8-1955

The Normal Curve As The Limiting Form Of The Binomial Distribution

Joseph Robert Phillips

THE NORMAL CURVE AS THE LIMITING FORM OF THE BINOMIAL DISTRIBUTION

PHILLIPS

1955

JOSEPH ROBERT PHILLIPS

1923 Born in Port Arthur, Texas

1929 Entered elementary school

1941 Graduated from Lincoln High School, Port Arthur, Texas

1942 Received advanced certificate in Engineering Drawing from Prairie View A & M College

1943-45 Served in Army

1949 Graduated from Prairie View A & M College with B.S. degree (Mechanical Engineering Major)

1949 to present Instructor of Mathematics, Daule High School, Cuero, Texas

1949-55 Summers Relief Engineer at Prairie View

DIVISION OF GRADUATE STUDY

PRAIRIE VIEW AGRICULTURAL AND MECHANICAL COLLEGE

FINAL EXAMINATION

of

JOSEPH ROBERT PHILLIPS

For the Degree of

MASTER OF SCIENCE

Friday, August 12, 1955 2:00 p.m. Ag 204

COMMITTEE IN CHARGE:

E. W. Haskins	Chairman and Professor of Physics
A. I. Thomas	Director of Industrial Education Division
A. V. Stewart	Professor of Mathematics
J. E. Mosby, Jr.	Visiting Professor of Industrial Education
A. W. Randall	Head, Department of Mathematics

LIST OF GRADUATE COURSES

Major Field: Mathematics

Math 423	
Math 563	Methods of Teaching Mathematics
Math 573	Higher Plane Curves
Math 713	Advanced Calculus
Math 743	Mathematical Statistics
Math 803	Numerical Solutions of Equations
Math 893	Complex Variables

Minor Field: Industrial Education

Ed 543	Audio-Visual Aids
Ed 553	Driver Education
Ed 673	High School Methods
Ind Ed 723	Administration and Supervision of Vocational-Industrial Education

BRIEF OF THESIS

THE NORMAL CURVE AS THE LIMITING FORM OF THE BINOMIAL DISTRIBUTION

It was the purpose of this study:

1. Since most skew curves can be obtained by plotting the terms of the expansion of the binomial $(p+q)^n$,

   a. What happens to such curves if "n" is allowed to increase without limit?

   b. What is the mathematical justification?

Findings:

1. That a variable which is the resultant of several independent, equally potent causes, has a binomial distribution. This binomial distribution is discrete, and may be represented by a histogram. The limiting form of this histogram as the number of these independent, equally potent, causes becomes infinitely great, is the normal curve.

THE NORMAL CURVE AS THE LIMITING
FORM OF THE BINOMIAL DISTRIBUTION

By

Joseph Robert Phillips

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Master of Science

In The

Graduate Division

of

Prairie View Agricultural and Mechanical College
Prairie View, Texas

August, 1955

Approved By:

Ethelbert W. Haskins Date Aug. 12, 1955
Advisor

A. W. Randall Date 8/19/55
Head, Department of Mathematics

ACKNOWLEDGEMENTS

The writer wishes to express thanks and appreciation to Mr. E. Haskins, my advisor, for his collaborative assistance and meticulous checking of details. The writer is also grateful to the staff of the W. R. Banks Library and to Mrs. Emma C. Green for being so cooperative.

DEDICATION

To my family, because theirs like mine is a service of people, rather than things.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Normal distributions are very useful as mathematical models for many frequency distributions found in nature and industry. Thus, a great many measurements made on manufactured articles possess distributions that can be approximated well by normal distributions. The same is also true of many biological measurements. Even if there were few natural distributions of this form, the normal distribution would still be extremely important because of its place in theoretical work. As may be readily inferred, there are many curves which are like the normal curve in that they have at most one mode, but differ from the normal curve in that they are not symmetrical. Such curves are often referred to as skew curves.[1] Most of these types of curves are presented by the various graphs obtained by plotting the terms of the expansion of the binomial $(p+q)^n$, where p represents the probability of the occurrence of a certain event, q the probability of its failure and n the number of trials. Lest one should think that the graphs of the terms of the expansion would be rectangular histograms, it is well to say that it is the frequency curve which correspond to these histograms to which we refer as representative types of frequency curves. When, as a special case, $p=q=\frac{1}{2}$, the corresponding graph is symmetrical - that is, skewness is absent or zero - and is essentially a form of the normal curve. To show the mathematical justification of this statement, is the writer's purpose for making this investigation and writing this thesis.

---

[1]Forsyth, C. H., An Introduction To The Mathematical Analysis Of Statistics, John Wiley and Sons, Inc. New York, N. Y., 1924.

The curve which we now call "normal" is the curve of normal probability or curve of error which attracted the interest of most of the great mathematicians and astronomers of the first half of the nineteenth century. More than two centuries ago, Abraham De Moivre (1667 - 1754), a refugee mathematician living in London, making his living partly by solving problems for wealthy gamblers, recognized that the random variation in the number of heads appearing on throws of n coins corresponds to the terms of the binomial expansion of $(.5 + .5)^n$, and that as n becomes larger this distribution approaches a definite form. In 1733 De Moivre derived the equation for this curve and presented it privately to some friends. To him it was only a mathematical exercise, utterly unconnected with any sort of application to empirical data.

At that time there were no collection of empirical data at hand for study. As yet no one had made any measurements of any large number of individuals. A Swiss mathematician named Jacques (or James) Bernoulli (1654 - 1705, eldest of the very remarkable family of mathematicians by that name) had suggested that the theory of probability might have useful applications in economic and moral affairs, but he, himself, was too near death to investigate the applications, and, moreover, he had no numerical data which he could have used for that purpose. In 1713, his great book on probability, was published posthumously under the editorship of his nephew Nicolas. Nearly a century passed before any scientific worker began to gather large masses of concrete data and to study the properties of distributions. The application of the normal curve in studies of concrete data begins with the work of the great mathematical astronomers who lived at the beginning of the nineteenth century, chiefly Laplace (1749-

1827) in France, and Gauss (1777 - 1855) in Germany, each of whom derived the law independently and presumably without any knowledge of De Moivre's derivation.

The probability curve is often called the Gaussian Curve, because until recently it was supposed that Gauss had been the first person to make use of its properties. However, in 1924, Karl Pearson discovered an unknown derivation by Abraham De Moivre.

The idea that this curve could be used to describe data other than errors of observation in the physical sciences seem to have originated with the great Belgian statistician, Adolphe Quetelet (1796 - 1874), who first popularized the idea that statistical method was a fundamental discipline adaptable to any field of human interest in which mass data were to be found. He was convinced that the measurement of mental and moral traits waited only for the collection of sufficient and trustworthy data, and was so sure that when such measurement was feasible, the distribution of these traits would be found to be in accordance with the "law of error" that he talked about "the normal curve of error".

Most people are familiar in a general way with this bell-shaped, symmetrical frequency distribution which has an unlimited range and which approaches the base line more closely as the absolute value of the abscissa becomes larger. Most people have heard so much about this curve that they have an exaggerated idea of its universality, and imagine that every distribution which has a peak near the center is normal. Actually the normal curve is a mathematical abstraction, a sort of mathematical model. Some concrete data have distributions which approximate this form, but it is not in any sense the expected pattern to which observed distributions conform.

CHAPTER II

A PRIORI PROBABILITY

A variable which is the resultant of several independent, equally potent causes, each as likely to be present as absent in any given instance, has a binomial distribution. This binomial distribution is discrete, and may be represented by a histogram. If the number of independent, equally potent causes increases, the number of intervals in the histogram increases. When the number of intervals becomes very great, the histogram begins to suggest a smooth curve. The limiting form of this histogram as the number of these independent, equally potent causes becomes infinitely great, is the normal curve.

Thus if several unbiased coins are tossed simultaneously, the number of heads appearing on a single throw is a variable which depends upon the way each of the several coins fall. These coins are independent; each is, presumably, as likely to fall one way as to fall the other; and each has the same effect upon the total. In the long run, the proportion of N throws which would be expected to show no heads, 1 head, 2 heads, etc. is as follows:

Number of pennies tossed	Proportion expected to show given number of heads					
	0	1	2	3	4	5
1	1/2	1/2				
2	1/4	2/4	1/4			
3	1/8	3/8	3/8	1/8		
4	1/16	4/16	6/16	4/16	1/16	
5	1/32	5/32	10/32	10/32	5/32	1/32

If histograms of these proportions were drawn, one would readily see that as the number of coins becomes very large, these histograms would approach the normal distribution.

Let us consider a mathematical illustration of the binomial expansion and acquire greater familiarity with the significance of the terms of the expansion. If n coins were tossed up, the first term of the following expansion

$$(\tfrac{1}{2} + \tfrac{1}{2})^n = (\tfrac{1}{2})^n + N\,(\tfrac{1}{2})^{n-1}(\tfrac{1}{2}) + \frac{n(n-1)}{2}\,(\tfrac{1}{2})^{n-2}(\tfrac{1}{2})^2 + \text{etc.},$$

is the probability that all coins will be "heads". Similiarly, the second term is the probability that all but one will be "heads", since $(\frac{1}{2})^{n-1}(\frac{1}{2})$ is the probability that one particular coin will be "tails" and the rest "heads" and this one coin can be chosen in n ways. Likewise, the third term is the probability that all but two will be "heads" and so on for the rest of the terms.

In this experiment the writer chose n to be four and varied the number of throws. As a model with $n = 4$, the proportion expected to show the given number of heads, for any single toss can be shown by expanding $(p + q)^4$, thus obtaining $p^4 + 4p^2q + 6p^2q^2 + 4pq^3 + q^4$. Letting "H" represent heads and "T" for tails the various combinations are as follows:

1. H H H H	7. H T T H	13. T T H T
2. H H H T	8. H T H T	14. T H T T
3. H H T H	9. T T H H	15. T H H H
4. H T H H	10. T H T H	16. T T T T
5. T H H H	11. T H H T	
6. H H T T	12. H T T T	

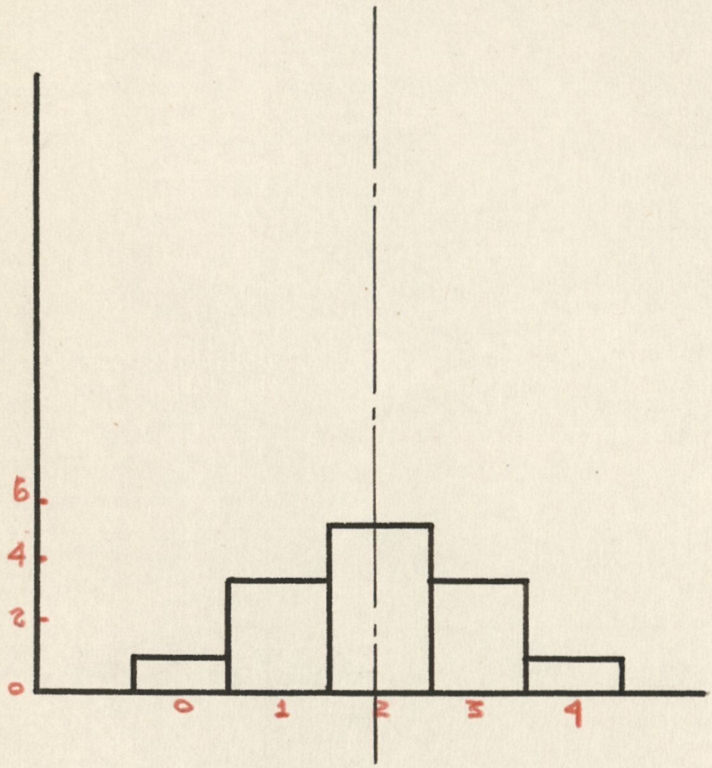

Figure 1. Binomial distribution: proportion expected to to show given number of heads when 4 coins are tossed one time.

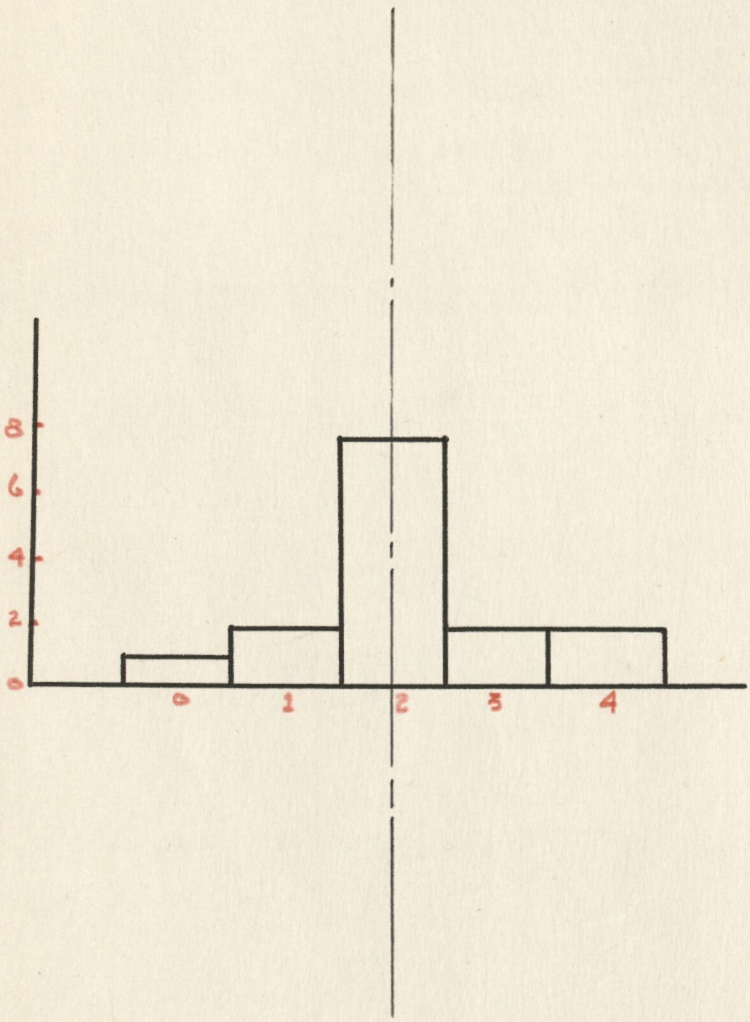

Figure 2. Binomial distribution: showing given number of heads when 4 coins were tossed 16 times.

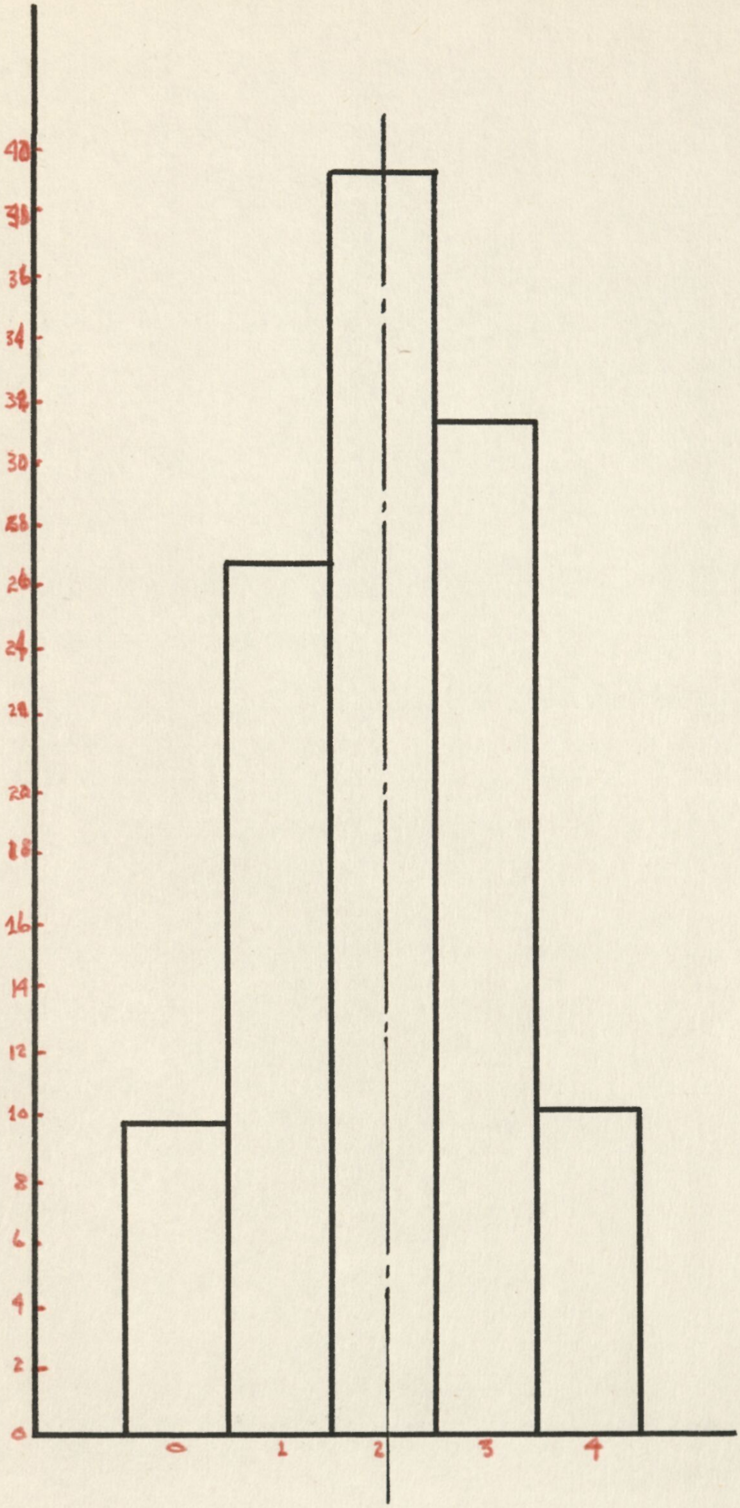

Figure 3. Binomial distribution showing given number of heads when 4 coins were tossed 160 times.

Since the total number combinations is sixteen, one needs but to look at the above expansion and see that the expected values for 0, 1, 2, 3, and 4 heads is 1, 4, 6, 4, and 1 respectively. The histogram of this distribution is shown in figure 1 and was taken as the model to measure the experimental data. Figure 2 is the histogram of data recorded when four unbiased coins (pennies) were tossed 16 times and figure 3, when the same four coins were flipped 160 times. It will be observed that they both have a bell shape, but that the one in figure 3 is more nearly symmetrical than the one in figure 2. Also the histogram in figure 3 has a slightly broader hump than the one in figure 2.

CHAPTER III

THE NORMAL CURVE: LIMITING FORM OF THE BINOMIAL DISTRIBUTION

Proof Number I

Let us first consider the binomial distribution and determine mathematically what happens as n is allowed to increase without limit.

Let r denote the number of successes in n trials, and let y (r) denote the probability of exactly r successes. Then, from the equation.

$$p\ (r) = P\left[x = r\right] = C\ (n,r)\ p^r\ q^{n-r} = \frac{n!\ p^r\ q^{n-r}}{r!\ (n-r)!}$$ one has

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

The slope of the line segment joining the midpoint of the top of the bar which represent r successes to the midpoint of the top of the bar representing r + 1 successes, as in figure 4, is

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

$$\frac{n!\ p^r\ \ q^{n-r}}{r!\ (n-r)!} = \frac{N!\ p^r\ q^{n-r}}{r!\ (n-r)!}\left[\frac{P\ (n-r) - 1}{q(r+1)}\right]$$

The factor $\frac{N!\ p^r\ q^{n-r}}{r!\ (n-r)!}$ is Y (r), so that one can write

$$(1)\quad \frac{1}{y(r)}\ \frac{\Delta y\ (r)}{\Delta r} = \frac{np - rp - rq - q}{q\ (r+1)} = \frac{-\ NP - r - q,}{q\ (r+1)}$$

where Δr denotes the change in r passing from one bar to the next, $\Delta y\,(r) = y\,(r + 1) - y\,(r)$, and use is made of the fact that $p + q = 1$.

Since the scales of the histogram must be changed to standard scales, let x denote the abscissa in the standard scale. The standard horizontal scale has $E\,[r] = np$ for its zero point and $\sigma = npq$ for its unit. Then the relation between x and r is,

$$(2) \qquad x = \frac{r - np}{\sqrt{npq}},$$

or, solved for r,

$$(2') \qquad r = x\sqrt{npq} + np.$$

The origin of the vertical scale is not changed, and a change of the vertical unit does not change the ratio $\frac{\Delta y}{y}$. Therefore the slope of the line joining the midpoint of the top of one bar to the midpoint of the top of the next, denoted by $\frac{\Delta y}{\Delta x}$ in the standard scales, satisfies an equation obtained from (1) after r has been replaced by its value in terms of x as

$$(2'') \qquad \Delta r = \Delta x\sqrt{Npq},$$

so that equation (1) becomes

$$\frac{1}{y}\,\frac{\Delta y}{\Delta x\sqrt{npq}} = \frac{-x\sqrt{npq} - q}{q(x\sqrt{npq} + Np + 1)}\,.$$

after multiplying through by $\sqrt{npq}$ this becomes

$$\frac{1}{y}\;\frac{\Delta y}{\Delta x} = -\,\frac{npqx + q\sqrt{Npq}}{npq + xq\sqrt{npq} + q} = -\,\frac{x + \frac{q}{\sqrt{npq}}}{1 + \frac{xq}{\sqrt{npq}} + \frac{q}{npq}}\,.$$

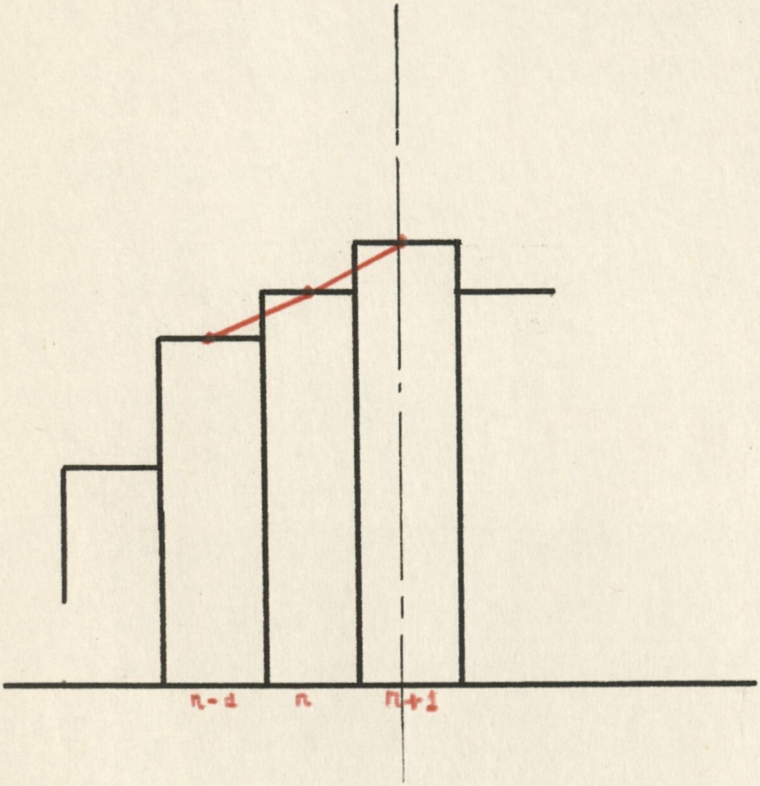

Figure 4.

As n increases without bound, Δr remains equal to 1. So x approaches zero, by equation (2"); $\frac{\Delta y}{\Delta x}$ approaches the derivative of y with respect to x; and the right-hand member approaches $-x$. Hence the result of taking the limit of both members of the above equation is

$$\frac{1}{y} \frac{dy}{dx} = -x.$$

then, by integration,

$$\ln y = -\frac{x^2}{2} + \ln c,$$

or

$$y = Ce^{\frac{-x^2}{2}} \tag{3}$$

The constant of integration is determined from the fact that this is a probability function and must satisfy the equation

$$\int_c^d p(x)\,dx = P\left[C \leq x \leq d\right] = 1.$$

thus

$$P\left[-\infty < x < \infty\right] = \int_{-\infty}^{\infty} y\,dx = C\int_{-\infty}^{\infty} e^{\frac{-x^2}{2}}\,dx = 1$$

To evaluate this integral, set

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

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

This double integral represents the volume above the first quadrant of the xy - plane under the surface $Z = e^{-\frac{x^2 + y^2}{2}}$.

To evaluate the integral I^2, a change to polar coordinates may be employed. Let x + y $= \rho$, and keep in mind that in the transformation the differentials of x + y becomes the Jacobian J of x and y with respect to ρ an θ times the differentials of ρ and θ; that is

$$dxdy = J \, d\rho \, d\theta$$

To evaluate the Jacobian, we use

$$x^2 + y^2 = \rho^2$$
$$x = \rho \cos \theta$$
$$y = \rho \sin \theta$$

$$J = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -\rho \sin \theta \\ \sin \theta & \rho \cos \theta \end{vmatrix} = \rho \cos^2 \theta + \rho \sin^2 \theta = \rho$$

When $y = o$ then $\theta = o$ and as $y \to \infty$ then $\theta = \frac{\pi}{2}$

$x = o$ then $\rho = o$ and as $x \to \infty$ then $\rho \to \infty$

$$\therefore I^2 = \int_0^{\frac{\pi}{2}} \int_0^{\infty} e^{-\frac{\rho^2}{2}} \rho \, d\rho \, d\theta = \int_0^{\frac{\pi}{2}} \left[-e^{-\frac{\rho^2}{2}} \right]_0^{\infty} d\theta$$

$$= \int_0^{\frac{\pi}{2}} d\theta = \frac{\pi}{2}$$

Since 2c I = 1, $C = \frac{1}{\sqrt{2\pi}}$. substituting this value of C in

equation (3), we obtain the equation of the Normal Probability Curve, namely,

$$y = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Moment - Generating Function Method

Thus for the fact that the binomial distribution can be approximated well for large n by the normal distribution with $m = np$ and $\sigma = \sqrt{npq}$ has been made plausible by proof number 1. Now consider the verification of this fact by means of the moment-generating function. Here it is convenient to use the variable

$$\frac{x - np}{\sqrt{npq}} = \frac{x - m}{\sigma} = t$$

From the properties[1]

(1) $M_{cG}(\theta) = M_G(C\theta)$

(2) $M_{G+C}(\theta) = e^{C\theta} M_G(\theta)$

(3) $M_x(\theta) = (q + pe^{\theta})^n$

it follows (respectively) that

$$M_t(\theta) = M_{x-m}\left(\frac{\theta}{\sigma}\right)$$

$$M_t(\theta) = e^{\frac{-M\theta}{\sigma}} M_x\left(\frac{\theta}{\sigma}\right)$$

$$= e^{\frac{-M\theta}{\sigma}} (q + p e^{\frac{\theta}{\sigma}})^n$$

---

[1]Hoel, Paul G., Introduction to Mathematical Statistics, John Wiley and Sons, Inc., New York, New York., 1947.

Taking the logarithms of both sides to the base e gives

$$\log M_t(\theta) = -\frac{M\theta}{\sigma} + N \log (q + pe^{\frac{\theta}{\sigma}})$$

Expanding $e^{\frac{\theta}{\sigma}}$ and replacing q + p by 1 yields

$$\log M_t(\theta) = -\frac{M\theta}{\sigma} + N \log \left\{ 1 + p\left[\left(\frac{\theta}{\sigma}\right) + \frac{1}{2!}\left(\frac{\theta}{\sigma}\right)^2 + \frac{1}{3!}\left(\frac{\theta}{\sigma}\right)^3 + \ldots \right] \right\}$$

The logarithm on the right may be treated as of the form $\log \left| 1 + Z \right|$.

If $\left| \theta \right|$ is chosen sufficiently small, the expansion

$$\log \left\{ 1 + Z \right\} = Z - \frac{Z^2}{2} + \frac{Z^3}{3} - \frac{Z^4}{4} + \ldots$$

may be applied to give

$$(4) \quad \log M_t(\theta) = -\frac{M\theta}{\sigma}$$

$$+ N \left\{ p\left[\left(\frac{\theta}{\sigma}\right) + \frac{1}{2!}\left(\frac{\theta}{\sigma}\right)^2 + \ldots \right] - \frac{P^2}{2}\left[\left(\frac{\theta}{\sigma}\right) + \frac{1}{2!}\left(\frac{\theta}{\sigma}\right)^2 + \ldots \right]^2 + \ldots \right\}$$

Collecting terms in powers of θ gives

$$\log M_t(\theta) = \left(-\frac{M}{\sigma} + \frac{NP}{\sigma}\right)\theta + N\left(\frac{P}{\sigma^2} - \frac{P^2}{\sigma^2}\right)\frac{\theta^2}{2!} + \ldots$$

But, since np = m and = npq, the coefficient of θ vanishes and the coefficient of $\frac{\theta^2}{2!}$ reduces to 1.

$$\left(-\frac{M}{\sigma} + \frac{NP}{\sigma}\right)\theta + N\left(\frac{P}{\sigma^2} - \frac{P^2}{\sigma^2}\right)\frac{\theta^2}{2!} + \ldots$$

$$\left(-\frac{NP}{\sigma} + \frac{NP}{\sigma}\right)\theta + \left(\frac{NP}{NPQ} - \frac{NP^2}{NPQ}\right)\frac{\theta^2}{2!}$$

$$(0)\theta + \left(\frac{1-P}{q}\right)\frac{\theta}{2!}$$

but $1 - P = q$

consequently

$$\log M_t(\theta) = \frac{\theta^2}{2!} + \text{terms in } \theta^k, \quad K = 3, 4, \ldots$$

From an inspection of (4), which shows how terms in θ^k arises, it is clear that all terms in θ^k contains $\frac{N}{k}$ as a common factor. Since this other factor does not involve n and since

$$\frac{N}{\sigma^k} = \frac{N}{(npq)^k}$$

with $K \geq 3$ here, all such terms will approach zero as n becomes infinite. This implies that

$$\lim_{N \to \infty} \log M_t(\theta) = \frac{\theta^2}{2}$$

Which in turn implies that

$$\lim_{N \to \infty} M_t(\theta) = e^{\frac{\theta^2}{2}}$$

This is a form of the Normal Curve.

CHAPTER IV

SUMMARY AND CONCLUSION

The binomial distribution discussed in the previous chapters is an important example of a probability function of a discrete random variable. The most important probability function of a continuous random variable is the one which is the subject of the following theorem:

If the histograms of the probability functions, with p and q fixed are drawn in standard scales, then, as n increases without bound, the histograms converge to the curve where equation is

$$y = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$$

This curve is called the Normal Probability Curve.

The importance of the normal probability function rests on the fact that many families of probability functions converge to the normal probability function. It was shown in the previous chapter how the normal curve is the limit of the family of binomial distributions. The binomial distribution can be used to solve many practical problems related to repeated trials of an event. This distribution is applicable to a discrete variable that assumes only one or the other of two values. For situations in which more than two values are possible and desirable, a generalization of the binomial distribution is needed.

BIBLIOGRAPHY

Cramer, Harald, Mathematical Methods of Statistics. Princeton University Press, 1946.

Fry, T. C., Probability and Its Engineering Uses. Van Nostrand, New York, 1939.

Fryer, H. C., Elements of Statistics. New York and London, 1954.

Gavett, Irving G., A First Course in Statistical Methods, New York and London, 1937.

Hoel, Paul G., Introduction to Mathematical Statistics, John Wiley and Sons, Inc., New York, 1947.

Jacoby, Oswald, How to Figure the Odds, Gordon City, New York, 1953.

Kenney, J. F., Mathematics of Statistics, Van Nostrand, New York, 1939.

Scarborough, J. B . and R. W. Wagner, Fundamentals of Statistics, Gin and Company, B oston, 1948.

Uspensky, J. V., Introduction to Mathematical Probability. McGraw - H ill, New York, 1937.

Walker, Helen M., Elementary Statistical Methods. Holt, New York, 1943.