left\{\left|\frac{Sn}{n} - p\right| < E\right\} \to 1$$

i.e., As n increases, the probability that the average number of successes deviates from p by more than any preassigned E tends to 0.

This law serves as a basis for the intuitive notion of probability as a measure of relative frequencies — without this law, probability theory would lose its intuitive foundation.

CHAPTER FIVE

THEORETICAL RESULTS CONCERNING THE BINOMIAL DISTRIBUTION

In many practical problems, the values of n and k are very large, and a direct use of the Binomial Distribution formula $b(k;n,p)$ becomes impossible as the binomial coefficients are difficult to evaluate for large n and k. In such situations, two approximations to the Binomial Distribution are available: one the Poisson distribution, derived by S. P. Poisson and bearing his name, and the other, the Normal distribution.

1. The Poisson Approximation:

In the Binomial Distribution, if n is large and p is small so that the mean np is of moderate magnitude, say, of the order of unity in any given application, the Binomial Distribution is then approximated by the Poisson's Law, which states that

$$\lim_{n \to \infty} \binom{n}{k} p^k q^{n-k} = \frac{e^{-np}(np)^k}{k!} = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$\text{if } \lambda = np; \quad k = 0, 1, 2, \ldots\ldots \qquad (5.1.1)$$

As an approximation to the Binomial, the Poisson Approximation is useful in a class of binomial problems in which neither p nor n are known, but their product λ is known or can be estimated. The applications of this approximation are many and varied: they range from the number of articles lost in subways to the frequency of comets. Before the appearance of elaborate tables of the Binomial Distribution, a successful application of the Poisson approximation to the Binomial was to sampling inspection of industrial product. The probability p of a defective unit of product is typically small, and the number of units inspected n is often fairly large.

As a law itself, many random phenomena obey the Poisson law. Among the usual ones are deaths resulting from horse-kicks; occurrence of accidents; errors and breakdowns. In physics, the random emission of electrons from the filament of a vacumn tube, and the spontaneous decomposition of radioactive atomic nuclei lead to phenomean obeying the Poisson law. This law can also be applied in the field of operation research and management science.

As an example of the Poisson approximation, We quote the experimental data of Rutherford and Geiger showing the number of alpha particles emitted from a radioactive speciman is 2,608 periods of time each of 7½ seconds.

The Rutherford-Geiger Data, and corresponding Poisson frequencies:

No. of Emssions Xi	Observed frequency fi	Poisson frequency
0	57	54
1	203	210
2	383	407
3	525	525
4	532	508
5	408	394
6	273	254
7	139	140
8	45	68
9	27	29
10	10	11
11	4	4
12	0	1
13	1	1
14	1	1
	2608	2607

The parameter λ of the Poisson approximation is the Arithmetic mean of the Poisson variable.

$$\bar{x} = \frac{x_1f_1+x_2f_2+\ldots x_{15}f_{15}}{\Sigma fi} = \frac{10,097}{2,608} = 3.87$$

The Poisson frequencies are calculated from (5.1.1).

For $\lambda = np$ estimated by 3.87, we need not determine n and p here as we are approximating a binomial distribution.

Since the Poisson frequencies are fairly close to Observed frequencies, the conditions underlying the Poisson approximation may be satisfied, conditions being that we have Bernoulli trials with small p and large n, for it may be argued that the probability p of an atom emitting an alpha particle is small, that the number of atoms in the specimen available to emit — i.e., the number of independent trials — is very large, and that p is constant from trial and trial, i.e., the various atoms have the same chance of emitting particles.

<u>Derivation of Poisson Approximation to the Binomial Distribution.</u>

Let $\lambda = np$ so that $p = \frac{\lambda}{n}$ and $p \longrightarrow 0$ for $n \longrightarrow \infty$.

$$\binom{n}{k} p^k q^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n} \right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{n(n-1)\ldots.(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{k}\right)^n x \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= 1\left[\left(1-\frac{1}{n}\right)\ldots.\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\right]\frac{\lambda^k}{k!} \times$$

$$\left(1-\frac{\lambda}{n}\right)^n \longrightarrow \frac{\lambda^k}{k!} e^{-\lambda} \text{ for } n \longrightarrow \infty$$

Since the factors in the square brackets converge to 1 and so the product also converges to 1 when the number of factors is finite.

$$\left(1 - \frac{\lambda}{n} \right)^n \longrightarrow e^{-\lambda}$$

This can be arrived at by the application of the Taylor expansion and the use of logarithms, viz:-

$$b(k;n,p) = \binom{n}{k} p^k q^{n-k}$$

$$b(0;n,p) = \binom{n}{0} p^0 q^{n-0} = q^n = (1-p)^n$$

when $\lambda = np$, we have $p = \frac{\lambda}{n}$.

$$b(0;n,p) = \left(1 - \frac{\lambda}{n}\right)^n$$

$$\log b(0;n,p) = n \log\left(1 - \frac{\lambda}{n}\right) = -\lambda - \frac{\lambda^2}{2n} \cdots$$

so that for large n,

$$b(0;n,p) = \left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}$$

$$\therefore \lim_{n \to \infty} \binom{n}{k} p^k q^{n-k} = \frac{e^{-\lambda}\lambda^{-k}}{k!} \quad \text{if } \lambda = np;\ k = 0, 1, \ldots$$

TABLE 3

The Binomial Distribution $b(k;n,p)$ for $n=5$, 10, 20, 50 and 100 & $p=\frac{1}{n}$ (i.e., $np=1$) together with the Poisson Distribution.

n	5	10	20	50	100	∞	
p	0,2	0,1	0,05	0,02	0,01	0	
np	1	1	1	1	1	1	Difference
npq	0,8	0,9	0,95	0,98	0,99	1	
$\sqrt{npq}$	0,894	0,949	0,975	0,990	0,995	1	
k	b(k;5,0,2)	b(k;10,0,1)	b(k;20,0,05)	b(k;50,0,02)	b(k;100,0,01)	Poisson Distribution	
(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(vi - vii)
0	0.3277	0.3487	0.3585	0.3642	0.3660	0.3679	-0.0019
1	0.4096	0.3874	0.3774	0.3716	0.3697	0.3679	0.0018
2	0.2048	0.1937	0.1887	0.1858	0.1849	0.1839	0.0010
3	0.0512	0.0574	0.0596	0.0607	0.0610	0.0613	-0.0003
4	0.0064	0.0112	0.0133	0.0146	0.0149	0.0153	-0.0004
5	0.0003	0.0015	0.0023	0.0027	0.0029	0.0031	-0.0002
6		0.0001	0.0003	0.0004	0.0005	0.0005	0.0000
7			0.0000	0.0001	0.0001	0.0001	0.0000

From Table 3, we note that the Poisson Distribution is a good approximation to the Binomial Distribution, also for small values of n, when p is sufficiently small. We generally apply the Poisson Distribution as approximation to the Binomial Distribution when $p<0.1$.

2. The Normal Approximation.

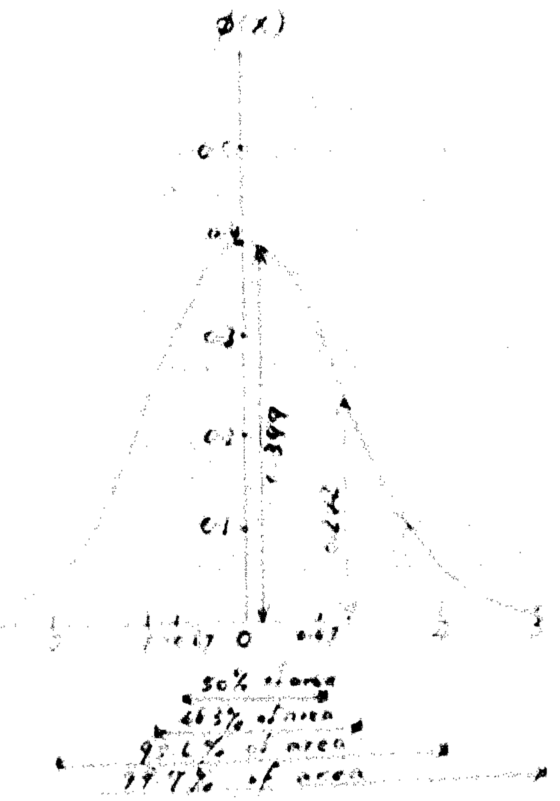
The Normal Distribution law is the limiting form of the Binomial Distribution Law, when both n and k are large, so large that $\frac{1}{X} = \frac{1}{np}$ and $\frac{1}{npq}$ are both negligible. This implies that p and q are numbers not greatly different from unity.

The Normal Distribution function, denoted by $\emptyset(x)$ is defined as the integral of the normal density function, denoted by $\emptyset(x) = \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}x^2}$

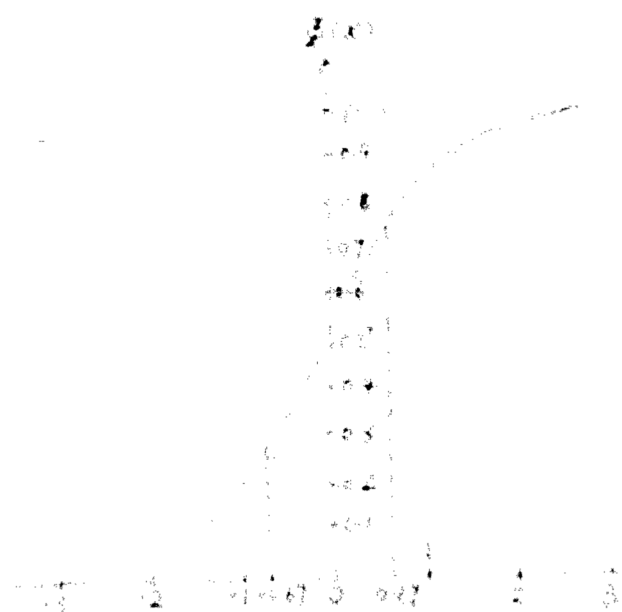
Hence

$$\emptyset(x) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\, dy \qquad (5.2.1)$$

The graph of $\emptyset(x)$ is a symmetric, bell -shaped curve, as shown below:-



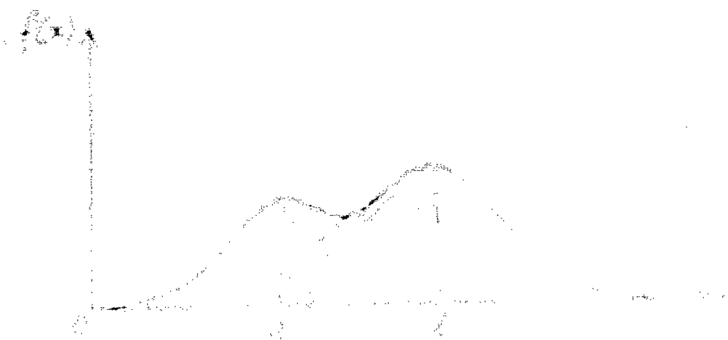
The Normal Density function



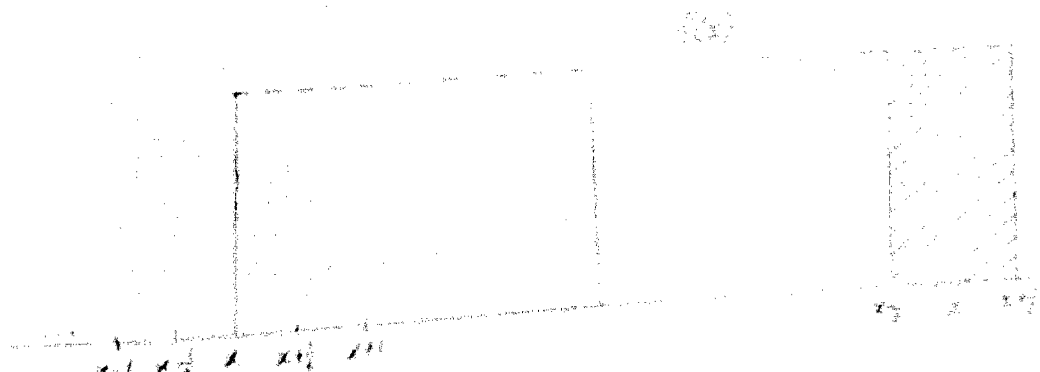
The Normal Distribution function

We study the Normal Approximation to the Binomial Distribution since, with exact Binomial Distribution, we are restricted by the extent of values available in the Binomial Distribution Tables. The largest tables supply us only values of n up to 1000, with large gaps of values of n in between. Furthermore the Normal Distribution formula sometimes offers a more manageable expression for a Binomial probability than does a complicated summation.

To study the limiting behaviour of the binomial family, i.e. to study the Normal approximation, we need to bear in mind that since the Binomial is a discrete distribution and the Normal a continuous one, the probabilities represented by Binomial ordinates need to be replaced by areas, as areas are used to represent probabilities in continuous distributions.



The probability that the random variable S takes a value between a and b is represented by the area of the shaded part of the figure. Shaded area gives $P(a \leq S \leq b)$. To fit a Binomial distribution by a continuous probability function, we replace each ordinate of a Binomial distribution by centering at x a rectangle whose width is one unit and whose height equals that of the original Binomial ordinate. The area of the rectangle has the same numerical measure as the height of the ordinate. To illustrate, the area over the interval from $x-\frac{1}{2}$ to $x+\frac{1}{2}$ in the figure below (b) has the same numerical value as the height of the ordinate at x in figure (a).



To study the Normal Approximation, we also need a change of scale for Sn, the random variable representing the total number of successes in n Bernoulli trials. Since the standard normal distribution has mean 0 and standard deviation 1, we standardized the random variable Sn into Sn*

$$Sn^* = \frac{Sn - u}{\sigma} = \frac{Sn - np}{(npq)^{1/2}} \qquad (5.2.2)$$

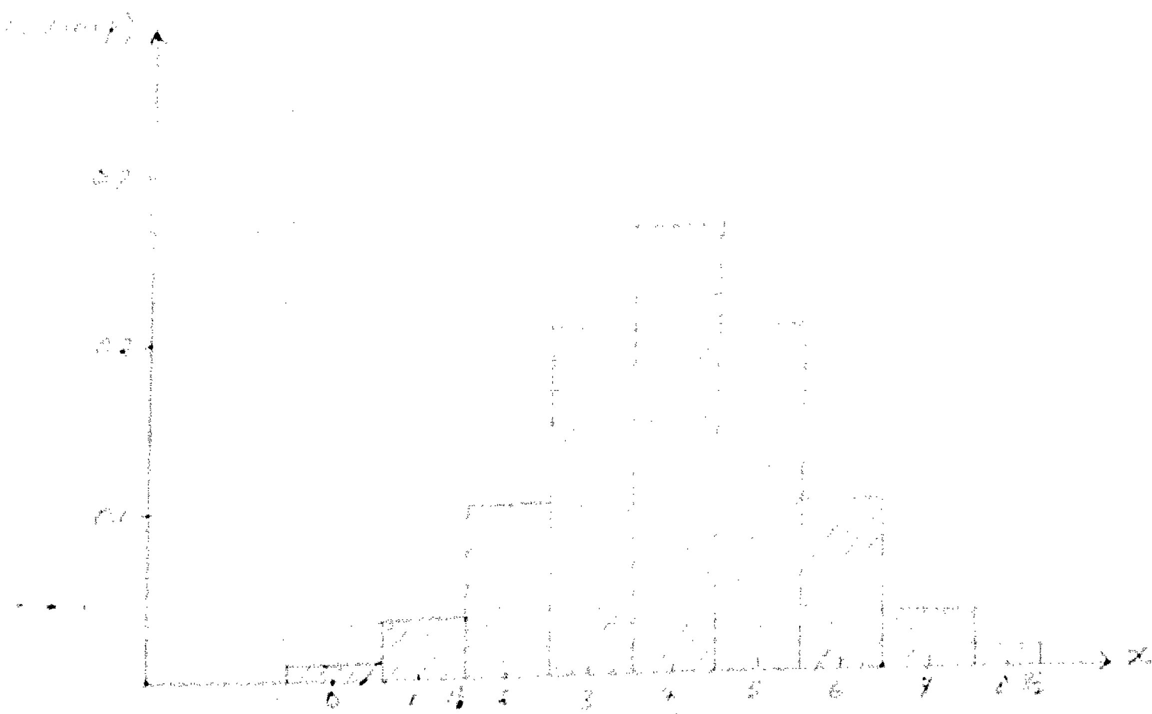
because in a binomial distribution, $u = np$ & $\sigma = (npq)^{1/2}$

Let us now use the Normal tables for Binomial Problems.

Given the Binomial Distribution with $n=8$, $p=\frac{1}{2}$, we have

$u = np = 8 \times \frac{1}{2} = 4$; $\sigma = (npq)^{1/2} = \sqrt{8 \times \frac{1}{2} \times \frac{1}{2}} = \sqrt{2} = 1.41$

The graph is shown below:-



If we use the areas of the figure to evaluate the probability of 2 or more successes, we need to include all the areas in the rectangles above the x-axis to the right of $x=1\frac{1}{2}$. If we use only the area to the right of $x=2$, we will leave half of P(2) behind.

Therefore, we take for x the value 1½ as the left hand boundary.

$$S^* = \frac{S - np}{(npq)^{½}} = \frac{1½ - 4}{2} = \frac{-2.5}{1.414} = -1.768$$

$P(S \geq 1½) = P(S^* \geq -1.768) = 0.9617$ from Normal tables.

An important tool for studying the limit of the Binomial Distribution is the DeMoivre-Laplace Theorem which enables us to compute approximate probabilities for sums using the Normal Distribution without ever knowing the exact distribution of the sum. The Theorem tells us that:

Let $S_1, S_2, \ldots\ldots S_n$ be a sequence of random variables where S_n is the number of successes in a binomial experiment with n trials, each with probability of success p, where p is non-negative. Let Sn^*, $n=1,2,\ldots\ldots$ be the corresponding sequence of adjusted random variables, where $Sn^* = \frac{Sn - np}{(npq)^{½}}$ and let a be a constant. Then as $n \rightarrow \infty$, $P(Sn^* \geq a)$ approaches the area to the right of a for the standard normal distribution.

The result of the above-mentioned theorem says, in practice, that for large values of n,

$$\begin{aligned} P(Sn \geq s) &= P(Sn > s - ½) \\ &= P\left(\frac{Sn-np}{[npq]^{½}} \geq \frac{s-½-np}{[npq]^{½}}\right) \\ &= P\left(Sn^* \geq \frac{s-½-np}{[npq]^{½}}\right) \end{aligned}$$

where Sn^* is a standard normal random variable.

Alternatively, the probability that a random variable obeying the binomial probability law with parameters n and p will have an observed value lying between a and b, inclusive, for any integers a and b, is given approximately by

$$\sum_{k=a}^{n} \binom{n}{k} p^k q^{n-k} = \Phi\left(\frac{b+½-np}{[npq]^{½}}\right) - \Phi\left(\frac{a-½-np}{[npq]^{½}}\right) \qquad (5.2.3$$

For the proof of the DeMoivre-Laplace Theorem, we need the following two lemma:-

Lemma 1.: The domain bounded by the graph of the normal density function $\phi(x)$ and the x-axis has unit area, that is

$$\int_{-\infty}^{\infty} \phi(x)dx = 1$$

Lemma 2.: As $x \to \infty$

$$1 - \Phi(x) \sim \frac{1}{(2\pi)^{1/2}x} e^{-1/2 x^2}$$

or more precisely,

$$\frac{1}{(2\pi)^{1/2}} e^{-1/2 x^2} \left\{ \frac{1}{x} - \frac{1}{x^3} \right\} < 1-\Phi(x) < \frac{1}{(2\pi)^{1/2}} e^{-1/2 x^2} \cdot \frac{1}{x}$$

Proof of the DeMoivre-Laplace theorem

According to the Binomial Distribution, $P\{S_n=k\} = b(k;n,p)$ where S_n stands for the number of successes in n Bernoulli trials with probability p for success. To evaluate the probability of the event that the number of successes lies between 2 pre-assigned limits, say α and β (α & β being integers and $\beta > \alpha$), we have

$$P\{\alpha \le S_n \le \beta\} = b(\alpha;np)+b(\alpha+1;n,p)+\dots$$
$$\dots+b(\beta;n,p)$$

As this sum may involve many terms, we derive approximations to $P\{\alpha \le S_n \le \beta\}$, assuming that n is large.

We need to prove that if α and β vary so that $hx_\alpha^3 \to 0$ and $hx_\beta^3 \to 0$, then

$$P\{\alpha \le S_n \le \beta\} \sim \Phi(x_{\beta+1/2})-\Phi(x_{\alpha-1/2}) \qquad (5.2.4)$$

where $h=(npq)^{-\frac{1}{2}}$ and $x_t=(t-np)h$

We get a simpler form of this DeMoivre-Laplace Theorem by introducing Sn^*

$$Sn^* = \frac{Sn - np}{(npq)^{\frac{1}{2}}}$$ where np is the mean and $(npq)^{\frac{1}{2}}$ is the standard deviation of Sn.

The inequality $\alpha \leq Sn \leq \beta$ is the same as $X_\alpha \leq Sn^* \leq X_\beta$ and (5.2.4) states that for arbitrary fixed X_α X_β,

$$P\left\{X_\alpha \leq Sn^* \leq X_\beta\right\} \sim \Phi(X_\beta + \frac{h}{2}) - \Phi(X_\alpha - \frac{h}{2})$$

where $h=(npq)^{-\frac{1}{2}}$

Now $h \rightarrow 0$ as $n \rightarrow \infty$ and the right-hand side tends to $\Phi(X_\beta) - \Phi(X_\alpha)$

Thus we have the following corollery to the Theorem:

For every fixed $a < b$,[1]

$$P\left\{a \leq Sn^* \leq b\right\} \rightarrow \Phi(b)-\Phi(a) \qquad (5.2.5)$$

For large n, the probability on the left is practically independent of p.

The limit and approximations are only valid if the number n of trials is fixed in advance independently of the outcome of the trials.

---

1. See W. Feller, op. cit., pages 168-172, for a full account of the proof of this theorem.

TABLE 4

The Binomial Distribution b(k;n,p) for p=0.1 and n=5, 10, 20, 50, 100, with the Normal Distribution for mean = 10 and Standard Deviation 3.

n	5	10	20	50	100		
np	0.5	1.0	2.0	5.0	10.0	Mean=10.0	Difference
npq	0.45	0.90	1.80	4.50	9.0	σ^2 =9.00	
$\sqrt{npq}$	0.67	0.95	1.34	2.12	3.0	σ =3.00	
k	b(k;5,0.1)	b(k;10,0.1)	b(k;20,0.1)	b(k;50,0.1)	b(k;100,0.1)	Normal Distribution	
(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(vi - vii)
0	0.5905	0.3487	0.1216	0.0052	0.0000	0.0005	- 0.0005
1	0.3281	0.3874	0.2702	0.0286	0.0003	0.0015	- 0.0012
2	0.0729	0.1937	0.2852	0.0779	0.0016	0.0039	- 0.0023
3	0.0081	0.0574	0.1901	0.1386	0.0059	0.0089	- 0.0030
4	0.0005	0.0112	0.0898	0.1809	0.0159	0.0183	- 0.0024
5		0.0015	0.0319	0.1849	0.0339	0.0334	0.0005
6		0.0001	0.0089	0.1541	0.0596	0.0549	0.0047
7			0.0020	0.1076	0.0889	0.0807	0.0082
8			0.0004	0.0643	0.1148	0.1062	0.0086
9			0.0001	0.0333	0.1304	0.1253	0.0051
10				0.0152	0.1319	0.1324	- 0.0005
11				0.0061	0.1199	0.1253	- 0.0054
12				0.0022	0.0988	0.1062	- 0.0074
13				0.0007	0.0743	0.0807	- 0.0064
14				0.0002	0.0513	0.0549	- 0.0036
15				0.0001	0.0327	0.0334	- 0.0007
16					0.0193	0.0183	0.0010
17					0.0106	0.0089	0.0017
18					0.0054	0.0039	0.0015
19					0.0026	0.0015	0.0011
20					0.0012	0.0005	0.0007
21					0.0005	0.0002	0.0003
22					0.0002	0.0000	0.0002
23					0.0001	0.0000	0.0001
24					0.0000	0.0000	0.0000

CHAPTER SIX

CALCULATIONS WITH THE BINOMIAL DISTRIBUTION -- TABLES

The Binomial Distribution Function

Let $b(k;n,p)$ be the probability that n Bernoulli trials with probabilities p for success and $(1-p)$ for failure result in k successes and n-k failures ($0 \leq k \leq n$).

Then,

$$b(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}; \quad k=0,1,\ldots.n$$

The distribution function $b(k;n,p)$ is discontinous, as it is only defined for $k=0,1,\ldots..n$.

Now let

$$f_n\{k\} = \binom{n}{k} p^k (1-p)^{n-k}$$

$$f_n\{k+1\} = \binom{n}{k+1} p^{k+1} (1-p)^{n-k-1}$$

Taking the ratio of these two terms, we have

$$g(k) = \frac{f\{k+1\}}{f\{k\}} = \frac{n-k}{k+1} \cdot \frac{p}{1-p}; \quad k=0,1,\ldots.n-1. \qquad (6.1)$$

As it is simple to tabulate $g(k)$, the distribution function may be tabulated by computing a single value of $f\{k\}$, and the other values may be obtained by successive multiplication and division by $g(k)$ according to the formulae:-

$$f\{k+1\} = f\{k\} g(k) \qquad (6.2)$$

$$f\{k-1\} = \frac{f\{k\}}{g(k-1)} \qquad (6.3)$$

The single value of $f\{k\}$ may, for example, be calculated by means of logarithms as

$$\log f\{k\} = \log \binom{n}{k} + k \log p + (n-k) \log (1-p)$$

and $$\log \binom{n}{k} = \log n! - \log k! - \log (n-k)!$$ [2]

Generally, we choose the starting value of $f\{k\}$ near the maximum of the distribution.

<u>Example:</u> Compute the binomial distribution for $n=100$ and $p=0.1$

$$f\{k\} = \binom{100}{k} 0.1^{k}\, 0.9^{100-k} \qquad k=0,1,\ldots\ldots,100.$$

The computation of this by the ordinary method will be very cumbersome; so we use the method discussed in this chapter.

Since $np=10$, an integer, the mode is $m=np=10$ and the maximum value is

$$f\{10\} = \binom{100}{10} 0.1^{10}\, 0.9^{90}$$

From Table XIV we obtain $\log\binom{100}{10}=13.2383$, so that

$$\log f\{10\} = \log\binom{100}{10} + 10 \log 0.1 + 90 \log 0.9$$

$$= 13.2383 - 10.0000 + 0.8818 - 5$$

$$= 0.1291 - 1$$

$$f\{10\} = 0.1319$$

For comparing this binomial distribution with a corresponding normal distribution, we make use of the more accurate value $f\{10\} = 0.13187$ as our starting value.

Formula (6.1) leads to

---

2. For $n \leq 100$, $\log_{10}\binom{n}{k}$ has been tabulated in Table XIV of A. Hald: <u>Statistical Tables and Formulas</u>, John Wiley & Sons, N. Y., 1952. Other Tables referred to in this chapter, are also from A. Hald, except otherwise stated.

$$g(k) = \frac{n-k}{k+1} \cdot \frac{p}{1-p} = \frac{100-k}{k+1} \cdot \frac{0.1}{0.9} ; \qquad k=0,1,\ldots,99.$$

The following table illustrates the computation of g(k) and f{k} according to the formulae (6.1), (6.2) and (6.3)

TABLE 5

Computation of the distribution function f{k} and the Cumulative distribution function P{k} for the Binomial Distribution with n=100 and p=0.1

k	g(k)	f{k}	p{k}
0	11.1	.00003	.00003
1	5.500	.00029	.00032
2	3.630	.00162	.00194
3	2.6944	.00589	.00783
4	2.1333	.01588	.02371
5	1.7593	.03387	.05758
6	1.4921	.05958	.11716
7	1.29167	.08890	.20606
8	1.13580	.11483	.32089
9	1.01111	.13042	.45131
10	.909091	.13187	.58318
11	.824074	.11988	.70306
12	.75214	.09879	.80185
13	.69048	.07430	.87615
14	.63704	.05130	.92745
15	.59028	.03268	.96013
16	.54902	.01929	.97942
17	.51235	.01059	.99001
18	.4795	.00543	.99544
19	.4500	.00260	.99804
20	.4233	.00117	.99921
21	.399	.00050	.99971
22	.377	.00020	.99991
23	.36	.00008	.99999
24	.34	.00003	1.00002
25	.32	.00001	1.00003
26	-	.00000	1.00003

Extensive tables for the Binomial Distribution have been prepared; some of the well-known ones are:

Tables of the Binomial Probability Distribution, National Bureau of Standards, Applied Mathematics Series, Vol.6, 1950.

H. C. Romig, 50-100 Binomial Tables, John Wiley & Sons Inc., 1953.

Tables of the Cumulative Binomial Probability Distribution, Annals of the Computation Laboratory of Harvard University, Vol. XXXV, Harvard University Press, 1955.

Here we reproduce part of the Cumulative Probability Table showing

$$P(S_n \geq r) = b(r;n,p)+b(r+1;n,p)+\ldots\ldots$$
$$+\ldots\ldots+b(n;n,p).$$

TABLE 6

Cumulative Binomial Probabilities

The entry is $P(S_n \geq r) = \sum_{k=r}^{n} b(k;n,p)$. Missing entries are $< .0005$

n	r	p=.01	p=.05	p=.10	p=.20	p=.30	p=.40	p=.50
1	1	.010	.050	.100	.200	.300	.400	.500
2	1	.020	.098	.190	.360	.510	.640	.750
	2		.002	.010	.040	.090	.160	.250
3	1	.030	.143	.271	.488	.657	.784	.875
	2		.007	.028	.104	.216	.352	.500
	3				.008	.027	.064	.125
4	1	.039	.185	.344	.590	.760	.870	.938
	2	.001	.014	.052	.181	.348	.525	.688
	3			.004	.027	.084	.179	.312
	4				.002	.008	.026	.062
5	1	.049	.226	.410	.672	.832	.922	.969
	2	.001	.023	.081	.263	.472	.663	.812
	3		.001	.009	.058	.163	.317	.500
	4				.007	.031	.087	.188
	5					.002	.010	.031
6	1	.059	.265	.469	.738	.882	.953	.984
	2	.001	.033	.114	.345	.580	.767	.891
	3		.002	.016	.099	.256	.456	.656
	4			.001	.017	.070	.179	.344
	5				.002	.011	.041	.109
	6					.001	.004	.016
7	1	.068	.302	.522	.790	.918	.972	.992
	2	.002	.044	.150	.423	.671	.841	.938
	3		.004	.026	.148	.353	.580	.773
	4			.003	.033	.126	.290	.500
	5				.005	.029	.096	.227
	6					.004	.019	.062
	7						.002	.008

n	r	p=.01	p=.05	p=.10	p=.20	p=.30	p=.40	p=.50
8	1	.077	.337	.570	.832	.942	.983	.996
	2	.003	.057	.187	.497	.745	.894	.965
	3		.006	.038	.203	.448	.685	.855
	4			.005	.056	.194	.406	.637
	5				.010	.058	.174	.363
	6				.001	.011	.050	.145
	7					.001	.009	.035
	8						.001	.004
9	1	.086	.370	.613	.866	.960	.996	.998
	2	.003	.071	.225	.564	.804	.929	.980
	3		.008	.053	.262	.537	.768	.910
	4		.001	.008	.086	.270	.517	.746
	5			.001	.020	.099	.267	.500
	6				.003	.025	.099	.254
	7					.004	.025	.090
	8						.004	.020
	9							.002
10	1	.096	.401	.651	.893	.972	.994	.999
	2	.004	.086	.264	.624	.851	.954	.989
	3		.012	.070	.322	.617	.833	.945
	4		.001	.013	.121	.350	.618	.828
	5			.002	.033	.150	.367	.623
	6				.006	.047	.166	.377
	7				.001	.011	.055	.172
	8					.002	.012	.055
	9						.002	.011
	10							.001
20	1	.182	.642	.878	.988	.999	1.000	1.000
	2	.017	.264	.608	.931	.992	.999	1.000
	3	.001	.075	.323	.794	.965	.996	1.000
	4		.016	.133	.589	.893	.984	.999
	5		.003	.043	.370	.762	.949	.994
	6			.011	.196	.584	.874	.979
	7			.002	.087	.392	.750	.942
	8				.032	.228	.584	.868
	9				.010	.113	.404	.748
	10				.003	.048	.245	.588
	11				.001	.017	.128	.412

TABLE 6 (continued)

n	r	p=.01	p=.05	p=.10	p=.20	p=.30	p=.40	p=.50
20	12					.005	.057	.252
	13					.001	.021	.132
	14						.006	.058
	15						.002	.021
	16							.006
	17							.001
	18							
	19							
	20							

We illustrate the use of this table in the following examples:-

Example 1

Among the integers from 1 to 10 inclusive, there are four members that are prime. If a number is chosen at random from the integers from 1 to 10, the probability that it is a prime is 0.4. Suppose 10 numbers are chosen at random in this way, each choice being made independently from the full set from 1 to 10. What is the probability

(a) that 5 or more are primes

(b) that 4 or more are primes

(c) that 4 of them are primes

& (d) that 3 or fewer are primes?

The answer to (a) is supplied by the entry in the table for $n=10$, $p=0.4$ and $r=5$. The probability that at least five of the numbers are prime is <u>0.367</u>.

The answer to (b) is supplied by the entry in the table for the same values of n and p, with $r=4$. The probability that at least four of the numbers are prime is <u>0.618</u>.

The answer to (c) is given by the difference between the answers to (b) and (a)

$$0.618 - 0.367 = \underline{0.251}$$

To find $P(S_{10}=4)$, we use the idea of $P(S_{10}=4) = P(S_{10}\geq 4) - P(S_{10}\geq 5)$, since the event $(S_{10}\geq 4)$ is the union of the mutually exclusive events $(S_{10}=4)$ and $(S_{10}\geq 5)$.

The event "three or fewer are primes" is the complement of the event "four or more are primes". So the answer to (d) is found by subtracting the answer to (b) from 1

$$1 - 0.618 = 0.382$$

i.e., to find $P(S_{10} \leqslant 3)$ when $p=0.40$, we write

$$P(S_{10} \leqslant 3) = 1-P(S_{10} \geqslant 4) = 1 - 0.618 = \underline{0.382}$$

Example 2

Among the integers from 1 to 10, there are 7 numbers that are not divisible by 3, i.e., these 7 numbers are prime to 3. If 6 numbers are chosen at random, each from the full set of integers from 1 to 10, what is the probability that at least 5 of them are prime to 3?

In this problem, we are asked to find the probability of at least 5 successes in 6 Bernoulli trials with $p=0.7$. As there are no entries for $p=0.7$ in our table, we compute instead the equal probability of at most 1 failure in 6 trials, but now entering the table with the probability appropriate to a failure, namely $p=0.3$

$P(S_6 \leqslant 1)$ when $p=0.3$

$= 1 - P(S_6 \geqslant 2)$

$= 1 - 0.580$

$= \underline{0.420}$

CHAPTER SEVEN

APPLICATIONS OF THE BINOMIAL DISTRIBUTION

The Binomial Distribution based on the notion of Bernoulli Trials, is applicable to many areas: in the True-false test, in working out the probability of winning a series of games, in industrial quality control, in power supply, in vaccine test, in random walk problem and in Mendelian Hereditary theory etc. We discuss in the following pages some of the applications of the distribution.

(1) True-false Test

In a 10-question true-false examination, suppose a student tosses a fair coin to determine his answer to each question. If the coin falls heads, he answers "true"; if it falls tails, he answers "false". Six correct answers are needed to pass the examination. What is the probability that he passes the examination?

Solution:

If we assume that the probability p is the same for all trials (giving the correct answer) and that the trials are independent, then a Bernoulli process serves as a Mathematical model.

The probability of his passing the examination is

$$P(S_{10}=6) = \binom{10}{6}\left(\frac{1}{2}\right)^6\left(\frac{1}{2}\right)^4$$

$$= \underline{0.377} \quad \text{where } p=\tfrac{1}{2}$$

(2) Winning in a Series of Games[1].-

The Table-tennis champions of two schools are competing for a prize. The prize will be awarded to the one who wins a majority of the game in a series of games. Suppose one player is known to be superior to others, with probability of winning 0.6. What is the probability that the better player will win, assuming that all games in a series are played, if the series consists of 3 games, 5 games and 7 games?

Solution:

Let us first find the probability that the poorer player will win. For a 3, 5 and 7 game series. the probabilities of the poorer player winning are

$b(2;3,0.4)$; $b(3;5,0.4)$ & $b(4;7,0.4)$ which are 0.352, 0.317 & 0.290 respectively.

So the probabilities that the better player wins a 3, 5 and 7 game series are 0.648, 0.683 and 0.710

Hence the longer the series is, the higher the likelihood that the better player will win.

(3) Operation on Patients

Suppose a risky operation used for patients with no other hope of survival has a survival rate of 80%. What is the probability that exactly 80% of the next 5 patients operated upon survive?

Solution:

Denote the probability of survival by p. Then p=0.8. Since 80% of the 5 patients to be operated on

---

1. See also S. Goldberg: Probability, An Introduction, Prentice Hall, New York, pp.261-263, where the author discusses the probability of the better team winning the series in the National League Base-ball game, the American League and the World Competition Series.

will survive, the number surviving is 4.

Probability of 4 surviving out of 5 operated on

$$b(4;\ 5,\ 0.8) = {}^{5}C_{4}\ (0.8)^{4}\ (0.2)^{1}$$

$$= \underline{0.4096}$$

i.e., Probability of survival is about 41%.

(4) <u>Production of Metal Parts</u>

One percent of the metal parts produced by a machine are defective, the other 99% are good. How many parts must be produced in order for the probability of at least one defective to be ½ or more?

<u>Solution:</u>

We assume the production of parts to be a Bernoulli process for which each trial (producing one part) results in a success (defective part) or failure (good part). The probability p for success on any trial is given as $p = 0.01$. We look for the smallest integer n such that $P(Sn \geqslant 1) \geqslant \frac{1}{2}$.

$$P(Sn \geqslant 1) = 1 - P(Sn = 0)$$

$$= 1 - b(0;n,0.01)$$

$$= 1 - \binom{n}{0}\ (0.01)^{0}\ (0.99)^{n}$$

$$= 1 - (0.99)^{n}$$

$$1 - (0.99)^{n} \geqslant \tfrac{1}{2}$$

whence $\geqslant \underline{68.4}$

Hence, to have an even chance or better of finding at least 1 defective part in the lot, at least 69 parts must be produced.

(5) Industrial Quality Control

In an industrial process producing a large number of parts, we generally have a certain amount of defective output. Let the ~~process be called satisfactory~~ if the proportion of defective output is $\leq p_1$, and unsatisfactory if the proportion of defective output is $\geq p_2$ (and we have production of intermediate quality if the proportions of defective output fall between p_1 and p_2.) Suppose a manufacturer takes and inspects a sample size n from the process and find that some parts are defective. He will decide to accept or reject the production process as satisfactory or unsatisfactory according to his decision rule (n, b), in which n denotes the number of parts taken from the process and inspected and b indicates the maximun allowable number of defective parts in the sample of n for the process to be called satisfactory.

From the above we see that the process with proportion defective p_1 or less, being satisfactory, should be accepted and the process with proportion defective p_2, being unsatisfactory, should be rejected. Suppose Type 1 and Type 2 errors occur. Let the probability of rejecting the worst of the satisfactory processes (one with proportion defective p_1) be α; let the probability of accepting the best of the unsatisfactory process be β.

Suppose we are given p_1, p_2, α and β, determine the sampling plan and decision rule (n, b).

Solution:

Assuming that the sample of size n constitutes n Bernoulli trials, we have

$$\alpha = \sum_{k=b+1}^{n} {}^{n}C_k \, p_1^{\,k} \, (1-p_1)^{n-k}$$

$$\& \quad \beta = \sum_{k=0}^{h=b} {}^{n}C_k \, p_2^{\,k} \, (1-p_2)^{n-k}$$

Given p_1, p_2, α & β, we can solve for n & b by "manouvering" in the Binomial Tables.

Example:

If $p_1 = 0.02$, $p_2 = 0.07$, $\alpha = 0.05$ & $\beta = 0.10$ we find $n = 130$ & $b = 5$.

With this sampling plan and decision rule (130, 5), the risk of indicting a satisfactory process where $p = 0.02$ is 0.05 and the risk of approving an unsatisfactory process where $p = 0.07$ is 0.10.

In this problem, we apply the binomial distribution law to arrive at a certain decision rule[2], assuming that the manufacturing process is a Bernoulli process. Note, however this process is a mathematical idealization of the actual production process. From the point of view of quality control, it is desirable that the process conforms to the Binomial scheme, as with continuous control, noticeable departures can be used as an indication of impending trouble.

(6) Acceptance and Rejection: Operating Characteristic Curve.

In order to decide whether to accept or reject a very large lot of items ordered for sale, the buyer takes a sample of 20 items at random from the lot and tests them. If at most one defective item is found, he accepts the entire lot; if more than one defective item is discovered in the sample, he rejects the lot.

(a) Find the probability that the buyer accepts the lot if in fact it contains a proportion of defectives equal to p, where p assumes the value 0.01; 0.05; 0.10; 0.20; 0.30; 0.40; 0.50.

(b) Graph the probability that the buyer accepts the lot against the proportion of defectives, showing the probability of acceptance on the vertical axis.

---

2. Samuel Goldberg discusses in detail an example of testing a statistical hypothesis and illustrates the application of the Binomial Distribution in a problem of Statistical Inference. See Samuel Goldberg, *op. cit.*, pp.272-283.

(c) Draw the operating characteristic curve for the following alternative single-sample decision rule: a sample of only ten items is drawn at random from the lot tested. The lot is accepted if no defectives are found and rejected otherwise.

Solution:

The buyer takes a sample of 20 items at random. If at most 1 defective is found, he accepts the entire lot. If more than 1 defective is found, he rejects the lot.

Let p be the proportion of defectives. n here is equal to 20.

(a) Probability of accepting the lot

$= P(S_{20} \leq 1)$

$= 1 - P(S_{20} \geq 2)$

$= 1 - 0.017 = 0.983$ for $p=0.01$;

$= 1 - 0.264 = 0.736$ for $p=0.05$;

$= 1 - 0.608 = 0.392$ for $p=0.10$;

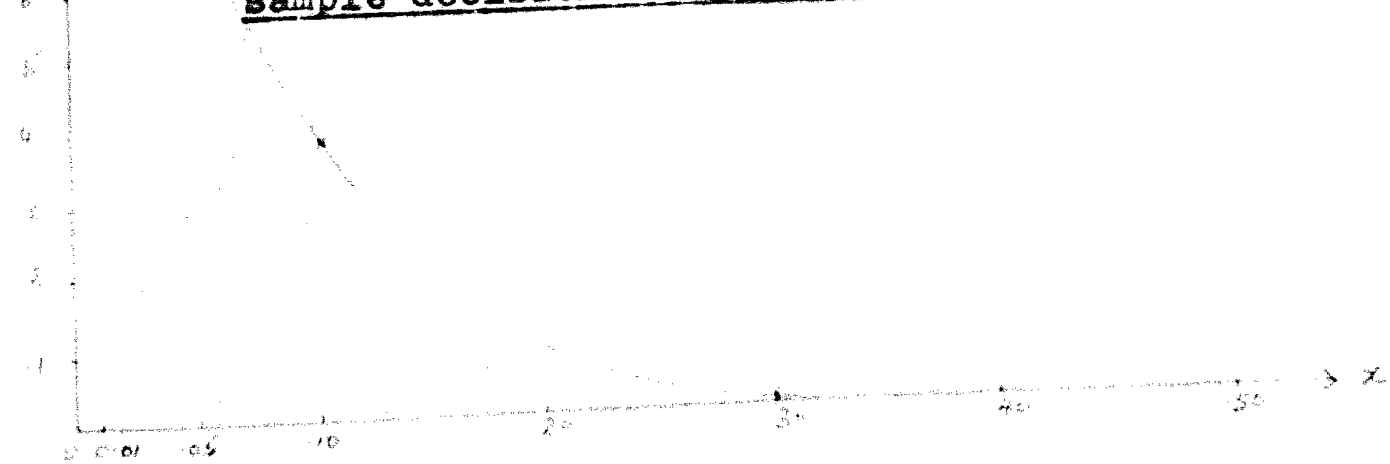
$= 1 - 0.931 = 0.069$ for $p=0.20$;

$= 1 - 0.992 = 0.008$ for $p=0.30$;

$= 1 - 0.999 = 0.001$ for $p=0.40$;

$= 1 - 1.000 = 0.000$ for $p=0.50$.

(b) Operating Characteristic Curve for the single-sample decision rule adopted by the buyer.



Proportion of Defectives

(c) Probability of accepting the lot

$= P(S_{10} \leqslant 0)$

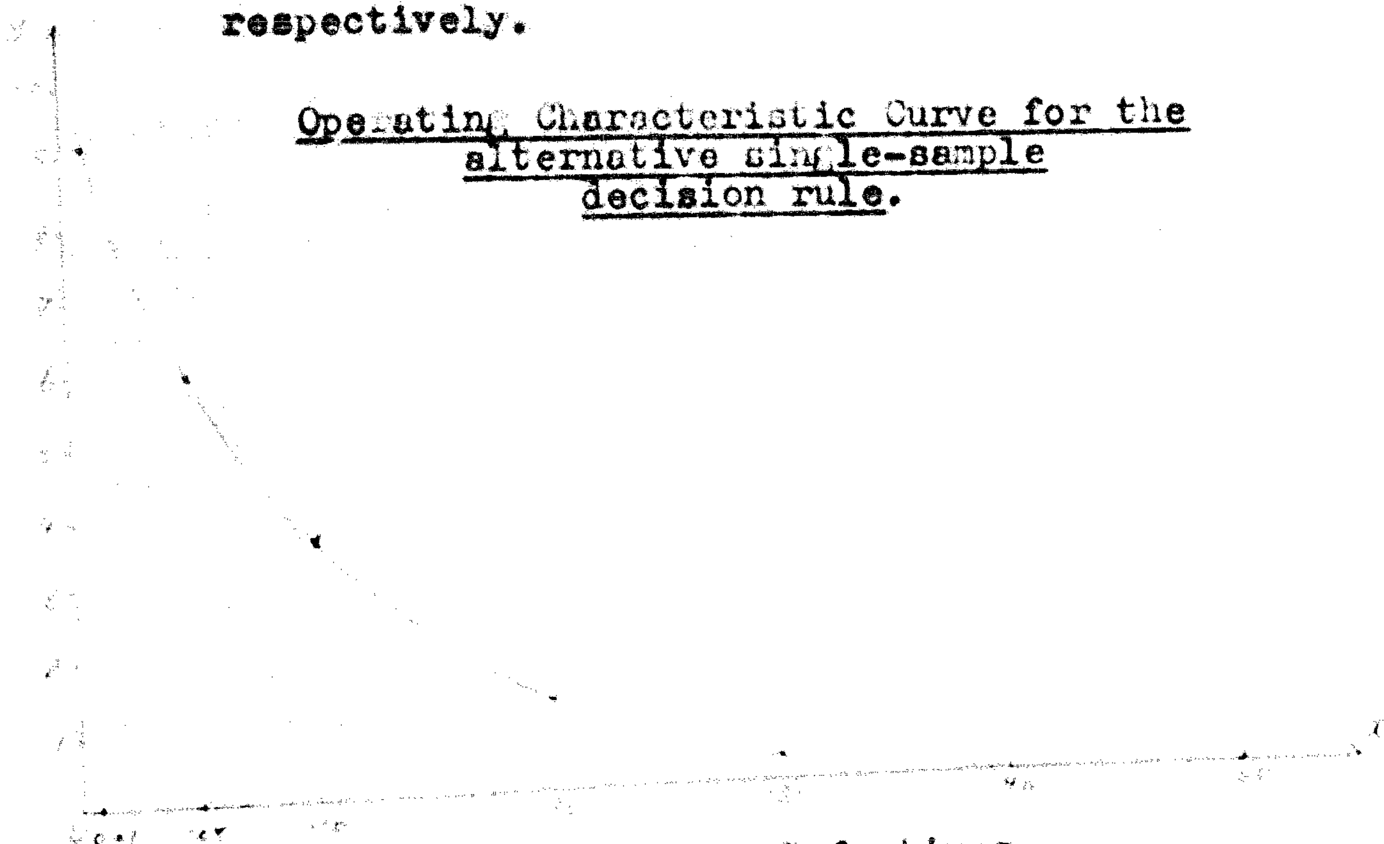
$= 1 - P(S_{10} \geq 1)$

$= 0.904;\ 0.599;\ 0.349;\ 0.107;\ 0.028;\ 0.006;$
$\&\ 0.001$

for $p = 0.01;\ 0.05;\ 0.10;\ 0.20;\ 0.30;\ 0.40;\ \&\ 0.50$

respectively.

<u>Operating Characteristic Curve for the alternative single-sample decision rule.</u>



Proportion of Defectives

(7) Testing a Statistical Hypothesis

The production manager of a company submits a report recommending hiring of additional repairmen. His conclusions are based on the assumption that, on the average, 20% of the machines in the shop will require maintenance on any given day. The president of the company is interested in testing this assumption, since the conclusions of the report will be defferent if the assumed 20% is either too high or too low. Suppose that only 20 machine-days are observed and the president is willing to take at most a 10% risk of rejecting the assumption if it is true, i.e., taking $\alpha = 0.10$ level of significence.

Solution:

To test the assumption, the president needs to formulate a null and alternative hypotheses and determine a reasonable decision rule for testing the null hypotheses.

Let the event of observing a machine for a day be a Bernoulli trial. This trial may result in success — machine needs repair or failure — machine does not require repair.

Let p be the probability of a success.

Null hypothesis: $p = 0.20$; Alternative hypothesis: $p \neq 0.20$.

The mean number of success is $np = 20 \times 0.2 = 4$ if the null hypothesis is true; we reject the null hypothesis if X, the number of successes observed, is either too much larger or too much smaller than four.

Let d denote the smallest deviation from the mean that makes X "too much larger" or "too much smaller" than the mean. Then we reject the null hypothesis if $X \leq 4-d$ or $X \geq 4+d$.

The number d is determined by requiring the probability of Type 1 error to be no larger than 0.10 (α) but as close to 0.10 as possible. This error probability is $P(X \leq 4-d) + P(X \geq 4+d)$, for $p = 0.20$

If $d=3$, $P(X \leqslant 1)+P(X \geqslant 7) > 0.10$; if $d=4$, $P(X \leqslant 1)+P(X \geqslant 7) < 0.10$

Therefore, the president will reject the null hypothesis if $X=0$ or $X \geqslant 8$.

If, of the 20 machine-days observed, seven required services of a repairman. We wish to find out the descriptive level of significance of the event.

We note that the probability that X deviates from its mean in either direction by at least as much as the observed value does is $P(X \leqslant 1)+P(X \geqslant 7)$ which is equal to $0.069+0.087=0.156$. Since 0.156 0.10, it is not significant at 0.10 level and we accept the null-hypothesis at this level.

(8) <u>Tests of Significance for Differences in samples</u>:

Sometimes we conduct a statistical investigation by selecting two groups of elementary units from the universe by a random process, designating one group the "control" group and the other the "experimental" group. The samples chosen may be independent or related. In tests of significance for differences in samples, we require as far as possible the same number of elementary units in the control group as there are in the experimental group. In related samples the samples need to be matched and paired. For instance, in consumer surveys, market researchers sometimes select two groups of families that are matched so that the two families in each pair are as nearly as possible alike in, for example, ages of the husband and wife, their level of educational attainment, number of children in family and income of the family, etc. In the case of educational experiments, the paired students are matched in characteristics like age, sex, I. Q., interests and sptitude, etc.

In this section we shall discuss the sign test of significance which is really an application of the binomial probability law. Apart from being simple and flexible, the significance-test does not require a pre-knowledge of the shape of distribution of the universe from which samples are taken and that the data need not be in the form of a truly quantitative classification.

Suppose that some perfectly matched pairs of elementary units have been selected for an experiment, and that the experiment results in two scores or two measurements for each matched pair, one measurement indicating the condition of one member of the matched pair and the other the condition of the other member.

Matched Pair	Score of one member under condition 1	Score of other member under condition 2	Difference for matched pair	Sign for Matched Pair
A_1 B_1	$x_1=63$	$y_1=68$	$y_1-x_1=+5$	+
A_2 B_2	$x_2=67$	$y_2=67$	$y_2-x_2=0$	
A_3 B_3	$x_3=69$	$y_3=64$	$y_3-x_3=-5$	−
A_4 B_4	$x_4=61$	$y_4=61$	$y_4-x_4=0$	
A_5 B_5	$x_5=64$	$y_5=66$	$y_5-x_5=+2$	+
A_6 B_6	$x_6=65$	$y_6=62$	$y_6-x_6=-3$	−
A_7 B_7	$x_7=65$	$y_7=68$	$y_7-x_7=+3$	+
A_8 B_8	$x_8=68$	$y_8=73$	$y_8-x_8=+5$	+

The premise upon which the significance test is based is that if the two conditions are equivalent, plus and minus signs would be equally likely to occur and, if it were not for chance in random sampling, one-half of the signs would be plus and the other half would be minus, i.e., the sign test is based on the binomial probability distribution $(\frac{1}{2} + \frac{1}{2})^n$ where n is the total number of plus and minus signs for the matched pairs under consideration. In the sign test, we assume that the underlying variable is continuous, and continuity implies that, theoretically, the matched pair could have tied scores or measurements; thus matched pairs that produce zero differences are dropped from the analysis.

In the table, we have two minus signs and four plus signs. Let $n_1=2$, the smaller number of signs. Let $n_2=4$. Then $n=n_1+n_2=2+4=6$. The probability of obtaining not more than two minus signs in a random sample of 6 signs if $p=q=½$ is found by adding the first three terms of the binomial expansion $(½+½)^6$.

Thus we have

$$(½)^6+{}^6C_1(½)^1(½)^5+{}^6C_2(½)^2(½)^4 = 0.344$$

Thus the probability for one tail test is 0.344. If we conduct our test at $\alpha=0.10$ level of significance, since Probability=0.344 is greater than $\alpha=0.10$, we cannot reject the null hypothesis.

This example is equally applicable to a "Before-and-after" experiment. For instance, a random sample of people is selected from a universe and the people are classified according to whether they are in favour of or opposed to a proposal. These people are then exposed to a publicity campaign with the purpose of influencing them to develop a favourable attitude toward the proposal. Finally, they are re-classified according to whether they are now in favour of or opposed to the proposal. A plus sign is assigned to those who change in a desirable way and a minus sign otherwise. The null hypothesis for the investigation is that the publicity campaign will have no effect on the people in the universe from which the sample is drawn, the alternative hypothesis is that the publicity campaign causes more people to change in a favourable way. We thus have a one-tail test and the conclusion is the same as the one stated above.

In the table, the two conditions stated may be:-

(i) two different methods of on-the-job trainin in a factory;

(ii) two different medical treatments for a certain complaint; or

(iii) two different ways of teaching a subject to school-children.

Let us now consider the following hypothetical case where the sign test is applicable.

In an experiment to determine ways to improve the working relationships in its factories, a large corporation selected 25 foremen at random from the several hundred in its plants. Then, without the foremen or their men knowing it, a record was kept of all complaints made to management by the workmen about the 25 foremen during a period of six months. After that, the 25 foremen were sent to a management training school maintained by the corporation; there they attended for four weeks a programme of lectures and discussion meetings dealing with human relations. For the next six months after the 25 foremen returned to their jobs a record was again kept of all the complaints made about them to management by the workmen. A comparison of the records of each of the 25 foremen for the six months prior to the training period and the six months following the training period shows the following results:-

Foreman	No. of Complaints after training	Foreman	No. of Complaints after training
A	Less	N	Same
B	Same	O	Less
C	More	P	More
D	Less	Q	Less
E	Less	R	Less
F	Less	S	More
G	Same	T	More
H	Less	U	Less
I	More	V	Less
J	Less	W	Less
K	Less	X	Same
L	Less	Y	Less
M	More		

Assuming that all other conditions of work remained the same during the two six-month work periods, test the null hypothesis that education in human relations has no effect on the ability of foremen to get along with the men who work for them. Use as alternative the statement that foremen who have attended a training course in human relations tend to get along better with their men. Make the test at the $\alpha = 0.05$ level of significance.

We assign a plus sign to those foremen who get a less number of complaints after training from the workers and a minus sign for more complaints. Therefore, we have six minus signs and fifteen plus signs.

Let $n_1 = 6$, the smaller number of signs. Let $n_2 = 15$, then $n = n_1 + n_2 = 21$.

The probability of obtaining not more than six minus signs is found by adding the first seven terms of the binomial distribution $(\frac{1}{2}+\frac{1}{2})^{21}$.

The sum is

$$(\tfrac{1}{2})^{21} + {}^{21}C_1(\tfrac{1}{2})^1(\tfrac{1}{2})^{20} + {}^{21}C_2(\tfrac{1}{2})^2(\tfrac{1}{2})^{19} + \dots {}^{21}C_6(\tfrac{1}{2})^6(\tfrac{1}{2})^{15}$$

$$= 0.039$$

Since $0.039 < 0.05$, we reject the null hypothesis in this one-tail test.

In other words, the foremen who have attended a human-relationship course tend to get along better with their workers.

(9) <u>Power Supply</u>

Suppose that $n = 10$ workers are to use intermittently electric power, and we wish to estimate the total load to be expected. For rough approximation, imagine that at any given time each worker has the same probability p of requiring a unit of power. If they work independently, the probability of exactly k workers requiring power at the same time should be $b(k;n,p)$. If on the average, a worker uses power for 24 minutes per hour, we put $p = \frac{24}{60} = 0.4$. The probability of 7 or more workers requiring current at the same time is then $P(S_{10} \geqq 7)$ Given $p = 0.4$, $P(S_{10} \geqq 7) = 0.055$. In other words, if the supply is adjusted to 6 power units, an overload has probability 0.055, and should be expected for one in 18 minutes approximately. The probability of eight or more workers requiring current at the same time is 0.012, about 4½ times less.

On the other hand, if on the average, the consumption of power by a worker is reduced by half; i.e., using power for 12 minutes per hour only, we have $p = 0.2$. For the probability of 7 or more workers requiring current at the same time is $P(S_{10} \geqq 7) = 0.00086$ (with $p = 0.2$), i.e., an overload has probability 0.00086, assuming that the supply is also adjusted to six power units, and should be expected for about one minute in 1157 or one minute in 20 hours. The probability of eight or more workers requiring current at the same time is 0.0000779, about eleven times less.

(10) Testing Sera or Vaccines

Assume that the normal rate of infection on a certain disease in cattle is 25%. To test a newly discovered serum, we injected n healthy animals with it. How are we to evaluate the result of the experiment?

If the serum is completely worthless, the probability that exactly k of the n test animals remain free from infection is $b(k;n,0.75)$. Now assume that we inject all the $n=10$ healthy animals with the serum. That is, $k=n=10$. Probability of 10 animals out of 10 tested remaining free from infection, i.e., out of 10 test animals, probability of none of the animals catching infection

$$= \binom{n}{k} p^k q^{n-k} = \binom{10}{10} (0.75)^{10} (0.25)^0$$

$$= 0.056 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

Next assume that we inject all the n=12 animals, i.e., k=n=12.

Probability of 12 animals out of 12 tested remaining free from infection, or in other words, out of 12 test animals the probability of none of the animals catching infection

$$= \binom{12}{12} (0.75)^{12} (0.25)^{0} = 0.032 \quad \ldots\ldots\ldots\ldots \text{(ii)}$$

If there is no serum, the probability that out of 17 animals at most one catches infection is

$$\binom{17}{0} (0.25)^{0} (0.75)^{17} + \binom{17}{1} (0.25)^{1} (0.75)^{16}$$

$$= 0.0501 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iii)}$$

Comparing probabilities (i) & (iii), since $0.0501 < 0.056$

We conclude that there is stronger evidence in favour of the serum.

For n=23, the probability of at most two animals catching infection is about 0.0492 and thus 2 failures out of 23 is again better evidence for the serum than one out of 17 or none out of 10.

(11) <u>A Rocket Designer's Problem</u> (<u>Prendergast</u>)

F, the rocket designer, has come to B, the reliability expert, with a problem:

"The vehicle is designed. We can use two large motors or four small motors and get the same thrust and the same weight. However, we know that the motors are subject to catastrophic failure and we have designed so that we will still get into orbit if half the motors fail. Now if you will tell me the probability of a motor failing in time required to get into orbit, I can decide to use two or four."

B replied, "We have analyzed the test data on the motors, and have found that the large and small motors have the same probability of failing in a given time. I can assure you that it makes no difference whether you use two or four motors. However, this failure probability is classified top secret and I cannot give it to you."

F said, "Never mind. From what you've just told me, I can calculate the failure probability for myself, for a motor and for the rocket."

What is the failure probability for a motor and for the rocket?

Solution:

Let the event "motor failing" be considered as a success and the probability of the small motor failing be p. The probability of 3 or 4 small motors failing (and thereby, of the rocket failing) is

$$^4C_3\ p^3\ q^1 + {}^4C_4\ p^4\ q^0 = 4\ p^3\ q + p^4$$

The probability of 2 large motors failing (and thereby of the rocker failing) is

$$^2C_2\ p^2\ q^0 = p^2$$

Since the large and small motors have the same probability of failing in a given time, we have

$$4\ p^3\ q + p^4 = p^2$$

$$p^4 + 4p^3q - p^2 = 0$$

$$p^2(p^2+4pq-1) = 0$$

$$p^2\left\{p^2+4p(1-p)-1\right\} = 0 \quad \text{since } p+q=1.$$

$$p^2-4p^2+4p-1 = 0$$

$$3p^2-4p+1 = 0$$

$$(3p-1)\ (p-1) = 0$$

$$p = \frac{1}{3}$$

(12) One- and Two-engine planes, etc.

Suppose that in flight, aeroplane engines fail with probability q, independently from engine to engine and that a plane makes a successful flight if at least half of its engines run. For what values of q is a single-engine plane to be preferred to a two-engine plane? Assume that the probability of an engine not failing is p=1-q.

Solution:

In the case of a single engine plane, the flight will be successful if one engine runs.

Probability of successful flight = p = 1-q.

For the two-engine plane, the flight is successful if 1 or 2 engines run.

Probability of successful flight = 1 - Probability of 2 engines failing = $1 - q^2$

Since $q > q^2$, $-q^2 > -q$ and $1-q < 1-q^2$, so for all $p \neq 0, 1$, the two-engine plane is preferable.

For instance p=0.3. Probability of successful flight for two-engine plane is $1-(0.7)^2=0.51$ and that for one-engine plane is only 0.3.

Now let us compare the performance of two-engine plane with that of a four-engine plane, using the same assumption as those of the above example.

Let X be the number of engines that do not fail.

For the two-engine plane, the probability of successful flight is

$$P(X \geq 1) = 1-P(0) = 1-q^2$$

For the four-engine plane, the corresponding probability is

$$\ldots 2) = 1 - P(0) - P(1)$$

$$= 1 - {}^4C_0\, p^0\, q^4 - {}^4C_1\, p^1\, q^3$$

$$= 1 - q^4 - 4pq^3$$

$$= 1 - q^4 - 4(1-q)q^3.$$

$$= 1 - 4q^3 + 3q^4.$$

For a two-engine plane to be preferable to a four-engine one, we should have

$$1-q^2 > 1-4q^3+3q^4$$

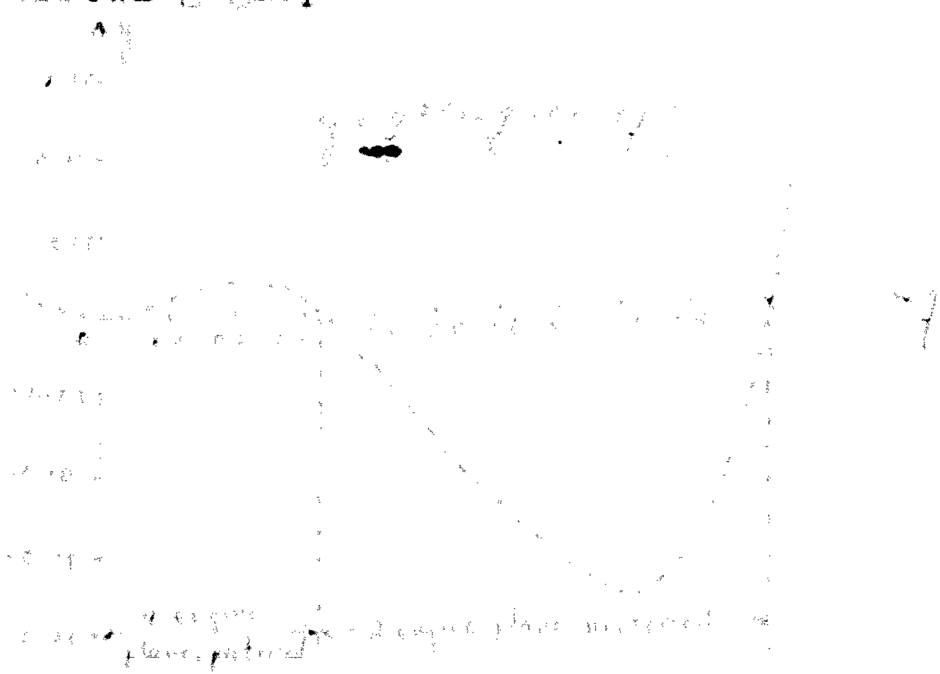
$$-3q^4+4q^3+q^2 > 0$$

$$3q^4-4q^3+q^2 < 0$$

$$q^2\ (1-q)\ (1-3q) < 0$$

For $q=0$, 1, or $\frac{1}{3}$, the probabilities for each kind of aeroplane are the same.

For values of q between $\frac{1}{3}$ & 1, however, a two-engine plane is preferable, as shown by the following graph:



$$P(X \geq 2) = 1 - P(0) - P(1)$$

$$= 1 - {}^4C_0\, p^0\, q^4 - {}^4C_1\, p^1\, q^3$$

$$= 1 - q^4 - 4pq^3$$

$$= 1 - q^4 - 4(1-q)q^3$$

$$= 1 - 4q^3 + 3q^4.$$

For a two-engine plane to be preferable to a four-engine one, we should have

$$1-q^2 > 1-4q^3+3q^4$$

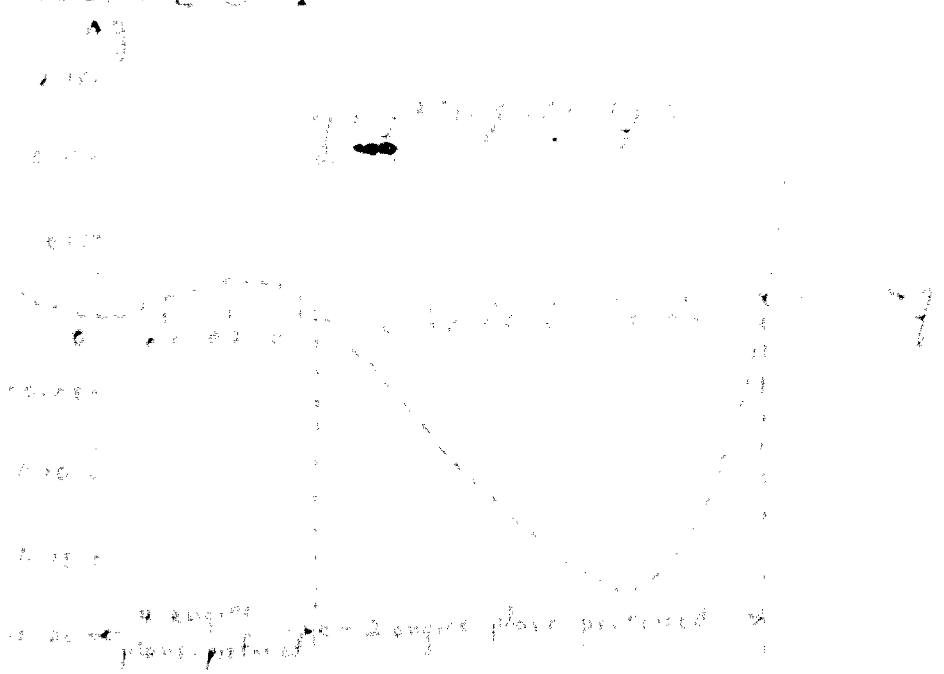
$$-3q^4+4q^3+q^2 > 0$$

$$3q^4-4q^3+q^2 < 0$$

$$q^2\,(1-q)\,(1-3q) < 0$$

For $q=0$, 1, or $\frac{1}{3}$, the probabilities for each kind of aeroplane are the same.

For values of q between $\frac{1}{3}$ & 1, however, a two-engine plane is preferable, as shown by the following graph:



Using the same assumptions, for what values of q is a two-engine plane to be prefered to a 3-engine plane?

Working out the probabilities of successful flight for the one-engine plane and three engine plane and for the two-engine plane to be preferrable we have

$$1 - q^2 > 1 - {}^3C_0\, p^0 q^3 - {}^3C_1\, p^1 q^2$$

i.e.,

$$1 - q^2 > 1 - q^3 - 3p^1 q^2$$

$$q^3 - q^2 + 3p^1 q^2 > 0$$

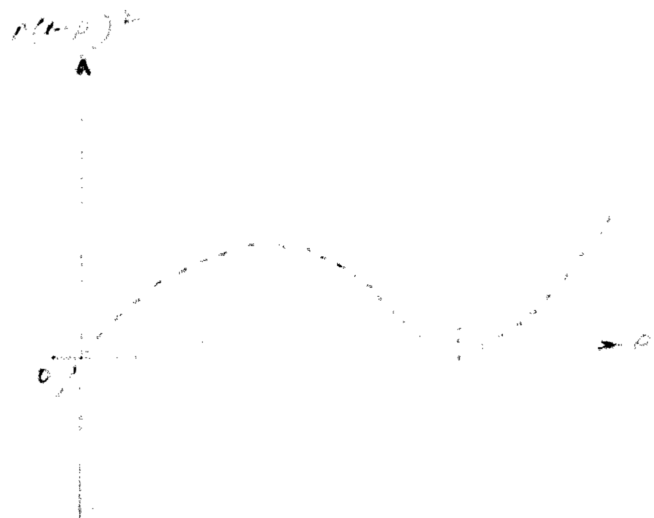
$$q^2\,(q-1+3p) > 0$$

$$q^2\,(-p+3q) > 0$$

$$2p\,(1-p)^2 > 0$$

or

$$p\,(1-p)^2 > 0$$



Thus the two-engine plane is preferred for all values except 0 and 1 where we are indifferent between the planes.

(15) Random Walk Problem

The Binomial Distribution theorem states that the probability of exactly k successes in n Bernoulli trials is

(i) $^nC_k p^k q^{n-k}$ where p denotes the probability of success and q that of failure.

In terms of random variables, this theorem can be restated as follows:-

If $X_1, X_2 \ldots\ldots X_n$ are stochastically independent random variables each of which assumes the values 0 and 1 with probability q and p respectively, then

$$Sn = k = \sum_{i=1}^{n} Xi$$

is a random variable with probability function:

(ii) $P(Sn=k) = {}^nC_k p^k q^{n-k}$

We will make use of this in solving random walk problems.

Suppose a point starts from the origin and moves along the X-axis in jumps of 1 unit each. The point may move forward or backward 1 unit. We assume that at each step the probability for each direction is ½, and that each jump is independent of all the others. After n jumps, the point may be at any one of the points in the range −n to +n. We wish to find the probability of its being at each of the possible points in the range −n to +n.

If we let Xi (i=1, 2, ... n) be the displacement on the ith jump, the Xi's, being independent random variables, will each have the following probability function

Xi	−1	+1
P(Xi)	½	½

The net displacement after n jumps is the sum of the n individual displacements; and this sum is the same as the abscissa of the point. Let the abscissa be X

$$X = \sum_{i=1}^{n} Xi$$

These variables do not fit (ii) above; but variables

$$Zi = \frac{Xi + 1}{2} \quad \text{do fit.}$$

From (ii), we have

$$P\left\{\sum_{i=1}^{n} Zi = k\right\} = {}^{n}C_{k} \, (½)^{k} \, (½)^{n-k}$$

$$= {}^{n}C_{k} \, (½)^{n}$$

$$= \frac{{}^{n}C_{k}}{2^{n}}$$

Note that

$$\sum_{i=1}^{n} Zi = \sum_{i=1}^{n} \frac{Xi+1}{2} = \frac{X+n}{2}$$

X=a means $\sum Zi = \frac{a+n}{2}$

$$P\left\{X = a\right\} = \frac{{}^{n}C_{(a+n)/2}}{2^{n}}$$

Thus the probability function for X is, in general case,

$$\frac{{}^{n}C_{(x+n)/2}}{2^{n}}$$

In the case of 2 dimensional random walk problem, where the point may more forward, backward, up or down 1 unit, with probability $\frac{1}{4}$, the point probability function for the abscissa and ordinate of the moving point after n steps is

$$\frac{{}^{n}C_{(x+n)/2} \cdot {}^{n}C_{(y+n)/2}}{4^{n}} \quad ^{3}$$

(14) <u>Application in Genetics: The Mendelian Hereditary Theory</u>

The Mendelian Theory of heredity provides an interesting illustration of the application of the Binomial distribution.

Heritable characters depend on genes. These genes, which lie on the chromosomes, appear in pairs. The chromosomes, visible in the cells of an organism, appear in pairs too and paired genes occupy the same position on paired chromosomes. In the simplest case, each gene of a particular pair can assume two forms G and g. Three different pairs can be formed, and, with respect to this particular pair, the organism belongs to one of the three genotypes GG, Gg and gg. Each pair of genes determines one heritable factor, though the majority of observable properties of organisms depend on several factors.

In this section, we discuss genotypes and inheritance for only one particular pair of genes.

The reproductive cells or gametes, are formed by a splitting process and receive one gene only. Organisms of the pure GG and gg-genotypes (or homozygotes) produce therefore gametes of only one kind, but Gg-organisms (hybrids or heterozygotes) produce G- and g-gametes in equal numbers. New organisms receive their genes from two parental gametes, each pair including a paternal and a maternal gene.

---

3. See M. E. Munroe: <u>Theory of Probability</u>, McGraw-Hill Book Company, Inc., N. Y., 1951, pp.68-69.

In his book[4], J. Neyman formulates a few axioms for solving probabilistic problems in genetics. As we are interested in the probability of the progeny inheriting a specified combination of genes, given that the parents or the grandparents, etc., possess some particular genetical composition, the following axioms are necessary:-

<u>AXIOM 1</u>:

An even number, say 2n, of reproductive cells contained in each of the parental organisms are to be fertilized. The genes carried in the reproductive cells, produced by the parental organisms, are those present in the parental organisms alone and no others -- mutation being ignored. If the parental organism carries two identical genes, then the reproductive cells carry the same gene. If the parental organism carries two different genes, for example g and G, then one of the reproductive cells carries gene g and the other gene G.

---

4. J. Neyman: <u>First Course in Probability & Statistics</u>, Holt, Rinehart & Winston, Inc., N. Y.

The discussion here is based on Neyman's work.

Note the special notation used by J. Neyman:-

Let M stand for mother, F for father, C for the child, ξ for a specified combination of genes, g, G, h, H for some two pairs of genes, X and Y the maternal and paternal reproductive cells which combine to produce the child C.

M: ξ denotes that M possesses the particular combination of genes ξ.

P{C: gG, hH/(M:gG,hh)(F:gg,hH)} stands for the probability that the child will inherit the combination gG, hH given that the mother and the father have the combinations gG, hH and gg, hH respectively.

C: gG=(X:g)(Y:G)+(X:G)(Y:g) indicates that for the child to be a hybrid gG, it is necessary (and sufficient) that one reproductive cell carries the recessive gene and the other the dominant gene.

AXIOM 2:

Fertilization is random. Suppose the number of reproductive cells contained by M the mother be 2n' and that contained by F the father be 2n". The forthcomeing organism C selects one reproductive cell from M and one from F. The probability of selecting a cell from M and F will be $\frac{1}{2n'}$ and $\frac{1}{2n''}$ respectively. The two selected reproductive cells combine to produce the first cell of C.

AXIOM 3:

The genetical composition of the reproductive cell selected by C from M is independent of the genetical composition of the reproductive cell selected from F.

Inheritence of a Single Pair of genes

Suppose an organism carries single pair of genes g and G. There are three possible combinations: gg, gG and GG and an organism will carry one of these three combinations. Let ξ, η and ζ each stand for any of these combinations. We are interested to compute probability of the child C inheriting the composition ξ given that the genetical composition of the mother M and father F is η and ζ respectively.

i.e., $P\left\{C: \xi/(M:\eta)(F:\zeta)\right\}$

As

$$C:gg = (X:g)(Y:g)$$
$$C:gG = (X:g)(Y:G) + (X:G)(Y:g)$$
$$C:GG = (X:G)(Y:G)$$

Here X and Y denote the reproductive cell in Mother M and Father F, and applying Axioms 1, 2 and 3 and the Addition and Multiplication principles, we arrive at the following table:

Table 7

Probabilities of Inheritance of a single pair of genes

Mother \ Father	gg	gG	GG
gg	$P\{C:gg\} = 1$	½	0
	$P\{C:gG\} = 0$	½	1
	$P\{C:GG\} = 0$	0	0
gG	$P\{C:gg\} = ½$	¼	0
	$P\{C:gG\} = ½$	½	½
	$P\{C:GG\} = 0$	¼	½
GG	$P\{C:gg\} = 0$	0	0
	$P\{C:gG\} = 1$	½	0
	$P\{C:GG\} = 0$	½	1

Example:

Let us see how we arrive at $P\{C:gg\} = ¼$, given that both M and F have the genetical composition gG.

$$C:gg = (X:g)(Y:g) \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (7.1)$$

whatever the genetical composition η and ζ of the mother and father,

$$P\{C:gg/(M:\eta)(F:\zeta)\}$$
$$= P\{(X:g)(Y:g)/(M:\eta)(F:\zeta)\} \quad \ldots\ldots\ldots\ldots \quad (7.2)$$

Using (a) Axiom 2 stating that the genetical composition of C is determined by the genes carried in the reproductive cell X selected from M and in the reproductive cell Y selected from F,

and (b) Axiom 1 asserting that a reproductive cell may carry either gene g or gene G but not both of them.

Now formula (7.2) becomes

$$P\{C:gg/(M:\eta)(F:\xi)\} = P\{X:g/M:\eta\}\,P\{Y:g/F:\xi\}$$

$$\ldots\ldots\ldots\ldots\ldots \quad (7.3)$$

since Axiom 3 states that the genetical composition of the male reproductive cell is independent of the genetical composition of the female reproductive cell. This independence has enabled us to apply the multiplication principle.

In our example, we are given that M:gG. Axiom 1 asserts that one-half of the reproductive cells of M will carry gene g and the other half gene G. Axiom 3 implies that $P\{X:g/M:gG\} = ½$. Similarly $P\{Y:g/F:gG\} = ½$.

Therefore, from (7.3)

$$\begin{aligned} &P\{X:g/M:\eta\}\,P\{Y:g/F:\xi\} \\ &= ½ \times ½ \\ &= ¼ \end{aligned}$$

$$\underline{P\{C:gg\} = ¼ \text{ if } M:gG \text{ and } F:gG.}$$

We can similarly compute the probabilities of C:gG and C:GG for any of the possible combination of genetical compositions of M and F.

Note that

$$C:gG = (X:g)(Y:G) + (X:G)(Y:g)$$

and $P\{C:gG/(M:\eta)(F:\xi)\}$

$$\begin{aligned} &= P\{(X:g)(Y:G) + (X:G)(Y:g)/(M:\eta)(F:\xi)\} \\ &= P\{(X:g)(Y:G)/(M:\eta)(F:\xi)\} + P\{(X:G)(Y:g)/(M:\eta)(F:\xi)\} \\ &= P\{X:g/M:\eta\}\,P\{Y:G/F:\xi\} + P\{X:G/M:\eta\}\,P\{Y:g/F:\xi\} \end{aligned}$$

using the axioms and the Addition and Multiplication Principles.

Note also that

$$P\{X:g/M:gg\} = P\{X:G/M:GG\} = 1$$

$$P\{X:g/M:gG\} = P\{X:G/M:gG\} = ½$$

and $$P\{X:g/M:GG\} = P\{X:G/M:gg\} = 0$$

Successive generations

Let us now consider problems on the distribution of various genetical types among the individuals forming successive generations which reproduce under a system of random (Mendelian) mating, technically called Panmixa. The mates are selected independently of their hereditary characteristics. If r descendants, for instance, in the first filial generation are chosen at random, then their parents form a random sample of size r, with possible repetitions from the aggregate of all possible parental pairs. In other words, each descendant is to be regarded as the product of a random selection of parents, and all selections are mutually independent.

Successive generations under Panmixia with no selection -- case of one pair of genes.

Consider a single pair of genes g, G and a sequence of successive generations.

$$II_0; \quad \pi_1, II_1; \quad \pi_2, II_2; \; \ldots\ldots \; ; \quad \pi_n, II_n$$

where II_0, II_1, ... indicate original generation born, first generation born ..... with probabilities of distribution of genetical types in II_0, II_1 denoted by

P_{01}, P_{02}, ... P_{11}, P_{12}, ...

and

π_1 π_2 ..... represent groups of individuals in first generation mating, second generation mating .... with probabilities of genetical types in π_n denoted by P_{n1}, P_{n2}, ...

Since there are only 3 possible genetical types (where n=1, 2.) denote GG, gG and gg, let Pn, Qn and Rn the probability that an individual of the n^{th} generation born Π_n be a dominant, a hybrid and a recessive respectively.

For computing Pn, Qn and Rn, we make the following assumptions:-

(i) the probabilities of the three genetical types in each generation born are the same for males and for females;

(ii) each generation mating π_n, beginning with n=1, is obtained from the preceding generation born without selection;

(iii) the probabilities in π_1 of the three different genetical types are as follows:

Types:	GG	gG	gg	
Probabilities in	p'	q'	r'	for females
	p"	q"	r"	for males

(iv) the mating in all generations π_1, π_2 ..... is panmixia.

Before coming to the general case, Pn, Qn and Rn, we first compute P_1, Q_1 and R_1. It can be shown that[5]

$$\left.\begin{aligned} P_1&=(p'+\tfrac{1}{2}q')(p''+\tfrac{1}{2}q'') \\ Q_1&=(p'+\tfrac{1}{2}q')(r''+\tfrac{1}{2}q'')+(r'+\tfrac{1}{2}q')(p''+\tfrac{1}{2}q'') \\ R_1&=(r'+\tfrac{1}{2}q')(r''+\tfrac{1}{2}q'') \end{aligned}\right\} \text{— (7.4)}$$

If the distribution of genetical types in π_1 among fathers is the same as that among mothers, so that p'=p"=p, q'=q"=q and r'=r"=r where p, q and r are arbitrarily assigned constants, then

5. See Appendix 2.

$$\left.\begin{aligned} P_1 &= (p+\tfrac{1}{2}q)^2 \\ Q_1 &= 2(p+\tfrac{1}{2}q)(r+\tfrac{1}{2}q) \\ R_1 &= (r+\tfrac{1}{2}q)^2 \end{aligned}\right\} \text{ -- } (7.5)$$

As $p+q+r=(p+\tfrac{1}{2}q)+(r+\tfrac{1}{2}q)=1$, we have

$$\left.\begin{aligned} P_1 &= \left[1-(r+\tfrac{1}{2}q)\right]^2 = (1-\sqrt{R})^2 \\ Q_1 &= 2\sqrt{P_1}\sqrt{R_1} = 2(1-\sqrt{R_1})\sqrt{R_1} \end{aligned}\right\} \text{ -- } (7.6)$$

Since there is no selection, the probabilities P_n, Q_n and R_n must be connected with P_{n-1}, Q_{n-1} and R_{n-1} by the same relation (7.5) which connect P_1, Q_1 and R_1 with p, q and r.

$$\therefore \quad P_2 = (P_1+\tfrac{1}{2}Q_1)^2$$

$$Q_2 = 2(P_1+\tfrac{1}{2}Q_1)(R_1+\tfrac{1}{2}Q_1)$$

$$R_2 = (R_1+\tfrac{1}{2}Q_1)^2$$

Thus we can deduce the following:-

$$P_2 = P_3 = P_4 \cdots\cdots\cdots P_n$$

$$Q_2 = Q_3 = Q_4 \cdots\cdots\cdots Q_n$$

$$R_2 = R_3 = R_4 \cdots\cdots\cdots R_n$$

The distribution P_2, Q_2 and R_2 in the second generation depends on whether or not, in the first generation mating, the distribution of the genetical types among the fathers is the same as that among the mothers.

In the general case,

$$\left.\begin{aligned} P_2 &= (P_1+\tfrac{1}{2}Q_1)^2 = \left[\frac{p'+p''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right]^2 \\ Q_2 &= 2(P_1+\tfrac{1}{2}Q_1)(R_1+\tfrac{1}{2}Q_1) \\ &= 2\left[\frac{p'+p''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right]\left[\frac{r'+r''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right] \\ R_2 &= (R_1+\tfrac{1}{2}Q_1)^2 = \left[\frac{r'+r''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right]^2 \end{aligned}\right\} \quad (7.7)$$

Thus, if the distribution is the same, i.e., if $p'=p''$, $q'=q''$ and $r'=r''$, then $P_2=P_1$, $Q_2=Q_1$ and $R_2=R_1$. Otherwise, the distribution of genetical types in the second generation born need not be the same as in the first.

Successive generations under Panmixia and mass selection against recessive.

J. Neyman defines this as the selection of the n^{th} generation mating π_n out of the preceding generation born Π_{n-1}, which consists in including in π_n all the dominants and all the hybrids present in Π_{n-1}, but none of the recessives.

We assume the following:

(a) The original generation Πo is born out of panmixia with respect to a single pair of genes g, G which are not linked with sex. Therefore, by formula (7.5), the distribution of the 3 genetical types in Πo is determined by the probabilities:

$$P\{GG/\Pi o\} = Po = (1-\sqrt{Ro})^2,$$

$$P\{Gg/\Pi o\} = Qo = 2(1-\sqrt{Ro})\sqrt{Ro},$$

$$P\{gg/\Pi o\} = Ro$$

(b) Out of each generation born, all the recessives are removed and the remaining dominants and hybrids mate according to panmixia.

We are interested in finding the probabilities Pn, Qn and Rn. As each generation mates under panmixia and as there is no linkage with sex, we have

$$\left.\begin{aligned} Pn &= (1-\sqrt{Rn})^2 \\ Qn &= 2(1-\sqrt{Rn})\sqrt{Rn} \quad \text{for } n=1,2,\ldots \end{aligned}\right) \quad \text{--} \quad (7.8)$$

We need to solve for Rn, for n=1,2, ... This Rn is connected with p_n, q_n and r_n which show the probabilities that a member of the n^{th} generation mating will be a dominant, hybrid or recessive. Here $r_n=0$, but in formula (7.4), we note that $R_1=(r+\frac{1}{2}q)^2$. Therefore, $R_n=(\frac{1}{2}q_n)^2$ for each n. By applying the theorem on relative probability, we have

$$Rn = (\tfrac{1}{2}q_n)^2 = \frac{R_{n-1}}{(1+\sqrt{R_{n-1}})^2} \qquad (7.9)$$

By successive substitution and by Mathematical induction, we get

$$Rn = \frac{Ro}{(1+n\sqrt{Ro})^2} \qquad (7.10)$$

This formula shows that, as the number of successive applications of mass selection is increased, the proportion Rn of recessives in the generation born becomes smaller and smaller.

We append herewith two worked examples in probabilistic problems of genetics.

(1) <u>Example illustrating successive generations under panmixia with no selection.</u>

In a population the distribution of genetical types is as follows:-

Types	GG	gG	gg
Distribution of Females	0.1	0.5	0.4
Distribution of Males	0.6	0.3	0.1

Compute the distribution of genetical types in the two successive generations which follow π under panmixia and without selection.

As mating in generations Π_1, Π_2, ... is panmixia with respect to the genes g and G and there is no selection, we have, by formula (7.4)

$$P_1=(p'+\tfrac{1}{2}q')(p''+\tfrac{1}{2}q'')$$

$$Q_1=(p'+\tfrac{1}{2}q')(r''+\tfrac{1}{2}q'')+(r'+\tfrac{1}{2}q')(p''+\tfrac{1}{2}q'')$$

$$R_1=(r'+\tfrac{1}{2}q')(r''+\tfrac{1}{2}q'')$$

where P_1, Q_1 and R_1 denote the probability that an individual of the first generation born Π_1 will be a dominant GG, a hybrid gG and a recessive gg respectively, and p', q', r' and p'', q'', r'' represent distribution of types for females and males respectively.

$$\begin{aligned} P_1 &=(0.1+\tfrac{1}{2}\times 0.5)(0.6+\tfrac{1}{2}\times 0.3)\\ &= 0.35 \times 0.75\\ &= \underline{0.2625} \end{aligned}$$

$$\begin{aligned} Q_1 &=(0.1+\tfrac{1}{2}\times 0.5)(0.1+\tfrac{1}{2}\times 0.3)+(0.4+\tfrac{1}{2}\times 0.5)(0.6+\tfrac{1}{2}\times 0.3)\\ &= 0.35 \times 0.25 + 0.65 \times 0.75\\ &= 0.0875 + 0.4875\\ &= \underline{0.5750} \end{aligned}$$

$$\begin{aligned} R_1 &=(0.4+\tfrac{1}{2}\times 0.5)(0.1+\tfrac{1}{2}\times 0.3)\\ &= 0.65 \times 0.25\\ &= \underline{0.1625} \end{aligned}$$

Let P_2, Q_2 and R_2 stand for the probability that an individual of the second generation born will be GG, gG and gg respectively. We have, by formula (7.7),

$$P_2 = \left[\frac{p'+p''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right]^2$$

$$= \left[\frac{0.1+0.6}{2} + \frac{1}{2}\times\frac{0.5+0.3}{2}\right]^2$$

$$= \left[\frac{0.7}{2} + \frac{1}{2}\times\frac{0.8}{2}\right]^2$$

$$= \underline{0.3025}$$

$$Q_2 = 2\left[\frac{p'+p''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right]\left[\frac{r'+r''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right]$$

$$= 2\,(0.55)(0.45)$$

$$= \underline{0.4950}$$

$$R_2 = \left[\frac{r'+r''}{2} + \frac{1}{2}\times\frac{q'+q''}{2}\right]^2$$

$$= \left[\frac{0.4+0.1}{2} + \frac{1}{2}\times\frac{0.5+0.3}{2}\right]^2$$

$$= \left[0.25 + 0.2\right]^2$$

$$= \underline{0.2025}$$

(2) Example illustrating successive generations under panmixia and mass selection against recessives.

In a population Π the distribution of three genetical types with respect to genes g and G is the same among males and females. The proportion of recessives in the population II born out of panmixia in Π is equal to 0.81. How many times must the process of mass selection be applied to reduce the proportion of recessives to something less than one percent?

$$R_n = \frac{R_0}{(1+n\sqrt{R_0})^2}$$ shows that the proportion R_n of recessives in the generations born becomes smaller and smaller as the number of successive applications of mass selection is increased. To determine the generation $II_{n(\alpha)}$, such that, beginning with this generation and in all that follow, the probability of a recessive will be smaller than , we need to solve the inequality

$$R_n = \frac{R_0}{(1+n\sqrt{R_0})^2} < \alpha$$

In our problem, we have $R_0 = 0.81$ and $\alpha = 0.01$ using

$$\frac{R_0}{(1+n\sqrt{R_0})^2} < \alpha$$, we have

$$n > \frac{1}{\sqrt{\alpha}} - \frac{1}{\sqrt{R_0}} = \frac{1}{\sqrt{0.01}} - \frac{1}{\sqrt{0.81}}$$

$$= \frac{1}{0.1} - \frac{1}{0.9}$$

$$= 10 - 1.11$$

$$n > 8.89$$

The process of mass selection must be applied <u>9 times</u>.

(19) <u>A Parking Problem</u>

Suppose we are interested to find out the least number of car parking lots required for a large number n cars at a Research Institute, so that any motorist will find one parking lot immediately available at a high level of probability, say 95% or 99%.

To solve the problem, we use the Binomial Probability law and regard the question as one involving independent Bernoulli trials. We suppose that for the i^{th} car, chosen at random, there is a probability p_i that a parking lot is required. We can estimate p_i by observing in the course of a day the number of hours x a parking lot is occupied, and estimating p_i by $\frac{x}{8}$ on the assumption that a day consists of 8 working hours. If the average length of parking time is 1 hour 15 minutes, we have then $\frac{1.25}{8} \approx \frac{1}{10} = 0.1$. In order to have repeated Bernoulli trials, we assume $p_1 = p_2 = \dots = p_n = p$. We also assume independence of events in which A_i is the event that the i-th car requires a parking lot. The probability that exactly k parking lots are required at a given moment is $\binom{n}{k} p^k (1-p)^{n-k}$.

Suppose that K denotes the number of parking lots required, then the probability that a car driver in need of a parking lot will find one immediately available is the same as the probability that the number of parking lots demanded is less than or equal to K. This probability is

$$\sum_{k=0}^{K} \binom{n}{k} p^k (1-p)^{n-k}$$

If the Poisson approximation to the Binomial distribution is applicable, the above summation will be equal to

$$F_p(n, \lambda) = \sum_{k=0}^{K} \frac{e^{-\lambda} \lambda^k}{k!} \qquad \text{where } \lambda = np.$$

If the Normal Distribution holds, the summation will be equal to

$$\Phi\left(\frac{K - np + \frac{1}{2}}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{-np - \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

where $\sqrt{np(1-p)}$ is the standard deviation of the binomial distribution and $+\frac{1}{2}$ and $-\frac{1}{2}$ are the correction factors.

We know that for the Binomial Distribution, subject to the fact there is a correction for continuity so that the discrete distribution is changed into a continuous one, the area under the curve represents proability ($0 \leq p \leq 1$). If we let $u(P)$ denote the P - percentile of the Normal Distribution function, then

$$P = \Phi(\, u(P)\,) = \int_{-\infty}^{u(P)} \phi(X)\,dx$$

To obtain a probability greater than a pre-assigned level Po, the least number K or parking lots required, so that a motorist will find one parking lot immediately available, is

$$F_p\,(K;\, np) \geq Po \qquad \text{for the Poisson approximation}$$

$$\text{and} \quad K \quad u(Po) \quad np(1-p) + np - \tfrac{1}{2}^{**} \qquad \text{for the Normal Distribution}$$

From the Normal Tables, we have

$u\,(0.90) = 1.282$

$u\,(0.95) = 1.645$

$u\,(0.99) = 2.326$

Assuming that the Normal Approximation holds, we have, for $n=90$; $p=0.1$ and $Po=95\%$

$$K = 1.645\sqrt{90 \times 0.1 \times 0.9} + 90 \times 0.1 - \tfrac{1}{2} = 13.2$$

i.e., 14 parking lots are required for 90 cars at 95% probability level. Similarly, 16 parking lots are required for 90 cars at 99% level.

** This can be obtained as follows:-

$$P = \Phi(\, u(P)\,) = \Phi\left(\frac{K-np+\frac{1}{2}}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{-np-\frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

If $npq \geq 25$, $\sqrt{npq} \geq 5$. Since $p+q=1$, $q \leq 1$

$$np \geq npq. \quad \frac{np}{\sqrt{npq}} \geq npq \geq 5.$$

$$\Phi\left(\frac{-np-\frac{1}{2}}{\sqrt{npq}}\right) = \Phi\left(-\left(\frac{np}{\sqrt{npq}} - \frac{\frac{1}{2}}{\sqrt{npq}}\right)\right) = 0 \quad \text{when } \frac{np}{\sqrt{npq}} \geq 5.$$

We can work out K similarly for $n=900$, $p=0.1$, $Po=95\%$ and for $n=900$, $p=0.1$, $Po=99\%$.

K will be 104.3 and 110.4 respectively.

If the Poisson Approximation applies, calculation of K from $F_p(K;\ np) \geq Po$ will give us $k=14$ and 17 assuming that $n=90$, $p=0.1$, $Po=95\%$ and $n=90$, $p=0.1$, $Po=99\%$.

From these figures, we can deduce that for 90 cars with probability of parking = 0.1, only 16 or 17 parking lots are required in order that a motorist will find one parking space available almost 100% certain.

<u>APPENDIX 1</u>

MATHEMATICAL DERIVATION OF VARIANCE OF A RANDOM VARIABLE

Suppose X is a random variable with distribution $\{f(x_j)\}$ and suppose $r \geq 0$ is an integer. If the expectation of the random variable X^r, i.e., $E(X^r) = \sum x_j^r f(x_j)$ exists, then it is called the r^{th} moment of X about the origin. The r^{th} moment only exists if the series converges absolutely. Since $|X|^{r-1} \leq |X^r| + 1$, whenever the r^{th} moment exists, so does the $(r-1)^{st}$ and hence all preceding moments.

If the 2nd moment exists, so does the mean $u = E(X)$.

Instead of the random variable X, let us introduce the deviation from the mean, X-u. Since $(x-u)^2 \leq 2(x^2+u^2)$, we see that the second moment of X-u exists whenever $E(x^2)$ exists. We find

$$
\begin{aligned}
E\left((X-u)^2 \right) &= \sum_j (x_j^2 - 2ux_j + u^2)\, f(x_j) \\
&= \sum_j x_j^2\, f(x_j) - 2u \sum_j x_j f(x_j) + u^2 \sum_j f(x_j) \\
&= E(X^2) - 2uX(E) + u^2 \\
&= E(X^2) - 2u^2 + u^2 \\
&= E(X^2) - u^2.
\end{aligned}
$$

<u>Example</u>:

If X is the number of points scored with a symmetric die, the Var(X)

$$
\begin{aligned}
&= \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) - \left(\frac{7}{2}\right)^2 \\
&= \frac{35}{12}
\end{aligned}
$$

APPENDIX 2 (A)

THEORM ON CENTRAL TERM

Proof: Let us compare the term b(k;n,p) and the preceding term.

$$\frac{b(k;n,p)}{b(k-1;n,p)} = \frac{\binom{n}{k} p^k q^{n-k}}{\binom{n}{k-1} p^{k-1} q^{n-k+1}} \tag{A 2.1}$$

$$= \frac{\frac{n!}{k!(n-k)!} p^k q^{n-k}}{\frac{n!}{(k-1)!(n-k+1)!} p^{k-1} q^{n-k+1}}$$

$$\frac{n!}{k!(n-k)!} \times \frac{(k-1)!(n-k+1)!}{n!} \times p^{k-k+1} q^{n-k-n+k-1}$$

$$= \frac{(n-k+1)p}{kq}$$

$$= 1 + \frac{(n+1)p-k}{kq}$$

If $(n+1)p$ is not an integer,

the term $b(k;n,p) > b(k-1;n,p)$ if $k < (n+1)p$

and $b(k;n,p) < b(k-1;n,p)$ if $k > (n+1)p$

$\therefore$ As k goes from 0 to n, b(k;n,p) increases up to a maximum value which occurs for k=m and then decreases.

If $(n+1)p=m$ happens to be an integer, then $b(m;n,p) = b(m-1;n,p)$

$\therefore$ b(k;n,p) increases up to b(m-1;n,p) which is equal to (m;n,p) and then decreases.

$\therefore$ There exists m the unique integer for which

$$(n+1)p-1 < m \leqslant (n+1)p. \tag{A 2.2}$$

APPENDIX 2 (B)

THEOREM ON TAILS

The ratio in formula (A 2.1) above [See Appendix 2(a)] decreases monotonically as k increases.

When $k=r+1$,

$$\frac{b(k;n,p)}{b(k-1;n,p)} = \frac{(n-r)p}{(r+1)q}$$

When $k > r+1$ e.g. $k=r+2$

$$\frac{b(k;n,p)}{b(k-1;n,p)} < \frac{(n-r)p}{(r+1)q}$$

When $k \geqslant r+1$,

(A 2.3) $$\frac{b(k;n,p)}{b(k-1;n,p)} \leqslant \frac{(n-r)p}{(r+1)q}$$

Set herein $k=r+1$, $r+2$, ... $r+v$ and multiply the v inequalities to obtain

(A 2.4) $$\frac{b(r+v;n,p)}{b(r;n,p)} \leqslant \left\{\frac{(n-r)p}{(r+1)q}\right\}^{v}$$

[For example, if $v=3$, then $k=r+1$; $k=r+2$; $k=r+3$,

then $$\frac{b(k;n,p)}{b(k-1;n,p)} = \frac{b(r+1;n,p)}{b(r;n,p)} \times \frac{b(r+2;n,p)}{b(r+1;n,p)} \times \frac{b(r+3;n,p)}{b(r+2;n,p)}$$

$$= \frac{b(r+3;n,p)}{b(r;n,p)}$$

$$\therefore \frac{b(r+v;n,p)}{b(r;n,p)} \leqslant \left\{\frac{(n-r)p}{(r+1)q}\right\}^{v}$$]

For $r \geqslant np$, the fraction within the braces is less than unity, and summation over N leads to a finite geometric series with ratio $\frac{(n-r)p}{(r+1)q}$

We conclude, for $r \geq np$,

$$\sum_{v=0}^{n-r} b(r+v;n,p) \leq b(r;n,p) \frac{(r+1)q}{(r+1)-(n+1)p}$$

On the left we have the right tail of the binomial distribution, namely the probability of at least r successes.

Using the relationship $b(k;n,p) = b(n-k;n,q)$ and

$$\binom{n}{s} = \binom{n}{n-s} \qquad \& \qquad \binom{n}{r} = \binom{n}{n-r},$$

We can derive

$$\sum_{\emptyset=0}^{s} b(\emptyset;n,p) \leq b(s;n,p) \frac{(n-s+1)}{(n+1)p-s}$$

APPENDIX 3

ON HEREDITARY LAWS

The probabilities in π_1 of different genetical types GG, gG and gg are given as p', q', r' and p'', q'', r'' for females and for males respectively.

Let M, F and C stand for mother, father and child in a family with parents from $_1$. If C' denotes any specified genetical composition ^{1}C, then

$$
\begin{aligned}
P\{C:C'\} = P\{&(M:gg)(F:gg)(C:C') + (M:gg)(F:gG)(C:C') \\
&+ M(gg)(F:GG)(C:C') + (M:gG)(F:gg)(C:C') \\
&+ M(gG)(F:gG)(C:C') + (M:gG)(F:GG)(C:C') \\
&+ (M:GG)(F:gg)(C:C') + (M:GG)(F:gG)(C:C') \\
&+ (M:CC)(F:GG)(C:C')\}
\end{aligned}
$$

i.e., C having the property C' is represented by the sum of 3x3=9 mutually exclusive properties. Applying the addition priciple, we note that $p\{C:C'\}$ is the sum of nine probabilities of the type

$P\{(M:M')(F:F')(C:C')\}$ when M' and F' represent some specific combination of the genes g, G.

Applying the multiplication principle, we have

$$
\begin{aligned}
&P\{(M:M')(F:F')(C:C')\} \\
&= P\{M:M'\}\,P\{F:F'/M:M'\}\quad P\{C:C'/(M:M')(F:F')\}
\end{aligned}
$$

and since the mating is under panmixia,

$$P\{F:F'/M:M'\} = P\{F:F'\}$$

$$\therefore P\{(M:M')(F:F')(C:C')\}$$

$$= P\{M:M'\}\,P\{F:F'\}\quad P\{(C:C'/(M:M')(F:F')\}$$

The value of $P\{M:M'\}$ is, as given in the problem, either p', q' or r'. Similarly p'', q'' or r'' for $P\{F:F'\}$

$$\therefore P_1 = p'p'' + p'q''\frac{1}{2} + p'r'' \times 0$$
$$+ q'p''\frac{1}{2} + q'q''\frac{1}{4} + q'r'' \times 0$$
$$+ r'p'' \times 0 + r'q'' \times 0 + r'r'' \times 0$$

where P_1 denotes the probability that an individual of the 1st generation born II_1 will be a <u>DOMINANT.</u>

$$\therefore P_1 = (p' + \frac{1}{2}q')(p'' + \frac{1}{2}q'')$$

Similarly,

$$Q_1 = (p' + \tfrac{1}{2}q')(r'' + \tfrac{1}{2}q'') + (r' + \tfrac{1}{2}q')(p'' + \tfrac{1}{2}q'')$$

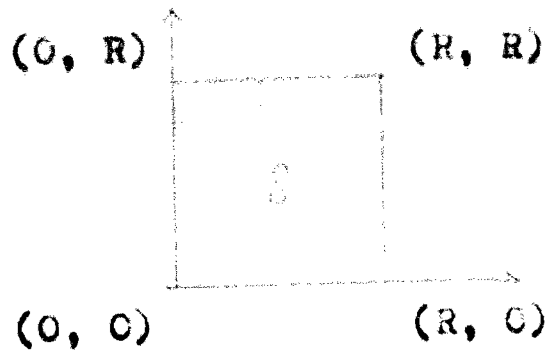
$$R_1 = (r' + \tfrac{1}{2}q')(r'' + \tfrac{1}{2}q'')$$

AREA UNDER NORMAL DISTRIBUTION CURVE

To prove that $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} \, dx = 1$

We prove this by the method of double integration, transforming the variables x, y into polar co-ordinates.

Take $\iint_S e^{-\frac{x^2+y^2}{2}} \, dx \, dy$



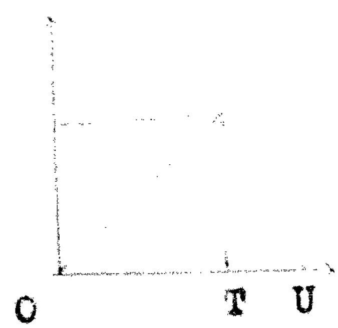
and integrate it with respect to y.

Thus

$$I = \iint_S e^{-\frac{x^2+y^2}{2}} \, dx \, dy = \int_0^R \left[\int_0^R e^{-\frac{x^2+y^2}{2}} \, dy \right] dx$$

$$= \int_0^R e^{-\frac{y^2}{2}} \, dy \int_0^R e^{-\frac{x^2}{2}} \, dx$$

$$= \left[\int_0^R e^{-\frac{x^2}{2}} \, dx \right]^2$$



$$\iint_T \leqslant I \leqslant \iint_U$$

Since $x = r\cos\theta$ & $y = r\sin\theta$

$x^2 + y^2 = r^2$

In polar co-ordinates, we have

$$\iint_T e^{-\frac{r^2}{2}} r\,dr\,d\theta \leq I \leq \iint_u e^{-\frac{r^2}{2}} r\,dr\,d\theta$$

$$= \int_0^{\frac{\pi}{2}} \int_0^{R\sqrt{2}} e^{-\frac{r^2}{2}} r\,dr\,d\theta$$

$$= \int_0^{\frac{\pi}{2}} \left[-e^{-\frac{r^2}{2}} \right]_0^{R\sqrt{2}} d\theta$$

$$= \left(1 - e^{-R^2} \right) \times \frac{\pi}{2}$$

On the Left-Hand-Side, the limits are 0 to R and 0 to $\frac{\pi}{2}$

$$\text{L. H. S.} \iint_T = \left(1 - e^{-\frac{R^2}{2}} \right) \times \frac{\pi}{2}$$

Thus we have

$$\left(1 - e^{-\frac{R^2}{2}} \right) \times \frac{\pi}{2} \leq I \leq \left(1 - e^{-R^2} \right) \times \frac{\pi}{2}$$

As $R \to \infty$, $\frac{1}{e^{R^2}} \to 0$

$$\therefore \frac{\pi}{2} \leq I_\infty \leq \frac{\pi}{2}$$

$$\therefore I_\infty = \frac{\pi}{2}$$

In other words, we have shown that

$$\left[\int_0^{\infty} e^{-\frac{x^2}{2}}\,dx\right]^2 = \frac{\pi}{2}$$

$$\therefore \quad \int_0^{\infty} e^{-\frac{x^2}{2}}\,dx = \sqrt{\frac{\pi}{2}}$$

and for limits from $-\infty$ to $+\infty$

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = 2\sqrt{\frac{\pi}{2}} = \sqrt{2\pi}$$

$$\therefore \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\,dx = \frac{1}{\sqrt{2\pi}} \times \sqrt{2\pi} = 1$$

BIBLIOGRAPHY

1. Feller, W.: An Introduction to Probability Theory and its Applications, 2nd edition, John Wiley and Sons, Inc., 1957.

2. Parzen, E.: Modern Probability Theory and its Applications, John Wiley and Sons, Inc., 1960.

3. Goldberg, S.: Probability; An Introduction, Prantice-Hall, Inc., 1960.

4. Munroe, M. E.: Theory of Probability, McGraw-Hill Book Company, Inc., 1951.

5. Neyman, J.: First Course in Probability and Statistics, Henry Holt and Company, Inc., 1950.

6. Freeman, F.: Introduction to Statistical Inference, Addison-Wesley Publishing Co., Inc., 1963.

7. Hald, A.: Statistical Theory with Engineering Applications, John Wiley and Sons, Inc., 1957.

TAMAT

0 MM 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1 0 CM