

On the Convergence of Negative Binomial Distribution

Subhash Bagui^{1,*}, K. L. Mehra²

¹Department of Mathematics and Statistics, The University of West Florida, Pensacola, FL, United States

²Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB, Canada

Abstract This paper offers four different methods of proof of the convergence of negative binomial $NB(n, p)$ distribution to a normal distribution, as $n \rightarrow \infty$. All these methods of proof may not be available together in a book or in a single paper in literature. The reader should find the presentation enlightening and worthwhile from a pedagogical viewpoint. The article should of interest to teachers and undergraduate seniors in probability and statistics courses.

Keywords Negative binomial distribution, Central limit theorem, Moment generating function, Ratio method, Stirling's approximations

1. Introduction

The negative binomial (NB) distribution was first initiated by Pascal (1679), although its earliest concrete formulation and introduction was due to Montmort (1741); see Todhunter (1865). Montmort derived the distribution of number of tosses of a coin required to obtain a specified number of heads. Student (1907) in an empirical study employed NB to model countings on haemocytometer data. Bartko (1961) published an excellent review article on many aspects of the NB distribution. The NB is an often used distribution in statistical modeling. The inverse binomial sampling scheme is modeled with NB distribution. This scheme is described as: in an infinite sequence of independent Bernoulli trials continue to select items until a fixed number of successes (or failures) are captured. For example, suppose we are conducting a wildlife survey and wanting to catch n of a certain type of restricted birds. In the sampling process we capture birds at random until we have bagged n of these restricted birds (and an unknown random number of other types of birds). Thus, the resulting sample size would be more than n . The NB models the distribution of either the resulting total sample size or the unknown random number of other types of birds.

Thus, there are a couple of variations of the NB distribution. The first version deals with the total number of trials (say, Y_n) necessary to obtain n successes. With this version, the probability mass function of Y_n is given by

$P(Y_n = y) = \binom{y-1}{n-1} p^n q^{y-n}$, for integer $y \geq n$. Here p is the probability of success and $q = 1 - p$ is the probability of failure, with $0 < p < 1$. If we let X_n denote the number of failures before the n th success, then $X_n = Y_n - n$. The second version counts the number of failures before the n th success. In this version, the probability mass function of X_n is given by

$$P(X_n = x) = \binom{n+x-1}{x} p^n q^x, \text{ for integer } x = 0, 1, 2, \dots \quad (1.1)$$

It is well known that the mean of X_n is $E(X_n) = \mu = (nq)/p$ and the variance of X_n is $\text{Var}(X_n) = \sigma^2 = (nq)/p^2$. A little algebra shows that $\sigma^2 = \mu/p = \mu(p+q)/p = \mu + \mu^2/n$. Thus, the variance is always larger than the mean for NB distribution. For the data which points to a larger variance than the mean, the Poisson distribution is unsuitable for modeling as it requires the mean and variance to be equal. In such cases, the negative binomial seems more suitable. The NB distribution has become increasingly popular as a flexible alternative to the Poisson, especially when the underlying variance seems greater than the mean and when independence of the counts is also doubtful. Among others, Arbous and Kerrich (1951), Greenwood and Yule (1920), and Kemp (1970) have applied NB to model accident statistics. Furry (1937) and Kendall (1949) have shown its applications in birth-and-death processes. For more applications see Johnson et al. (1993) and also Feller (1957).

Relationship with Poisson distribution.

Suppose $X | \Theta = \theta$ follows Poisson distribution with Θ following a Gamma (n, β) ; then

* Corresponding author:

sbagui@uwf.edu (Subhash Bagui)

Published online at <http://journal.sapub.org/ajms>

Copyright © 2019 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

$$P(X=x) = \frac{1}{\beta^n \Gamma(n)} \int_0^\infty \frac{e^{-\theta} \theta^x}{x!} \theta^{n-1} e^{-\theta/\beta} d\theta$$

$$= \binom{n+x-1}{n-1} \left(\frac{1}{1+\beta} \right)^n \left(\frac{\beta}{1+\beta} \right)^x.$$

So the marginal distribution of X is negative binomial with parameters n and $p = 1/(1+\beta)$. This result was due to Greenwood and Yule (1920). They used it to model “accident proneness”. The parameter $\Theta = \theta$ is the expected number of accidents for an individual, which is presumed to vary from person to person.

Another important formulation is due to Luders (1934) and Quenouille (1949). In this formulation the NB arises as the distribution of the sum of N independent random variables each having the same logarithmic distribution and N having a Poisson distribution. This has application in entomology where the counts of larvae over the plots in a field are observed. The larvae are hatched from egg masses which appear at random over the field. If the number of egg masses on a plot follow a Poisson distribution and the survivors from the egg masses follow a Logarithmic distribution, then the resulting distribution of larvae on plots will be a Negative Binomial (See Gurland 1959).

For various methods of approximating the NB probability, see Bartko (1965). It is well-known that as n increases indefinitely, the NB converges to a normal distribution. Accordingly, in statistical practice the NB probabilities are approximated using the appropriate normal density for large n . There are multiple ways one can show the convergence of the NB to the normal. But all these proofs may be not be found together in a single book or an article. The main aim of this article is to the present four different methods of proof of this convergence, as $n \rightarrow \infty$. The first proof is based on the well-known Stirling’s formula, and the other three methods are the Ratio method, the Method of Moment Generating Functions (mgf’s), and lastly that of the Central Limit Theorem. These contrasting methods of proof are useful from a pedagogical standpoint. Bagui and Mehra (1917) dealt with similar proofs for showing the convergence of Binomial to the limiting normal, as $n \rightarrow \infty$.

The paper is organized as follows: In Section 2, we list some preliminary results that will be used in the subsequent sections for proving these convergences. The details of various poofs of convergence are provided in Section 3. Some concluding remarks are given in Section 4.

2. Preliminaries

In this section we state a few useful definitions, formulas, Lemmas, and Theorems which we shall employ in detailing a number of proofs in Section 3.

Formula 2.1. For large n , the Stirling’s formula for approximating $n! = n(n-1)(n-2)\cdots(3)(2)(1)$ is given by

$$n! \approx \sqrt{2\pi n} (n/e)^n \quad (\text{namely, that } \left[\frac{n!}{\sqrt{2\pi n} n^n e^{-n}} \right] \rightarrow 1$$

$$\text{as } n \rightarrow \infty); \quad (2.1)$$

(\approx stands for “approximately equals to”, in the above sense, for large n)

Formula 2.2. The following equations hold:

$$(i) \quad \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots = \sum_{i=1}^{\infty} (-1)^{i-1} \frac{x^i}{i} \quad \text{for } -1 \leq x \leq 1;$$

and

$$(ii) \quad \ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots = -\sum_{i=1}^{\infty} \frac{x^i}{i} \quad \text{for } -1 \leq x \leq 1.$$

Definition 2.1. Let X be a random variable. The moment generating function (mgf) of the r.v. X is defined by $M_X(t) = E(e^{tX})$ provided it is finite for all $|t| < h$, for some $h > 0$.

We say then that the mgf $M_X(t)$ of X exists. If it exists, it is associated with a unique distribution. That is, there is a one-to-one correspondence between the pdf’s (or pmf’s) of X and the above defined mgf’s.

Lemma 2.1. Let Z be a random variable with density $f_Z(z) = (1/\sqrt{2\pi}) e^{-z^2/2}$, $-\infty < z < \infty$; that is, the r.v. $Z \sim N(0,1)$, the standard normal distribution. Then the mgf of Z is given by $M_Z(t) = e^{t^2/2}$.

Proof. From the above definition, the mgf of Z evaluates to

$$M_Z(t) = E(e^{tZ}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-z^2/2} dz$$

$$= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(1/2)(z-t)^2} dz = e^{t^2/2}.$$

Lemma 2.2. Suppose $\{\lambda(n), n \geq 1\}$ is a sequence of real numbers such that $\lim_{n \rightarrow \infty} \lambda(n) = 0$. Then

$\lim_{n \rightarrow \infty} \left(1 + \frac{\alpha}{n} + \frac{\lambda(n)}{n} \right)^{\beta n} = e^{\alpha\beta}$, as long as α and β do not depend on n .

Theorem 2.1. Suppose $\{X_n\}$ is a sequence of r.v.’s with mgf’s $M_{X_n}(t)$ for $|t| < h$ and $n = 1, 2, 3, \dots$. Suppose the r.v. X has mgf $M_X(t)$ for $|t| < h$. If $\lim_{n \rightarrow \infty} M_{X_n}(t) = M_X(t)$ for $|t| < h$, then $X_n \xrightarrow{d} X$, as $n \rightarrow \infty$.

(The symbolization $X_n \xrightarrow{d} X$ means that the distribution of the r.v. X_n converges to the distribution of the r.v. X , as $n \rightarrow \infty$).

Theorem 2.2. Let X_1, X_2, \dots, X_n be a random sample of independent and identically distributed observations from a population that has a finite mean μ and a finite variance

$\sigma^2 > 0$. Define $S_n = \sum_{i=1}^n X_i$ and $\bar{X}_n = (S_n/n)$. Then

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0,1), \text{ the}$$

standard normal distribution, as $n \rightarrow \infty$.

Theorem 2.2 is often referred to as the basic Central Limit Theorem (CLT).

Big O and Small o notations

The Big O notation $h(n) = O(g(n))$ implies that the ratio $|h(n)/g(n)|$ stays bounded, as $n \rightarrow \infty$; that is, there exists a positive constant $C < \infty$ such that $|h(n)/g(n)| < C$ for all n , however large the n may be. For example, if $|g(n)| \rightarrow 0$, as $n \rightarrow \infty$, $h(n) = O(g(n))$ implies that $|h(n)| \rightarrow 0$ at the same or higher rate than that of $|g(n)|$.

The small o notation $h(n) = o(g(n))$ implies that the ratio $|h(n)/g(n)| \rightarrow 0$, as $n \rightarrow \infty$; that is, given an $\varepsilon > 0$, however small, there exists an $n_0 = n_0(\varepsilon)$ such that $|h(n)/g(n)| < \varepsilon$ for all $n \geq n_0$. Here for example, if $|g(n)| \rightarrow 0$, as $n \rightarrow \infty$, $h(n) = o(g(n))$ implies that $|h(n)| \rightarrow 0$ at a higher rate than that of $|g(n)|$.

3. Multiple Proofs

In this section, we offer four different methods of proof for showing the convergence of negative binomial to a normal distribution, as $n \rightarrow \infty$.

3.1. Stirling Approximation Formula Method

First we rewrite the negative binomial pmf given in (1.1) as

$$P(X_n = x) = \frac{(n+x-1)!}{(n-1)!x!} p^n q^x = \left(\frac{n}{n+x}\right) \frac{(n+x)!}{n!x!} p^n q^x. \quad (3.1)$$

Now substitute Stirling's approximation formula given by (2.1) in (3.1). After some algebraic simplifications, we have

$$\begin{aligned} P(X_n = x) &\approx \left(\frac{n}{n+x}\right) \frac{\sqrt{2\pi(n+x)}(n+x)^{n+x} e^{-(n+x)}}{\sqrt{2\pi n}(n)^n e^{-n} \sqrt{2\pi x}(x)^x e^{-x}} p^n q^x \\ &= \left(\frac{n}{n+x}\right)^{1/2} \frac{(n+x)^{n+x}}{\sqrt{2\pi} \sqrt{x}(n)^n x^x} p^n q^x \\ &= \left(\frac{n}{n+x}\right)^{1/2} \frac{(n+x)^{n+x}}{\sqrt{2\pi} \sqrt{x}(n)^{n+x} (x/n)^x} p^n q^x \\ &= \frac{1}{\sqrt{2\pi} (\sqrt{nq/p})} \left(\frac{n+x}{n}\right)^{n+x-1/2} \frac{p^{n-1}}{(x/nq)^{x+1/2}} \\ &= C \left(\frac{np+xp}{n}\right)^{n+x-1/2} \left(\frac{xp}{nq}\right)^{-(x+1/2)}, \end{aligned} \quad (3.2)$$

where $C = 1/[\sqrt{2\pi}(\sqrt{nq/p})]$. Taking natural logarithms on both sides of (3.2), we get

$$\begin{aligned} \ln P(X_n = x) &= \ln C + (n+x-1/2) \ln \left(\frac{np+xp}{n}\right) \\ &\quad - (x+1/2) \ln \left(\frac{xp}{nq}\right). \end{aligned} \quad (3.3)$$

Note now that the mean μ_n and variance σ_n^2 of the NB(n, p) r.v. are given by $\mu_n = (np/q)$ and $\sigma_n^2 = (nq/p^2)$, respectively. Suppose we set $Z_n = \frac{X_n - (np/q)}{\sqrt{nq/p}}$ and

$$z = \frac{x - (np/q)}{\sqrt{nq/p}} = \frac{xp - nq}{\sqrt{nq}}. \text{ The terms in the last equation lead}$$

$$\text{to: } xp = nq + z\sqrt{nq}, \text{ so that } \frac{xp}{nq} = \left(1 + \frac{z}{\sqrt{nq}}\right), \text{ and}$$

$$np + xp = n + z\sqrt{nq}, \text{ so that } \frac{np+xp}{n} = \left(1 + \frac{z\sqrt{q}}{\sqrt{n}}\right). \text{ Using the}$$

these simplifications we re-write (3.3) as

$$\begin{aligned} \ln P(Z_n = z) &= \ln C + [(n/p) + z(\sqrt{nq}/p) - 1/2] \\ &\quad \ln[1 + z(\sqrt{q}/\sqrt{n})] \\ &\quad - [(nq/p) + z(\sqrt{nq}/p) + 1/2] \ln[1 + z/\sqrt{nq}] \\ &= \ln C + I_1(z, p, q) + I_2(z, p, q), \end{aligned} \quad (3.4)$$

where

$$\begin{aligned} I_1(z, p, q) &= [(n/p) + z(\sqrt{nq}/p) - 1/2] \ln[1 + z(\sqrt{q}/\sqrt{n})] \\ &= [(n/p) + z(\sqrt{nq}/p) - 1/2] \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (\sqrt{q/n})^k z^k \\ &= (n/p) \\ &\quad \left[z\sqrt{q/n} - \frac{z^2}{2} (\sqrt{q/n})^2 + z^3 (\sqrt{q/n})^3 \sum_{k=3}^{\infty} \frac{(-1)^{k-3}}{k} (\sqrt{q/n})^{k-3} z^{k-3} \right] \\ &\quad + (z\sqrt{nq}/p) \left[z\sqrt{q/n} - z^2 (\sqrt{q/n})^2 \sum_{k=2}^{\infty} \frac{(-1)^{k-2}}{k} (\sqrt{q/n})^{k-2} z^{k-2} \right] \\ &\quad - \frac{z}{2} \sqrt{q/n} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (\sqrt{q/n})^{k-1} z^{k-1} \\ &= \frac{z\sqrt{nq}}{p} - \frac{z^2 q}{2p} + \frac{z^3 q^{3/2}}{p\sqrt{n}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+3)} (\sqrt{q/n})^k z^k + \frac{z^2 q}{p} \\ &\quad - \frac{z^3 q^{3/2}}{p\sqrt{n}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+2)} (\sqrt{q/n})^k z^k \\ &\quad - \frac{zq^{1/2}}{2\sqrt{n}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)} (\sqrt{q/n})^k z^k \\ &= \frac{z\sqrt{nq}}{p} + \frac{z^2 q}{2p} + O(1/\sqrt{n}), \end{aligned} \quad (3.5)$$

where the last order term $O(1/\sqrt{n})$ follows since the

absolute values of the three infinite sums above can be shown to remain bounded, as $n \rightarrow \infty$, by applying Formulas 2.2 to each of them individually. Similarly,

$$\begin{aligned} I_2(z, p, q) &= -[(nq)/p + z\sqrt{nq}/p + 1/2] \ln[1 + z/\sqrt{nq}] \\ &= -(nq/p) \left[\frac{z}{\sqrt{nq}} - \frac{z^2}{2nq} + \frac{z^3}{(\sqrt{nq})^3} \sum_{k=3}^{\infty} \frac{(-1)^{k-3}}{k} \frac{z^{k-3}}{(\sqrt{nq})^{k-3}} \right] \\ &\quad - (z\sqrt{nq}/p) \left[\frac{z}{\sqrt{nq}} - \frac{z^2}{(\sqrt{nq})^2} \sum_{k=2}^{\infty} \frac{(-1)^{k-2}}{k} \frac{z^{k-2}}{(\sqrt{nq})^{k-2}} \right] \\ &\quad - \frac{z}{2\sqrt{nq}} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \frac{z^{k-1}}{(\sqrt{nq})^{k-1}} \\ &= -\frac{z\sqrt{nq}}{p} + \frac{z^2}{2p} - \frac{z^3}{pq^{1/2}\sqrt{n}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+3)} \frac{z^k}{(\sqrt{nq})^k} \\ &\quad - \frac{z^2}{p} - \frac{z^3}{pq^{1/2}\sqrt{n}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+2)} \frac{z^k}{(\sqrt{nq})^k} \\ &\quad - \frac{z}{2\sqrt{nq}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)} \frac{z^k}{(\sqrt{nq})^k} \\ &= -\frac{z\sqrt{nq}}{p} - \frac{z^2}{2p} + O(1/\sqrt{n}), \end{aligned} \quad (3.6)$$

the last order term $O(1/\sqrt{n})$ in (3.6) following again by an application of Formulas 2.2, as done for (3.5) above.

Now substituting (3.5) and (3.6) in (3.4), we have

$$\begin{aligned} \ln P(Z_n = z) &\approx \ln C + \frac{z\sqrt{nq}}{p} + \frac{z^2 q}{2p} - \frac{z\sqrt{nq}}{p} - \frac{z^2}{2p} + \\ &\quad O(1/\sqrt{n}) \\ &= \ln C - \frac{z^2(1-q)}{2p} + O(1/\sqrt{n}) = \ln C - \frac{z^2}{2} + O(1/\sqrt{n}). \end{aligned} \quad (3.7)$$

Hence, for large n , from (3.7) we get

$$P(Z_n = z) \approx C e^{-z^2/2} \Delta_n = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, \quad (3.8)$$

where $dz \approx \Delta_n = p/\sqrt{nq} = (1/\sigma_n)$. The above equation (3.8) may also be written as

$$P(X_n = x) \approx P(Z_n = z) \approx \frac{1}{\sqrt{2\pi} (\sqrt{nq}/p)} e^{-\frac{(x-nq/p)^2}{2(nq)/p^2}}. \quad (3.9)$$

This completes the proof.

3.2. The Ratio Method [5]

The ratio method uses the ratio of two successive probability terms of the pmf. The ratio of two consecutive probability terms of the negative binomial pmf given in (1.1) is leads to

$$\begin{aligned} \frac{P(X_n = x+1)}{P(X_n = x)} &= \frac{p^n \binom{n+x}{n-1} q^{x+1}}{p^n \binom{n+x-1}{n-1} q^x} \\ &= \frac{(n+x)!}{(n-1)!(x+1)!} \frac{(n-1)!x!}{(n+x-1)!} q = \frac{n+x}{x+1} q \end{aligned} \quad (3.10)$$

Let $z = [(x - [nq/p]) / (\sqrt{nq}/p)]$, so that $x = [nq/p] + z(\sqrt{nq}/p)$; substituting this expression for x into (3.10), we obtain the equation

$$\begin{aligned} \frac{P[X_n = [nq/p] + z(\sqrt{nq}/p) + 1]}{P[X_n = [nq/p] + z(\sqrt{nq}/p)]} \\ = \frac{np + nq + z\sqrt{nq}}{nq + z\sqrt{nq} + p} (q) = \frac{n + z\sqrt{nq}}{n + z(\sqrt{nq}/q) + (p/q)}. \end{aligned} \quad (3.11)$$

By setting $Z_n = (X_n - nq/p) / (\sqrt{nq}/p)$ and $\Delta = p/\sqrt{nq}$, we can rewrite (3.11) as

$$\frac{P(Z_n = z + \Delta)}{P(Z_n = z)} = \frac{1 + zq(\Delta/p)}{1 + z(\Delta/p) + \Delta^2/p}. \quad (3.12)$$

Now assume that there exists a smooth pdf $f(z)$ such that for large n , $P(Z_n = z) \approx f(z)dz$ and therefore $P(Z_n = z + \Delta) \approx f(z + \Delta)dz$. Under this broad assumption, the left hand side of (3.12) can be rewritten as

$$\frac{f(z + \Delta)}{f(z)} \approx \frac{1 + zq(\Delta/p)}{1 + z(\Delta/p) + \Delta^2/p}. \quad (3.13)$$

Upon taking logarithms on both sides of (3.13), dividing by Δ and taking limits as $n \rightarrow \infty$, or equivalently $\Delta \rightarrow 0$, we have

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \left[\frac{\ln f(z + \Delta) - \ln f(z)}{\Delta} \right] &= \lim_{\Delta \rightarrow 0} \left[\frac{\ln(1 + zq(\Delta/p))}{\Delta} \right] \\ &\quad - \lim_{\Delta \rightarrow 0} \left[\frac{\ln(1 + z(\Delta/p) + \Delta^2/p)}{\Delta} \right]. \end{aligned} \quad (3.14)$$

We simplify right hand side of (3.14) by applying L'Hopitals rule. Since the left hand side of (3.14) is evidently the derivative of $[\ln f(z)]$, as a consequence we obtain the following differential equation

$$\begin{aligned} \frac{d \ln f(z)}{dz} &= \lim_{\Delta \rightarrow 0} \left[\frac{z(q/p)}{1 + z(q/p)\Delta} \right] - \lim_{\Delta \rightarrow 0} \left[\frac{z/p + (2\Delta/p)}{1 + (z/p)\Delta + (\Delta^2/p)} \right] \\ &= \frac{zq}{p} - \frac{z}{p} = -\frac{z(1-q)}{p} = -z. \end{aligned} \quad (3.15)$$

Integrating on both sides of equation (3.15) with respect to z , we obtain $\ln f(z) = -\frac{z^2}{2} + c$, where c is the constant of

integration. We can rewrite this equation as $f(z) = ke^{-z^2/2}$ with $k = 1/\sqrt{2\pi}$ to make $f(z)$, $-\infty < z < \infty$, a valid density. We can conclude thus that the r.v. $Z_n = (X_n - nq/p)/(\sqrt{nq}/p)$ converges in distribution, as $n \rightarrow \infty$, to a standard normal $N(0,1)$ r.v., or equivalently, that the negative-binomial $NB(n, p)$ r.v. X_n follows approximately, for large n , the normal distribution with mean $\mu_n = (nq)/p$ and $\sigma_n^2 = (nq)/p^2$ as the variance.

3.3. The MGF Method [4]

Let X_n be a negative binomial r.v. with pmf given in (1.1). Then the mgf of X_n is derived as

$$\begin{aligned} M_{X_n}(t) &= E(e^{tX_n}) = \sum_{x=0}^{\infty} e^{tx} \binom{n+x-1}{x} p^n q^x \\ &= \left[p/(1-qe^t) \right]^n. \end{aligned} \quad (3.16)$$

Let $Z_n = [(X_n - nq/p)/(\sqrt{nq}/p)] = [(pX_n)/\sqrt{nq}] - \sqrt{nq}$. The mgf of Z_n then evaluates to

$$\begin{aligned} M_{Z_n}(t) &= E(e^{tZ_n}) = E\left[e^{t(pX_n/\sqrt{nq} - \sqrt{nq})} \right] \\ &= e^{-(\sqrt{nq})t} E\left[e^{(p/\sqrt{nq})tX_n} \right] = e^{-(\sqrt{nq})t} M_{X_n}\left(\frac{pt}{\sqrt{nq}}\right) \\ &= e^{-(\sqrt{nq})t} \left[p/(1-qe^{pt/\sqrt{nq}}) \right]^n \\ &= e^{-(q/\sqrt{nq})tn} \left(\frac{1}{p} - \frac{q}{p} e^{pt/\sqrt{nq}} \right)^{-n} \\ &= \left(\frac{1}{p} e^{(q/\sqrt{nq})t} - \frac{q}{p} e^{(t/\sqrt{nq})} \right)^{-n}. \end{aligned} \quad (3.17)$$

The Taylor series expansion for $e^{(q/\sqrt{nq})t}$ gives us

$$\begin{aligned} \frac{1}{p} e^{(q/\sqrt{nq})t} &= \frac{1}{p} \left[1 + \frac{qt}{\sqrt{nq}} + \frac{q^2 t^2}{(2!)nq} + \frac{q^3 t^3}{(3!)(nq)^{3/2}} e^{\xi(n)} \right] \\ &= \frac{1}{p} + \frac{qt}{p\sqrt{nq}} + \frac{qt^2}{p(2n)} + \frac{(q)q^2 t^3}{p(3!)(nq)^{3/2}} e^{\xi(n)}. \end{aligned} \quad (3.18)$$

where the sequence $\xi(n)$ lies between 0 and (tq/\sqrt{nq}) and $\xi(n) \rightarrow 0$ as $n \rightarrow \infty$.

Similarly, we obtain

$$\begin{aligned} \frac{q}{p} e^{t/\sqrt{nq}} &= \frac{q}{p} \left[1 + \frac{t}{\sqrt{nq}} + \frac{t^2}{(2!)nq} + \frac{t^3}{(3!)(nq)^{3/2}} e^{\zeta(n)} \right] \\ &= \frac{q}{p} + \frac{qt}{p\sqrt{nq}} + \frac{t^2}{p(2n)} + \frac{qt^3}{p(3!)(nq)^{3/2}} e^{\zeta(n)}, \end{aligned} \quad (3.19)$$

where the sequence $\zeta(n)$ here lies between 0 and (t/\sqrt{nq}) and $\zeta(n) \rightarrow 0$ as $n \rightarrow \infty$.

Substituting equations (3.18) and (3.19) in the last expression for $M_{Z_n}(t)$ in (3.17), after some algebraic simplifications, we have

$$\begin{aligned} M_{Z_n}(t) &= \left[\left(\frac{1}{p} - \frac{q}{p} \right) - \frac{t^2}{2n} \left(\frac{1}{p} - \frac{q}{p} \right) \right. \\ &\quad \left. + \frac{q}{p} \frac{t^3}{(3!)(nq)^{3/2}} (q^2 e^{\xi(n)} - e^{\zeta(n)}) \right]^{-n} \\ &= \left[\frac{1-q}{p} - \frac{t^2}{2n} \frac{1-q}{p} + \frac{t^3}{(n)p(3!)\sqrt{nq}} (q^2 e^{\xi(n)} - e^{\zeta(n)}) \right]^{-n} \\ &= \left[1 - \frac{t^2}{2n} + \frac{1}{n} \frac{t^3}{(3!)\sqrt{nq}} (q^2 e^{\xi(n)} - e^{\zeta(n)}) \right]^{-n}. \end{aligned} \quad (3.20)$$

The preceding equation (3.20) may be written as

$$M_{Z_n}(t) = \left(1 - \frac{t^2}{2n} + \frac{\psi(n)}{n} \right)^{-n}, \text{ where } \psi(n) = (t^3/(6)\sqrt{nq}) (q^2 e^{\xi(n)} - e^{\zeta(n)}).$$

Since both $\xi(n), \zeta(n) \rightarrow 0$ as $n \rightarrow \infty$, it implies that $\lim_{n \rightarrow \infty} \psi(n) = 0$ for every fixed t . Hence, by Lemma 2.2 we have

$$\lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{t^2/2} = M_Z(t), \text{ where } Z \sim N(0,1), \quad (3.21)$$

for all real t . Hence, by Theorems 2.1, we can conclude that the r.v. $Z_n = (X_n - nq/p)/(\sqrt{nq}/p)$ has the limiting standard normal distribution, as $n \rightarrow \infty$. Alternatively, we can state that the NB r.v. X_n has approximately a normal distribution with mean $\mu_n = [nq/p]$ and variance $\sigma_n^2 = [nq/p^2]$, for large n .

3.4. The CLT Techniques

Consider an infinite sequence of Bernoulli trials with success probability p , $0 < p < 1$, and failure probability $q = 1 - p$. Define the r.v. Y to be the number of failures before the first success. Then Y is said to have the Geometric distribution with parameter p . The pmf of Y can be stated as

$$f_Y(y) = \begin{cases} pq^y & \text{for } y = 0, 1, 2, \dots \\ 0 & \text{otherwise.} \end{cases} \quad (3.4.1)$$

Let Y_1, Y_2, \dots, Y_n be a sequence of independent and identically distributed r.v.'s from the above Geometric distribution. Define $X_n = Y_1 + Y_2 + \dots + Y_n$. Then X_n denotes total number of failures before the n th success. The sum $X_n = \sum_{k=1}^n Y_k$ can be easily seen to have the negative-binomial distribution with parameters n and p . We

can verify it by the mgf technique: The mgf of each Y_k is

$$M_{Y_k}(t) = \sum_{y=0}^{\infty} e^{ty} p q^y = p \sum_{y=0}^{\infty} (e^t q)^y = p/(1 - qe^t),$$

$k = 1, 2, \dots, n$, so that the mgf of $X_n = \sum_{k=1}^n Y_k$ is obtained as

$$M_{X_n}(t) = E(e^{tX_n}) = [M_Y(t)]^n = \left[p/(1 - qe^t) \right]^n.$$

This is exactly the mgf of the negative-binomial $NB(n, p)$ r.v. derived directly in (3.16). Hence, X_n follows a negative-binomial with parameters n and p . Thus, the negative binomial r.v. X_n can be viewed as the sum of n i.i.d. Geometric r.v.'s, Y_i , $i = 1, 2, \dots, n$, with mean $[q/p]$ and variance $[q/p^2]$. Clearly therefore, the mean of X_n is $E(X_n) = \mu = nq/p$ and variance $\sigma^2 = nq/p^2$. Accordingly, by the CLT Theorem 2.2, we can conclude that $Z_n = [(X_n - n\mu)/(\sqrt{n}\sigma)] = [(X_n - (nq)/p)/(\sqrt{nq}/p)] \xrightarrow{d} N(0,1)$, as $n \rightarrow \infty$ or equivalently, that the r.v. X_n follows approximately the normal distribution with mean $\mu = [(nq)/p]$ and variance $\sigma^2 = [(nq)/p^2]$ for large n .

4. Concluding Remarks

This article offers four different methods of proof for the convergence of negative-binomial $NB(n, p)$ distribution to a limiting normal distribution. The first one is due to DeMoivre which uses Stirling's Approximation formula. The second one is the Ratio Method which is based on and utilizes the ratio of two successive probability terms of the pmf and, thereby, circumvents the use of Stirling's approximation formula. The method requires only a basic knowledge of Calculus, viz., limits, derivatives, Taylor series expansions and simple integration, along with basic probability concepts. This method also makes a broad assumption which is not always verifiable. Accordingly, the method - strictly speaking - may be regarded only as a technique rather than a complete proof. The other two methods are, respectively, the MGF method based on Laplace transforms and lastly that of the Central Limit Theorems based on Fourier transforms and the complex analysis, [3].

This article may serve as a useful teaching reference paper. The contents of this article should be of pedagogical interest to teachers, and can be discussed in senior level probability courses. It should be of reading interest for undergraduate students in probability or mathematical statistics. The teachers may gainfully assign these different methods to students as class projects.

ACKNOWLEDGEMENTS

The authors would like to thank an anonymous referee for careful reading and constructive suggestions on the earlier version of the paper.

REFERENCES

- [1] Arbous, A. G. and Kerrich, J. E. (1951). Accident statistics and the concept of accident proneness. *Biometrics*, 7, 340-432.
- [2] Bagui, S. C., Bagui, S. S., and Hemasinha, R. (1913). Nonrigorous proofs of Stirling's formula, *Mathematics and Computer Education*, 47, 115-125.
- [3] Bagui, S. C. and Mehra, K. L. (1917). Convergence of Binomial to Normal: Multiple proofs, *International Mathematical Forum*, 12, 399-422.
- [4] Bagui, S. C. and Mehra, K. L. (1916). Convergence of Binomial, Poisson, Negative Binomial and Gamma to normal distribution: Moment generating functions technique, *American Journal of Mathematics and Statistics*, 6, 115-121.
- [5] Bagui, S. C. and Mehra, K. L. (1917). Convergence of known distributions to limiting normal or non-normal distributions: An Elementary ratio technique, *The American Statistician*, 71(3), 265-271.
- [6] Bartko, J. J. (1961). The Negative binomial distribution: A review of properties and applications, *Virginia Journal of Science*, 12, 18-37.
- [7] Bartko, J. J. (1965). Approximating the Negative Binomial, *Technometrics*, 8(2), 345-350. Feller, W., (1957). *An introduction to probability theory and its applications*, 2nd edition, Wiley, New York, p. 263.
- [8] Furry, W. H. (1937). On fluctuation phenomena in the passage of high energy electrons through lead, *Physical Review*, 52, 569-581.
- [9] Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, *Journal of the Royal Statistical Society, A*, 83, 255-279.
- [10] Gurland, John (1959). Some applications of the negative binomial and other contagious distributions, *American Journal of Public Health*, 49(10), 1388-1389.
- [11] Johnson, N. L., Kotz, S., and Kemp, A. W. (1993). *Univariate Discrete Distributions*, Wiley, New York.
- [12] Kemp, A. W. (1970). "Accident proneness" and discrete distribution theory, *Random counts in Scientific work*, 2: *Random counts in Biomedical and Social Sciences*, G. P. Patil (editor), 41-65. University Park: Pennsylvania State University Press.
- [13] Kendall, D. G. (1949). Stochastic processes and population growth, *Journal of the Royal Statistical Society, B*, 11, 230-282.

- [14] Luders, R. (1934). Die statistik der seltenen ereignisse, *Biometrika*, 26, 108-128. Montmort, P. R. (1741). *Essai d'analyse sur les jeux de hazards*. Pascal, B. (1679). *Varia opera Mathematica D. Petri de Fermat*. Tolossae.
- [15] Quenouille, M. H. (1949). A relation between the logarithmic, Poisson, and negative binomial series, *Biometrics*, 5, 162-164.
- [16] Student (1919). On the error of counting with a haemocytometer, *Biometrika*, 5, 351-360.
- [17] Todhunter, I., (1865). *A history of the mathematical theory of probability from the time of Pascal to that of Laplace*, MacMillan and Co, p.97.