

**DEVELOPMENT AND APPLICATION OF ZERO-TRUNCATED
PROBABILITY MASS FUNCTION OF THE COM-BINOMIAL
DISTRIBUTION.**

By

ADEROJU, SAMUEL ADEWALE

Matriculation Number: 07/55EG101

B.Sc.; M.Sc. (Ilorin)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS IN PARTIAL
FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF
DOCTOR OF PHILOSOPHY (Ph.D.) IN STATISTICS, FACULTY OF PHYSICAL
SCIENCES, UNIVERSITY OF ILORIN,

ILORIN, NIGERIA

MAY, 2018

CERTIFICATION

This is to certify that this thesis is an original work carried out by Samuel Adewale ADEROJU with Matriculation Number 07/55EG101 in the Department of Statistics, Faculty of Physical Science, University of Ilorin. The Thesis has been read and approved as meeting the requirements for the Award of Doctor of Philosophy in Statistics of the University of Ilorin, Ilorin, Nigeria.

.....
Prof. E.T. Jolayemi
(Supervisor)

.....
Date

.....
Dr. G. M. Oyeyemi
(Departmental PG Coordinator)

.....
Date

.....
(Internal Examiner from related Department)

.....
Date

.....
Prof. B. L. Adeleke
(Head of Department)

.....
Date

.....
(External Examiner)

.....
Date

DEDICATION

This thesis is dedicated to all my family members who had in one way or the other contributed to the success of my career.

DECLARATION

I, ADEROJU, Samuel Adewale hereby declare that this thesis entitled the development and application of zero-truncated probability mass function of the com-binomial distribution is a record of my research work. It has neither been presented nor accepted in any previous application for a higher degree. All sources of information have been specifically acknowledged.

ADEROJU, S. A

Date

ACKNOWLEDGEMENT

I am most grateful to God Almighty for His mercy and goodness in my life. He is my all in all, my comforter and my defense. He has brought me from grass to grace! He is the qualifier of the unqualified.

My profound gratitude goes to my loving and passionate supervisor, Prof. E.T. Jolayemi for his fatherly advice, correction, guidance and dealing with me right from my undergraduate days.

I similarly, appreciate the efforts of all my able and loving lecturers especially Prof. R.A. Ipinyomi, Prof. B.A. Oyejola and others for their assistance, contributions and encouragement.

I would also like to appreciate Prof. Bayo Lawal for his sincere support and encouragement in the course of the study. I appreciate Mr. I. Adeniyi and Dr. K Issah also for their technical support during the study.

I appreciate the sponsorship extended to me by my Institution (through the Staff Development Award) by the authority of the Kwara State University, Malete, which has enabled me to bring this research to completion.

I also wish to acknowledge all members of staff of the Department of Statistics and Mathematical Sciences, Kwara State University, Malete namely: Dr. A. Abdulraheem, Dr. T. Aliu, Dr. K. A Dauda, Mr A. O Ibrahim, Mr. K. O. Olorede, Mr. J. B Olaifa, Mrs. S. F Olaniran and others who contributed in one way or the other in the course of my research. I equally appreciate Mr. S. Adepoju (Geology). You are all warmly appreciated.

I appreciate my darling wife and lovely children (the Olorires) for their unmeasurable support. Moreover, my precious and profound gratitude goes to my parents; Mr. &Mrs. Moses Aderoju as well as my Daddies and Mummies; Pastor and Mrs A.O Aderoju, Pastor & Mrs G. Alamu,

Mr&Mrs J.O Aderoju, Mr&Mrs Femi Aderoju and Sis. (Nurse) Olubukola for their moral, spiritual, financial and various supports. I cannot but appreciate my lovely niece and nephews for their moral and various supports.

I also wish to express my profound gratitude and appreciation to Engr.(Dr.) &Mrs. I. Megbonwon, Engr.&Mrs S. Megbowon, Engr.&Mrs W. Megbowon, Dr. & Mrs Ebenezer Megbowon and Dr. Funke Yahya for their moral, financial and spiritual supports. I appreciate Mrs. O. Shotade and my caring friends; Mr. Michael Falade, Mr. Dayo (the doctor) and others.

My sincere appreciation and thanks go to Pastor Oladimeji (the State coordinator, DLCF, Kwara State.), Pastor & Mrs. O. Orowole, Pastor & Mrs Taiwo Ige, Pastor (Dr.) &Mrs. Tope Adebayo, Pastor & Mrs Ajayi (Honeycomb), Pastor&Mrs Opataye, Pastor Gbade (Ibadan), Pastor Ibrahim (Ibadan) and all my Delightful, Lovely, Caring and Faithful (DLCF) brethren at sango, Ajegunle and other locations; Engr. A.O. Olawuwo and others.

TABLE OF CONTENTS

Title page	i
Certification	ii
Dedication	iii
Declaration	iv
Acknowledgements	v
Table of Contents	vii
Abstract	xii
1 GENERAL INTRODUCTION	1
1.0 Introduction	1
1.1 Modeling count data	1
1.2 Generalized Linear Models	2
1.2.1 Modeling Zero-truncated count data	4
1.3 Statement of Problem	5
1.4 Justification of the Study	7
1.5 Aim and Objectives of the Study	7
1.6 Limitation of the Study	8
1.7 Summary by chapter	8
2 LITERATURE REVIEW	9
2.0 Introduction	9

2.1	Derivation of the Poisson from the Binomial distribution	10
2.2	Issues in modeling count data	11
2.3	Overview of some distributions	17
2.3.1	Binomial distribution.	17
2.3.2	Beta-binomial distribution.	19
2.3.3	Generalized Beta-Binomial distribution.	20
2.3.4	Kumaraswamy-Binomial distribution.	22
2.3.5	Multiplicative Binomial distribution.	24
2.3.6	Com-Poisson-Binomial distribution.	25
2.4	Application of the Binomial based models	26
2.5	Overview of some zero-truncated models	29
2.5.1	Zero-truncated binomial distribution	30
2.5.2	Zero-truncated Poisson distribution	31
2.5.3	Zero-truncated Multiplicative Binomial distribution	32
2.6	Summary and Research Gaps	33
3	RESEARCH METHODOLOGY	34
3.1	Zero-truncation Com-Binomial distribution.	35
3.2	Properties of Zero-truncation Com-Binomial distribution	37
3.3	Methods of Estimating the Parameter	47
3.4	Overall Goodness-of-fit statistics	48

4	APPLICATION OF ZERO-TRUNCATED COM-BINOMIAL DISTRIBUTION TO REAL LIFE DATASETS	50
4.1	Example 4.1 [immunogold assay data]	50
4.2	Example 4.2 [counts of flower heads]	53
4.3	Example 4.3: [Mortality data]	56
4.4	Example 4.4: [Mortality data]	59
4.5	Example 4.5: [Mortality data]	62
4.6	Example 4.4: [Mortality data_NHS2008]	65
4.7	General Remark	68
4.8	Summary of the Analyses	68
5	CONCLUSION AND RECOMMENDATION.	70
5.1	Conclusion	70
5.2	Recommendation	71
5.3	Contribution to Knowledge	71
5.4	Suggestion for further studies	72
	Reference	73
	Appendices	81
	Appendix A	81
	Appendix B	85

List of tables

Table 2.1 Result of the analysis of dice's data.	27
Table 2.2 The expected and observed number of counts of flower heads with fly eggs.	28
Table 4.1 The expected and observed number of counts from immunogold assay data.	50
Table 4.2 The expected and observed number of counts of flower heads with fly eggs.	53
Table 4.3 The expected and observed number of mothers of the rural area having at least one live birth and one neonatal death.	56
Table 4.4 The expected and observed number of literate mothers with at least one live birth by the number of infant deaths.	59
Table 4.5 The expected and observed number of mothers having at least one neonatal death.	62
Table 4.6 The expected and observed number of mothers of the complete fertility that experienced at least one child death.	65
Table 4.7 Summary of comparative of goodness of fit of ZTCB and other selected existing distributions.	68

List of figures

Figure 3.1 The histogram of pmf of ZTCB distribution with $\pi = p = 0.5$ and at various values of v .	39
3.2 The pmf of ZTCB distribution with $\pi = 0.25$ and $n = 7$ at various values of v .	40
3.3 The pmf of ZTCB distribution with $\pi = 0.5$ and $n = 7$ at various values of v .	41
3.4 The pmf of ZTCB distribution with $\pi = p = 0.5$ and $n = 7$ at various values of v	42
3.5 The pmf of ZTCB distribution with $\pi = p = 0.25$ and $n = 15$ at various values of v	43

3.6	<i>The pmf of ZTCB distribution with $\pi = p = 0.5$ and $n = 15$ at various values of v</i>	44
3.7	<i>The pmf of ZTCB distribution with $\pi = p = 0.75$ and $n = 15$ at various values of v</i>	45
4.1a	The plot of observed and expected frequency obtained by each distribution for the immunogold assay data.	51
4.1b	The joint plot of observed and expected frequency obtained by all the distributions.	52
4.2a	The plot of observed and expected frequency obtained by each distribution.	54
4.2b	The joint plot of observed and expected frequency obtained by each distribution.	55
4.3a	The plot of observed and expected frequency obtained by each distribution.	57
4.3b	The plot of observed and expected frequency obtained by each distribution.	58
4.4a	The plot of observed and expected frequency obtained by each distribution.	60
4.4b	The joint plot of observed and expected frequency obtained by each distribution.	61
4.5a	The plot of observed and expected frequency obtained by each distribution.	63
4.5b	The plot of observed and expected frequency obtained by each distribution.	64
4.6a	The plot of observed and expected frequency obtained by each distribution.	66
4.6b	The plot of observed and expected frequency obtained by each distribution.	67

ABSTRACT

The major problem in modeling count data is over-dispersion, which occurs as a result of long tail, too many occurrences of zeros or none occurrence than would normally be expected by Poisson and binomial distributions. The zero-truncated distributions arise in those situations where there is no zero by nature of the data (structurally). The existing zero-truncated distributions are not flexible enough to capture peculiar characteristics such as long tail as well as over-dispersion at the same time. The aim of this research was to develop and implement a more generalized zero-truncated distribution that come from the mixture of distributions for non-zero count data to include most of the characteristics which existing ones do not have. The specific objectives were to: (i) examine the issues zero-truncated count data; (ii) investigate the effectiveness of selected binomial-based distributions for count data with various dispersions; (iii) develop zero-truncated Com-binomial distribution; (iv) identify some properties of the newly developed distribution and (v) demonstrate the use and assess the performance of the distribution against some existing ones using real life datasets.

Following mixture distribution methodology, the Zero-Truncated Com-Binomial (ZTCB) was derived from the mixture of Conway-Maxwell-Poisson generalization distribution to the Binomial distribution. The first two moments via probability generating function were also derived. The Maximum Likelihood Estimation of the parameters were also obtained by direct maximization of the log-likelihood function using “*optim*” routine in R software.

The findings of the study were:

- (i) the challenge in modeling count data, with a long tail and the problem arising from zeros (truncated or excess) were resolved;

- (ii) the Com-Binomial (CB) distribution outperformed other mixture distributions in handling either under-dispersion or over-dispersion situation;
- (iii) the probability mass function (pmf) of ZTCB distribution was derived and the first two moments were derived via probability generating function;
- (iv) the ZTCB distribution peaks around the mean to have two tails to resemble the normal distribution or *peaks around 1 or n* to resemble exponential distribution for carefully chosen parameter v . It is the generalization of the Zero-Truncated Com-Poisson, Zero-Truncated Poisson and Zero-Truncated Binomial distributions; and
- (v) the ZTCB distribution is more flexible to handle all levels of dispersion than Zero-Truncated Multiplicative Binomial distribution.

The study concluded that the ZTCB distribution which is characterized by two parameters is more flexible than other distributions. . It is recommended that when modeling structurally zero-truncated data, ZTCB distribution be used to obtain a robust result. This study therefore provides useful alternative to the existing distributions.

CHAPTER ONE

GENERAL INTRODUCTION

1. Introduction

In modeling count data, there are so many distributions possible for the response variable. The choice of the appropriate one is a challenge. However, the choice should be made based on the available knowledge of the variable. If we are to model a dichotomous variable; each outcome is Bernoulli. The binomial distribution should be used when the number of repetitions has been fixed and the variable is sum of a kind. If the data are counts as above but the sample size is large, then the Poisson distribution is an option. Moreover, the above two distributions are the based distribution when it comes to modeling count data. Nevertheless, there are so many mixed distributions derived from them as improvements when modeling count data of different kind.

1.1 Modeling Count Data

When discussing modeling of count data, it is imperative to clarify what is count and count data as well as count variable. The word “count” is typically used as verb meaning to enumerate units, items, or events, (Hilbe, 2014). As examples, it is possible to count the number of children to a woman within a particular period of time, or the number of patients that died at a particular hospital within 24 hours of having cardiovascular arrest.

“Count Data” refers to observations made about events, or items that are enumerated. Count data, in statistics, refer to observations that have only nonnegative integer values ranging from zero to some greater undetermined value. Theoretically, count data can range from zero to infinity, but they are always limited to some less distinct value – generally the maximum value of the count data being modeled. A “count variable” is a specific list or array of count data, (Hilbe, 2014).

1.2 Generalized Linear Models

Generalized Linear Models (GLMs) is a covering algorithm allowing for the estimation of a number of otherwise distinct statistical models within a single framework/family. First developed by Nelder and Wedderburn in 1972, the algorithm and overall GLM methodology has proved to be of substantial value to statisticians in terms of the scope of models under its domain as well as the number of accompanying model statistics facilitating an analysis of fit. In the early days of statistical computing (1972 to 1990), the GLM estimation algorithm also provided a substantial savings of computing memory compared to what was required using standard maximum likelihood techniques. Prior to Nelder and Wedderburn's studies, GLMs were typically estimated using a Newton-Raphson type full maximum likelihood method, with the exception of the Gaussian/Normal model. Commonly known as linear regression, which is usually estimated using a least squares algorithm. GLMs, is a generalization of ordinary least

squares (OLS) regression, employing a weighted least squares algorithm that iteratively solves for parameter estimates and standard errors.

Nelder (in 1974) coordinated a project to develop a specialized statistical application called Generalized Linear Interactive Modeling (GLIM). Sponsored by the Royal Statistical Society and Rothamsted Experimental Station, GLIM provided the means for statisticians to easily estimate GLMs, as well as other more complicated models which could be constructed using the GLM framework. GLIM soon became one of the most used statistical applications worldwide, and was the first major statistical application to fully exploit the PC environment in 1981. However, it was discontinued in 1994. Presently, nearly all leading general purpose statistical packages offer GLM modeling capabilities; e.g. SAS, R, Stata, S-Plus, Genstat, and SPSS. If the errors do not follow a multivariate normal distribution, GLMs may be used to relax assumptions about response variable and error term. Here, the response variable is a count and the variability in counts is related to the magnitude of the count. The value of the response variables are also constrained to be non-negative. The assumption of discrete distribution error structure will be most appropriate for these types of data (Lawal and Famoye 2013).

The Generalized Linear Models (GLMs) are a large class of statistical models for relating responses to linear combinations of predictor variables, including many commonly encountered types of dependent variables and error structures as special cases. In addition to regression models for continuous dependent variables, models for rates and proportions, binary, ordinal and counts variables can be handled as Generalized Linear

Models. The approach is attractive because it provides a general theoretical framework for many commonly encountered statistical models. It as well simplifies the implementation of these different models in statistical software, since essentially the same algorithm can be used for estimation, inference and assessing model adequacy for all Generalized Linear Models. There are considerable ranges of application of generalized linear models.

1.2.1 Modeling Zero-truncated count data.

In a situation we are to model count data, which structurally exclude zero counts. An example of that is the hospital length-of-stay data. When a patient first enter the hospital, the count starts at registration, with the length-of-stay given as one. Hence, there can be no zero days; except we are talking of patients who do not enter the hospital let alone register. Binomial (with random variable say $y = 0, 1, 2, \dots, n$) and Poisson (with random variable say $y = 0, 1, 2, \dots$) distributions include zeros. However, when data exclude zero counts structurally, the underlying probability distribution must be adjusted to exclude this outcome if the data are to be modeled properly. The Poisson and Binomial probability functions and their respective log-likelihood functions must be adjusted to rule-out zeros, and at the same time provide for all probabilities in the distribution to add to one. Models based on this kind of development are usually referred to as zero-truncated count models.

Referring to the Poisson distribution, the probability of a zero count, based on the log-likelihood function is $\exp(-\lambda)$. This value needs to be subtracted from 1 with the remaining probabilities rescaled (Hilbe, 2011). Then the log-likelihood function becomes:

$$\mathcal{L}(\lambda; y|y > 0) = \sum_{i=1}^n \{y_i \ln(\lambda) - \lambda - \ln(y_i!) - \ln[1 - \exp(-\lambda)]\} \quad \mathbf{1.0}$$

The scenario for other zero-truncated models is the same.

1.3 Statement of Problem.

Recall that over-dispersion can occur as a result of too many occurrences of zeros or none occurrence than would normally be expected by Poisson model. This may occur as; According to Lawal and Famoye (2013)

- i. **Structural Zeros**: this is the presence of zeros naturally as expected. For example, in a survey to determine the number of cigarettes consumed by respondents per day, there would be individuals in the sample who do not smoke at all. Such people will have the number recorded as zeros. Obviously, we naturally do not expect them to have a count.
- ii. **Sample zeros**: for instance, if in the above survey, an individual does smoke but he/she did not smoke a single stick of cigarette during the survey period, then such individuals would have a count of zeros called sample zeros.

The zero-truncated models arise in those situations where zero cannot be realized as a count.

In modeling the relationship between a response variable and a set of explanatory variables, regression is usually on the focus. When the response variable of interest is a count independent, the Ordinary Least Squares (OLS) regression is not applicable. However, the Poisson regression is the basis model for the count data sets. Unfortunately, the model has a strict assumption that most count data violate; the assumption of equi-dispersion; mean and variance identical! In most cases, count data exhibit over-dispersion which may yield wrong inference on the significance of the explanatory variables. Moreover, if the data exhibit zero-truncation, there is need to account for the structurally missing zero.

Several attempts directed at circumventing this common problem in count data resulted to the development of many model techniques to account for the dispersion parameter. Among the earlier methods developed for this, resulted into distributions such as Zero-truncated Poisson, Zero-truncated Binomial, Zero-truncated Multiplicative Binomial among many others.

Expectedly, some of the existing methods perform well to account for the missing zeros, but unfortunately they fail in some areas.

The prime goal of this work therefore, is to develop a new flexible approach that would efficiently account for extreme over-dispersion and yet preserve zero truncation in count data.

Our method shall be evaluated on some count data sets while its general performance relative to those provided by some of the few selected existing methods shall be examined.

1.4 Justification of the Study

The main contributions of this research work include, but not limited to, the attempt to develop a flexible distribution that performs well in the presence of a very heavy tail. This, it is hoped, shall be helpful in modeling zero-truncated count data as applicable in various fields.

1.5 Aim and Objectives

Main Aim:

To develop and implement a more generalized zero-truncated distribution that come from the mixture of distributions for non-zero count data that would account for a heavy tail.

Specific Objectives were to:

- i. examine the issues zero-truncated count data;
- ii. investigate the effectiveness of selected binomial-based distributions for count data with various dispersions;
- iii. develop zero-truncated Com-binomial distribution;

- iv. identify some properties of the newly developed distribution and
- v. demonstrate the use and assess the performance of the distribution against some existing ones using real life datasets.

1.6 Limitation of the Study

There are so many limitations to the study, right from the beginning. The major one is the problem of real life data. Hospital count data are difficult to obtain. Inevitably, survey such as National Health Service (NHS) had to be used.

1.7 Summary by Chapters

The outline of this thesis is as follows: Chapter two provides an overview of literature on count data and statistical models for count data. Methodology and the performance of the Com-Binomial distribution in comparison with other distributions are in chapter three. In chapter four, the Zero-truncated Com-binomial (ZTCB) distribution is derived and some mathematical properties are provided and compared with three other common ones in analyzing some real life data. Chapter five summarizes the main findings of this research. It also documents possible future work directions.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

Poisson and binomial are the base count response regression models. Other count models deal with data that violate the assumptions of Poisson and Binomial distributions, (Hilbe, 2011). Since the distributions play such a pivotal role in count response modeling, we would first discuss them and their relationship, derivation and structure.

If n independent trials are made and in each trial there is probability p that the outcome E will occur, then the number of trials in which E occurs can be represented by a random variable Y having the binomial distribution with parameters n, p (or π , as we used them interchangeably in this study). This scenario occurs when a sample of fixed size n is taken from an infinite population where each element in the population has an equal and independent probability p of possession of a specified attribute.

The binomial distribution, as one of the oldest to have been the subject of study, was derived by James Bernoulli (in his treatise *Ars Conjectandi*, published in 1713), Johnson et al (2005). The early history of the distribution is discussed, among others, by Boyer (1950), Stigler (1986), Edwards (1987), and Hald (1990).

Similarly, Poisson (1837) published a derivative of a distribution that bears his name. He approached the distribution by considering the limit of a sequence of binomial

distribution in which n tends to infinity and p tends to zero while np remains finite equal to λ .

2.1 Derivation of the Poisson from the Binomial Distribution.

There are several ways mathematicians have derived the Poisson from the binomial distribution. Hilbe (2011) however, presented a standard method;

Consider a binomial pmf

$$P(Y = y) = f(y|p, n) = \binom{n}{y} p^y (1 - p)^{n-y}, \quad 0 \leq p \leq 1 \text{ and } y = 0, 1, \dots, n \quad (2.1)$$

The $\binom{n}{y}$ function may be expanded and incorporated into the equation as;

$$f(y|p, n) = \frac{n!}{y! (n - y)!} p^y (1 - p)^{n-y} \quad (2.2)$$

The binomial mean, $\lambda = np \rightarrow p = \lambda/n$

So that

$$f(y|p, n) = \frac{n!}{y! (n - y)!} \left(\frac{\lambda}{n}\right)^y \left(1 - \frac{\lambda}{n}\right)^{n-y} \quad (2.3)$$

We can reformulate the binomial distribution as, letting n become very large and p small;

$$\lim_{n \rightarrow \infty} f(y|p, n) = \frac{n(n-1) \dots (n-y+1)}{n^y} \frac{\lambda^y}{y!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y}$$

$$= \frac{n(n-1) \dots (n-y+1) \lambda^y}{n^y y!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-y} \quad (2.4)$$

Recall from basic Mathematics, $e_{\lim x \rightarrow \infty} = \left(1 + \frac{1}{x}\right)^x$, and $e_{\lim x \rightarrow \infty}^k = \left(1 + \frac{k}{x}\right)^x$. With the left term cancelling to equal 1, $\left(1 - \frac{\lambda}{n}\right)^n$ being the form for $e^{-\lambda}$, and with n being very large in the final term, resulting in $e_{\lim x \rightarrow \infty} \left(1 - \left(\frac{\lambda}{n} \rightarrow 0\right)\right) = 1$, equation 2.04 becomes; (Hilbe, 2011)

$$\lim_{n \rightarrow \infty} f(y|p, n) = (1) \left(\frac{\lambda^y}{y!}\right) (e^{-\lambda}) (1) \quad (2.5)$$

Hence,

$$f(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, y = 0, 1, 2, \dots \quad (2.6)$$

Equation (2.6) is the standard form of the Poisson probability distribution.

The Poisson distribution gives the probability of y successes in the series, given the series is large and p is small.

2.2 Issues in Modeling Count Data

In a situation where over-dispersion is present the standard distributions such as the binomial or the Poisson is not absolutely appropriate. Failure to take over-dispersion into

consideration and account for it, the result will be less precise than the Binomial or Poisson based model.

One way of coping with this problem is to use mixture models such as beta-binomial and Com-binomial among others.

Let us review some of the works done on the distributions/models for the binary data sets and their behaviours:

William (1975) described a parametric method for analyzing binary response data after Nelder and Wedderburn (1972) developed Generalized Linear Models (GLMs), which covers algorithm allowing for the estimation of a number of distinct statistical models within a single framework.

Altham (1978) derived generalized extensions of binomial distribution such as multiplicative and additive generalized binomial distributions. The pmf of the multiplicative binomial distribution is a multiplication of its pmf by a factor. It makes the variance greater or less than the corresponding binomial variance depending on the values of the factors.

Kupper and Haseman (1978) proposed a type of correlated binomial model (for analysis of certain toxicological experiments) and is shown to have certain theoretical advantage over a beta-binomial model proposed by Williams (1975)

Pack (1986) discovered that the Beta-binomial (BB) is superior to alternative models such as the correlated binomial of Kupper and Haseman (1978).

Paul (1982), in analyzing the proportions of affected fetuses in teratological experiments, observed that the Beta-binomial model is superior to the multiplicative binomial model.

Kupper, et al (1986) studied Binomial and Beta-binomial as applicable to the impact of litter effect on dose –response modeling in teratology. They concluded that the failure to take over-dispersion into account can lead to under-estimation of variances associated with the parameter estimates, so the use of binomial model does not seem advisable. They advocate the use of Beta-binomial distribution.

Tripathi, et al (1994) discussed some alternative methods for estimating the parameters in the beta-binomial and truncated beta-binomial models.

Lindsey (1995) and Luceno & Ceballos (1995) proposed a generalized binomial distribution which was discussed in details by Diniz et al (2010). They applied a Bayesian approach is applied to the correlated binomial model, proposed by Luceño and Ceballo (1995). The data augmentation scheme is used in order to overcome the complexity of the mixture likelihood. MCMC methods, including Gibbs sampling and Metropolis within Gibbs, are applied to estimate the posterior marginal for the probability of success p and for the correlation coefficient ρ . The sensitivity of the posterior is studied taking into account several reference priors and it is shown that the posterior characteristics appear not to be influenced by these prior distributions.

Lindsey and Altham (1998) discussed the beta-binomial and other models for binomial over-dispersion, analyzing sex-ratio data.

Altham (2000) discussed beta-binomial modeling in relation to the data set given in the 1998 BMJ paper by Spiegelhalter and Marshall on success rate of 52 in vitro fertilization clinics in the UK. According to him the beta-binomial model shows a better fit over binomial model for over-dispersed binomial data.

Shmueli et al (2005) introduced Conway-Maxwell-Poisson-Binomial distribution by adding a dispersion parameter to the binomial distribution. Nevertheless, they did not study the mathematical properties of their proposed family of distributions.

Li, et al, (2011) studied Kumarasway-Binomial (KB) with some real life data sets. They observed that KB models perform better than Beta-Binomial model in some occasions but not often.

Wijekoon and Yapa (2013) Introduced a new Binomial mixture distribution called the McDonald Generalized Beta-Binomial distribution (McGGB). The results of the real data shows that the new McGGB mixture model provides a better fit and good improvement in the goodness of fit tests than BB and KB models. From the results of the simulation study, it is also evident that the McGGB model is superior to its nested models for some parameter combinations. Hence the proposed McGGB distribution has a great potential for handling Binomial over-dispersion.

Borges, et al (2014) studied the probability distribution's moments and how the dispersion parameter affect the asymptotical approximation of the Com-Binomial (Conway-Maxwell-Poisson-Binomial) distribution to the Com-Poisson distributions and discovered that both distributions (Beta-binomial and Com-Binomial yield significant

improvement over the binomial distribution) give almost identical fit to the data but Com-Binomial gives slightly better fit than the Beta-binomial.

Similarly, here are some of the works done on the modeling zero-truncated count data sets with different characteristics and the performance of each model.

Finney (1949) developed Zero-truncated Binomial distribution. He described its properties as well as parameter estimation by standard method based on score and information.

Sampford (1955) discussed the methods for estimation of the parameters of the Zero-truncated Negative Binomial distribution. The corresponding problem of estimation of the zero-truncated Poisson distribution has been discussed by David & Johns-on (1952). They suggest that the use of estimates of less than maximum efficiency is justifiable only if they are directly obtainable as explicit solutions of easily constructed equations. In discussing the truncated negative binomial, therefore, they do not consider estimates based on the first two sample moments, which do not provide explicit solutions, but confine their attention to a method using certain ratios of the first three moments.

Boswell and Patil (1970) gave a genesis of the *zero-truncated negative binomial distribution* as a distribution for the sizes of groups (a group size distribution). They also showed that it can be derived as a mixture of zero-truncated Poisson distributions

Gart (1974) studied the exact moments of the statistic for the variance homogeneity test for samples from the zero truncated Poisson distribution and gave a number of references to earlier work on the distribution.

Various methods of estimation for the beta-binomial and zero-truncated beta-binomial distributions are explored in Tripathi, et al (1994). They are compared with maximum likelihood on the basis of asymptotic relative efficiency. Examples of their use and recommendations are provided

Bayesian inference concerning the size of the zero class given data from the zero-truncated Poisson distribution was studied by Scolnik (1997).

Ghitany et al (2007) introduced the zero-truncated Poisson-Lindley distribution. In particular, the method of moments and maximum likelihood estimators of the distribution's parameter are compared in small and large samples. Application of the model to real data is given. The main limitation of this distribution is its inability to converge.

Hilbe (2011) discussed a situation which gives rise to overdispersion on both Poisson and negative binomial (NB2) models. This relates to when there is no possibility of a zero count in the data. He studied and compared zero-truncated Poisson and zero-truncated negative binomial models among others.

Elamir (2011) derived the zero-truncated multiplicative-binomial distribution. In particular, they obtained the maximum likelihood estimators of the parameters and their

asymptotic variance-covariance matrix. The simulation procedures for generating data from zero-truncated multiplicative binomial distribution using R software were given. Applications of the model to real data are illustrated. However, he did not compare its performance with other existing distributions.

Aderoju et al (2017) developed zero-truncated generalized Poisson-Lindley distribution. They studied the properties and applications of the distribution. They as well observed that the distribution performs better in many real life data than the zero-truncated Poisson-Lindley distribution and Zero-truncated Poisson distribution, though it has the same limitation of convergence.

2.3 Overview of Some Distributions

2.3.1 Binomial Distribution

Several response variables are binary, which represent the success and failure outcomes, usually denoted by 1 and 0 respectively. For Bernoulli trial specifies probabilities $\Pr(Y=1) = \pi$ and $\Pr(Y=0) = 1 - \pi$, for which $E(Y) = \pi$.

The probability mass function is

$$P(Y = y) = \pi^y (1 - \pi)^{1-y} \quad 0 \leq \pi \leq 1 \text{ and } y = 0 \text{ and } 1 \quad (2.7)$$

This is a special case of Binomial distribution when $n = 1$.

This is in the natural exponential family, identifying θ with π . $a(\pi) = (1 - \pi)$, $b(y) = 1$ and

$Q(p) = \log\left(\frac{\pi}{1-\pi}\right)$. The natural parameter $Q(\pi)$ is the log odds of the response 1, the logit of π . This is the canonical link.

In several cases, data are form of proportions, y/n where y is a non-negative count and is bounded above by the positive integer n . When n is assumed fixed and known, Y can be modeled as binomial $(n; \pi)$; i.e., view Y as the sum of n independent Bernoulli random variables, y_i ($i = 1, \dots, n$), with $\pi = E(y_i)$. If some correlation existed among the y_i , then Y would no longer be distributed as binomial.

Let $Y \sim B(n; \pi)$, then the pmf is;

$$\begin{aligned} P(Y = y) &= \binom{n}{y} \pi^y (1 - \pi)^{n-y} \quad 0 \leq \pi \leq 1 \text{ and } y = 0, 1, \dots, n \\ &= \binom{n}{y} (1 - \pi)^n \left(\frac{\pi}{1 - \pi}\right)^y \end{aligned} \quad (2.8)$$

Having correlated binary observations, the sum of those observations may not fit the binomial model.

In cases with strong evidence of extra-binomial variability, the mixed binomial models are preferable to the binomial model. Testing for departure from the binomial distribution has been discussed by Cochran (1954) and Tarone (1979) inter alia. Risko and Margolin (1996) provide a review and commentary on these methods. In contrast to the binomial distribution, however, formal testing of whether the beta-binomial distribution fits over-dispersed proportion data would be discussed as thoroughly in the study.

2.3.2 Beta-Binomial Distribution

A common question in the analysis of binary count data is how to deal with over-dispersion. One widely advocated sampling distribution for over-dispersed binary count data is the beta-binomial model.

If π is beta distributed and $(Y; n, \pi)$ follows binomial distribution with parameters n and π , then the marginal density of Y given n is beta-binomial (Williams, 1975; Haseman and Kupper, 1979).

When it is obvious that there is strong evidence of over-dispersion in a binary data, the Beta-binomial model is preferable to the binomial model.

Let Y_i be the binary response variable, that is, each observation has one of dichotomous outcomes, denoted by 1 and 0; such that Y_i are independent binomial random variables.

$$(Y_i; n_i, \pi_i) \sim B(n_i, \pi_i) \quad i = 1, 2, \dots, n$$

while $\pi_i \sim \text{Beta}(\alpha, \beta)$ and π_i are independent.

Therefore,

$$f(\pi; \alpha, \beta) = B(\alpha, \beta)^{-1} \pi^{\alpha-1} (1 - \pi)^{\beta-1} \quad 0 \leq \pi \leq 1, \quad (2.9)$$

where α and β are unknown positive constant and $B(\cdot, \cdot)$ is the beta function

$$\text{such that } B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

The distribution of Y_i is expressed as the Beta-binomial probability mass function,

$$P(Y = y; n, \alpha, \beta) = \binom{n}{y} \frac{B(\alpha + y, \beta + n - y)}{B(\alpha, \beta)} \quad y = 0, 1, 2, \dots \quad (2.10)$$

$\lambda = \frac{\alpha}{\alpha + \beta}$ and let $k = (\alpha + \beta)^{-1}$ as suggested by Williams (1975), then the mean and

variance of Y_i can be expressed by

$$E(Y_i; n_i, \lambda, k) = n_i \lambda \quad \text{and}$$

$$Var(Y_i; n_i, \lambda, k) = n_i \lambda (1 - \lambda) (1 + k)^{-1} (1 + n_i k)$$

$\lambda = \text{mean parameter of the marginal proportions.}$

$k = \text{dispersion parameter}$

$0 < \lambda < 1$ and when $k > 0$, then the data are said to be over-dispersed.

In some circumstances, the variance of Y_i may appear smaller than that of Binomial distribution, which implies under-dispersion (Engel and Brake, 1993). However, this is not common in real life data.

The maximum likelihood estimator (MLE) of λ and k can be determined numerically. As k approaches 0, the variance of Y_i monotonically decreases to $n_i \lambda (1 - \lambda)$ and $(Y_i; n_i, \lambda, k)$ converges to a Binomial random variable.

2.3.3 Generalized Beta-Binomial (GBB) Distribution

McDonald's Generalized Beta distribution is a Binomial mixture distribution called the McDonald Generalized Beta-Binomial distribution as introduced by *Manoj et al* (2013). The distribution is obtained through an integration approach (see *Manoj et al* 2013 for detail).

Suppose $Y_i \sim \text{Bin}(n, \pi)$ $i = 1, 2, \dots, n$

The unconditional probability mass function of Y can be obtained by evaluating the well-known integral;

$$f(y) = \int P_{y|\pi}(y) f_{\pi}(\pi|\theta) d\pi \quad \text{for } y = 0, 1, \dots, n \quad (2.11)$$

The Θ is the parameter space of the mixing distribution.

$$\therefore Y \sim GBB(n, \alpha, \beta, \gamma)$$

$$\text{iff } Y|\pi \sim \text{Bin}(n, \pi) \quad \text{and } \pi \sim \text{GB1}(\alpha, \beta, \gamma), \quad \text{where } \alpha, \beta, \gamma > 0.$$

The probability mass function of the GBB is

$$f(y; n, \alpha, \beta, \gamma) = \binom{n}{y} \frac{\gamma}{B(\alpha, \beta)} \sum_{i=0}^{\infty} (-1)^i \binom{\beta-1}{i} B(y + \alpha\gamma + \gamma i, n - y + 1) \quad (2.12)$$

where $y = 0, 1, \dots, n$ and $\alpha, \beta, \gamma > 0$

the rearranged pmf of the distribution was given as;

$$f(y; n, \alpha, \beta, \gamma) = \binom{n}{y} \frac{1}{B(\alpha, \beta)} \sum_{j=1}^{n-y} (-1)^j \binom{n-y}{j} B\left(\frac{y}{\gamma} + \alpha + \frac{j}{\gamma}, \beta\right), \quad (2.13)$$

where $y = 0, 1, \dots, n$ and $\alpha, \beta, \gamma > 0$

The r^{th} moment of the distribution is given as $E(Y^r) = n \frac{B(\alpha, \beta, \frac{r}{\gamma})}{B(\alpha, \frac{r}{\gamma})}$

Taken $\pi = \frac{B(\alpha, \beta, \frac{1}{\gamma})}{B(\alpha, \frac{1}{\gamma})}$ and

$$\rho = \frac{\left(\frac{B\left(\alpha + \beta, \frac{2}{\gamma}\right)}{B\left(\alpha, \frac{2}{\gamma}\right)} \right) - \left(\frac{B\left(\alpha + \beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)} \right)^2}{\left(\frac{B\left(\alpha + \beta, \frac{2}{\gamma}\right)}{B\left(\alpha, \frac{2}{\gamma}\right)} \right) - \left(\frac{B\left(\alpha + \beta, \frac{1}{\gamma}\right)}{B\left(\alpha, \frac{1}{\gamma}\right)} \right)^2}$$

$$E(Y) = n\pi \text{ and } Var(Y) = n\pi(1 - \pi)\{1 + (n-1)\rho\}$$

The ρ can be termed as over-dispersion parameter of the Generalized Beta-Binomial distribution.

2.3.4 Kumaraswamy-Binomial Distribution

When we consider a Binomial experiment, say $Y \sim Bin(n, \pi)$, while n is the number of trials and π is the probability of success, which is expected to be constant. Similarly, all trials are mutually independent and $0 \leq \pi \leq 1$.

However, π often varies from individual to individual as a parameter in some scenarios. Therefore, it is inevitable and profitable to regard π as a realization of a continuous random variable π with density function $f(\pi)$ when the over-dispersion should be taken into account. Beta distribution, say $B(\alpha, \beta)$, is very flexible in modeling various variable in the unit interval. So, it is often the first alternative to be considered for the distribution of π where, $\pi \sim B(\alpha, \beta)$; then Y is well-known to be Beta-Binomial random variable.

According to Kumaraswamy (1980), the kumaraswamy distribution is on the interval $[0,1]$. The distribution also assumes a wide variety of shapes and hence is as versatile as Beta distribution. It also has close-form probability density function and cumulative distribution function, which make it much simpler to be utilized.

Taking over-dispersion into account in Binomial model where all independent trials have a common probability of success π ; we get simple one parameter mixture model (Li et al 2011).

$$Y \sim \text{Bin}(n, \pi) \text{ and } \pi \sim K(1, \beta)$$

Therefore, the probability mass function of the KB1 is;

$$f(y) = \binom{n}{y} \beta B(y+1, n-y+\beta), \quad \text{where } \beta > 0 \text{ and } y = 0, 1, \dots, n \quad (2.14)$$

$B(\cdot)$ is beta function. The mean and variance were derived to be

$$E(Y) = n\beta B(2, \beta)$$

$$\text{Var}(Y) = n^2\beta[B(3, \beta) - \beta B^2(2, \beta)] + n\beta[B(2, \beta) - B(3, \beta)]$$

2.3.5 Multiplicative Binomial Distribution

One of the distribution that allows for over-dispersion and under-dispersion (related to standard binomial) is Multiplicative Binomial (MB). The distribution could be very useful to model count data for these situations (Elamir, 2011).

Let z_i be a random variable with two possible outcomes for a sample of n units and $Y = \sum_{i=1}^n z_i$ be the sample frequency of successes. Lovison (1998) proposed the pmf of the Multiplicative Binomial as

$$P(Y = y) = \frac{\binom{n}{y} \pi^y (1 - \pi)^{n-y} k^{y(n-y)}}{\sum_{t=0}^n \binom{n}{t} \pi^t (1 - \pi)^{n-t} k^{t(n-t)}} \quad y = 0, 1, \dots, n \quad (2.15)$$

where $0 < \pi < 1$ and $k > 0$ are the parameters when $k = 1$, the standard binomial is derived.

Elamir (2011) gave the moments of Multiplicative Binomial as follow;

$$\text{Let } n_{(i)} = n(n-1) \dots (n-i+1), \quad n_{(0)} = 1,$$

$$\mu_i = \pi^i \frac{k_{n-1}(\pi, k)}{k_n(\pi, k)}$$

and

$$k_{n-a}(\pi, k) = \sum_{y=0}^{n-a} \binom{n-a}{y} \pi^y (1 - \pi)^{(n-a-y)} k^{(n-a-y)(y+a)} \quad (2.16)$$

The first moment is:

$$\mu'_1 = n_{(1)}\theta_1 = E(Y) = \mu$$

and σ^2 was given as;

$$\sigma^2 = n_{(1)}\theta_1 + n_{(2)}\theta_2 - n_{(1)}^2\theta_1^2$$

2.3.6 Com-Binomial Distribution

The Binomial and Poisson are well known model for aggregative binary outcome and count data as the building blocks. The two distributions have been generalized in many ways to handle the problem of dispersion commonly in the analysis of discrete data.

The Com-Binomial (CB) distribution arises as the conditional distribution of a Com-Poisson variable (Conway and Maxwell, 1962; Borges et al, 2013), given a sum of two Com-Poisson variables with the same dispersion parameter. Obviously, the CB distribution belongs to the exponential and weighted Poisson family (Castillo & Pérez-Casany, 1998, 2005).

The CB's probability mass function, (Shmueli et al 2005) referred to as Compoisson-Binomial, is given by;

$$f(Y = x; n, \pi, v) = \frac{\binom{n}{y}^v \pi^x (1 - \pi)^{n-y}}{\sum_{i=0}^n \binom{n}{i}^v \pi^i (1 - \pi)^{n-i}} , \quad y = 0, 1, \dots, n \quad (2.17)$$

for $n \in Z^+$, $\pi(0,1)$ and $v \in \mathbb{R}$

where, Z^+ is a set of unknown non – negative integers

NOTE:

when $v = 1$, we have the Binomial distribution,

$v > 1$, we have under – dispersion,

$v < 1$, we have over – dispersion with respect to the Binomial distribution

2.4 Application of the Binomial Based Models

In this section we demonstrate how each model behaves in modeling over-dispersed binomial outcome data.

Example 2.1:

Four identical six-sided dice are tossed 200 times. At each tossing, a record is made of the number of dice whose score on the uppermost face is even. The result are as follows:

Table 2.1: Result of the analysis of dice's data

Incidents	Observed frequency	Expected					
		<i>B</i>	<i>BB</i>	<i>GBB</i>	<i>CB</i>	<i>MB</i>	<i>KB2</i>
0	10	8.20125	10.29765	10.29310	10.32995	10.29340	10.53878
1	41	40.09500	40.81597	40.79154	40.66937	40.79000	40.49452
2	70	73.50750	68.86585	68.87045	69.06938	68.86957	68.43975
3	57	59.89500	58.67584	58.70577	58.53341	58.71734	58.99391
4	22	18.30125	21.34468	21.33915	21.39789	21.32969	21.53304
Total	200	200	200	200	200	200	200
G ²		1.4028	0.0963	0.0982	0.0831	0.0996	0.1478
Df		3	2	1	2	2	2
P value		0.7578	0.9849	0.9847	0.9864	0.9846	0.929163
X ²		1.1797	0.1518	0.1530	0.1411	0.1540	0.1469423
AIC		580.462	581.1556	581.1576	581.1424	581.1588	581.2072
Loglik		- 289.231	- 288.5778	-288.5788	- 288.5712	-288.5794	-288.6036
Parameter estimates		$\pi = 0.55$	$\alpha = 15.72$ $\beta = 12.87$	$\alpha = 8.395$ $\beta = 13.505$ $\gamma = 1.627$	$\pi = 0.545$ $\nu = 0.857$	$\pi = 0.545$ $k = 0.938$	$\beta = 34.32$ $\gamma = 6.73$

The results Likelihood Ratio and other statistics are obtained as shown in the table above.

The mixed-binomial distributions give almost identical fits to the data. The Deviance values of the CB, BB, GBB, MB and KB2, distributions are 0.083, 0.0963, 0.0982, 0.0996 and 0.1478, respectively. These values show that the four distributions give fit the data well but the binomial distribution yields significant improvements over the mixed-binomial distributions as expected, since there is no evidence of significant over-dispersion. According to the principle of parsimony, the simple distribution that has the least number of parameter performs best. However, among the mixed-binomial distributions, CB distribution gives slightly better fit and good improvement in the

goodness of fit tests than others. Note that CB distribution has only two additional parameters, have the same flexibility in modeling binomial data.

Example 2.2

Considering a data set used by Borges et al (2014), which refer to the number of male children among the first 12 children of family size 13 in 6115 families taken from the hospital records in the nineteenth century Saxony (Sokal & Rohlf (1995), Lindsey (1995)). Thus, $n = 12$ is the family size and Y is the number of male children. The expected frequencies and estimates of the parameters under the binomial and other distributions are given in the table below. The log-likelihood, Deviance values and others are as well showed below:

Table 2.2: Results of the analysis number of male children

Incidents	Observed frequency	Expected Frequency					
		B	BB	GBB	CB	MB	KB2
0	3	0.932839	2.350686	2.356716	2.679739	2.34941	2.881745
1	24	12.08884	22.58587	22.6179	23.27579	22.58525	24.69309
2	104	71.80317	104.8487	104.9213	104.7038	104.8572	107.3972
3	286	258.4751	310.8895	310.9785	308.7439	310.8945	308.9284
4	670	628.055	655.6679	655.7376	653.5397	655.6278	646.645
5	1033	1085.211	1036.076	1036.142	1037.7	1036.012	1027.344
6	1343	1367.279	1257.8	1257.893	1262.358	1257.838	1259.977
7	1112	1265.63	1182.08	1182.147	1184.048	1182.27	1193.999
8	829	854.2466	853.6351	853.5775	850.877	853.8093	862.9805
9	478	410.0126	462.046	461.8729	458.6598	462.028	461.8689
10	181	132.8357	177.9797	177.8124	177.4812	177.8301	173.104
11	45	26.08246	43.82205	43.74143	45.01855	43.71029	40.66525
12	7	2.347273	5.217996	5.201568	5.91394	5.188828	4.516337
Total	6115	6115	6115	6115	6115	6115	6115
Pearson X^2		110.505	14.46901	14.5007	13.3597	14.5353	17.8997
df		11	10	9	10	10	10
P value		<0.0001	0.15	0.11	0.20	0.15	0.06

Loglik		-12534.17	-12492.87	-12492.9	-12492.35	-12492.9	-12494.5
G^2		97.0065	14.40504	14.4566	13.36482	14.46861	17.7446
AIC		25070.34	24989.74	24991.80	24988.70	24989.80	24993.08
Parameter estimates	-	$\pi = 0.519$	$\alpha = 34.05$ $\beta = 31.53$	$\alpha = 19.65$ $\beta = 32.99$ $\gamma = 1.51$	$\pi = 0.52$ $\nu = 0.84$	$\pi = 0.52$ $k = 0.97$	$\beta = 538.18$ $\gamma = 10.36$

The results of Deviances as well the P-value statistics are obtained as shown in the table above. The CB, BB, GBB and MB distributions give almost identical fits to the data. The G^2 values of the COMPB, BB, GBB and MB distributions are 13.36, 14.41, 14.46 and 14.47, respectively. These values show that the four models give significant improvements over the binomial distribution as expected. The CB distribution gives slightly better fit and good improvement in the goodness of fit tests than other distributions as in the case of the first data set. The KB2 does not perform well in the second example. This confirms Li, et al, (2011) finding that the KB models perform well in some occasions, not often.

2.5 Overview of Some Zero-Truncated Models.

The most common form of truncation is the omission of the zero class; this occurs if the observational apparatus becomes active only when at least one event occurs, (Johnson, 2005). This is usually called the zero-truncated or positive distribution.

2.5.1 Zero-Truncated Binomial (ZTB) Distribution

The ZTB was developed and studied by Finney (1949) and Thomas & Gart (1971) respectively and the pmf defined as:

$$P(Y = y) = \begin{cases} \frac{1}{1 - (1 - \pi)^{n-y}} \binom{n}{y} \pi^y (1 - \pi)^{n-y}, & \text{if } y = 1, 2, \dots, n \\ 0, & \text{e.w} \end{cases} \quad (2.18)$$

The moment generating function is obtained to be:

$$M_y(t) = \frac{(1 + \pi(e^t - 1))^n - (1 - \pi)^n}{1 - (1 - \pi)^n} \quad (2.19)$$

Moreover, the basic parameters such as the mean and variance respectively are:

$$M'_y(t = 0) = \frac{n\pi}{1 - (1 - \pi)^n}$$

and

$$Var(Y) = \frac{n\pi(1 - \pi - (1 - \pi + n\pi)(1 - \pi)^n)}{(1 - (1 - \pi)^n)^2}$$

The Likelihood function is:

$$L(\lambda; y_i | y_i > 0) = \prod_{i=1}^n \frac{1}{1 - (1 - \pi)^{n-y_i}} \binom{n}{y_i} \pi^{y_i} (1 - \pi)^{n-y_i} \quad (2.20)$$

And the log-likelihood function is:

$$\mathcal{L}(\lambda; y_i) = \sum_{i=1}^n \left\{ \ln \binom{n}{y_i} + y_i \ln(\pi) + (n - y_i) \ln(1 - \pi) - \ln[1 - (1 - \pi)^{n-y_i}] \right\} \quad (2.21)$$

For $y_i > 0$.

2.5.2 Zero-Truncated Poisson (ZTP) Distribution

The probability mass function (pmf) of a ZTP variable as derived and studied by David & Johnson (1952) and Johnson & Kotz (1969) respectively. The pmf was given as;

$$P(Y = y) = \begin{cases} \frac{\lambda^y}{y! (e^\lambda - 1)}, & \text{if } y = 1, 2, \dots \\ 0, & \text{e.w} \end{cases} \quad (2.22)$$

The difference with the standard Poisson distribution lies in the correction factor $(1 - e^{-\lambda})^{-1}$, which reflects the fact that a value of 0 cannot occur.

The moment generating function is obtained to be:

$$M_y(t) = \frac{e^{\lambda e^t} - 1}{e^\lambda - 1} \quad (2.23)$$

Moreover, the basic parameters such as the mean and variance respectively are:

$$M'_y(t = 0) = \frac{\lambda e^\lambda}{e^\lambda - 1} \quad \text{and} \quad Var(Y) = \frac{\lambda e^\lambda}{e^\lambda - 1} \left[1 - \frac{\lambda}{e^\lambda - 1} \right]$$

The likelihood function is:

$$L(\lambda; y_i | y_i > 0) = \prod_{i=1}^n \frac{\lambda^{y_i}}{y_i! (e^\lambda - 1)} \quad (2.24)$$

And the log-likelihood function is:

$$\mathcal{L}(\lambda; y_i | y_i > 0) = \sum_{i=1}^n \{y_i \ln(\lambda) - \ln(y_i!) - \ln[\exp(\lambda) - 1]\} \quad (2.25)$$

2.5.3 Zero-Truncated Multiplicative Binomial (ZTMB)

Distribution.

The pmf of a ZTMB variable as defined by Elamir (2011) is:

$$P(Y = y) = \begin{cases} \frac{\binom{n}{y} \pi^y (1 - \pi)^{n-y} \omega^{y(n-y)}}{\sum_{t=1}^n \binom{n}{t} \pi^t (1 - \pi)^{n-t} \omega^{t(n-t)}}, & \text{if } y = 1, 2, \dots, n; \omega > 0 \\ 0, & \text{e.w} \end{cases} \quad (2.26)$$

The ZTRMB contains the zero-truncated Binomial distribution as a special case for $\omega = 1$:

$$P_{\omega=1}(Y = y) = \frac{\binom{n}{y} \pi^y (1 - \pi)^{n-y}}{1 - (1 - \pi)^n} \quad (2.27)$$

Elamir derived the first four moments and other characteristics of the ZTMB (for detail, see Elamir, 2011).

The Likelihood function is:

$$L(\lambda; y_i | y_i > 0) = \prod_{i=1}^n \frac{\binom{n}{y_i} \pi^{y_i} (1 - \pi)^{n-y_i} \omega^{y_i(n-y_i)}}{\sum_{t=1}^n \binom{n}{t} \pi^t (1 - \pi)^{n-t} \omega^{t(n-t)}} \quad (2.28)$$

And the log-likelihood function is:

$$\mathcal{L}(\lambda; y_i) = \sum_{i=1}^n \left\{ \ln \binom{n}{y_i} + y_i \ln(\pi) + (n - y_i) \ln(1 - \pi) + y_i(n - y_i) \ln(\omega) - Q \right\} \quad (2.29)$$

$$\text{where, } Q = n \sum_{t=1}^n \ln \binom{n}{t} + t \ln(\pi) + (n - t) \ln(1 - \pi) + t(n - t) \ln(\omega)$$

2.6 Summary and Research Gaps

In modeling count data, mixed distributions had proven effective in handling under-, over- and equi-dispersion and other issues. When the count data is structurally zero-truncated, different zero-truncated distributions had been derived by different authors to handle such data. However, to the best of our knowledge, zero-truncated Com-Binomial had not been derived let alone to examine and compare with the existing ones. Therefore, this study was embark on to fill these gaps.

CHAPTER THREE

RESEARCH METHODOLOGY

In most count data mixture models have proven to yield better results compared to non-mixture models. Such models as Beta-Binomial, Poisson-Gamma and so on.

For instant, a distribution of Y can be generated by assuming that a parameter θ of a distribution for Y itself comes from a distribution. Assuming that, given θ , Y has conditional probability mass function $f(y|\theta)$ and marginally θ has probability distribution function $g(\theta)$. Then the marginal of Y is given by

$$f(y) = \begin{cases} \int f(y|\theta)g(\theta)d\theta, & \text{if } \theta \text{ is continuous} \\ \sum f(y|\theta)g(\theta), & \text{if } \theta \text{ is discrete} \end{cases} \quad (3.1)$$

The marginal distribution of Y is called a discrete mixture distribution or mixture probability mass function if θ is discrete.

However, when the data of interest is structurally zero-truncated, the distributions must be adjusted to account for the missing zeros. A typical example is a study of length of hospital stay (for patients). As soon as the patient enters the hospital, the length of stay is at least one. Another example is a study by the county traffic court on the number of tickets received by teenagers (Zuur et al, 2009). Only individuals who have received at least one citation are in the traffic court files.

Considering, Poisson, Negative Binomial, Geometric and Binomial distributions, they can be used for count data. Suppose we sample n sites, and at each site we count the number of animals, denoted by y_i . The value that we can measure are 0, 1, 2, . . . ,etc. For a given mean, the Poisson, Geometric and Negative binomial distributions specify the probability of having a count of 0, 1, 2, . . . , etc. For instance, if we use the Poisson distribution with mean = 3 and the probability of counting zero animals is 12.5%. Therefore, if we have a sample of size $n = 100$, we would expect to have a zero count approximately 13 times in our resulting data set. But we may have a variable that cannot take the value zero, that is, structurally exclude zeros. Poisson and Binomial, however, do not exclude this value, and this can be a problem for small mean values, Zuur et al (2009).

The remedy is to modify the distribution and exclude the probability of a zero observation, and this is referred to as zero-truncated distribution. We illustrate the process for the Zero-truncated Com-binomial distribution here.

3.1 Zero-Truncated Com-Binomial (ZTCB) Distribution

The probability mass function (pmf) of the Com-Binomial (CB) is defined as, (Shmuali et al, 2005);

$$f(Y = y; n, \pi, v) = \frac{\binom{n}{y}^v \pi^y (1 - \pi)^{n-y}}{\sum_{i=0}^n \binom{n}{i}^v \pi^i (1 - \pi)^{n-i}} , \quad y = 0, 1, \dots, n \quad (3.2)$$

for $n \in Z^+$, $\pi \in (0,1)$ and $v \in \mathbb{R}$

where, Z^+ is a set of unknown non – negative integers

NOTE:

when $v = 1$, we have the Binomial distribution,

$v > 1$, we have under – dispersion,

$v < 1$, we have over – dispersion with respect to the Binomial distribution

Note that (according to Borge et al, 2013):

- i. The CMB distribution can be interpreted as a sum of equi-correlated Bernoulli variables.
- ii. The Com-Poisson distribution (Conway and Maxwell, 1962) is approximation to the CMB distribution when n is getting large. (for details, see Borge et al 2013). For more on its parameters estimation, see Appendix A at the end of the Thesis.

However, the Zero-truncated version of the distribution, which we refer to as Zero-truncated Com-Binomial (ZTCB) can be derived as

$$P(y; n, \lambda, v) = \frac{f(Y = y; n, \pi, v)}{1 - f(Y = 0; n, \pi, v)} \quad y = 1, 2, 3, \dots, n \quad (3.3)$$

$$\text{Where, } f(Y = 0; n, \pi, v) = \frac{(1-\pi)^n}{\sum_{j=0}^n \binom{n}{j}^v \pi^j (1-\pi)^{n-j}}$$

and $f(Y = y; n, \pi, v)$ is the CB defined above.

Therefore, the pmf of the ZTCB distribution is derived as

$$\begin{aligned}
P(y; n, \lambda, v) &= \frac{\binom{n}{y}^v \pi^y (1 - \pi)^{n-y}}{\sum_{j=0}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j}} \div \left\{ 1 - \frac{(1 - \pi)^n}{\sum_{j=0}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j}} \right\} \\
&= \frac{\binom{n}{y}^v \pi^y (1 - \pi)^{n-y}}{\sum_{j=0}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j}} * \left[\frac{\sum_{i=0}^n \binom{n}{i}^v \pi^i (1 - \pi)^{n-i}}{\sum_{j=0}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j} - (1 - \pi)^n} \right] \\
&= \frac{\binom{n}{y}^v \pi^y (1 - \pi)^{n-y}}{\sum_{j=0}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j} - (1 - \pi)^n}, \quad y = 1, 2, 3, \dots, n
\end{aligned} \tag{3.4}$$

Hence, the pmf of the ZTCB distribution can be re-written as, (Aderoju and Jolayemi, 2018);

$$P(y; n, \pi, v) = \frac{\binom{n}{y}^v \pi^y (1 - \pi)^{n-y}}{\sum_{j=1}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j}}, \quad y = 1, 2, 3, \dots, n \tag{3.5}$$

3.2 Properties of ZTCB Distribution

The probability generating function of the ZTCB can be expressed as:

$$E(t^y) = \sum_{y=1}^n t^y \frac{\binom{n}{y}^v \pi^y (1 - \pi)^{n-y}}{\sum_{j=1}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j}} \tag{3.6}$$

Hence the first and second moment can be expressed respectively as:

$$\mu'_1 = \sum_{y=1}^n \frac{y \binom{n}{y}^v \pi^y (1-\pi)^{n-y}}{\sum_{j=1}^n \binom{n}{j}^v \pi^j (1-\pi)^{n-j}} \quad (3.7)$$

and

$$\mu'_2 = \sum_{y=1}^n \frac{y(y-1) \binom{n}{y}^v \pi^y (1-\pi)^{n-y}}{\sum_{j=1}^n \binom{n}{j}^v \pi^j (1-\pi)^{n-j}} \quad (3.8)$$

From the pmf of the ZTCB distribution defined above, when $v = 1$, the expected value of the random variable Y , $E(Y) = \frac{n\pi}{1-(1-\pi)^n}$.

However, when $v > 1$, suppose $r = v$ and $q = v - 1$, then the first two moment are derived respectively as:

$$\mu'_1 = \frac{n^p \pi T}{(\pi - 1) \left(1 - rFq \left[\{-n_1, -n_2, \dots, -n_r\}, \{1_1, 1_2, \dots, 1_q\}, \frac{\pi}{\pi - 1} \right] \right)}$$

and

$$\mu'_2 = \frac{n^r \pi (T - S)}{(\pi - 1) \left(1 - rFq \left[\{-n_1, -n_2, \dots, -n_r\}, \{1_1, 1_2, \dots, 1_q\}, \frac{\pi}{\pi - 1} \right] \right)}$$

where

$$T = rFq \left[\{(1-n)_1, (1-n)_2, \dots, (1-n)_r\}, \{2_1, 2_2, \dots, 2_q\}, \frac{\pi}{\pi - 1} \right]$$

$$S = rFq \left[\{(1-n)_1, (1-n)_2, \dots, (1-n)_r\}, \{1, 2_1, 2_2, \dots, 2_{q-1}\}, \frac{\pi}{\pi - 1} \right]$$

This is one of the major contribution of this word.

Figures below present the pmf of the ZTCB distribution for $n = 7$ at different values of π (as p) and v . For $v \rightarrow \infty$, the pmf is concentrated around the point $n\pi$ and it has two tails; for $v \rightarrow -\infty$, it concentrated around 1 or n and as well resemble exponential distribution

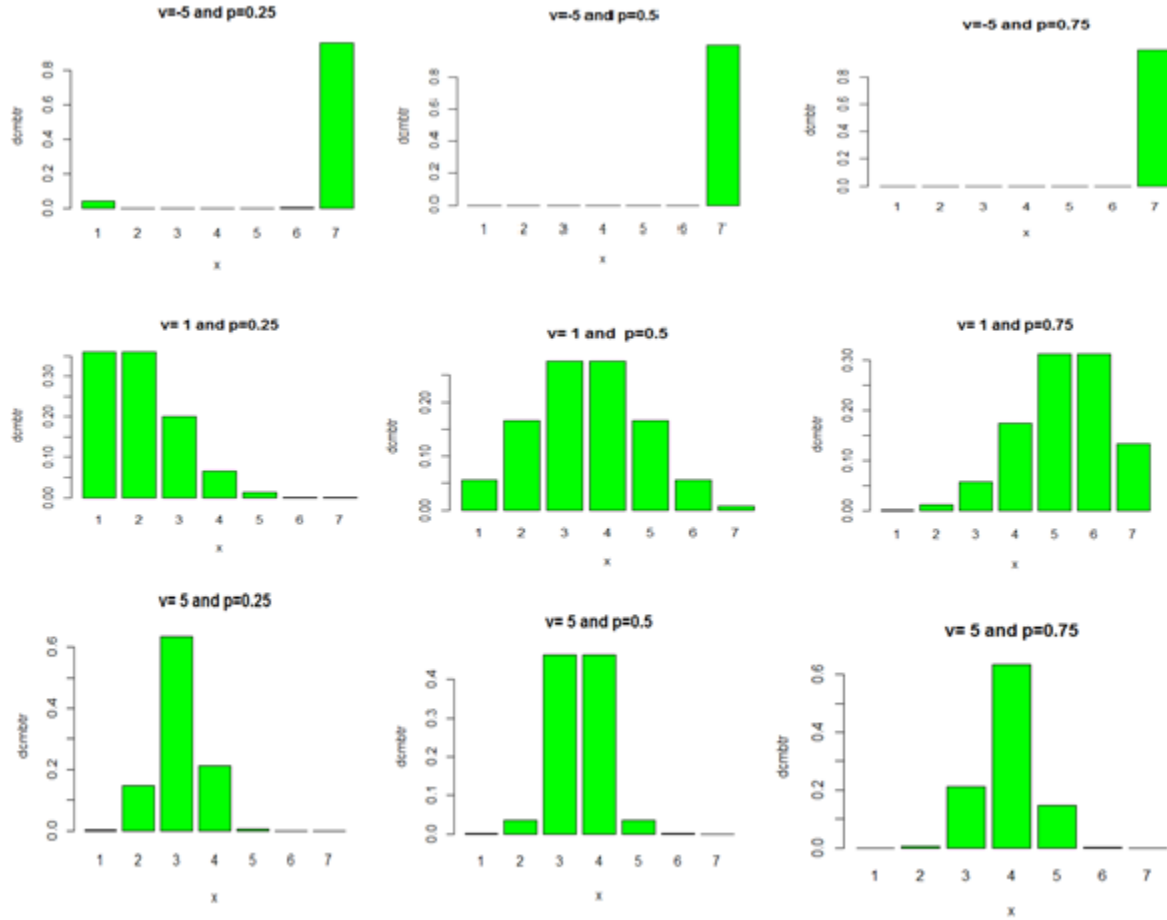


Figure.3.1: The histogram of pmf of ZTCB distribution with $\pi = p = 0.5$ and at various values of v .

Figures below present the pmf of the ZTCB distribution for $n = 7$ and $n = 15$ at different values of π (as p) and v . For $v \rightarrow \infty$, the pmf is peak around the point $n\pi$ and for $v \rightarrow -\infty$, it peaks around 1 or n .

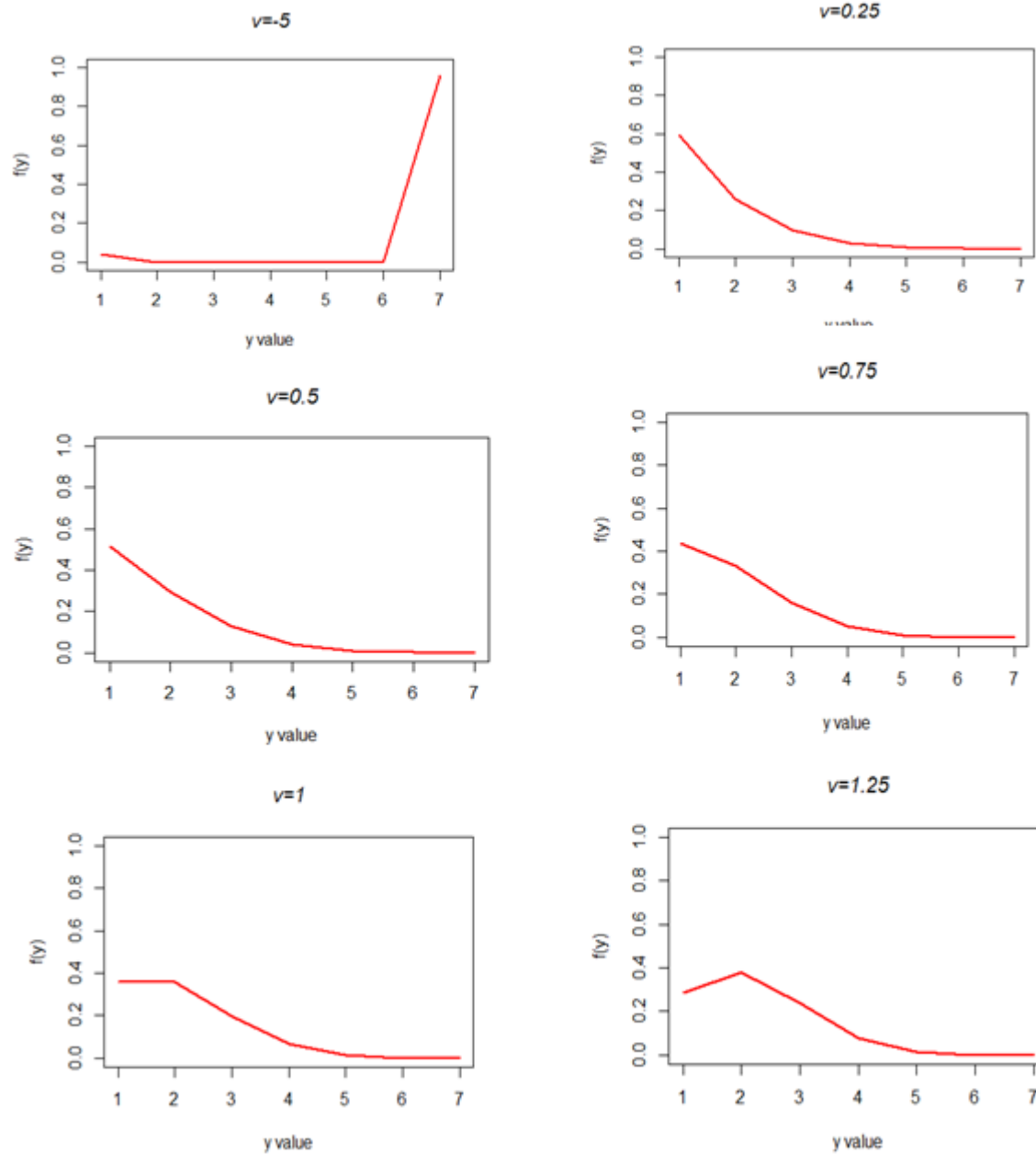


Figure 3.2: The pmf of ZTCB distribution with $\pi = p = 0.25$ and $n = 7$ at various values of v .

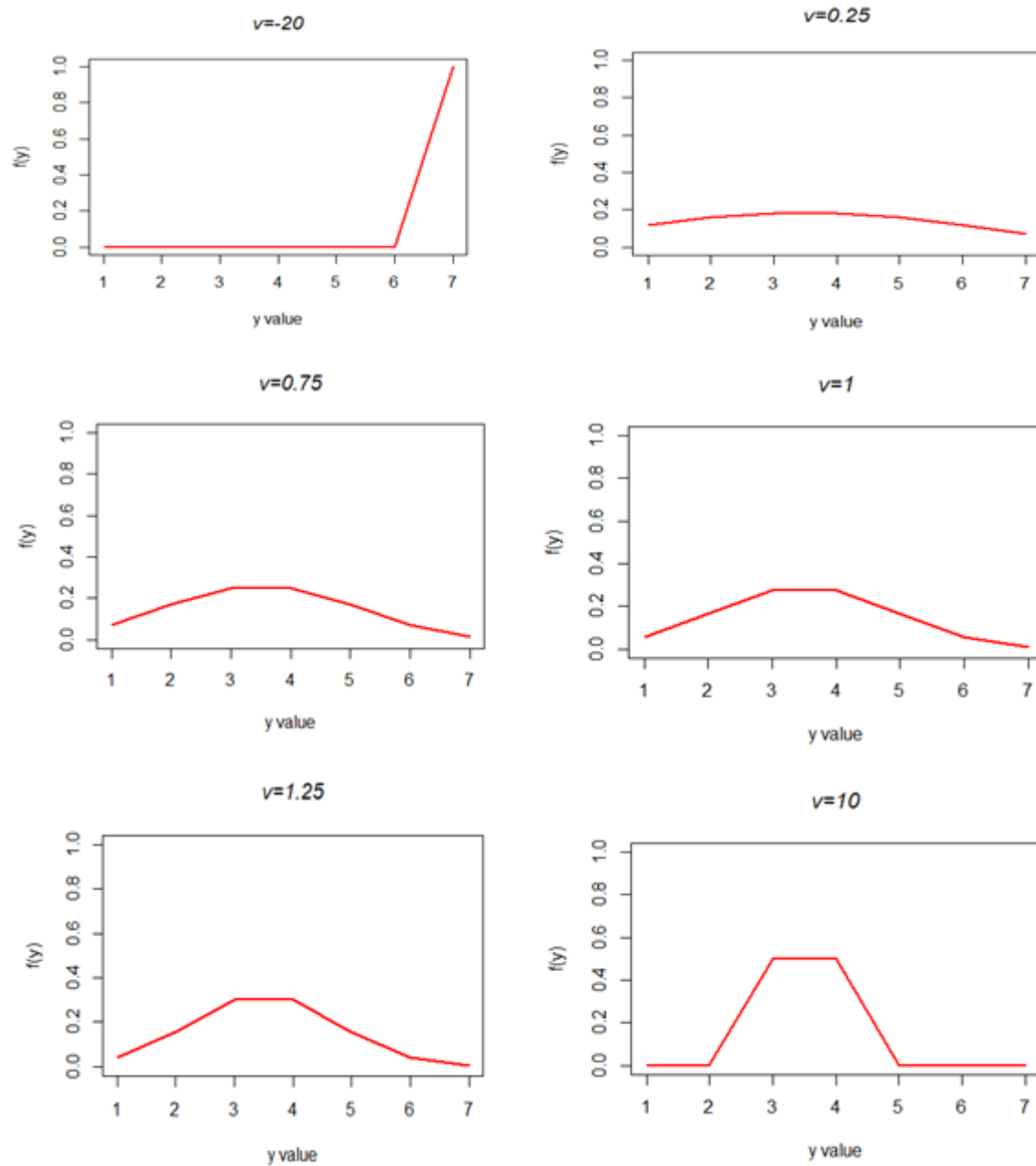


Figure 3.3: The pmf of ZTCB distribution with $\pi = p = 0.5$ and $n = 7$ at various values of v .

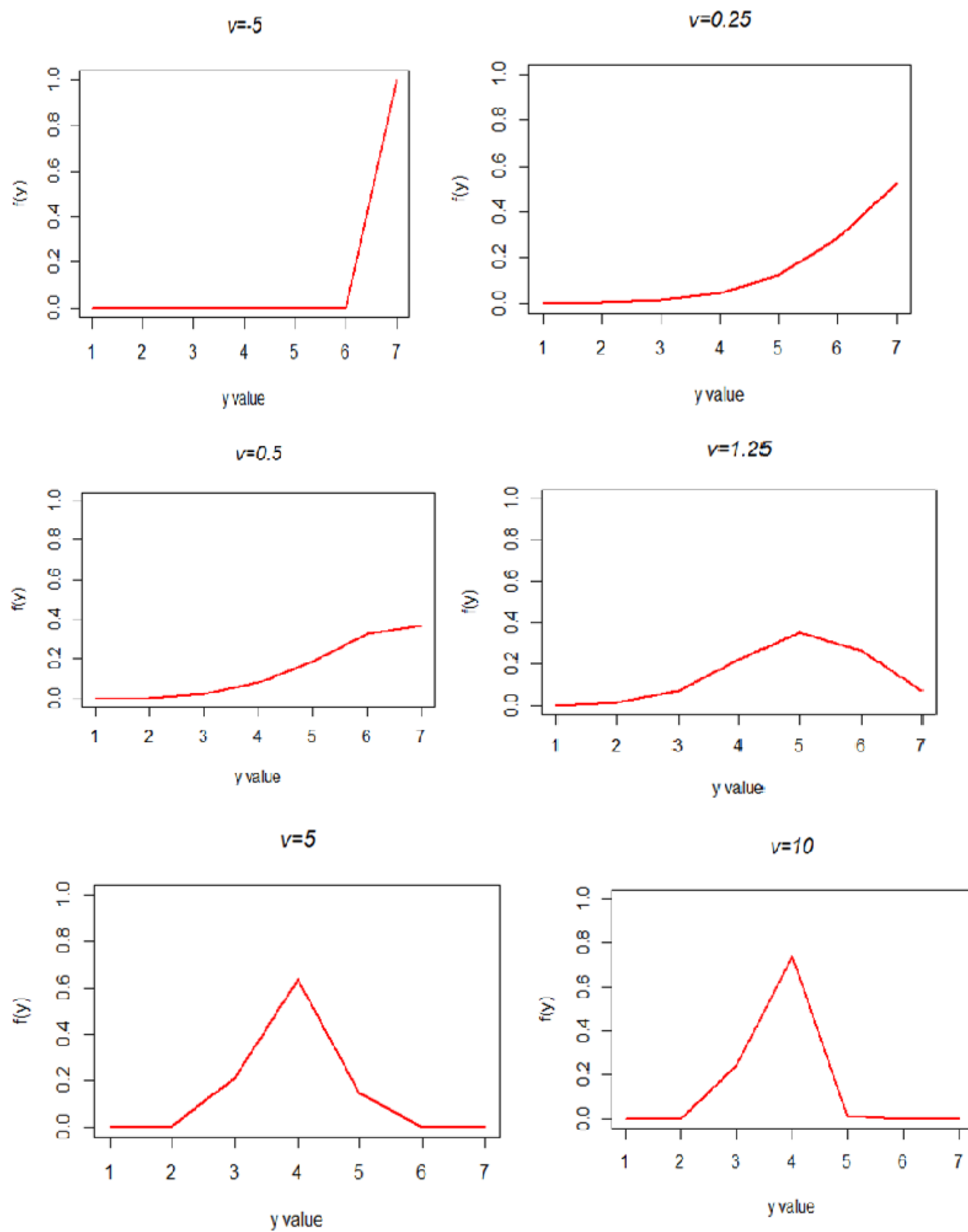


Figure 3.4: The pmf of ZTCB distribution with $\pi = p = 0.5$ and $n = 7$ at various values of v .

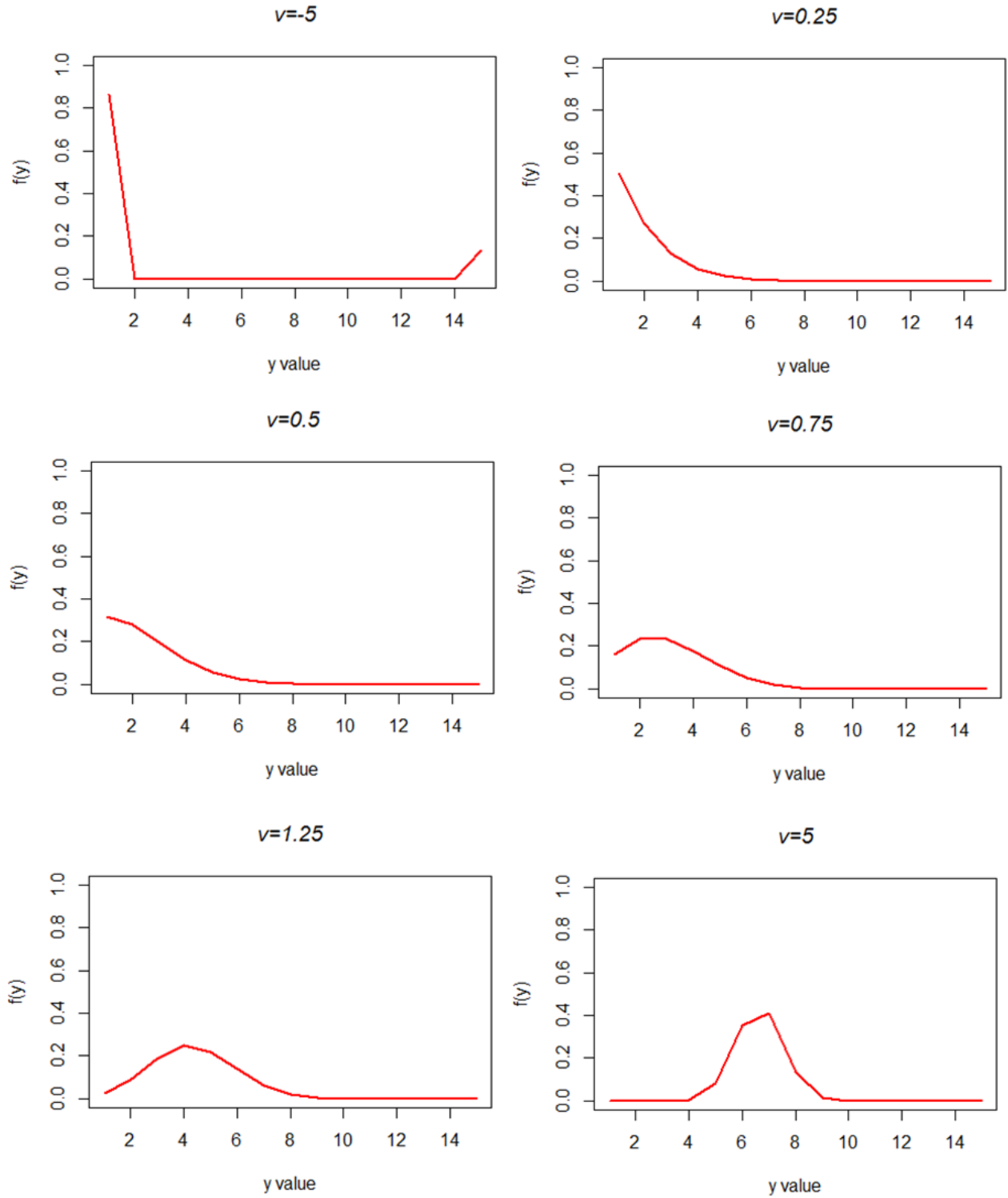


Figure 3.5: The pmf of ZTCB distribution with $\pi = p = 0.25$ and $n = 15$ at various values of v .

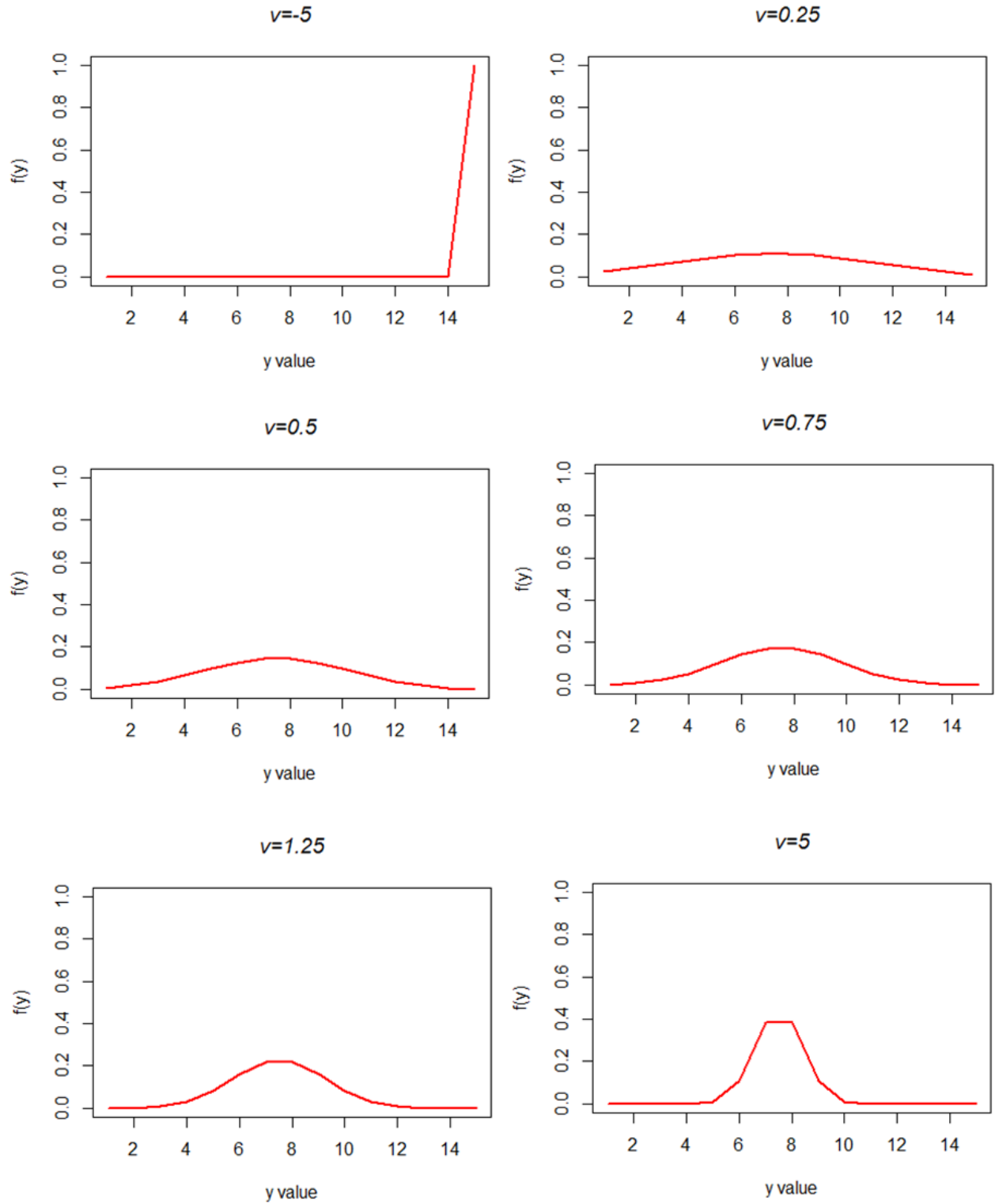


Figure 3.6: The pmf of ZTCB distribution with $\pi = p = 0.5$ and $n = 15$ at various values of v .

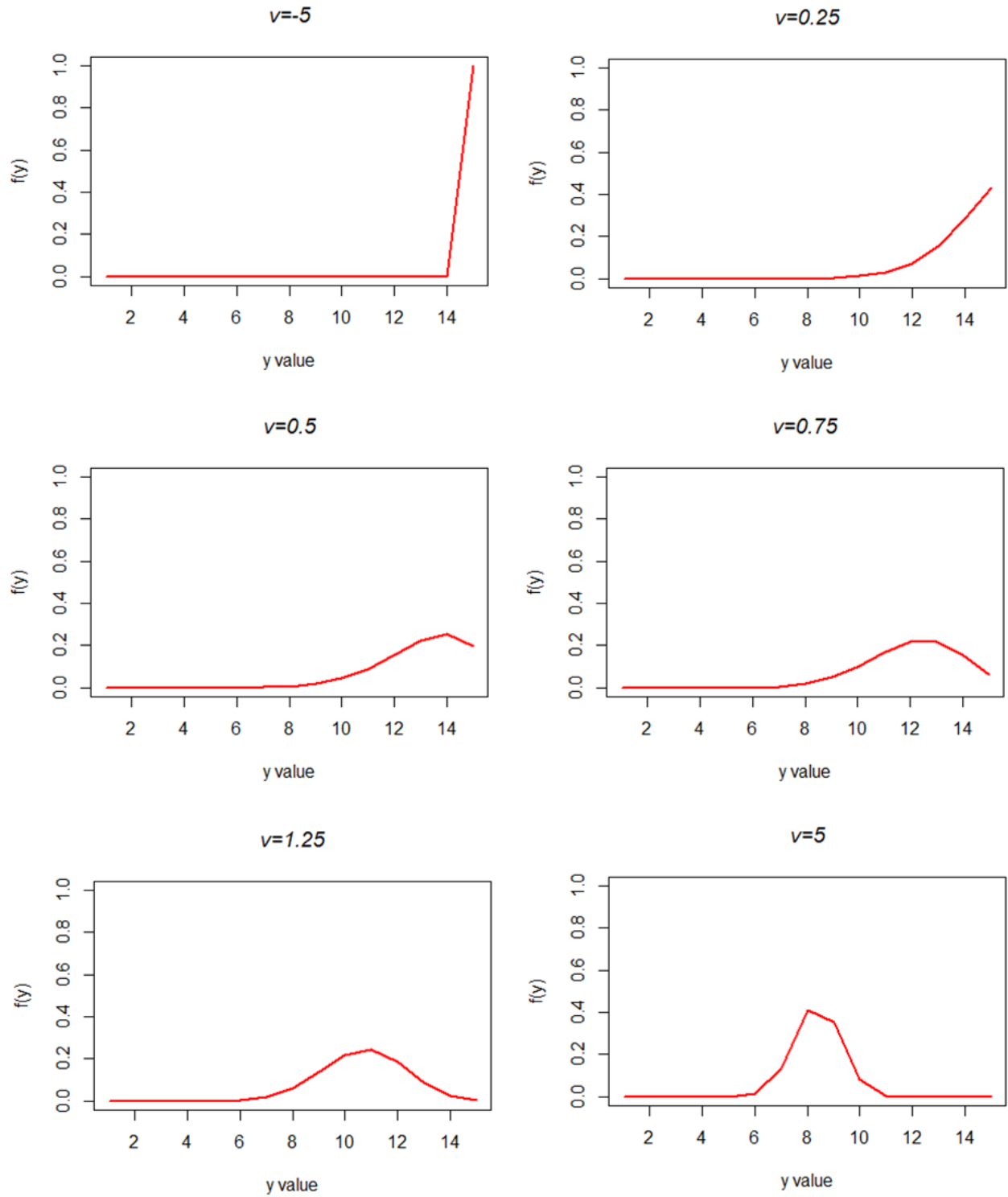


Figure 3.7: The pmf of ZTCB distribution with $\pi = p = 0.75$ and $n = 15$ at various values of v .

Remark: The Zero-truncated Com-Poisson (ZTCP) distribution is an approximation of the ZTCB when n is very large

$$\begin{aligned} \lim_{n \rightarrow \infty} P(y; n, \pi, v) &= \lim_{n \rightarrow \infty} \frac{\binom{n}{y_i}^v \left(\frac{\lambda}{n}\right)^{y_i} \left(1 - \frac{\lambda}{n}\right)^{n-y_i}}{\sum_{j=1}^n \binom{n}{j}^v \frac{\lambda^j}{n} \left(1 - \frac{\lambda}{n}\right)^{n-j}} \\ &= \frac{\lambda^y}{(y!)^v \sum_{k=1}^{\infty} \frac{\lambda^k}{(k!)^v}}, \quad y = 1, 2, \dots \end{aligned} \quad (3.9)$$

which is the pmf of the ZTCP distribution.

When $v = 1$, $\sum_{k=1}^{\infty} \frac{\lambda^k}{(k!)^v} = e^{\lambda} - 1$

Therefore, equation 3.9 above becomes

$$P(y | \lambda) = \frac{\lambda^y}{y! (e^{\lambda} - 1)}, \quad y = 1, 2, \dots \quad (3.10)$$

which is the pmf of the Zero-truncated Poisson (ZTP) distribution.

Hence, the ZTCB distribution is the generalization of the ZTCP, ZTP and ZTB distributions.

3.3 Methods of Estimating the Parameters

Generally, there are two methods usually used to estimate count response models: an iteratively re-weighted least squares (IRLS) algorithm founded on the method of Fisher scoring and maximum likelihood Newton-Raphson algorithm.

The likelihood function of the ZTCB is

$$L(\pi, v|y_i) = \prod_{i=1}^n \frac{\binom{n}{y_i}^v \pi^{y_i} (1 - \pi)^{n-y_i}}{\sum_{j=1}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j}} \quad (3.11)$$

While the log-likelihood function is

$$\mathcal{L} = \sum_{i=1}^n \left\{ v \log \binom{n}{y_i} + y_i \log \pi + (n - y_i) \log(1 - \pi) - \log Q \right\} \quad (3.12)$$

$$\text{where } Q = \sum_{j=1}^n \binom{n}{j}^v \pi^j (1 - \pi)^{n-j}$$

We are to find the first and second partial derivative of equation (3.12) with respect to each parameter and equate them to zero as:

$$\frac{\partial \mathcal{L}}{\partial \pi} = 0, \quad \frac{\partial^2 \mathcal{L}}{\partial \pi^2} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial v} = 0, \quad \frac{\partial^2 \mathcal{L}}{\partial v^2} = 0$$

However, the equations do not have closed form. Therefore, the maximum likelihood estimates (MLEs) of ZTCB cannot be solved analytically, an iterative methods such as

- i. Fisher Score Algorithm, or
- ii. *Newton-Raphson* (NR) iterative method, as implemented by Jolayemi (1990),

iii. *Iteratively reweighted least squares* (IRWLS) and so on, can be used.

We obtained the MLEs of the parameters by direct maximization of the log-likelihood function using “**optim**” routine of R software (R Development Core Team, 2016) with “**L-BFGS-B**” method (see Appendix B). This can as well be achieved by using **PROC NLMIXED** in SAS.

3.4 Overall Goodness-Of-Fit Statistics

Different comparative tests were used to assess the Goodness-Of-Fit (GOF) of the distributions. They include: the Pearson’s Chi-squared test, the Deviance and the log-likelihood value. Like the Pearson’s Chi-squared statistic (X^2), the Deviance (G^2) has approximately a Chi-squared distribution and the null hypothesis is rejected for a reasonable fit for large values of Deviance. The log-likelihood statistic ($LogLik$) was calculated by taking the logarithm of the estimated likelihood for each observation. Besides, given that the degree of freedom (df) for different distributions might differ within the same scenario, the value of Deviance divided by df (G^2/df) could also be used as an alternative for those three GOFs. The smaller the G^2/df , the better the fit. Akaike Information Criterion (AIC) was as well used. Those GOF statistics are given as:

- i. The $X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$
- ii. Deviance statistic has the form

$$G^2 = 2 \sum_{i=1}^n O_i \ln\left(\frac{O_i}{E_i}\right) \sim \chi^2_{n-p} \quad (3.13) .$$

This is asymptotically distributed as χ^2_{n-p}

The components of the deviance are residual measures of lack of fit.

where,

O_i is the observed frequency for the category of count equal to i ;

E_i is the expected frequency for the category of count equal to i ;

- iii. Akaike Information Criterion (AIC) – the Akaike Information Criteria (AIC) can be used to compare competing distributions. The higher values indicate worse fit; *the preferred model is the one with the minimum value*. Hence AIC not only rewards goodness of fit, but also penalizes over-fitting. The AIC attempts to select the model that best explains the data with a minimum of parameters (parsimony). The lower the AIC, the better the model.

$$AIC = -2L_m + 2(m + 1), \quad (3.14)$$

where L_m is log Likelihood of full model, Note that $m+1$ = the number of the parameters.

CHAPTER FOUR

4.0 Application of Zero-Truncated Com-Binomial Distribution to Real Life Datasets.

The newly introduced zero-truncated Com-Binomial is flexible and thus applicable in different datasets from health science, demography, biological sciences, social sciences and so on.

4.1 Example 4.1 [immunogold assay data]

Data from Mathews et al (1993), who gave counts of sites with 1, 2, 3, 4 and 5 particles from immunogold assay data. The sample mean and variance are 1.576 and 0.7897, respectively.

Table 4.1: The expected and observed number of counts from immunogold assay data.

Y	Obs.Freq	ZTP	ZTB	ZTMB	ZTCB
1	122	115.86	109.87	123.66	123.49
2	50	57.39	65.51	46.65	47.09
3	18	18.95	19.53	18.10	17.96
4	4	4.69	2.91	7.22	6.85
5	4	0.93	0.17	2.37	2.61
Total	198	198	198	198	198
MLE		$\lambda = 0.9906$	$p = 0.2297$	$\beta = 0.2795$ $k = 0.6972$	$\beta = 0.2761$ $v = 0.0001$
Loglik		205.9477	213.8822	203.7187	203.3796
X^2 (P-value)		2.329 (0.51)	5.517667 (0.06)	0.529 (0.91)	0.423 (0.94)
G^2		2.1154	5.7761	0.5375	0.4721
df		3	3	2	2
AIC		413.89	429.76	411.44	410.76

Comment on table 4.1: Mathews et al (1993) applied Zero-truncated Poisson (ZTP) and Elamir (2011) applied Zero-truncated Multiplicative-Binomial (ZTMB) distributions, which they observed fit the data well. The ZTMB performed better than ZTP. Nevertheless, our proposed ZTCB distribution outperforms ZTMB distribution.

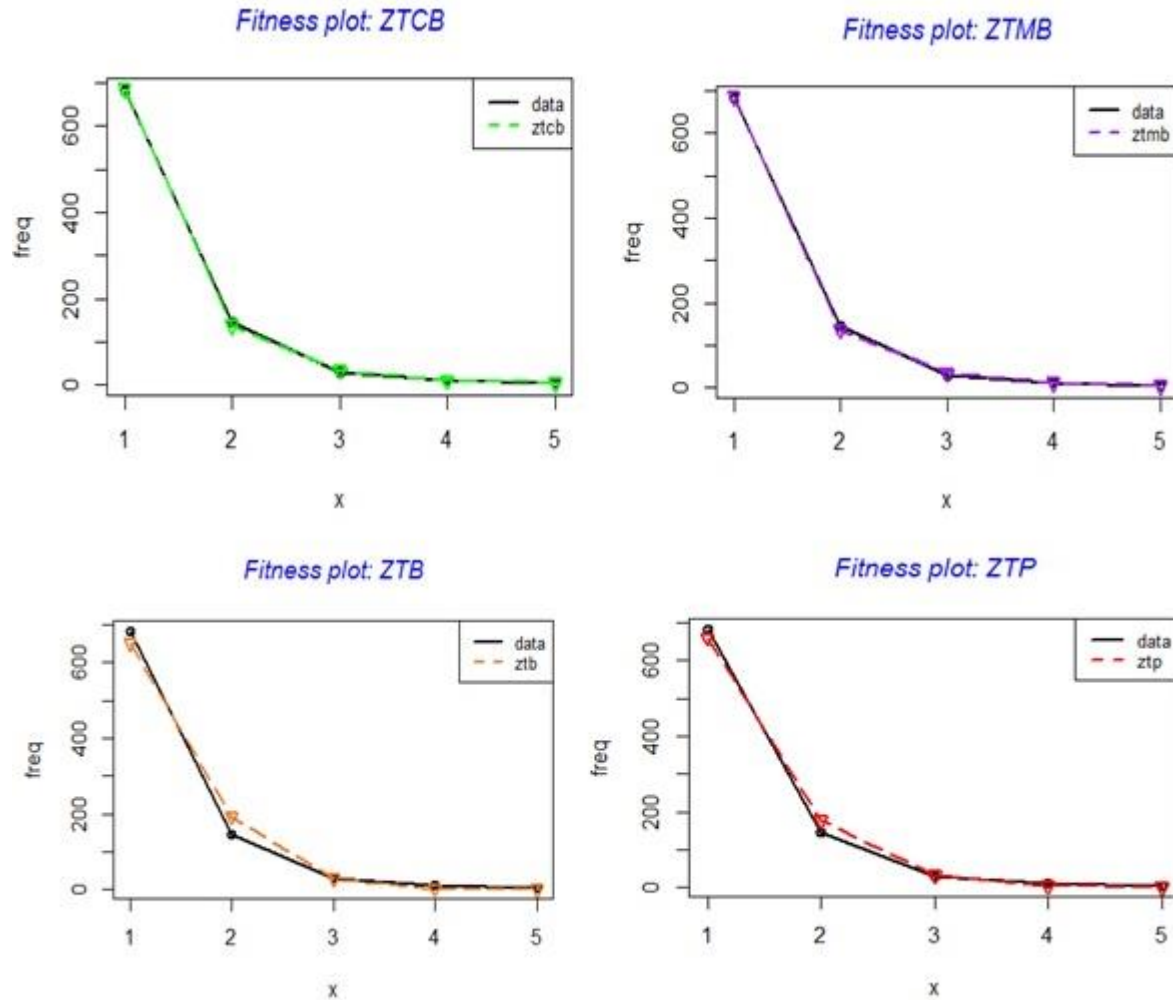


Figure 4.1a: the plot of observed and expected frequency obtained by each distribution for the immunogold assay data.

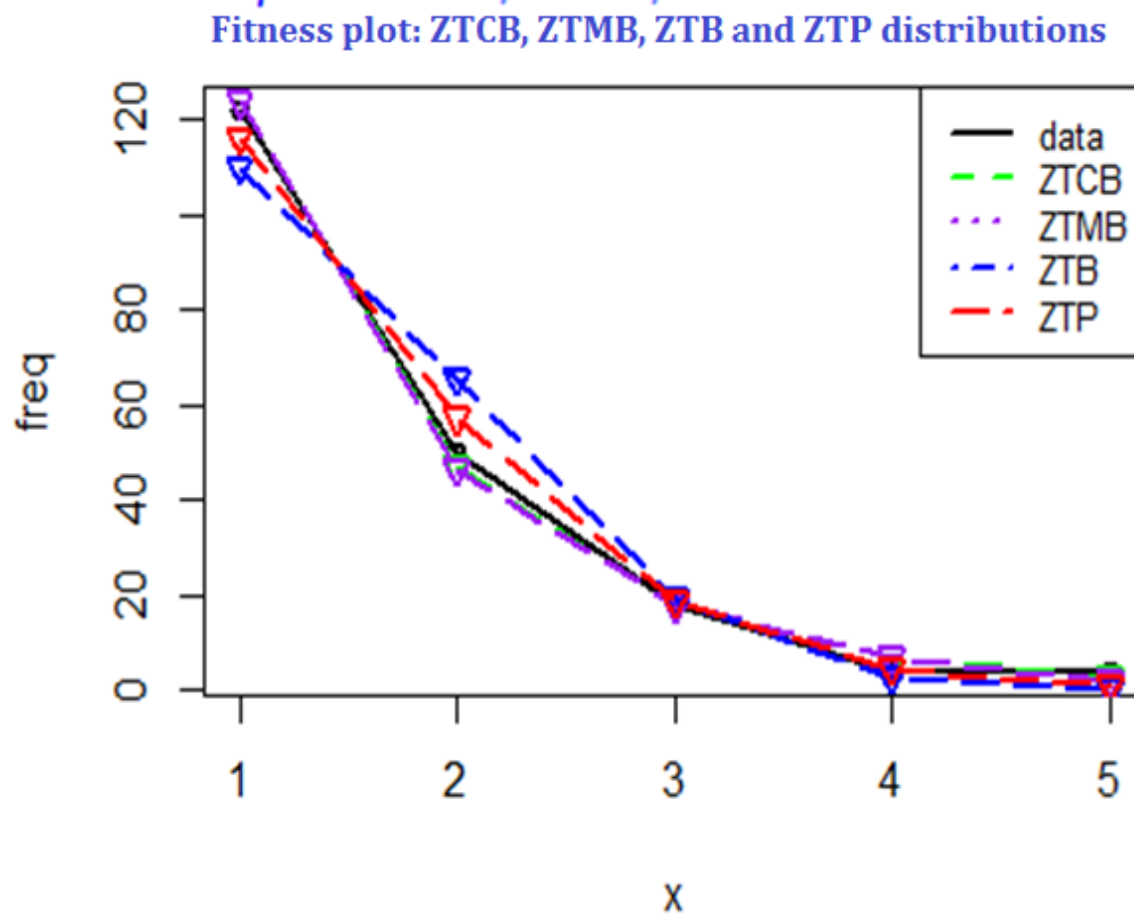


Figure 4.1b: the joint plot of observed and expected frequency obtained by all the distributions.

Figure 4.1a and figure 4.1b: clearly show the close association between the observed counts and the expected on the basis of ZTCB and ZTMB distributions. The fits are superior to that of ZTB and ZTP distributions.

4.2 Example 4.2 [counts of flower heads]

The data from Finney et al (1955), who gave counts of flower heads with 1, 2, . . . , 9 fly eggs. The sample mean and variance are 3.034 and 3.3056, respectively.

Table 4.2: : The expected and observed number of counts of flower heads with fly eggs.

Y	Obs.Freq	ZTP	ZTB	ZTMB	ZTCB
1	22	15.28	11.15	22.12	21.07
2	18	21.86	21.73	19.48	19.97
3	18	20.84	24.71	14.88	16.58
4	11	14.91	18.07	10.87	12.35
5	9	8.53	8.81	7.87	8.32
6	6	4.07	2.86	5.67	5.08
7	3	1.66	0.60	3.88	2.78
8	0	0.59	0.07	2.31	1.33
9	1	0.19	0.01	0.92	0.52
Total	88	88	88	88	88
MLE		$\lambda = 2.8604$	$\pi = 0.3277$	$\pi = 0.4212$ $k = 0.821$	$\pi = 0.4026$ $k = 0.2459$
Loglik		166.548	177.169	163.852	163.625
X^2 (P-value)		6.942 (0.23)	19.386 (<0.001)	1.524 (0.91)	0.568 (0.97)
G^2		6.6288	21.9502	1.5394	0.5758
df		4	4	3	3
AIC		335.10	356.34	331.704	331.25

Comment on table 4.2: Similarly, for the data in example two, Finney et al (1955) applied Zero-truncated Poisson (ZTP) and Elamir (2011) applied Zero-truncated Multiplicative-Binomial (ZTMB) distributions, which they observed fit the data well. The ZTMB performed better than ZTP. Again, our proposed ZTCB outperforms ZTMB.

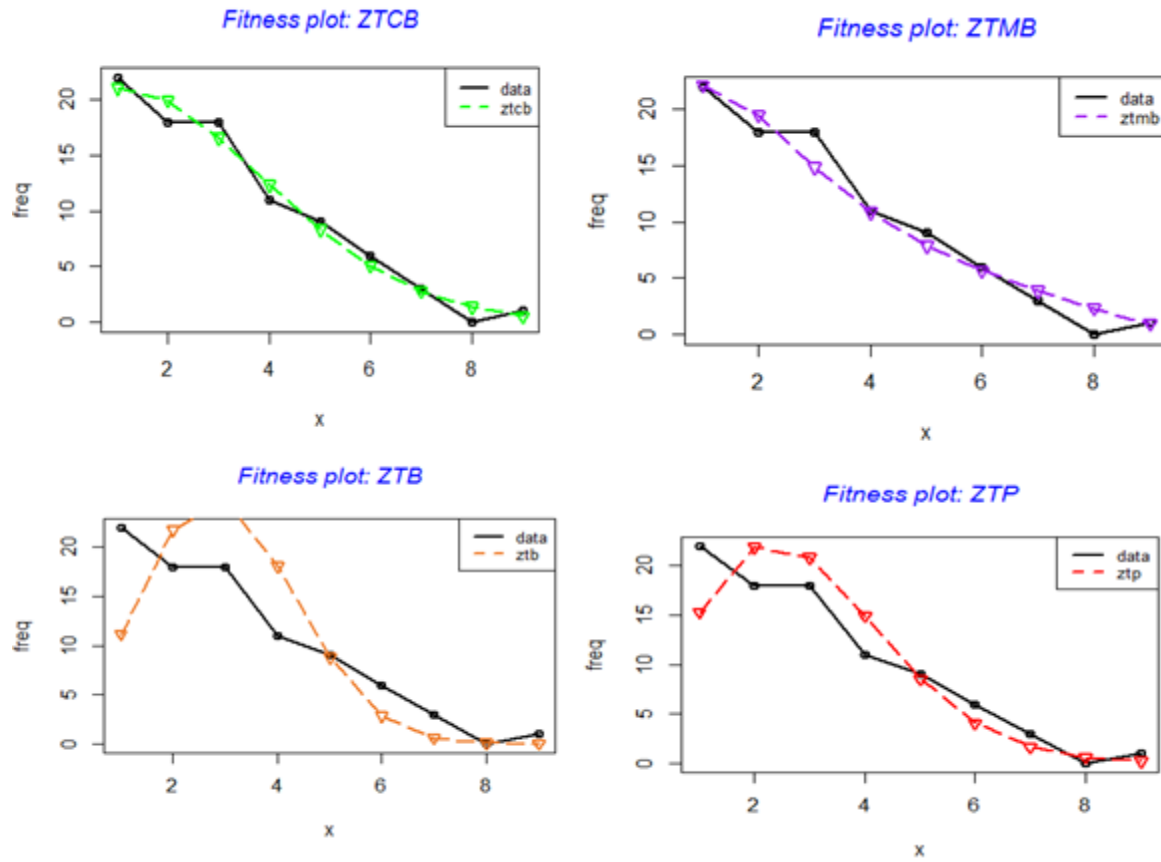


Figure 4.2a: the plot of observed and expected frequency obtained by each distribution.

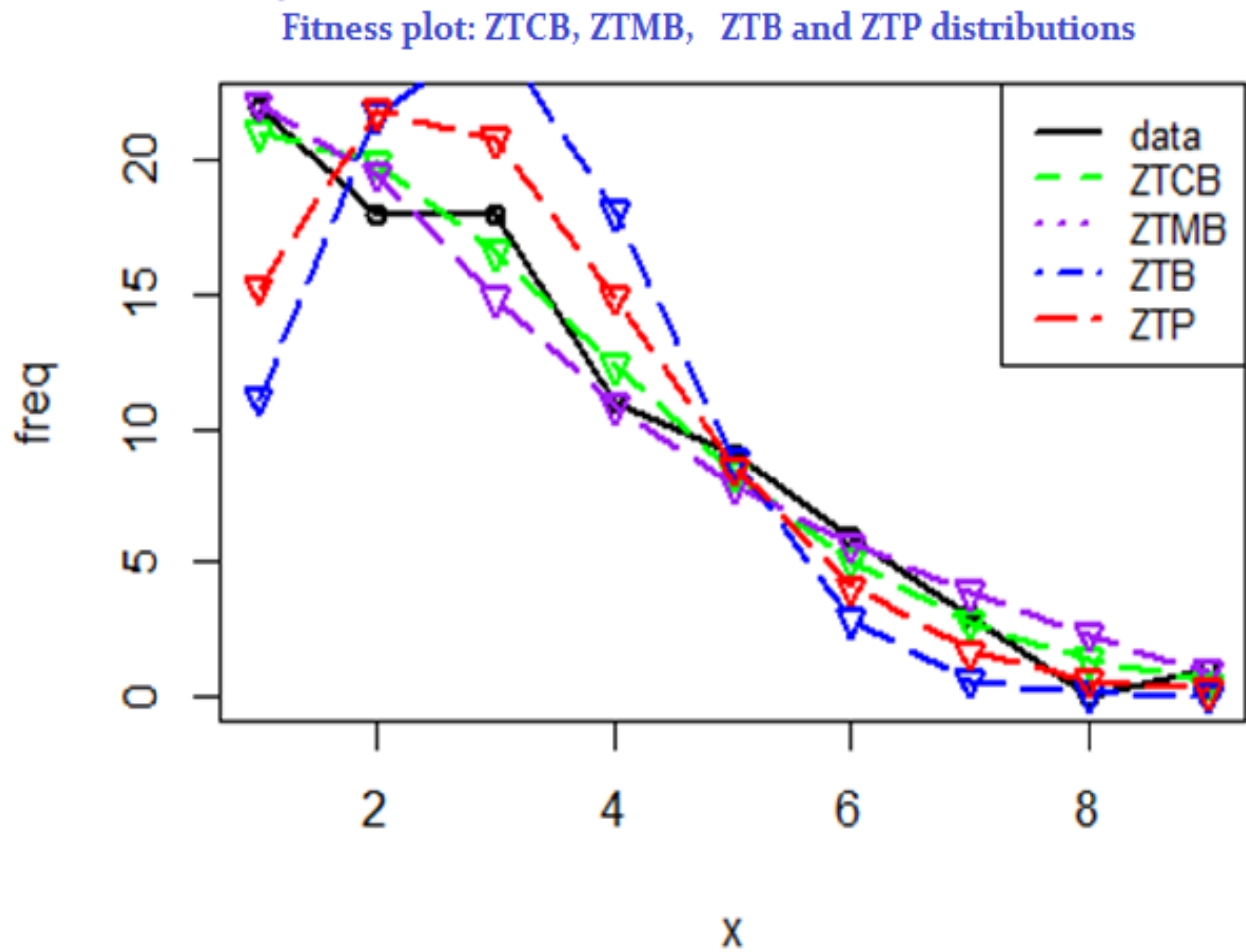


Figure 4.2b: the joint plot of observed and expected frequency obtained by each distribution.

Figures 4.2a and 4.2b clearly show the close association between the observed counts and the expected on the basis of ZTCB distribution. The fit is superior to that of ZTMB, ZTB and ZTP distributions.

4.3 Example 4.3: [Mortality data]

Data from Shanker et al (2015): counts of the number of mothers of the rural area having at least one live birth and one neonatal death. The sample mean and variance are 1.278 and 0.3546, respectively.

Table 4.3: The expected and observed number of mothers of the rural area having at least one live birth and one neonatal death

Y	Obs.Freq	ZTP	ZTB	ZTMB	ZTCB
1	409	399.70	394.76	409.27	409.05
2	88	102.30	110.62	87.19	87.69
3	19	17.50	15.50	19.78	19.58
4	5	2.20	1.09	4.78	4.56
5	1	0.30	0.03	0.98	1.12
Total	522	522	522	522	522
MLE		$\theta = 0.5120$	$\pi = 0.1229$	$\beta = 0.1849$ $k = 0.6853$	$\beta = 0.1825$ $v = -0.0589$
Loglik		352.11	358.25	348.92	348.93
χ^2 (P-value)		3.4653 (0.06)	9.3643 (<0.01)	0.0483 (0.82)	0.0372 (0.85)
G^2		3.5116	9.1386	0.0487	0.0362
df		1	1	1	1
AIC		706.21	718.50	701.84	701.86

Comment on table 4.3: here, Zero-truncated Binomial failed while Zero-truncated Poisson (ZTP) lightly fits the data. The Zero-truncated Multiplicative-Binomial (ZTMB) distribution fits the data well. However, our proposed ZTCB distribution outperforms ZTMB distribution.

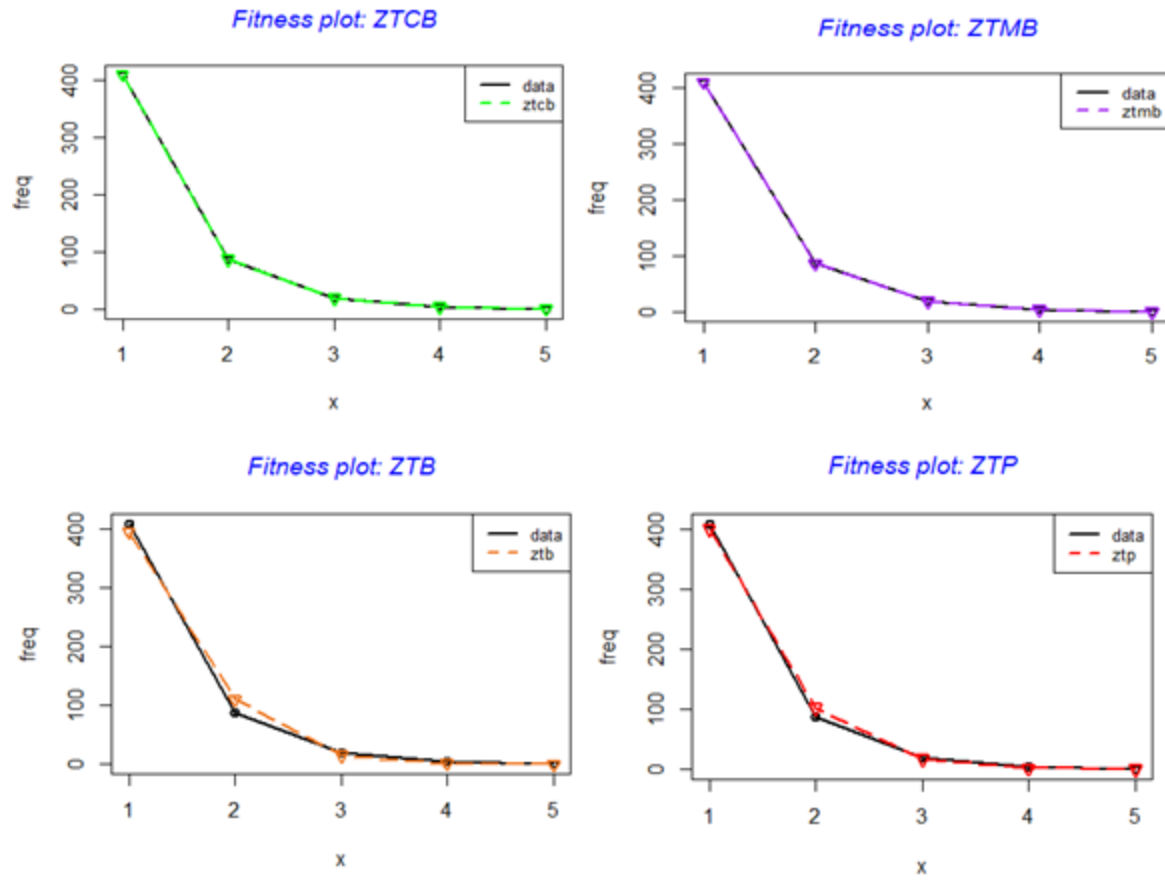


Figure 4.3a: the plot of observed and expected frequency obtained by each distribution.

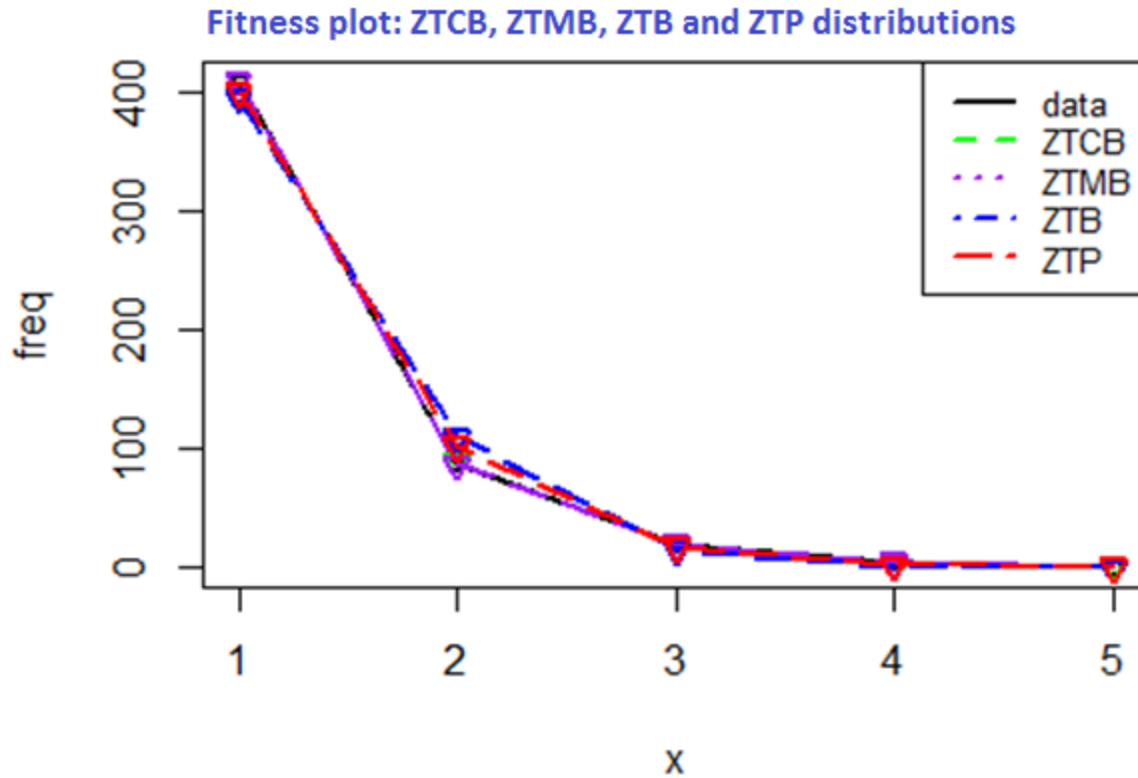


Figure 4.3b: the plot of observed and expected frequency obtained by each distribution.

In Figures 4.3a and 4.3b, it is clearly shown that the close association between the observed counts and the expected on the basis of ZTCB, ZTMB and ZTP distributions. The fits are superior to that of ZTB distributions.

4.4 Example 4.4: [Mortality data]

The data is taking from Shanker et al (2015): counts of the number of literate mothers with at least one live birth by the number of infant deaths. The sample mean and variance are 1.293 and 0.4185, respectively.

Table 4.4: The expected and observed number of literate mothers with at least one live birth by the number of infant deaths

Y	Obs.Freq	ZTP	ZTB	ZTMB	ZTCB
1	683	658.97	649.90	686.15	685.61
2	145	177.40	192.44	136.74	138.05
3	29	31.84	28.49	34.89	34.40
4	11	4.29	2.11	11.40	10.61
5	5	0.50	0.06	3.82	4.33
Total	873	873	873	873	873
MLE		$\theta = 0.5384$	$\pi = 0.1289$	$\beta = 0.2033$ $k = 0.6249$	$\beta = 0.1994$ $v = -0.3073$
Loglik		615.397	634.09	601.55	601.28
X^2 (P-value)		8.7063 (0.0128)	20.0876 (<0.01)	1.8900 (0.39)	1.3237 (0.52)
G^2		8.9621	20.3065	1.5985	1.3235
df		2	1	2	2
AIC		1232.79	1270.17	1207.11	1205.78

Comment on table 4.4: Zero-truncated Binomial and Zero-truncated Poisson (ZTP) failed to fits the data here. The Zero-truncated Multiplicative-Binomial (ZTMB) and ZTCB distributions fit the data well the proposed distribution outperforms ZTMB distribution.

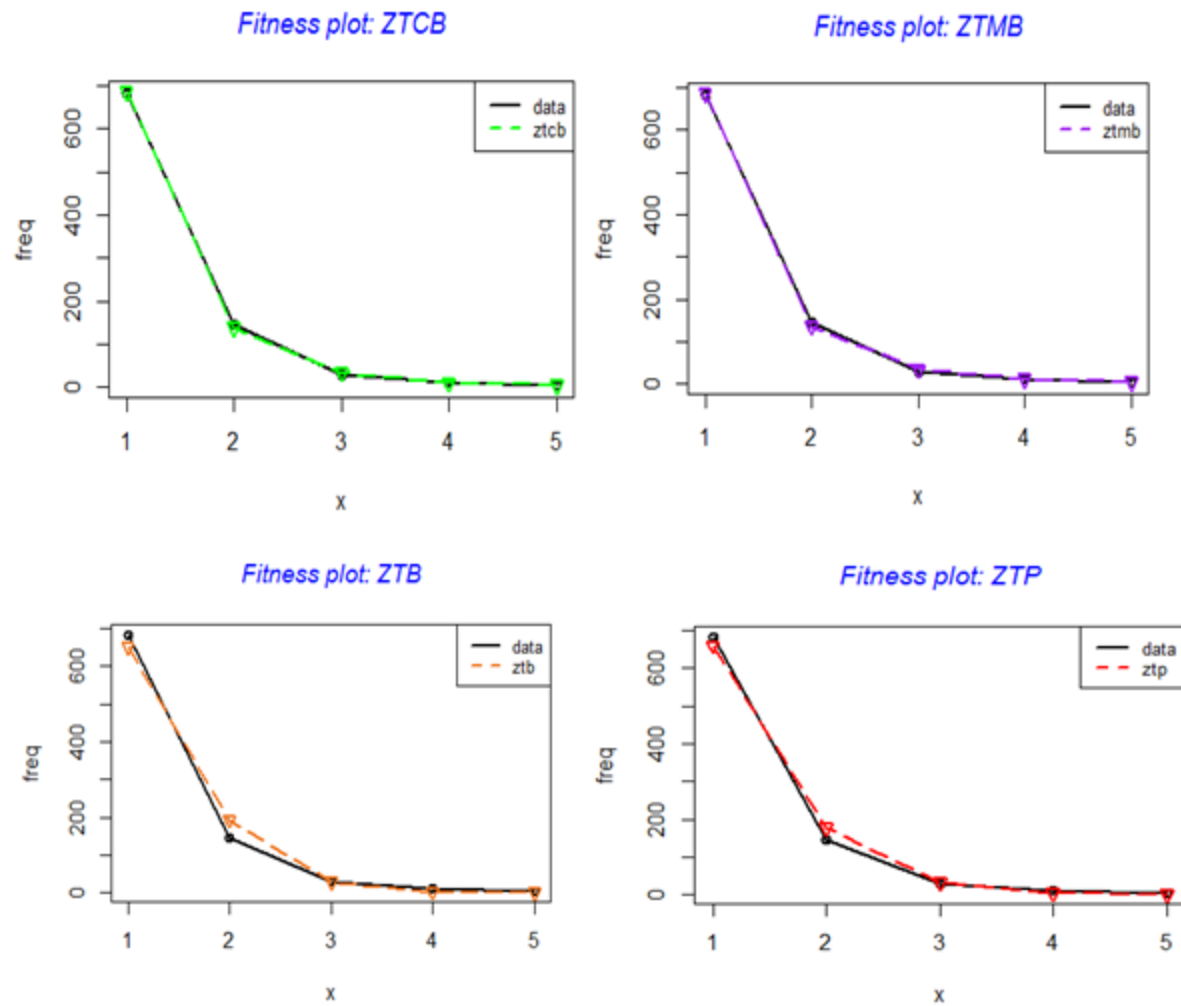


Figure 4.4a: the plot of observed and expected frequency obtained by each distribution.

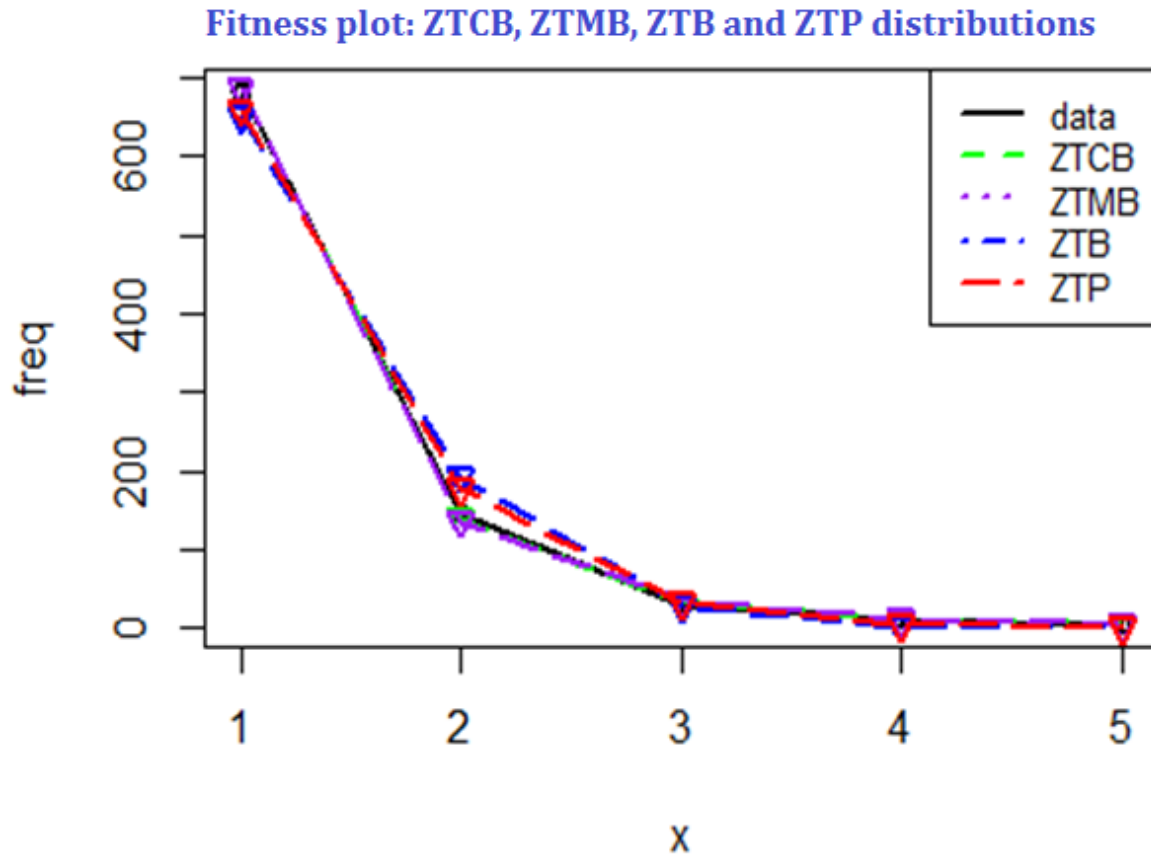


Figure 4.4b: the joint plot of observed and expected frequency obtained by each distribution.

Figure 4.4a and figures 4.4b clearly show that the close association between the observed counts and the expected on the basis of ZTCB distribution. The fit is superior to that of ZTMB, ZTB and ZTP distributions.

4.5 Example 4.5: [Mortality data]

The data is taking from Shanker et al (2015): counts of the number of mothers having at least one neonatal death. The sample mean and variance are 1.327 and 0.4647, respectively.

Table 4.5: The expected and observed number of mothers having at least one neonatal death.

Y	Obs.Freq	ZTP	ZTB	ZTMB	ZTCB
1	567	545.82	536.50	570.35	569.88
2	135	162.50	177.60	126.20	127.38
3	28	32.25	29.39	34.28	33.83
4	11	4.80	2.43	11.43	10.68
5	5	0.63	0.08	3.74	4.23
Total	746	746	746	746	746
MLE		$\theta = 0.5954$	$\pi = 0.1420$	$\beta = 0.2135$ $k = 0.6382$	$\beta = 0.2098$ $v = -0.2487$
Loglik		562.23	579.88	550.78	550.49
X^2		26.611	16.540	2.2221	1.6246
(P-value)		(<0.01)	(<0.01)	(0.33)	(0.44)
G^2		19.7798	16.9514	1.8921	1.6073
df		2	1	2	2
AIC		1126.46	1161.75	1105.57	1104.99

Comment on table 4.5: Zero-truncated Binomial and Zero-truncated Poisson (ZTP) failed to fits the data here. The Zero-truncated Multiplicative-Binomial (ZTMB) and ZTCB distributions fit the data well with the proposed distribution having an edge over ZTMB distribution.

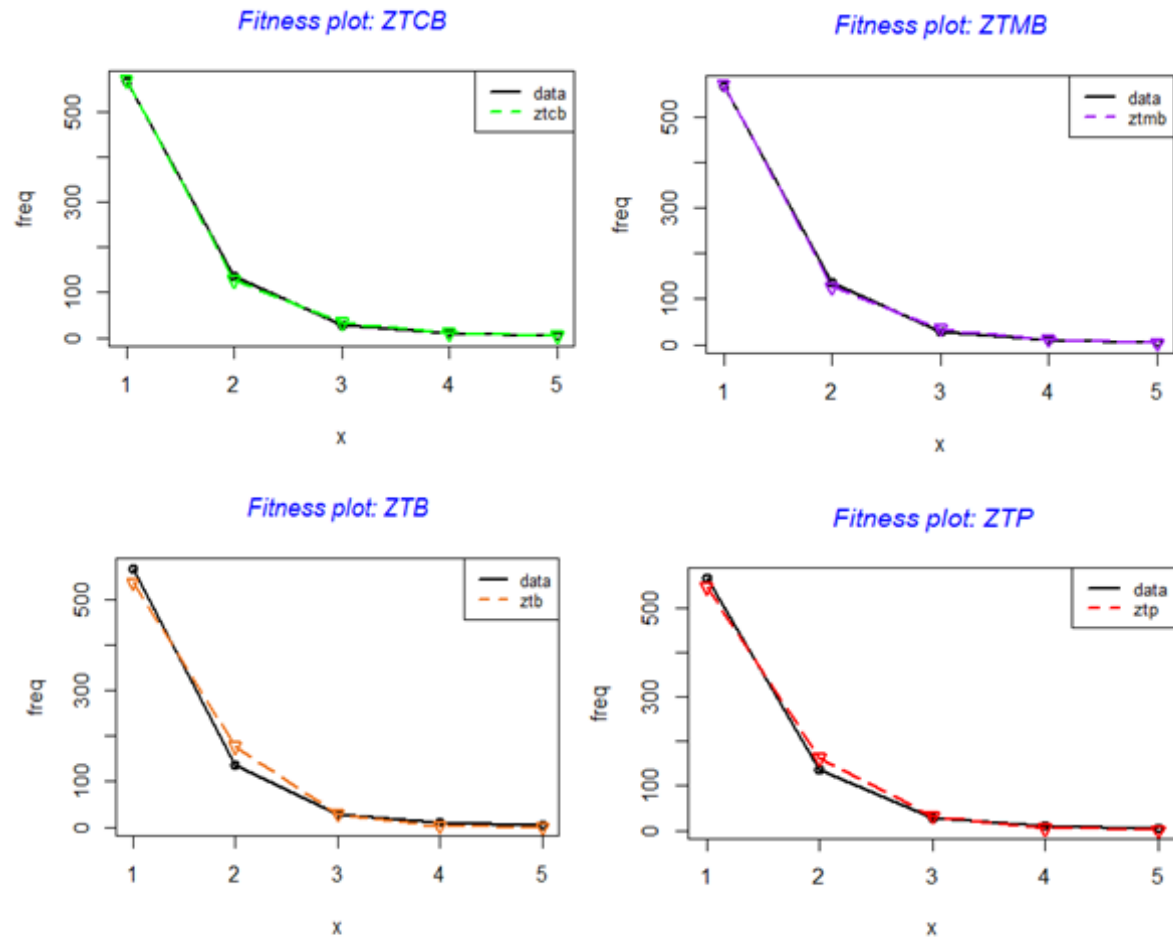


Figure 4.5a: the plot of observed and expected frequency obtained by each distribution.

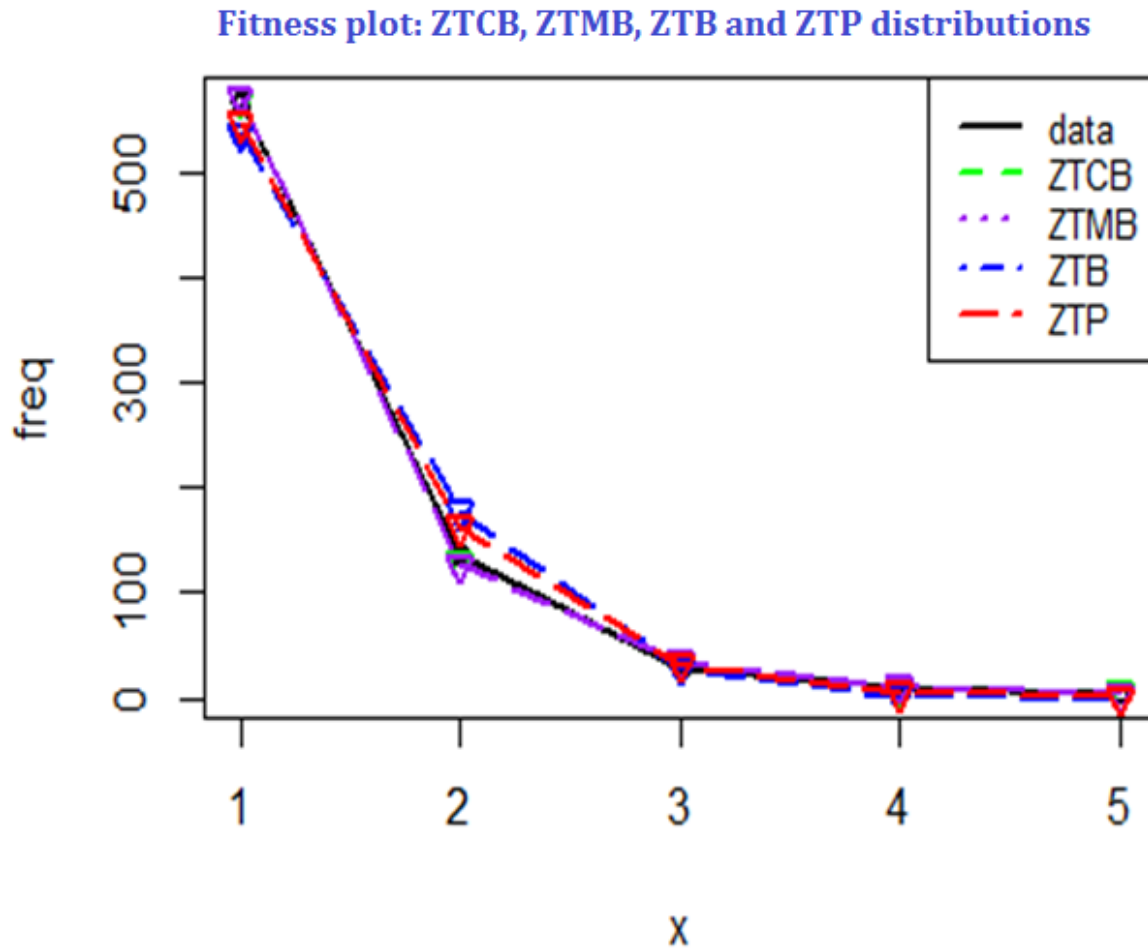


Figure 4.5b: the plot of observed and expected frequency obtained by each distribution.

In Figures 4.5a and 4.5b, it is clearly shown that the close association between the observed counts and the expected on the basis of ZTCB and ZTMB distributions. The fits are superior to that of ZTB and ZTP distributions.

4.6 Example 4.6: [Mortality data]

The data, Table 4.6, are reproduced from NHS 2008, Nigeria: representing counts of mothers that experienced at least one dead child within twelve months preceding data collection. The sample mean and variance are 2.061 and 2.145, respectively.

Table 4.6: The expected and observed number of mothers of the complete fertility that experienced at least one child death

Y	Obs.Freq	ZTP	ZTB	ZTMB	ZTCB
1	4770	3772.7	3603.7	4550.3	4724.5
2	2388	3159.5	3299.1	2664.3	2449.3
3	1237	1764.0	1869.7	1322.1	1264.6
4	677	738.6	733.6	621.8	650.9
5	350	247.4	211.1	293.6	334.1
6	177	69.1	46.0	143.8	171.1
7	97	16.5	7.7	74.3	87.4
8	37	3.5	1.0	40.9	44.6
9	20	0.6	0.1	24.0	22.7
10	12	0.1	0	14.9	11.5
11	1	0.0	0	9.5	5.8
12	3	0.0	0	6.2	2.9
13	2	0.0	0	3.8	1.5
14	0	0.0	0	1.9	0.7
15	1	0.0	0	0.6	0.4
Total	9772	9772	9772	9772	9772
MLE		$\theta = 1.6749$	$\pi = 0.1157$	$\beta = 0.3551$ $k = 0.8547$	$\beta = 0.3376$ $v = 0.0085$
Loglik		14623.28	15026.18	13990.38	13950.54
χ^2		2445.82	4375.83	86.90	8.91
(P-value)		(<0.01)	(<0.01)	(<0.01)	(0.35)
G^2		1264.258	1862.118	89.8209	9.2148
df		6	5	8	8
AIC		29248.56	30054.35	27984.76	27905.09

Comment on table 4.6: Only Zero-truncated Com-Binomial fits the data here. Other distributions failed to fit the data at all.

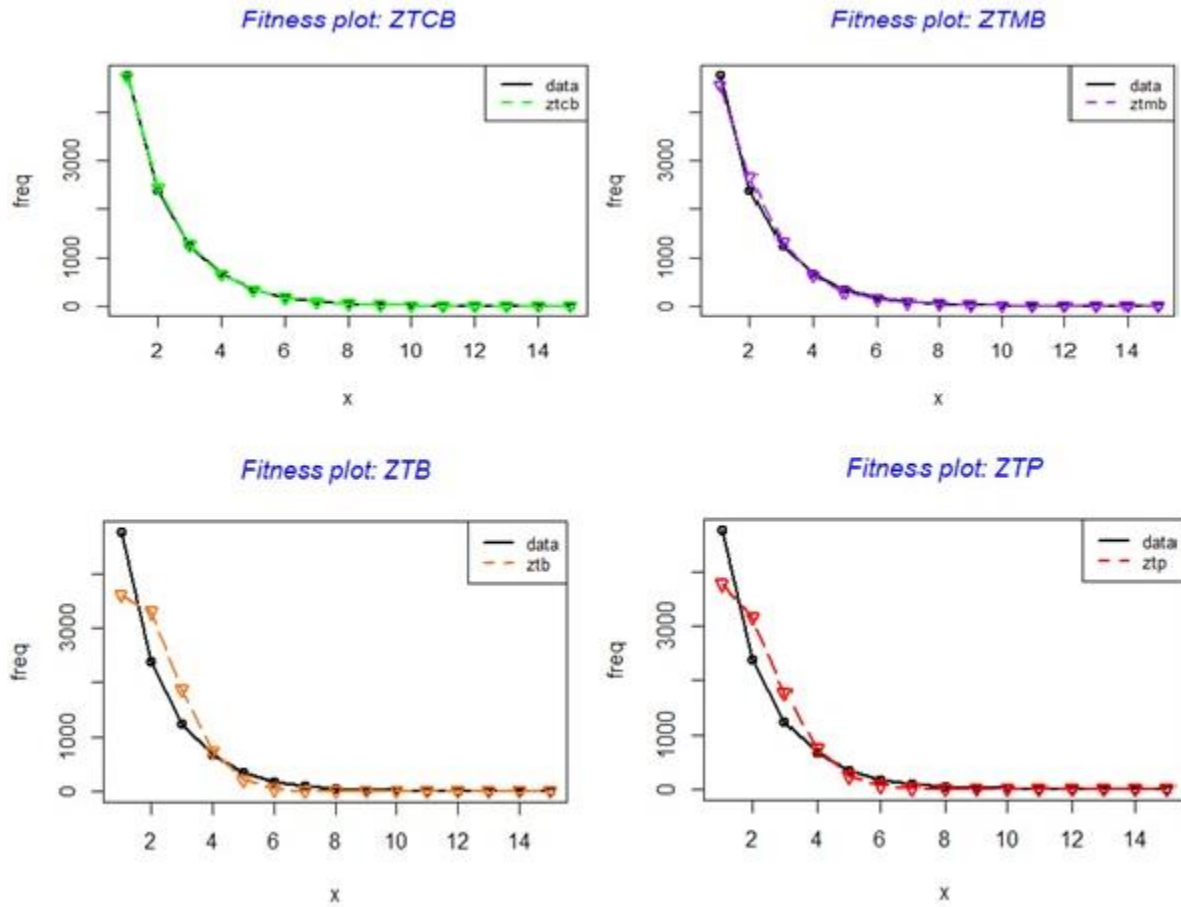


Figure 4.6a: the plot of observed and expected frequency obtained by each distribution.

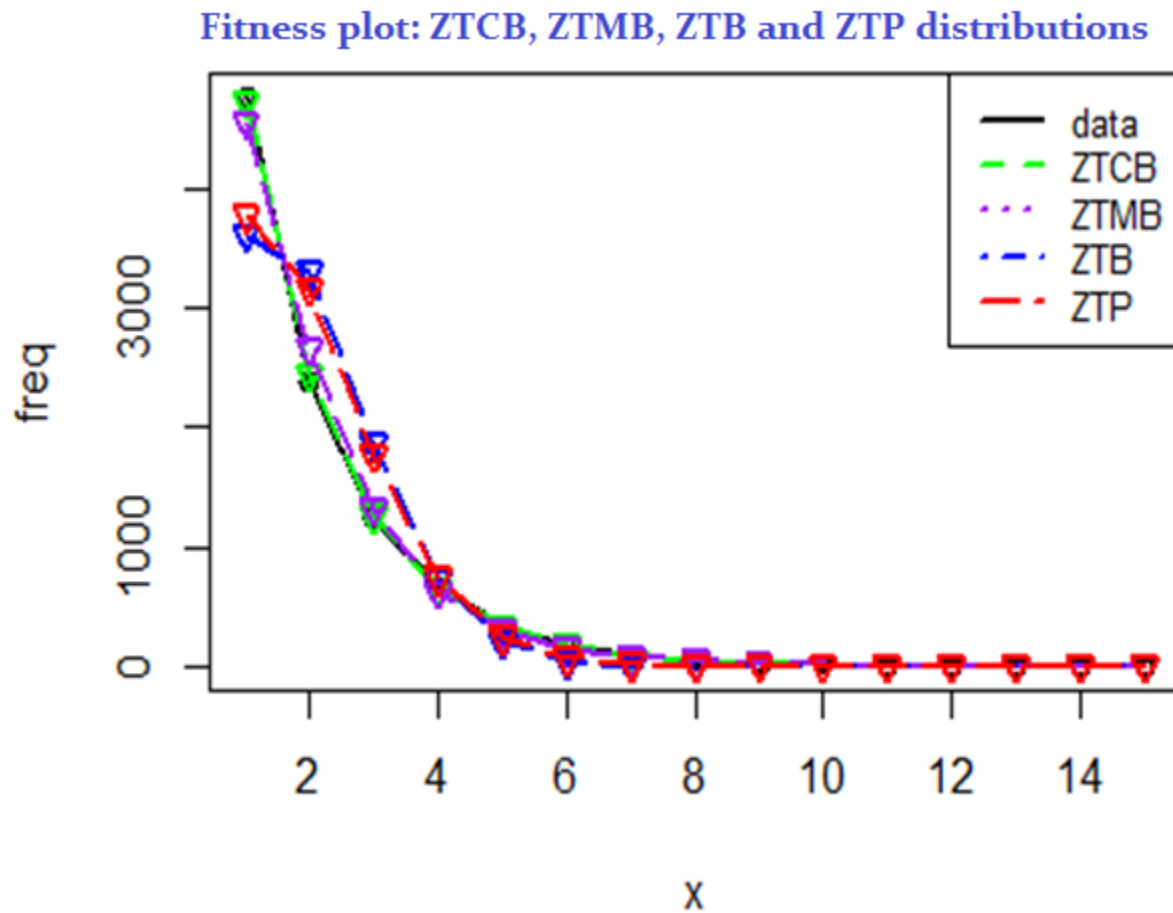


Figure 4.6b: the plot of observed and expected frequency obtained by each distribution.

In Figures 4.6a and 4.6b, it is clearly shown that the close association between the observed counts and the expected on the basis of ZTCB distributions. The fits are superior to that of ZTMB, ZTB and ZTP distributions.

4.7 General Remark

From the example 1 through to 6 (in table 1 to 6), the proposed ZTCB distribution fits the datasets from different fields well while the ZTMB distribution performs well in table 1 to 5 but fails to fit example 6 data. Obviously, in the datasets where ZTMB perform well, the variances are less than means. In example 6 where variance is not less than mean, it failed woefully! The examples illustrate the flexibility of the proposed ZTCB distribution, an edge on other existing distributions, which is one of the major contribution of this work.

4.8 Summary of the Analyses

Table 4.7: summary of comparative of goodness of fit of ZTCB and other selected existing distributions

	n	N	Log-lik	X ² (P-value)	G ²	Df	AIC
immunogold data	5	198					
ZTP			205.948	2.329 (0.51)	2.115	3	413.89
ZTB			213.882	5.518 (0.06)	5.776	3	429.76
ZTMB			203.719	0.529 (0.91)	0.540	2	411.44
ZTCB			203.380	0.423 (0.94)	0.472	2	410.76
counts of flower heads	9	88					
ZTP			166.55	6.942 (0.23)	6.629	4	335.1
ZTB			177.17	19.386 (<0.01)	21.950	4	356.34
ZTMB			163.85	1.524 (0.91)	1.539	3	331.7
ZTCB			163.63	0.568 (0.97)	0.576	3	331.25
Mortality data (ex3)	5	522					
ZTP			352.11	3.666 (0.06)	3.512	1	706.21
ZTB			358.25	9.360 (<0.01)	9.139	1	718.5
ZTMB			348.92	0.048 (0.82)	0.049	1	701.84
ZTCB			348.90	0.037 (0.85)	0.037	1	701.80
Mortality data (ex4)	5	873					
ZTP			615.4	8.706 (0.013)	8.962	2	1232.79

ZTB			634.09	20.088 (<0.01)	20.307	1	1270.17
ZTMB			601.55	1.890 (0.39)	1.599	2	1207.11
ZTCB			601.28	1.324 (0.52)	1.323	2	1205.78
Mortality data (ex5)	5	746					
ZTP			562.23	26.611 (<0.01)	19.780	2	1126.46
ZTB			579.88	16.540 (<0.01)	16.951	1	1161.75
ZTMB			550.78	2.222 (0.33)	1.892	2	1105.57
ZTCB			550.49	1.625 (0.44)	1.607	2	1104.99
Mortality data (ex6)	15	9772					
ZTP			14623.3	2445.82 (<0.01)	1264.258	6	29248.6
ZTB			15026.2	4375.83 (<0.01)	1862.118	5	30054.4
ZTMB			13990.4	86.90 (<0.01)	89.821	8	27984.8
ZTCB			13950.5	8.91 (0.35)	9.215	8	27905.1

The ZTCB distribution is characterized by two parameters. The maximum likelihood method is used to obtain the estimators of the parameters through R-software. The statistics, X^2 , G^2 and the AIC show that the proposed ZTCB distribution yields best result among the distributions under consideration as shown in the summary table above (table 4.7).

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS.

5.1 Conclusion

In this study we assessed the background of issues in modeling count data. Out of the challenges is the issue of zeros (truncated or excess). We also investigated the effectiveness of some binomial based distributions for modeling count data. Binomial mixture distributions; such as Beta-Binomial, Generalized Beta-Binomial, etc, were studied in handling Binomial over-dispersion. Through real life data, applications of the distributions thorough the maximum likelihood estimation show that the Com-Binomial (CB) distribution yields a better fit and good improvement in the goodness of fit tests than BB, GBB and MB distributions, which equally fit the data well in that order. Of the distribution, Multiplicative Binomial and Com-Binomial distributions are prominently stands out.

Moreover, we derived and discussed Zero-Truncated Com-Binomial distribution and its properties. For $\nu \rightarrow \infty$, the pmf is peak around the point $n\pi$ and it has two tails; for $\nu \rightarrow -\infty$, it peaks around 1 or n and it as well resembles exponential distribution. The probability generating function and the first two moments were derived and presented.

Finally, we examined the application of the Com-Binomial distribution to real life data. We as well compared its performance with other selected existing distributions.

Obviously, Zero-Truncated Com-Binomial is flexible to handle all level of dispersion. The ZTCB distribution is the generalization of the Zero-Truncated Com-Poisson, Zero-Truncated Poisson and Zero-Truncated Binomial distributions.

This proposed model is more appropriate than ZTB, ZTMB and other distributions as seen in the examples. Some of the datasets here have been studied previously with ZTP, ZTB and ZTMB and now introducing ZTCB. The goodness-of-fit statistics show that the proposed ZTCB distribution yields better results than the existing distributions under consideration.

5.2 Recommendation

This work is recommended particularly to model the structurally truncated count data with various dispersions.

5.3 Contribution to Knowledge

A flexible zero-truncated distribution have been developed to capture the dynamics of structurally zero-truncated count data. The distribution performs well in the presence of a very heavy tail count data. This work obviously provides useful alternative to the existing distributions.

5.4 Suggestion for Further Studies

The following areas could serve as objects for further studies:

- i. More attention is needed to investigate the mathematical properties of the ZTCB distribution extensively, such as other moments.
- ii. Developing a regression model with this distribution would be an important piece of research work too.

REFERENCE

- Aderoju, S.A and Jolayemi, E.T. (2018). Zero-Truncated mass function of Com-Binomial Distribution: Its development and application. *Journal of Global Research in Mathematical Archives (JGRMA)*, [Accepted]
- Aderoju, S.A, Jolayemi, E.T and Ibrahim, A.O. (2017). A Comparative Study on Zero-truncated Generalized Poisson-Lindley and Zero-truncated Poisson-Lindley Distributions. *International journal of Mathematical Archive (IJMA)* - 8(3), 92-98. Available online through www.ijma.info ISSN 2229 – 5046.
- Agresti, A., (2002). Categorical Data Analysis; John Wiley & Sons, Inc., Hoboken, New Jersey
- Altham, P. (1978). Two generalizations of the binomial distribution. *Applied Statistics*, 27, 162-167.
- Altham, P.M.E (2000). An aspect of discrete data analysis: fitting a beta-binomial distribution to the hospitals data. *Speech at “The design and analysis of potency Assays for Biotechnology Products” at the National Institute for Biological Standards and Control, UK.*
- Bernoulli, J. (1713). *Ars Conjectandi*, Basilea: Thurnisius. [11.2.19]
- Borges P, Rodrigues J and Balakrishnan N. (2014). A COM-Poisson type generalization of the Binomial distribution and its properties and applications. *Statistics and Probability Letters* 87. Elsevier B.V. pages158–166.

- Boswell, M. T., and Patil, G. P. (1970). Chance mechanisms generating the negative binomial distributions, *Random Counts in Scientific Work*, Vol. 1: *Random Counts in Models and Structures*, 3–22. University Park: Pennsylvania State University Press.
- Boyer, C. B. (1950). Cardan and the Pascal triangle, *American Mathematical Monthly*, **57**, 387–390.
- Castillo, J. and Perez-Casany, M. (1998). Weighted Poisson distributions for overdispersion and under-dispersion situations. *Annals of the Institute of Statistical Mathematics*, 50, 567-585.
- Castillo, J. and Perez-Casany, M. (2005). Over-dispersed and underdispersed Poisson generalizations. *Journal of Statistical Planning and Inference*, 134, 486-500.
- Cochran, W. G. (1954). Some methods for strengthening the common chi-square tests, *Biometrics*, 10, pp. 417-451.
- Conway, R. W. and Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12:132-6.
- David, F. N., and Johnson, N. L. (1952). The truncated Poisson distribution, *Biometrics*, **8**, 275–285.
- Diniz, C.A.R, Tutia, M.H and Leite, J.G (2010). Bayesian analysis of a correlated binomial model. *Brazilian Journal of Probability and Statistics* 2010, Vol. 24, No. 1, 68–77

- Edwards, A. W. F. (1987). *Pascal's Arithmetical Triangle*, London: Griffin.
- Elamir E.A.H, (2011). Multiplicative-Binomial Distribution: Some Results on Characterization, Inference and Random Data Generation. *Journal of Statistical Theory and Applications*, Vol. 12, No. 1,92-105
- Engel, B. and Brake, J. (1993) Analysis of embryonic development with a model for under-dispersion or over-dispersion relative to binomial variation, *Biometrics*, 49, pp. 269-279.
- Finney, D. J. (1949). The truncated binomial distribution. *Ann. Eugen.*14, 319
- Finney, D.J. and Varley, G.C. (1955). An example of the truncated Poisson distribution. *Biometrics*, 11, 387–394.
- Gart, J. J. (1974). Exact moments of the variance test for left-truncated Poisson distributions, *Sankhyā, Series B*, **36**, 406–416.
- Ghitany M.E, Al-Mutairi D.K and Nadarajah S. (2008) Zero-truncated Poisson-Lindley distribution and its Applications. *Mathematics and Computers in Simulation* 79(3): 279-287.
- Hald , A . (1990) . *A History of Probability and Statistics and Their Applications before 1750*, New York: Wiley.
- Hilbe, J.M. (2011). *Negative Binomial Regression (second edition)*. Cambridge University Press: Cambridge.
- Hilbe, J.M. (2014). *Modeling Count Data*. Cambridge University Press

- Johnson et al (2005), *Univariate discrete distribution*, John Wiley & Sons, Inc., Hoboken, New Jersey
- Johnson, N. L., and Kotz, S. (1969). *Discrete Distributions* (first edition), Boston: Houghton Mifflin.
- Jolayemi, E.T.(1990). Model selection for One-Dimensional Multinomials. *Biom.J.*32: 827-834.
- Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 46(1), 79-88.
[http://dx.doi.org/10.1016/0022-1694\(80\)90036-0](http://dx.doi.org/10.1016/0022-1694(80)90036-0)
- Kupper, L. L. and Haseman, J. K. (1979) Analysis of dichotomous response data from certain toxicological experiments, *Biometrics*, 35, pp. 281-293.
- Kupper, L.L. Portier, C. Hogan M.D and Yamamoto, E. (1986). The impact of litter effects on dose-response modeling in teratology, *Biometrics*, 42, 85-98.
- Lawal, B and Famoye, F (2013). *Applied Statistics: Regression and Analysis of variance*, University Pression of America, Inc.
- Li, X. H., Huang, Y. Y., and Zhao, X. Y. (2011). The Kumaraswamy Binomial Distribution. *Chinese Journal of Applied Probability and Statistics*, 27(5), 511-521
- Lindsey, J. (1995). *Modelling Frequency and Count Data*. Oxford: Oxford University Press.

- Lindsey, J.K and Altham, P.M.E. (1998): Analysis of the human sex ratio using over-dispersion models. *App Statist.* 47; 149-157
- Lovison, G. (1998). An alternative representation of Altham's multiplicative-binomial distribution. *Statistics & Probability Letters*, **36**, 415-420.
- Luceño, A. and Ceballos, F. (1995). Describing extra-binomial variation with partially correlated models. *Communications in Statistics. Theory and Methods* **24** 1637–1653.
- Manoj .C, Wijekoon P. and Yapa R.D. (2013). The McDonald Generalized Beta-Binomial Distribution: A New Binomial Mixture Distribution and Simulation Based Comparison with Its Nested Distributions in Handling Over-dispersion. *International Journal of Statistics and Probability*; Vol. 2, No. 2;
- Matthews, J.N.S. and Appleton, D.R. (1993). An application of the truncated Poisson distribution to immunogold assay. *Biometrics*, 49, 617–621
- Nelder, J., and Wedderburn. R.W, (1972). Generalized linear models. *J. Roy. Statist. Soc.Ser.A* **135**: 370384.
- Pack, S.E (1986). Hypothesis testing for proportion with over-dispersion, *Biometrics*, 42, 967-972.
- Paul, S.R (1982). Analysis of proportions of affected fetuses in teratology experiments, *Biometrics*, 38,361-370.
- Poisson, S. D. (1837). *Recherches sur la Probabilit´e des Jugements en Mati`ere*

Criminelle et en Matière Civile, Précédées des Regles Générales du Calcul des Probabilités, Paris: Bachelier, Imprimeur-Libraire pour les Mathématiques, la Physique, etc.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org>.

Risko, K.J and Margolin, B.H (1996). Some observations on detecting extra-binomial variability within the beta binomial model. In: B. J. T. MORGAN (Ed.), *Statistics in Toxicology*, pp. 57-65 (Oxford, Clarendon Press).

Sampford, M.R, (1955). The Truncated Negative Binomial Distribution: *Biometrika*, Vol. 42, pp. 58-69, URL:<http://www.jstor.org/stable/2333422>

Scollnik, D. P. M. (1997). Inference concerning the size of the zero class from an incomplete Poisson sample, *Communications in Statistics—Theory and Methods*, **26**, 221–236.

Shanker R, Hagos Fesshaye, Sujatha Selvaraj and Abrehe Yemane (2015), On Zero-Truncation of Poisson and Poisson-Lindley Distributions and Their Applications, *Biometrics & Biostatistics International Journal*, Vol 2, Issue 6.

Shmueli G., Borle S., and Boatwright P. (2005). A useful distribution for fitting discrete

- data: Revival of the Conway Maxwell-Poisson distribution. *Applied Statistics* ;54:127-42.
- Sokal, R. R. and Rohlf, F.J (1995). *Biometry: the principles and practice of statistics in biological research, 3rd ed.* Freeman: New York. 887 pp.
- Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA: Harvard University Press.
- Tarone, R.E. (1979). Testing the goodness of fit of the binomial distribution. *Biometrika*; 66 (3): 585-590. DOI: <https://doi.org/10.1093/biomet/66.3.585>
- Thomas D.G. and Gart J.J. (1971). Small sample performance of some estimators of the truncated binomial distribution. *Journal of the American Statistical Association*, 66, 169-175.
- Tripathi, R. C., Gupta, R. C., and Gurland, J. (1994). Estimation of parameters in the beta binomial model, *Annals of the Institute of Statistical Mathematics*, **46**, 317–331.
- Wijekoon, M.P and Yapa. R.D (2013). The McDonald Generalized Beta-Binomial Distribution: A New Binomial Mixture Distribution and Simulation Based Comparison with Its Nested Distributions in Handling Over-dispersion. *International Journal of Statistics and Probability*; Vol. 2, No. 2;
- Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity, *Biometrics*, 31, pp. 949-952.

Zuur A.F., Leno E.N., Walker N.J., Saveliev A. and Smith G.M. (2009). *Mixed effects models and extensions in ecology with R*. Statistics for Biology and Health, Springer Science 1-33.

APPENDIX A

Com-Poisson Model (Parameters) Estimation

Let $y_1, y_2, \dots, y_n$ be identically and independently distributed observations from Com-Poisson distribution, such that

$$f(y; \pi, v) = \frac{\pi^{y_i}}{(y_i!)^v Z(\pi, v)} \quad \pi > 0, \quad v \geq 0, \quad y_i = 0, 1, 2, \dots \text{ and } i = 1, 2, \dots, n$$

$$\text{where } Z(\pi, v) = \sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}$$

The likelihood function of Conway-Maxwell-Poisson probability distribution function is

$$L(\pi; y_i) = \prod_{i=1}^n \frac{\pi^{y_i}}{(y_i!)^v Z(\pi, v)}$$

Log-likelihood function, $\mathfrak{f}$, is

$$\mathfrak{f} = \ln(\pi) \sum_{i=1}^n y_i - v \sum_{i=1}^n \ln(y_i!) - \sum_{i=1}^n \ln Z(\pi, v) \quad \dots \dots (i)$$

Differentiating equation (i) partially with respect to the parameters we have

$$\frac{\partial \mathfrak{f}}{\partial \pi} = \frac{1}{\pi} \sum_{i=1}^n y_i - \frac{\partial}{\partial \pi} \sum_{i=1}^n \ln Z(\pi, v) = 0$$

However, let $\sum_{i=1}^n \ln Z(\pi, v)$, $Z = \sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}$, such that $\frac{\partial Q}{\partial \pi} = \frac{\partial Q}{\partial Z} \cdot \frac{\partial Z}{\partial \pi}$ (i.e chain rule)

Therefore, $\frac{\partial}{\partial \pi} \sum_{i=1}^n \ln Z(\pi, v) = \frac{\partial Q}{\partial \pi}$

$$= \frac{\partial Q}{\partial Z} \cdot \frac{\partial Z}{\partial \pi}$$

$$= \frac{1}{Z(\pi, v)} \cdot \sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v}$$

$$\text{So, } \frac{\partial f}{\partial \pi} = \frac{1}{\pi} \sum_{i=1}^n y_i - \frac{\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} = 0$$

$$\sum_{i=1}^n y_i = \pi \frac{\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \quad \dots \dots \dots (ii)$$

To obtain second derivative with respect to π , I differentiate equation (ii) above:

$$\frac{\partial^2 f}{\partial \pi^2} = -\frac{1}{\pi^2} \sum_{i=1}^n y_i - \frac{\partial}{\partial \pi} \left[\frac{\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \right]$$

Applying the quotient rule we have,

$$\frac{\partial^2 f}{\partial \pi^2} = -\frac{1}{\pi^2} \sum_{i=1}^n y_i - \left\{ \frac{\sum_{h=2}^{\infty} \frac{h(h-1)\pi^{h-2}}{(h!)^v} - \left[\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v} \right]^2}{\left[\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v} \right]^2} \right\}$$

$$\frac{\partial^2 f}{\partial \pi^2} = -\frac{1}{\pi^2} \sum_{i=1}^n y_i - \frac{\sum_{h=2}^{\infty} \frac{h(h-1)\pi^{h-2}}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} + \left[\frac{\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \right]^2 = 0$$

$$\sum_{i=1}^n y_i = \pi^2 \left\{ \left[\frac{\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \right]^2 - \frac{\sum_{h=2}^{\infty} \frac{h(h-1)\pi^{h-2}}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \right\} \quad \dots \dots \dots (iii)$$

Similarly, the first and second derivatives of the log-likelihood with respect to v are as follow:

$$\frac{\partial f}{\partial v} = - \sum_{i=1}^n \ln(y_i!) - \frac{\partial}{\partial v} \sum_{i=1}^n \ln Z(\pi, v) = 0$$

$$\text{Let } = \sum_{i=1}^n \ln Z(\pi, v), \quad Z = \sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}, \text{ such that } \frac{\partial Q}{\partial v} = \frac{\partial Q}{\partial Z} \cdot \frac{\partial Z}{\partial v} \text{ (i.e chain rule)}$$

$$\frac{\partial f}{\partial v} = - \sum_{i=1}^n \ln(y_i!) - \frac{1}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \cdot \sum_{h=2}^{\infty} \frac{-\pi^h \ln(h!)}{(h!)^v}$$

Equating to zero;

$$\sum_{i=1}^n \ln(y_i!) = \frac{\sum_{h=2}^{\infty} \frac{-\pi^h \ln(h!)}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \quad (\text{iv})$$

Furthermore, the second derivative is as follow;

$$\begin{aligned} \frac{\partial^2 f}{\partial v^2} &= - \frac{\partial}{\partial v} \left[\frac{\sum_{h=2}^{\infty} \frac{-\pi^h \ln(h!)}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \right] \\ &= - \left[\frac{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v} \left[\sum_{h=2}^{\infty} \frac{\pi^h \ln(h!)^2}{(h!)^v} \right] - \left[\sum_{h=2}^{\infty} \frac{-\pi^h \ln(h!)}{(h!)^v} \right]^2}{\left[\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v} \right]^2} \right] \\ &= \frac{\left[\sum_{h=2}^{\infty} \frac{-\pi^h \ln(h!)}{(h!)^v} \right]}{\left[\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v} \right]} - \frac{\sum_{h=2}^{\infty} \frac{\pi^h \ln(h!)^2}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \quad (\text{v}) \end{aligned}$$

Moreover, here is the partial derivative;

$$\begin{aligned}
\frac{\partial^2 \mathfrak{f}}{\partial v \partial \pi} &= -\frac{\partial}{\partial v} \left[\frac{\sum_{h=2}^{\infty} \frac{-\pi^h \ln(h!)}{(h!)^v}}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} \right] \\
&= \frac{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v} \left[\sum_{h=2}^{\infty} \frac{h\pi^{h-1}}{(h!)^v} \ln(h!) \right] - \sum_{h=2}^{\infty} \frac{\pi^h}{(h!)^v} \ln(h!) \left(\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v} \right)}{\left(\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v} \right)^2} \\
&= \frac{\sum_{h=2}^{\infty} \frac{h\pi^{h-1}}{(h!)^v} \ln(h!)}{\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v}} - \frac{\left(\sum_{h=1}^{\infty} \frac{h\pi^{h-1}}{(h!)^v} \right) \left(\sum_{h=2}^{\infty} \frac{\pi^h}{(h!)^v} \ln(h!) \right)}{\left(\sum_{h=0}^{\infty} \frac{\pi^h}{(h!)^v} \right)^2} \quad (\text{vi})
\end{aligned}$$

APPENDIX B

Some R commands/codes used

```
#####Zero-truncated Com-Binomial [MLE of parameters  $p$  and  $v$ ] #####
fn=function(par,x,n)
{
  p=par[1]
  v=par[2]
  sm=function(x,n,p,v)
  {
    sx=c()
    for(j in 1:length(x))
    {
      s=0
      for(i in 1:n)
      {
        temp=((choose(n,i)^v)*(p^i)*(1-p)^(n-i))
        s=s+temp
      }
      sx[j]=s
    }
    sx
  }

  ll=v*log(choose(n,x))+x*log(p)+(n-x)*log(1-p)-log(sm(x,n,p,v))
  -sum(ll)
}

rstr <- optim(c( , ),fn,method="L-BFGS-B",lower=c( , ),upper=c(1,Inf),x=fd,n=n)
cat(rstr$par,"= MLE of parameters  $p$  and  $v$  respectively\n")

AIC= -2*(-rstr$value)+2*p
BIC= -2*(-rstr$value)+2*log(n)

#### the probability density
dcmbr=function(x,n,par)
{
  p=par[1]
  v=par[2]
  sm=function(x,n,p,v)
```

```

    {
s=0
for(i in 1:n)
    {
temp=((choose(n,i)^v)*(p^i)*(1-p)^(n-i))
s=s+temp
    }
s
    }
ll=v*log(choose(n,x))+x*log(p)+(n-x)*log(1-p)-log(sm(x,n,p,v))
exp(ll)
    }

### Probability of the distribution
cg=c()
for(x in 1:n){cg[x]=dcmblr(x,n,c(rstr$par[1],rstr$par[2]))}
cat(cg,"= Prob(x= 1, 2, ..., n) respectively\n")
expcmb=cg*N
expcmb
cat(expcmb,"= Expected freq of (x= 1, 2, ..., n) respectively\n")
chisq <- sum(((f-expcmb)^2)/expcmb)
chisq
pval <- 1-pchisq(chisq,df)
pval

#####
exp=c()
ff=c()
cchisq <- sum(((ff-exp)^2)/exp)
cchisq
cpval <- 1-pchisq(cchisq,df)
cpval

D=2*sum(ff*log(ff/exp))
D

y=c(1:n)
exp=c()
da <-data.frame(y,f)
ztc=da.frame(y,exp)

```

```

plot_colors <- c("black","red","chocolate2","purple","green","royalblue","turquoise2")
plot(da, xlab="x", ylab="freq",type="o",pch=21, lty=1, lwd=2,col=plot_colors[1]);lines(z_tcb, type="o",
pch=25, lty=5,lwd=2, col=plot_colors[5]);title(main="Fitness plot: ZTCB", col.main="blue",
font.main=3);legend("topright", g_range[5],c("data","z_tcb"), cex=0.8, col=plot_colors[c(1,5)], lty=1:7, lwd=2)

```

###PROPERTIES OF ZTCB DISTRIBUTION#####

```

dcmbtr=function(x,n,par)
{
p=par[1]
v=par[2]

sm=function(x,n,p,v)
{
s=0
for(i in 1:n)
{
temp=((choose(n,i)^v)*(p^i)*(1-p)^(n-i))
s=s+temp
}
s
}
ll=v*log(choose(n,x))+x*log(p)+(n-x)*log(1-p)-log(sm(x,n,p,v))
exp(ll)
}
library(lattice)
require(grDevices)

n=7
x=c(1:n)
par=c(0.75, -20)
y1=dcmbtr(x,n,par)
plot(x, y1, ylim=c(0,1), type="l", lty=2, xlab="y value",ylab="f(y)", main=substitute(paste(italic("v=-
20'))));lines(x,y1, lwd=2, col="red")

x=c(1:n)

```

```

par=c(0.75, -5)
y1=dcmbr(x,n,par)
plot(x, y1, ylim=c(0,1), type="l", lty=2, xlab="y value",ylab="f(y)", main=substitute(paste(italic("v=-
5'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)

par=c(0.75, 0.25)
y1=dcmbr(x,n,par)
plot(x, y1,ylim=c(0,1), type="l", lty=2, xlab="y value",ylab="f(y)",
main=substitute(paste(italic("v=0.25'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)

par=c(0.75, 0.75)
y1=dcmbr(x,n,par)
plot(x, y1,ylim=c(0,1), type="l", lty=2, xlab="y value",ylab="f(y)",
main=substitute(paste(italic("v=0.75'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)

par=c(0.75, 0.5)
y1=dcmbr(x,n,par)
plot(x, y1,ylim=c(0,1), type="l", lty=2, xlab="y value",ylab="f(y)",
main=substitute(paste(italic("v=0.5'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)

par=c(0.75, 1)
y1=dcmbr(x,n,par)
plot(x, y1, ylim=c(0,1),type="l", lty=2, xlab="y value",ylab="f(y)",
main=substitute(paste(italic("v=1'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)

par=c(0.75, 1.25)
y1=dcmbr(x,n,par)
plot(x, y1, ylim=c(0,1),type="l", lty=2, xlab="y value",ylab="f(y)",
main=substitute(paste(italic("v=1.25'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)

par=c(0.75, 5)
y1=dcmbr(x,n,par)
plot(x, y1,ylim=c(0,1), type="l", lty=2, xlab="y value",ylab="f(y)",
main=substitute(paste(italic("v=5'))));lines(x, y1, lwd=2, col="red")

```

```

par=c(0.75, 10)
y1=dcmbr(x,n,par)
plot(x, y1,ylim=c(0,1), type="l", lty=2, xlab="y" value",ylab="f(y)",
main=substitute(paste(italic("v=10'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)
par=c(0.75, 20)
y1=dcmbr(x,n,par)
plot(x,y1,ylim=c(0,1), type="l", lty=2, xlab="y" value",ylab="f(y)",
main=substitute(paste(italic("v=20'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)
n=n
par=c(0.75, 25)
y1=dcmbr(x,n,par)
plot(x, y1,ylim=c(0,1), type="l", lty=2, xlab="y" value",ylab="f(y)",
main=substitute(paste(italic("v=25'))));lines(x, y1, lwd=2, col="red")

x=c(1:n)
n=n
par=c(0.75, 30)
y1=dcmbr(x,n,par)
plot(x, y1,ylim=c(0,1), type="l", lty=2, xlab="y" value",ylab="f(y)",
main=substitute(paste(italic("v=30'))));lines(x, y1, lwd=2, col="red")

```

```

# ##### ZERO-TRUNCATED Multiplicative binomial#####
#loglikelihood fn
fmbtr=function(par,x,n)
{
p=par[1]
k=par[2]

sm=function(x,n,p,k)
{

```

```

sx=c()
for(j in 1:length(x))
{
s=0
for(i in 1:n)
{
temp=((choose(n,i))*(p^i)*((1-p)^(n-i)))*k^(i*(n-i))
s=s+temp
}
sx[j]=s
}
sx
}

ll=log(choose(n,x))+x*log(p)+(n-x)*log(1-p)+(x*(n-x))*log(k)-log(sm(x,n,p,k))
-sum(ll)
}
#MLE of the parameters (optimizer) ZTMB
rsmblr=optim(c( , ), fnmbtr,method="L-BFGS-B",lower=c( , ),upper=c(1,Inf),x=fd,n=n)
rsmblr
cat(rsmblr$par,"= MLE of parameters p and k\n")

AIC= -2*(-rsmblr$value)+2*2
AIC

BIC= -2*(-rsmblr$value)+2*log(n)
BIC

#MB density fn
dmbtr=function(x,n,par)
{
p=par[1]
k=par[2]

sm=function(x,n,p,k)
{
s=0
for(i in 1:n)
{
temp=((choose(n,i))*(p^i)*((1-p)^(n-i)))*k^(i*(n-i))
s=s+temp

```

```

}
s
}
ll=log(choose(n,x))+x*log(p)+(n-x)*log(1-p)+(x*(n-x))*log(k)-log(sm(x,n,p,k))
exp(ll)
}

### Probability of the distribution
cg=c()
for(x in 1:n){cg[x]=dmbtr(x,n,rmbtr,$par)}
cat(cg,"= Prob(x= 1, 2, ..., n) respectively\n")
exp=cg*N
exp
cat(exp,"= expected freq of(x= 1, 2, ..., n) respectively\n")
chisq <- sum(((f-exp)^2)/exp)
chisq
pval <- 1-pchisq(chisq,df)
pval

D=2*sum(ff*log(ff/ cexp))
D

y=c(1:n)
cexp=c()
da <-data.frame(y,f)
ztmb=data.frame(y,cexp)

plot_colors <- c("black","red","chocolate2","purple","green","royalblue","turquoise2")
plot(da, xlab="x", ylab="freq",type="o",pch=21, lty=1, lwd=2,col=plot_colors[1]);lines(ztmb, type="o",
pch=25, lty=5,lwd=2, col=plot_colors[4]);title(main="Fitness plot: ZTMB", col.main="blue",
font.main=3);legend("topright", g_range[5],c("data","ztmb"), cex=0.8, col=plot_colors[c(1,4)], lty=1:n, lwd=2)

```

```

# ##### ZERO-TRUNCATED Binomial#####
fnztb=function(par,x,n)
{
p=par[1]
-sum(dbinom(x,n,par,log=T)-log(1-(1-p)^n))
}
rsztb <- optim(c(.),fnztb,method="Brent",lower=c(.),upper=c(1),x=fd,n)
rsztb

AIC= -2*(-rsztb$value)+2*1
AIC

BIC= -2*(-rsztb$value)+1*log(n)
BIC

#density function
dztb=function(x,n,par)
{
p=par[1]

ll=(dbinom(x,n,par,log=T)-log(1-(1-p)^n))
exp(ll)
}

## Probability of the distribution
cg=c()
for(x in 1:n){cg[x]=dztb(x,n,rsztb$par)}
cat(cg,"= Expected freq of (x= 1, 2, ..., n) respectively\n")
exp=cg*N
exp
cat(exp,"= Expected freq of (x= 1, 2, ..., n) respectively\n")
chisq <- sum(((f-exp)^2)/exp)
chisq
pval <- 1-pchisq(chisq,df)
pval

```

```

D=2*sum(ff*log(ff/ exp(xp)))
D

y=c(1:n)
exp=c()
da <-data.frame(y,f)
ztb=data.frame(y,exp)

plot_colors <- c("black","red","chocolate2","purple","green","royalblue","turquoise2")
plot(da, xlab="x", ylab="freq",type="o",pch=21, lty=1, lwd=2,col=plot_colors[1]);lines(ztb, type="o", pch=25,
lty=5,lwd=2, col=plot_colors[3]);title(main="Fitness plot: ZTB", col.main="blue",
font.main=3);legend("topright", g_range[5],c("data","ztb"), cex=0.8, col=plot_colors[c(1,3)], lty=1:5, lwd=2)

```

```

# ##### ZERO-TRUNCATED Poisson###
fnztp=function(par,x,n)
{
d=par[1]

ll=x*log(d)-lfactorial(x)-log(exp(d)-1)
-sum(ll)
}
rsztp <- optim(c(.),fnztp,method="L-BFGS-B",lower=c(.),upper=c(Inf),x=f(d,n)
rsztp

AIC= -2*(-rsztp$value)+2*1
AIC

BIC= -2*(-rsztp$value)+1*log(n)
BIC

#check mean
d=rsztp$par[1]
mnn= (d*exp(d))/(exp(d)-1) # for ZTP

```

```

#density function
dztp=function(x,n,par)
{
d=par[1]

ll=x*log(d)-lfactorial(x)-log(exp(d)-1)
exp(ll)
}

### Probability of the distribution
cg=c()
for(x in 1:n){cg[x]=dztp(x,n,rztp$par)}
cat(cg,"= Expected freq of (x= 1, 2, ..., n) respectively\n")
exp=cg*N
exp
cat(exp,"= Expected freq of (x= 1, 2, ..., n) respectively\n")
chisq <- sum(((f-exp)^2)/exp)
chisq
pval <- 1-pchisq(chisq,df)
pval

#ADJUSTED TEST
pexp=c()
ff=c()
cchisq <- sum((((ff-pexp)^2)/pexp)
cchisq
cpval <- 1-pchisq(cchisq,df)
cpval

D=2*sum(ff*log(ff/pexp))
D

y=c(1:n)

pexp=c()
#f=c()
da <-data.frame(y,f)
ztp=data.frame(y,pexp)

plot_colors <- c("black","red","chocolate2","purple","green","royalblue","turquoise2")

```

```
plot(da, xlab="x", ylab="freq", type="o", pch=21, lty=1, lwd=2, col=plot_colors[1]); lines(ztp, type="o", pch=25,
lty=5, lwd=2, col=plot_colors[2]); title(main="Fitness plot: ZTP", col.main="blue",
font.main=3); legend("topright", g_range[5], c("data", "ztp"), cex=0.8, col=plot_colors[c(1,2)], lty=1:5, lwd=2)
```

combined plot of the distributions

```
plot_colors <- c("black", "red", "blue", "purple", "green", "royalblue", "turquoise2")
plot(da, xlab="x", ylab="freq", type="o", pch=21, lty=1, lwd=2, col=plot_colors[1]); lines(ztcb, type="o",
pch=25, lty=5, lwd=2, col=plot_colors[5]); lines(ztmb, type="o", pch=25, lty=5, lwd=2,
col=plot_colors[4]); lines(ztb, type="o", pch=25, lty=5, lwd=2, col=plot_colors[3]); lines(ztp, type="o", pch=25,
lty=5, lwd=2, col=plot_colors[2]);
title(main="Fitness plot: ZTCB, ZTMB, ZTB and ZTP distributions", col.main="blue", font.main=3);
legend("topright", g_range[5], c("data", "ZTCB", "ZTMB", "ZTB", "ZTP"), cex=0.8,
col=plot_colors[c(1,5,4,3,2)], lty=1:7, lwd=2)
```