

REVIEW ARTICLE

A Note on the Asymptotic Convergence of Bernoulli Distribution

A. T. Adeniran¹, J. F. Ojo², J. O. Olilima³

¹Department of Mathematical Sciences, Augustine University Ilara-Epe, Nigeria, ²Department of Statistics, University of Ibadan, Ibadan, Nigeria, ³Department of Mathematical Sciences, Augustine University Ilara-Epe, Nigeria

Received: 10-03-2018; Revised: 15-04-2018; Accepted: 15-05-2018

ABSTRACT

This paper presents concepts of Bernoulli distribution, and how it can be used as an approximation of Binomial, Poisson, and Gaussian distributions with a different approach from earlier existing literature. Due to discrete nature of the random variable X , a more appropriate method of the principle of mathematical induction (PMI) is used as an alternative approach to limiting behavior of the binomial random variable. The study proved de Moivre–Laplace theorem (convergence of binomial distribution to Gaussian distribution) to all values of p such that $p \neq 0$ and $p \neq 1$ using a direct approach which opposes the popular and most widely used indirect method of moment generating function.

Key words: Bernoulli distribution, binomial distribution, Poisson distribution, Gaussian distribution, principle of mathematical induction, convergence, moment generating function

INTRODUCTION

If X is a random variable, the function $f(x)$ whose value is $P[X = x]$ for each value x in the range of X is called probability distribution.^[9] There are various probability distributions used in Sciences, Engineering, and Social Sciences. They are numerous to the extent that there is no single literature that can give their comprehensive list. Bernoulli distribution is one of the simplest probability distributions in literature. An experiment consisting of only two mutually exclusive possible outcomes, either success or failure, male or female, life or death, defective or non-defective, and present or absent is called a Bernoulli trial.^[7,9,13] These independent trials having a common success of probability p were first studied by the Swiss mathematician, Jacques Bernoulli in his book *Ars Conjectandi*, published by his nephew Nicholas 8 years after his death in 1713.^[1,13,17,19,20,21]

Definition 1.1 (The Bernoulli distribution) if X is a random variable that takes on the value 1 when the Bernoulli trial is a success (with probability

of success $p \in (0,1)$) and the value 0 when the Bernoulli trial is a failure (with probability of failure $q = 1 - p$), then X is called a Bernoulli random variable with probability function.

$$P[X = x] = \begin{cases} p^x (1-p)^{1-x}; & x = 0, 1; 0 \leq p \leq 1 \\ 0 & \text{elsewhere} \end{cases} \quad (1)$$

The function (1), where $0 < p < 1$ and $p + q = 1$, is called the Bernoulli probability function.

The distribution is rarely applied in real life situation due to its simplicity and because it has no strength of modeling a metric variable as it is restricted to whether an event occurs or not with probabilities p and $1-p$, respectively.^[6] Despite this limitation, Bernoulli distribution turns (converge) to many compounds and widely used robust distributions; n -Bernoulli distribution is binomial, limiting-value of binomial distribution gives Poisson distribution, binomial distribution under certain conditions yields Gaussian distribution, and so on.^[9]

Convergence of a particular probability distribution to another and other related concepts such as limiting value of probability function or proofing of central limit theorem and law of large number had been thoroughly dealt with in literature by many authors.^[2-4,10,22] All these studies adopted moment generating function (mgf) approach, perhaps

Address for correspondence:

T. Adeniran Adefemi,
E-mail: adefemi.adeniran@augustineuniversity.edu.ng

because it is easy to tract and the mathematics involved is less rigorous. Based on their approach (mgf approach), they were able to establish the convergence with valid evidence due to uniqueness property of the mgf which asserts that.

Definition 1.2 (uniqueness theorem) suppose F_x and F_y are two cumulative distribution functions whose moments exist, if the mgf's exist for the r.v.'s X and Y and $Mx(t) = My(t)$ for all t in $-h < t < h, h > 0$, then, $FX(u) = FY(u)$ for all u . That is, $f(x) = f(y)$ for all u . The limitation of this approach is that mgf does not exist for all distributions, yet the particular distribution approaches normal or converge to other distribution under some specified condition(s). For example, the lognormal distribution does not have a mgf, still, it converges to a normal distribution.^[12,16] A standard/direct proof of this more general theorem uses the characteristic function which is defined for any distribution.^[8,14] Bain and Engelhardt,^[5] Jeffrey and Richard,^[11] Inlow Mark,^[10] Bagui *et al.*,^[2] and Bagui and Mehra^[4] opined that direct proving of convergence of a particular probability density function (pdf) to another pdf as n increasing indefinitely is rigorous, since it is based on the use of characteristic functions theory which involves complex analysis, the study of which primarily only advanced mathematics majors students and professors in colleges and universities understand. In this paper, we employed the direct approach to proof convergence of binomial distribution to Poisson and Gaussian distributions with a lucid explanation and support our theorems with copious lemma to facilitate and enhance students understanding regardless of their mathematical background vis-a-vis their level.

The structure of the paper is as follows; we provide some useful preliminary results in section 2, these results will be used in section 3 where we give details convergence of Bernoulli to Binomial, Poisson, and Gaussian distributions. Section 4 contains some concluding remarks.

PRELIMINARIES

In this section, we state some results that will be used in various proofs presented in section 3.

Lemma 2.1 Show that $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$

Proof 2.1 the proof of this lemma is as follows: Using the binomial series expansion, $\left(1 - \frac{\lambda}{n}\right)^n$ can be expressed as follows;

$$\begin{aligned} \left(1 - \frac{\lambda}{n}\right)^n &= \binom{n}{0} 1^n \left(\frac{-\lambda}{n}\right)^0 + \binom{n}{1} 1^{n-1} \left(\frac{-\lambda}{n}\right)^1 \\ &+ \binom{n}{2} 1^{n-2} \left(\frac{-\lambda}{n}\right)^2 + \binom{n}{3} 1^{n-3} \left(\frac{-\lambda}{n}\right)^3 + \dots \\ &= 1 + n \left(\frac{-\lambda}{n}\right) + \frac{n(n-1)}{2!} \left(\frac{-\lambda}{n}\right)^2 \\ &+ \frac{n(n-1)(n-2)}{3n^2!} \left(\frac{-\lambda}{n}\right)^3 \\ &+ \frac{n(n-1)(n-2)(n-3)}{4!n^3} \left(\frac{-\lambda}{n}\right)^4 + \dots \\ &= 1 - \lambda + \frac{\lambda^2}{2!} \left(\frac{n-1}{n}\right) - \frac{\lambda^3}{3!} \frac{(n-1)(n-2)}{n^2} \\ &+ \frac{\lambda^4}{4!} \frac{(n-1)(n-2)(n-3)}{n^3} - \dots \\ &= 1 - \lambda + \frac{\lambda^2}{2!} \left(\frac{n-1}{n}\right) - \frac{\lambda^3}{3!} \frac{n-1}{n} \frac{n-2}{n} \\ &+ \frac{\lambda^4}{4!} \frac{n-1}{n} \frac{n-2}{n} \frac{n-3}{n} - \left(1 - \frac{\lambda}{n}\right)^n \\ &= 1 - \lambda + \frac{\lambda^2}{2!} \left(1 - \frac{1}{n}\right) - \frac{\lambda^3}{3!} \left(1 - \frac{1}{n}\right) \\ &\left(1 - \frac{1}{n}\right) + \frac{\lambda^4}{4!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right) - \dots \end{aligned}$$

Taking limit of the preceding equation as $n \rightarrow \infty$ gives

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n &= \lim_{n \rightarrow \infty} \left[\begin{aligned} &1 - \lambda + \frac{\lambda^2}{2!} \left(1 - \frac{1}{n}\right) \\ &- \frac{\lambda^3}{3!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right) \\ &+ \frac{\lambda^4}{4!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right) \\ &\left(1 - \frac{1}{n}\right) - \dots \end{aligned} \right] \\ &= 1 - \lambda + \frac{\lambda^2}{2!} \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) - \\ &\frac{\lambda^3}{3!} \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right) + \\ &\frac{\lambda^4}{4!} \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right) \left(1 - \frac{1}{n}\right) - \dots \\ &\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \\ &= 1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} - \dots \end{aligned}$$

The RHS of the preceding equation is the same with Maclaurin series expansion of $e^{-\lambda}$. Hence,

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda} \quad (2)$$

Lemma 2.2 Prove that $\lim_{n \rightarrow \infty} \frac{n}{(n-x)!n^x} = 1$

Proof 2.2

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n}{(n-x)!n^x} &= \\ \lim_{n \rightarrow \infty} \left\{ \frac{n(n-1)(n-2)\dots(n-x+1)(n-x)!}{(n-x)!n^x} \right\} &= \\ = \lim_{n \rightarrow \infty} \left\{ \frac{n(n-1)(n-2)\dots(n-x+1)}{n^x} \right\} &= \\ \lim_{n \rightarrow \infty} \frac{n}{(n-x)!n^x} &= \\ \lim_{n \rightarrow \infty} \left\{ \left(\frac{n}{n}\right) \cdot \left(\frac{n-1}{n}\right) \dots \left(\frac{n-x+1}{n}\right) \right\} &= \\ \lim_{n \rightarrow \infty} \frac{n}{(n-x)!n^x} &= \\ \lim_{n \rightarrow \infty} \left\{ 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{x-1}{n}\right) \right\} &= 1 \quad (3) \end{aligned}$$

Lemma 2.3 (Stirling approximation principle): Given an integer $\alpha; \alpha > 0$, the factorial of a large number can be replaced with the approximation

$$\alpha! \approx \sqrt{2\pi\alpha} \left(\frac{\alpha}{e}\right)^\alpha$$

Proof 2.3 This lemma can be derived using the integral definition of the factorial,

$$\alpha! = \Gamma(\alpha + 1) = \int_0^\infty x^\alpha e^{-x} dx \quad (4)$$

Note that, the derivative of the logarithm of the integrand can be written

$$\frac{d}{dx} \ln(x^\alpha e^{-x}) = \frac{d}{dx} (\alpha \ln x - x) = \frac{\alpha}{x} - 1$$

The integrand is sharply peaked with the contribution important only near $x = \alpha$. Therefore, let $x = \alpha + \delta$ where $\delta \leq \alpha$, and write

$$\begin{aligned} \ln(x^\alpha e^{-x}) &= \alpha \ln(\alpha + \delta) - (\alpha + \delta) \\ &= \alpha \ln \left[\alpha \left(1 + \frac{\delta}{\alpha}\right) \right] - (\alpha + \delta) \\ &= \alpha \left[\ln \alpha + \ln \left(1 + \frac{\delta}{\alpha}\right) \right] - (\alpha + \delta) \\ &= \alpha \left[\ln \alpha + \frac{\delta}{\alpha} - \frac{1}{2} \left(\frac{\delta^2}{\alpha^2}\right) + \dots \right] - (\alpha + \delta) \\ \ln(x^\alpha e^{-x}) &= \alpha \ln \alpha + \delta - \frac{1}{2} \left(\frac{\delta^2}{\alpha}\right) - \alpha - \delta + \\ \ln(x^\alpha e^{-x}) &= \ln \alpha^\alpha - \alpha - \frac{\delta^2}{2\alpha} + n \quad (5) \end{aligned}$$

Taking the exponential of each side of (5) gives

$$x^\alpha e^{-x} \approx e^{\ln \alpha^\alpha - \alpha - \frac{\delta^2}{2\alpha}} = e^{\ln \alpha^\alpha} e^{-\alpha} e^{-\frac{\delta^2}{2\alpha}} = \left(\frac{\alpha}{e}\right)^\alpha e^{-\frac{\delta^2}{2\alpha}} \quad (6)$$

Plugging (6) into the integral expression for $\alpha!$, that is, (4) gives

$$\alpha! \approx \int_0^\infty \left(\frac{\alpha}{e}\right)^\alpha e^{-\frac{\delta^2}{2\alpha}} d\delta = \left(\frac{\alpha}{e}\right)^\alpha \int_{-\infty}^\infty e^{-\frac{\delta^2}{2\alpha}} d\delta \quad (7)$$

From (7), suppose

$$I = \int_0^\infty \left(\frac{\alpha}{e}\right)^\alpha e^{-\frac{\delta^2}{2\alpha}} d\delta \quad \text{then } I^2 = \iint_{00}^\infty e^{-\frac{1}{2\alpha}(\delta^2+k^2)} \partial\delta\partial k$$

Transforming I^2 from algebra to polar coordinates yields $\delta = \rho \cos \theta$, $k = \rho \sin \theta$, and $\delta^2+k^2 = \rho^2$ Jacobian (J) of the transformation is

$$J = \begin{vmatrix} \frac{\partial\delta}{\partial\rho} & \frac{\partial\delta}{\partial\theta} \\ \frac{\partial k}{\partial\rho} & \frac{\partial k}{\partial\theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho$$

Hence,

$$I^2 = \iint_{00}^\infty e^{-\frac{\rho^2}{2\alpha}} \rho \partial\rho\partial\theta$$

$$\text{Let } u = \frac{\rho^2}{2\alpha} \Rightarrow d\rho = \frac{\alpha}{\rho} du$$

$$\therefore I^2 = \int_0^{2\pi} \int_0^\infty e^{-u} \rho \frac{\alpha}{\rho} du \partial\theta$$

$$= -\alpha \int_0^{2\pi} \left| e^{-u} \right|_0^\infty \partial \theta = \alpha \int_0^{2\pi} \partial \theta$$

$$= \alpha \left| \theta \right|_0^{2\pi} = 2\pi\alpha \Rightarrow I = \sqrt{2\pi\alpha}$$

Substituting for I in (7) gives

$$\alpha! \approx \sqrt{2\pi\alpha} \left(\frac{\alpha}{e} \right)^\alpha \tag{8}$$

CONVERGENCE OF BERNOULLI DISTRIBUTION TO OTHER PROBABILITY DISTRIBUTION

Bernoulli probability function to binomial probability distribution

Theorem 3.1 (Convergence of Bernoulli to Binomial) suppose that the outcomes are labeled success and failure in n Bernoulli trials and that their respective probabilities are p and $q = 1-p$. If $S_n = X$ is the total number of success in the n Bernoulli trials, what is the probability function of x ? In other word, what is $P[S_n = x]$ or $P[X = x]$ for $x = 0, 1, 2, \dots, n$?

Proof 3.1 To solve this problem, let us first consider the event of obtaining success on each of the first x Bernoulli trials followed by failures on each of the remaining $n-x$ Bernoulli trials. To compute the probability of this event, let s_i denote the event of success and f_i the event of failure on the i^{th} Bernoulli trial. Then, the probability we seek is $P(s_1 s_2 \dots s_x f_{x+1} f_{x+2} \dots f_n)$. Now, because we are dealing with repeated performances of identical experiments (Bernoulli experiments), the events $s_1 s_2 \dots s_x f_{x+1} f_{x+2} \dots f_n$ are independent. Therefore, the probability of getting x success and $n-x$ failures in the one specific order designated below is

$$P(s_1 s_2 \dots s_x f_{x+1} f_{x+2} \dots f_n) = P(s_1)P(s_2) \dots P(s_x)P(f_{x+1})P(f_{x+2}) \dots P(f_n) \\ P(s_1 s_2 \dots s_x f_{x+1} f_{x+2} \dots f_n) = p^x (1-p)^{n-x} \tag{9}$$

In fact, the probability of getting x success and $n-x$ failures in any one specific order is $p^x(1-p)^{n-x}$. Thus, the probability of getting exactly x successes in n Bernoulli trials is the number of ways in which successes can occur among the trials multiplied by the quantity $p^x(1-p)^{n-x}$. Now, the number of ways in which x successes can occur among the n trials is the same as the number of distinct arrangements of n items of which x are of one kind (success), and $n-x$ are of another kind (failure). In counting techniques, this number is the binomial coefficient

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \tag{10}$$

The probability function of S_n or X , the number of successes in Bernoulli trials, is the product of (9) and (10) which gives

$$P[X = x] = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}; & x = 0, 1, 2, \dots, n \\ 0 & \text{elsewhere} \end{cases} \tag{11}$$

This probability function is called the binomial probability function or the binomial probability distribution, or simply binomial distribution due to the role of the binomial coefficient in deriving the exact expression of the function. In view of its importance in probability and statistics, there is a special notation for this function, as we shall see in the definition given below.

Definition 3.1 (The Binomial probability function) Let $n \in \mathbb{Z}^+$, that is, be a positive integer, and let p and q be probabilities, with $p + q = 1$. The function (11)

$$b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, 2, \dots, n$$

is called the binomial probability function. A random variable is said to have a binomial distribution and is referred to as a binomial random variable if its probability function is the binomial probability function.

Binomial probability function to poisson probability function

In this section, we prefer to take different approach (Principle of Mathematical Induction) to show that as n number of Bernoulli trials tends to infinity, the binomial probability binom $[X = x; n; p]$ converges to the Poisson probability $\text{POI}[X = x; \lambda]$ provided that the value of p is allowed to vary with n so that np remains equal to λ for all values of n . This is formally stated in the following theorem, which is known as Poisson's Limit Law/Theorem.

Theorem 3.2 (Poisson's Limit Theorem)

If p varies with n so that $np = \lambda$, where λ is a positive constant, then

$$\lim_{n \rightarrow \infty} \text{binom}[X = x; n; p] = \text{POI}[X = x; \lambda]$$

Proof 3.2 Required to show that as $n \rightarrow \infty$, $p \downarrow 0$ and $np \rightarrow \lambda$ (a constant), one has $X(n; p) \rightarrow_D X$

Where X is Poisson with mean λ . Using (11), if $x = 0$ then

$$P[X = 0] = \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n$$

From Theorem (3.2), as $n \rightarrow \infty$, $p \downarrow 0$ and $np \rightarrow \lambda$ (a constant) implies $p = \frac{\lambda}{n}$

$$P[X = 0] = \left(1 - \frac{\lambda}{n}\right)^n$$

$$\therefore \lim_{n \rightarrow \infty} P[X = 0] = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}; \quad (\text{by lemma (2.1)})$$

Similarly; when $x=1$

$$P[X = 1] = \binom{n}{1} p^1 (1-p)^{n-1} = np(1-p)^{n-1} = \frac{np(1-p)^n}{(1-p)}$$

$$\lim_{n \rightarrow \infty} P[X = 1] = \lim_{n \rightarrow \infty} \frac{\lambda \left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^n} = \lambda \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \times \lim_{n \rightarrow \infty} \frac{1}{\left(1 - \frac{\lambda}{n}\right)}$$

$\lim_{n \rightarrow \infty} P[X=1] = \lambda e^{-\lambda} \times \lambda e^{(-\lambda)}$
Again, when $x=2$

$$P[X = 2] = \binom{n}{2} p^2 (1-p)^{n-2} = \frac{n!}{(n-2)!2!} p^2 \frac{(1-p)^n}{(1-p)^2} = \frac{np^2(n-1)(1-p)^n}{2!(1-p)^2} = \frac{(np)^2 \left(1 - \frac{1}{n}\right) (1-p)^n}{2!(1-p)^2}$$

$$\therefore \lim_{n \rightarrow \infty} P[X = 2] = \lim_{n \rightarrow \infty} \left[\frac{\lambda \left(1 - \frac{1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^n}{2! \left(1 - \frac{\lambda}{n}\right)^2} \right]$$

$$= \frac{\lambda^2}{2!} \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right) \lim_{n \rightarrow \infty} \frac{1}{\left(1 - \frac{\lambda}{n}\right)^2}$$

$$\lim_{n \rightarrow \infty} P[X = 2] = \frac{\lambda^2}{2!} \times 1 \times e^{-\lambda} \times 1 = \frac{\lambda^2 e^{-\lambda}}{2!}$$

For $x = k$

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

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

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

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

$$\lim_{n \rightarrow \infty} P[X = k] =$$

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

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

$$\lim_{n \rightarrow \infty} \left\{ \left(1 - \frac{\lambda}{n}\right)^n \right\} \lim_{n \rightarrow \infty} \left\{ \frac{1}{\left(1 - \frac{\lambda}{n}\right)^k} \right\}$$

$$\lim_{n \rightarrow \infty} P[X = k] = \frac{\lambda^k}{k!} \times 1 \times e^{-\lambda} \times 1 = \frac{\lambda^k e^{-\lambda}}{k!}$$

Therefore,

$$\lim_{n \rightarrow \infty} P[X = k] = \lim_{n \rightarrow \infty} \left\{ \binom{n}{k} p^k (1-p)^{n-k} \right\} = \frac{\lambda^k e^{-\lambda}}{k!} \tag{12}$$

Now for $x = k + 1$

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

Note: As $5 \times 4! = 5!$ so $(n-k)(n-k-1)! = (n-k)!$ and $(k+1)! = (k+1)k!$

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

$$P(k+1) = \frac{p}{1-p} \times \frac{(n-k)}{K+1} \left\{ \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \right\}$$

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

$$\therefore \lim_{n \rightarrow \infty} P(k+1) = \frac{\lambda}{k+1} \frac{\lim_{n \rightarrow \infty} \left(1 - \frac{k}{n}\right)}{\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)}$$

$$\left[\lim_{n \rightarrow \infty} \left\{ \binom{n}{k} p^k (1-p)^{n-k} \right\} \right]$$

$\lim_{n \rightarrow \infty} \left(1 - \frac{k}{n}\right) = 1$, $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right) = 1$ and from (12), the limit inside the square bracket is $\frac{\lambda^k e^{-\lambda}}{k!}$.

Hence,

$$\lim_{n \rightarrow \infty} P(k+1) = \frac{\lambda}{k+1} \times \frac{1}{1} \times \frac{\lambda^k e^{-\lambda}}{k!} = \frac{\lambda^{k+1} e^{-\lambda}}{k+1!} \quad (13)$$

Since this is true for $x=0,1,2,\dots,k$ and $k+1$, then it is true $\forall x \in Z^+$. That is,

$$P[X = x] = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!} & x = 1, 2, \dots, \\ 0 & \text{if } x < 0 \end{cases} \quad (14)$$

This is the required Poisson probability distribution function introduced by a French mathematician Simeon Denis Poisson (1781–1840).

Definition 3.2 (Poisson random variable) A random variable X , taking on one of the values $0,1,2,\dots$, (a count of events occurring at random in regions of time or space) is said to be a Poisson random variable with parameter λ if for some $\lambda > 0$,

$$P[X = x] = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

Equation (14) defines a probability mass function, since

$$\sum_{i=0}^{\infty} P[X = x] = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1 \quad (15)$$

Remarks 3.1 thus, when n is large and p is small, we have shown using principle of mathematical induction approach that Poisson with parameter λ is a very good approximation to the distribution of the number of success in “ n ” independent Bernoulli trials.

Binomial distribution to Gaussian distribution

In probability theory, de Moivre–Laplace theorem asserts that under certain conditions, the probability mass function of the random number of “successes” observed in a series of n independent Bernoulli trials, each having probability p of success, converges to the probability density function of the normal distribution with mean np and standard deviation $\sqrt{np(1-p)}$, as n grows large, assuming p is not 0 or 1. The theorem appeared in the second edition of *The Doctrine of Chances* by Abraham de Moivre, published in 1738.^[14,21] Although de Moivre did not use the term “Bernoulli trials,” he wrote about the probability distribution of the number of times “heads” appears when a coin is tossed 3600 times. He proved the result for $p = \frac{1}{2}$.^[18,20]

In this paper, we support the prove to all values of p such that $p \neq 0$ and $p \neq 1$ ”.

Theorem 3.3 (de Moivre–Laplace Limit Theorem) As n grows large, for x in the neighborhood of np we can approximate.

$$\binom{N}{x} p^{xq^{N-x}} \approx \frac{1}{\sqrt{2\pi Npq}} e^{-\frac{(x-Np)^2}{2Npq}},$$

$$p+q=1 \quad p, q > 0 \quad (16)$$

In the sense that the ratio of the left-hand side to the right-hand side converges to 1 as $N \rightarrow \infty, x \rightarrow \infty$ and p not too small and not too big.

Proof 3.3 From equation (11), the binomial probability function is

$$f(x; N, p) = \binom{N}{x} p^x (1-p)^{N-x}$$

$$= \frac{N!}{x!(N-x)!} p^x (1-p)^{N-x} \quad (17)$$

Here by lemma (2.3), we have

$$\begin{aligned}
 f(x; N, p) &\approx \frac{\sqrt{2\pi N} \left(\frac{N}{e}\right)^N}{\sqrt{2\pi x} \left(\frac{x}{e}\right)^x \sqrt{2\pi(N-x)} \left(\frac{N-x}{e}\right)^{N-x}} \\
 &\approx \frac{\sqrt{2\pi}}{\sqrt{2\pi}\sqrt{2\pi}} \frac{\sqrt{NN^N}}{\sqrt{x}\sqrt{N-x}x^x(N-x)^{N-x}} \frac{e^{-N}}{e^{(-x)}e^{-(N-x)}} \\
 &\approx \frac{1}{\sqrt{2\pi}} \frac{N^{N+\frac{1}{2}}}{x^{x+\frac{1}{2}}(N-x)^{N-x+\frac{1}{2}}} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \frac{N^{N+\frac{1}{2}}}{x^{x+\frac{1}{2}}(N-x)^{N-x+\frac{1}{2}}} p^x(1-p)^{N-x}
 \end{aligned}$$

Multiplying both numerator and denominator by $\sqrt{N} \equiv N^{\frac{1}{2}}$, we have;

$$\begin{aligned}
 f(x; N, p) &= \frac{1}{\sqrt{2\pi N}} \frac{N^{N+\frac{1}{2}} N^{\frac{1}{2}}}{x^{x+\frac{1}{2}}(N-x)^{N-x+\frac{1}{2}}} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \frac{N^{N+\frac{1}{2}+\frac{1}{2}+x-x}}{r^{r+\frac{1}{2}}(N-x)^{N-x+\frac{1}{2}}} p^x(1-p)^{N-x} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \frac{N^{x+\frac{1}{2}}}{x^{x+\frac{1}{2}}} \frac{N^{N-x+\frac{1}{2}}}{(N-x)^{N-x+\frac{1}{2}}} p^x(1-p)^{N-x} \\
 f(x; N, p) &\approx \frac{1}{\sqrt{2\pi N}} \left(\frac{N}{x}\right)^{x+\frac{1}{2}} \left(\frac{N}{N-x}\right)^{N-x+\frac{1}{2}} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \left(\frac{x}{N}\right)^{-x-\frac{1}{2}} \left(\frac{N-x}{N}\right)^{-N+x-\frac{1}{2}} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \left(\frac{x}{N}\right)^{-x-\frac{1}{2}} \left(\frac{N-x}{N}\right)^{-N+x-\frac{1}{2}} p^x(1-p)^{N-x}
 \end{aligned} \tag{18}$$

Change variables $x = Np + \varepsilon$, ε measures the distance from the mean of the Binomial Np , and the measured quantity x . The variance of a Binomial is $Np(1-p)$, so the typical deviation of x from Np is given by $\sqrt{Np(p-1)}$ terms of the form $\frac{\mu}{x}$ will,

therefore, be of order $\frac{1}{\sqrt{N}}$ and will be small.

Re-write 18 in terms of ε .

$$\begin{aligned}
 f(x; N, p) &\approx \frac{1}{\sqrt{2\pi N}} \left(\frac{Np + \varepsilon}{N}\right)^{-x-\frac{1}{2}} \\
 &\quad \left(\frac{N - Np - \varepsilon}{N}\right)^{-N+x-\frac{1}{2}} p^x(1-p)^{N-x} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \left(p + \frac{\varepsilon}{N}\right)^{-x-\frac{1}{2}} \left((1-p)\frac{\varepsilon}{N}\right)^{-N+x-\frac{1}{2}} \\
 &\quad p^x(1-p)^{N-x} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \left[p\left(1 + \frac{\varepsilon}{Np}\right)\right]^{-x-\frac{1}{2}} \\
 &\quad \left[(1-p)\left(1 - \frac{\varepsilon}{N(1-p)}\right)\right]^{-N+x-\frac{1}{2}} p^x(1-p)^{N-x} \\
 &\approx \frac{1}{\sqrt{2\pi N}} p^{-x-\frac{1}{2}} \left(1 + \frac{\varepsilon}{Np}\right)^{-x-\frac{1}{2}} (1-p)^{-N+x-\frac{1}{2}} \\
 &\quad \left(1 - \frac{\varepsilon}{N(1-p)}\right)^{-N+x-\frac{1}{2}} p^x(1-p)^{N-x} \\
 &\approx \frac{1}{\sqrt{2\pi N}} \left(1 + \frac{\varepsilon}{Np}\right)^{-x-\frac{1}{2}} \left(1 - \frac{\varepsilon}{N(1-p)}\right)^{-N+x-\frac{1}{2}} \\
 &\quad p^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}} \\
 f(x; N, p) &\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \left(1 + \frac{\varepsilon}{Np}\right)^{-x-\frac{1}{2}} \\
 &\quad \left(1 - \frac{\varepsilon}{N(1-p)}\right)^{-N+x-\frac{1}{2}}
 \end{aligned} \tag{19}$$

Rewrite (19) in exponential form to have

$$\begin{aligned}
 f(x; N, p) &\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \\
 &\quad \exp \left[\ln \left\{ \left(1 + \frac{\varepsilon}{Np}\right)^{-x-\frac{1}{2}} \left(1 - \frac{\varepsilon}{N(1-p)}\right)^{-N+x-\frac{1}{2}} \right\} \right]
 \end{aligned}$$

$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[\begin{array}{l} \ln \left(1 + \frac{\varepsilon}{Np} \right)^{-x-\frac{1}{2}} \\ \ln \left(1 - \frac{\varepsilon}{N(1-p)} \right)^{-N+x-\frac{1}{2}} \end{array} \right]$$

$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[\begin{array}{l} \left(-x - \frac{1}{2} \right) \ln \left(1 + \frac{\varepsilon}{Np} \right) \\ \left(-N + x - \frac{1}{2} \right) \ln \left(1 - \frac{\varepsilon}{N(1-p)} \right) \end{array} \right]$$

$$f(x; N, p) \approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[\begin{array}{l} \left(-Np - \varepsilon - \frac{1}{2} \right) \ln \left(1 + \frac{\varepsilon}{Np} \right) \\ \left(-N(1-p) + \varepsilon - \frac{1}{2} \right) \ln \left(1 - \frac{\varepsilon}{N(1-p)} \right) \end{array} \right] \quad (20)$$

Suppose $f(x) = \ln(1+x)$ using Maclaurin series
 $f(x) = x - \frac{x^2}{2} + n$ and similarly

$f(x) = \ln(1-x) \approx -x - \frac{1}{2}x^2$. Therefore,

$$\ln \left(1 + \frac{\varepsilon}{Np} \right) \approx \frac{\varepsilon}{Np} - \frac{1}{2} \left(\frac{\varepsilon}{Np} \right)^2 \quad (21)$$

$$\ln \left(1 - \frac{\varepsilon}{N(1-p)} \right) \approx \frac{-\varepsilon}{N(1-p)} - \frac{1}{2} \left(\frac{\varepsilon}{N(1-p)} \right)^2 \quad (22)$$

Putting (21) and (22) in (20) and simplify, we have

$$f(x; N, p) \approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[\begin{array}{l} \left(-Np - \varepsilon - \frac{1}{2} \right) \left(\frac{\varepsilon}{Np} - \frac{1}{2} \left(\frac{\varepsilon}{Np} \right)^2 \right) \\ \left(-N(1-p) + \varepsilon - \frac{1}{2} \right) \left(\frac{-\varepsilon}{N(1-p)} - \frac{1}{2} \left(\frac{\varepsilon}{N(1-p)} \right)^2 \right) \end{array} \right]$$

$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[\begin{array}{l} -Np \left(\frac{\varepsilon}{Np} - \frac{1}{2} \left(\frac{\varepsilon}{Np} \right)^2 \right) - \frac{\varepsilon^2}{Np} - N(1-p) \\ \left(\frac{-\varepsilon}{N(1-p)} - \frac{1}{2} \left(\frac{\varepsilon}{N(1-p)} \right)^2 \right) - \frac{\varepsilon^2}{N(1-p)} \end{array} \right]$$

$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[\begin{array}{l} -\varepsilon + \frac{\varepsilon^2}{2Np} - \frac{\varepsilon^2}{Np} + \varepsilon \\ + \frac{\varepsilon^2}{2N(1-p)} - \frac{\varepsilon^2}{N(1-p)} \end{array} \right]$$

$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[\begin{array}{l} \frac{\varepsilon^2}{Np} \left(\frac{1}{2} - 1 \right) \\ + \frac{\varepsilon^2}{N(1-p)} \left(\frac{1}{2} - 1 \right) \end{array} \right]$$

$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[-\frac{\varepsilon^2}{2Np} \left(\frac{1}{p} + \frac{1}{1-p} \right) \right]$$

$$\approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[-\frac{\varepsilon^2}{2Np(1-p)} \right]$$

$$f(x; N, p) \approx \frac{1}{\sqrt{2\pi Np(1-p)}} \exp \left[-\frac{\varepsilon^2}{2Np(1-p)} \right] \quad (23)$$

Recall that $x = Np + \varepsilon$ which implies that $\varepsilon^2 = (x - Np)^2$. From Binomial distribution $Np = \mu$ and $Np(1-p) = \sigma^2$ which also implies that $\sqrt{Np(1-p)} = \sigma$ making appropriate substitution of these in the equation (23) above provides;

$$f(x; N, p) \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right]$$

$$= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}; \text{ for } -\infty < x < \infty \quad (24)$$

We have easily derived equation (24) that is popularly and generally known anywhere in the whole world as normal or Gaussian distribution.

Definition 3.3 (Normal distribution) A continuous random variable has a normal distribution, is said

to be normally distributed, and is called a normal random variable if its probability density function can be defined as follows: Let μ and σ be constants with $-\infty < \mu < \infty$ and $\sigma > 0$. The function (24)

$$f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; \text{ for } -\infty < x < \infty$$

is called the normal probability density function with parameters μ and σ .

The normal probability density function is without question the most important and most widely used distribution in statistics.^[15] It is also called the ‘‘Gaussian curve’’ named after the Mathematician Carl Friedrich Gauss.

To verify that equation (24) is a proper probability density function with parameters μ and σ is to show that the integral

$$I = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx$$

Is equal to 1.

Change variables of integration by letting

$$z = \frac{x-\mu}{\sigma} \text{ which implies that } dx = \sigma dz. \text{ Then,}$$

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}} \sigma dz \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{z^2}{2}} dz = \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-\frac{z^2}{2}} dz \end{aligned}$$

Now,

$$I^2 = \left[\sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx \right] \left[\sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-\frac{y^2}{2}} dy \right]$$

Or equivalently

$$I^2 = \frac{2}{\pi} \int_0^{\infty} \int_0^{\infty} e^{-\frac{(x^2+y^2)}{2}} dx dy$$

Here, x, y are dummy variables. Switching to polar coordinate by making the substitutions $x = r \cos \theta$, $y = r \sin \theta$ produces the Jacobian of the transformation as

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \quad (25)$$

So,

$$I^2 = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta$$

Put $a = \frac{r^2}{2} \Rightarrow dr = \frac{da}{r}$. Therefore,

$$I^2 = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \int_0^{\infty} e^{-a} r \frac{da}{r} d\theta = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \left[e^{-a} \right]_0^{\infty} d\theta$$

$$I^2 = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} d\theta = \frac{2}{\pi} \left| \theta \right|_0^{\frac{\pi}{2}} = 1$$

Thus $I = 1$, indicating that (24) is a proper p.d.f.

CONCLUDING REMARKS

It is now well-known that a Binomial r.v. is the sum of i.i.d. Bernoulli r.v.s., Poisson r.v. arises from Binomial (n Bernoulli trial) with n increasing indefinitely, p reducing to 0 so that $np = \lambda$ (a constant), also, when n (number of trials in a Bernoulli experiment) increases without bound and $p \neq 0$ and $p \neq 1$ (that is, p is moderate), the resulting limiting distribution is Gaussian with mean (μ) = np and variance (σ^2) = $np(1-p)$. This our alternative technique provides a direct approach of convergence, this material should be of pedagogical interest and can serve as an excellent teaching reference in probability and statistics classes where only basic calculus and skills to deal with algebraic expressions are the only background requirements. The proofs are straightforward and require only an additional knowledge of Maclaurin series expansion, gamma function and basic limit concepts which were thoroughly dealt with in preliminaries section.

ACKNOWLEDGMENT

The authors are highly grateful to the editor and anonymous referees for their reading through the manuscript, constructive comments, and suggestions that helped in the improvement of the revised version of the paper.

REFERENCES

1. Baclawski K, Rota GC, Billey S. An Introduction to the Theory of Probability. Cambridge, MA: Massachusetts

- Institute of Technology; 1989.
2. Bagui SC, Bhaumik DK, Mehra KL. A few counter examples useful in teaching central limit theorem. *Am Stat* 2013a;67:49-56.
3. Bagui SC, Bagui SS, Hemasinha R. Non-rigorous proof's Stirling's formula. *Math Comput Educ* 2013b;47:115-25.
4. Bagui SC, Mehra L. Convergence of binomial, poisson, negative-binomial, and gamma to normal distribution: Moment generating functions technique. *Am J Math Stat* 2016;6:115-21.
5. Bain LJ, Engelhardt M. *Introduction to Probability and Mathematical Statistics*. 2nd ed. Belmont: Duxbury Press; 1992.
6. Billingsley P. *Probability and Measure*. In: Brion L, editor. 3rd ed. New York: Wiley; 1995.
7. Brion LM. *Asymptotic Relative Efficiency in Non-parametric Statistics*, Ph.D., Dissertation. Austin: University of Texas at Austin; 1990.
8. Feller W. *An Introduction to Probability Theory and Its Applications*. 3rd ed. Vol. 1. New York: John Wiley and Sons; 1973. p. 273, F3712.
9. Hogg RV, Tanis EA. *Probability and Statistical Inference*. 5th ed. Upper Saddle River, New Jersey, United State of America: Prentice-Hall, Inc., Simon and Schuster, a Viacom Company; 1997.
10. Inlow MA. Moment generating function proof of the Lindeberg-Levy central limit theorem. *Am Stat* 2010;64:228-30.
11. Jeffrey DB, Richard MR. Illustrating the law of large numbers. *Am Stat Assoc* 2003;1:51-7.
12. Lesigne E. Heads or tails: An introduction to limit theorems in probability. *Am Math Soc* 2005;28:150.
13. Peggy TS. *A First Course in Probability and Statistics with Applications*. 2nd ed. Washington, DC: Harcourt Brace Jovanovich; 1989.
14. Proschan MA. The normal approximation to the binomial. *Am Stat* 2008;62:62-3.
15. Ramana BV. *Higher Engineering Mathematics*. New Delhi: Tata McGraw Hill Publishing Company Limited; 2008.
16. Reed WJ. The Double Pareto-lognormal distribution: A new parametric model for size distributions. *Commun Stat Theory Methods* 2004;33:1733-53.
17. Reinhard V. *Probability and Statistics: Probability Theory, Stochastic Processes and Random Fields*. Oxford, United Kingdom: Vol. 1. Eolss Publishers Co. Ltd.; 2009.
18. Richards JI, Youn HK. *Theory of Distributions: A Non-technical Introduction*. Cambridge, MA: Published by Cambridge University Press; 1990.
19. Sefling RJ. *Approximation Theorems of Mathematical Statistics*. New York: Wiley; 1980.
20. Soong TT. *Fundamentals of Probability and Statistics for Engineers*. England: John Wiley and Sons Ltd.; 2004.
21. Stigler SM. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge, MA: The Belknap Press of Harvard University Press; 1986.
22. Walck C. *Hand-book on Statistical Distributions for Experimentalists*. Sweden: Particle Physics Group, Fysikum University of Stockholm; 2007.