Ain Shams University

Faculty of Science

Department of Mathematics

On the Probability Distributions Related to Markov-Bernoulli Sequences of Random Variables

A thesis

Submitted in Partial Fulfillment of the Requirements of the Award

Of the Ph. D. Degree in Mathematical Statistics

By

Khalid Ali Hamad Al-Ajmi

MSc. (2008)- Faculty of Science-Kuwait University

Supervisors

Prof. Dr. M. Gharib

Prof. of Mathematical statistics

Mathematics Department, Faculty

of Science, Ain Shams University.

Dr. M.M. Ramadan

Lecturer of Mathematical statistics

Faculty of Education,

Ain Shams University.

Cairo-2015

ACKNOWLEDGMENT

First of all, all gratitude is due to **ALLAH ALMIGHTY** who guided me and endowed me this inspiration and ability to introduce this work.

I would like to express my deepest appreciation and gratitude to **prof. Dr. Mohamad Gharib** Professor of Mathematical Statistics, Mathematics Department, Faculty of Science, Ain Shams University for his supervision, his permanent support his guidance and for giving me the chance to work on this interesting topic.

I'm grateful to my supervision, **Dr. Mahmoud Mohamed Ramadan,** Lecturer of Mathematical Statistics, Mathematics Department, Faculty of Education, Ain Shams University for his kind supervision and supporting me during the preparation of this thesis.

Many thanks are also due to the chairman and staff members of the Department of Mathematics, Faculty of Science, Ain Shams University for facilities kindly afford, continuous encouragement during the preparation of this thesis.

Last but never least; I thank my family for encouragement and understanding

during the work on this thesis.

ENGLISH SUMMARY

The uses of the binomial and multinomial distributions in statistical modelling and analyzing discrete data are very well understood, with a huge variety of applications and appropriate software, but there are plenty of real-life examples where these simple models are inadequate. Therefore, it seems wise to consider flexible alternative models to take into account the overdispersion or underdispersion (Hinde & Demetrio (1998)). Thus, the binomial and Poisson distributions have been generalized in several ways to handle the problem of dispersion inherent in the analysis of discrete data that may arise with the presence of aggregation of the individuals. The binomial distribution has been generalized in various ways. Rudolfer (1990), Madsen (1993) and Luceno & Ceballos (1995) have summarized most of these generalizations. Among these extensions, there are the generalized binomial distribution introduced by Edwards (1960) and the multiplicative and the additive generalized binomial distributions which were derived by Altham (1978).

As finite Markovian models are extensively used in varies application fields, the generalized Markov-binomial model (Markov- Bernoulli model MBM, also called Markov modulated Bernoulli process (Ozekici(1997))) introduced by Edwards (1960) have been studied by many researchers from the various aspects of probability, statistics and their applications, in particular the classical problems related to the usual Bernoulli model (Anis & Gharib (1982), Arvidsson & Francke (2007), Cmey et al (2008) , Cekanavicius & Vellaisamy (2010) , Gharib & Yehia(1987), Inal(1987), Maillart et al.(2008), Minkova & Omey (2011), Ozekici (1997), Ozekici et al (2003),Pacheco et al. (2009), Yehia & Gharib(1993) and others.). Further, due to the fact that the MBM operates in a random environment depicted by a Markov chain so that the probability of success at each trial depends on the state of the environment, this model represents an interesting application of stochastic processes, and thus used by numerous authors in stochastic modelling (see for example, Switzer (1967, 1971), Pedler (1980), Xekalaki & Panaretos (2004), Arvidsson, & Francke (2007), Pires & Diniz (2012)).

The present thesis is devoted to study the probability distributions related to the Markov- Bernoulli sequence of random variables (the MBM) from some aspects such as distributional properties, characterizations, limit theorems, generalizations, and throw the light on some applications.

The thesis consists of five chapters and an introduction. The introduction is devoted to show the actuality of the subject of study and to give a historical survey about it. **Chapter one** is devoted to give the basic definitions, properties and preliminary results concerning the Markov- Bernoulli sequence of random variables (MBM). Chapter one, also, throw light on some generalizations of MBM and some examples of its applications.

Chapter two is concerned with the Markov binomial and Markov negative binomial distributions, exploring their properties, characteristic functions and relations to other distributions. This chapter contains, also, a numerical study to specify the descriptive characteristics of the Markov binomial distribution, besides a detailed investigation for the generalized Markov-binomial distribution introduced by Xekalaki & Panaretos (2004).

Chapter three is devoted to investigate the properties of the Markov-Bernoulli geometric (MBG) distribution through characterizing it. It is worth mentioning that the results of both sections 3.2 and 3.3 are totally new and is published respectively, in Journal of Mathematics and Statistics (Vol. 10, No. 2, 186-191, 2014), and in International Journal of Statistics and Probability (Vol. 3, No. 3, 138-146, 2014).

Chapter four is devoted to investigating the limiting behavior of the sum of n- Markov-Bernoulli random variables. In section 4.1 a new prove is given for the central limit theorem using generating functions technique. In section 4.2 we discuss in details the results of Gharib et al. (1987) concerning uniform estimates of the rate of convergence in the central limit theorem. Section 4.3 is devoted to discussing the results of Gharib et al. (1991) concerning limit theorem in the space L_{π}, $(1 \leq \pi \leq \infty)$.

In **chapter five**, a new method is introduced for adding two parameters to an existing distribution. This new technique extends the methods of Edwards (1960) and Marshall and Olkin (1997) for adding a parameter to a family of distributions. The method is of direct relevance to the Markov-Bernoulli geometric distribution and is applied in particular, to a one parameter Burr XII distribution to yield a three parameter extended Burr XII distribution which may serve as a competitor to such commonly used three parameters families of distributions. The results of this chapter are totally new and are submitted for publication in an international specialized journal.

INTRODUCTION

The uses of the binomial and multinomial distributions in statistical modelling and analyzing discrete data are very well understood, with a huge variety of applications and appropriate software, but there are plenty of real-life examples where these simple models are inadequate. Therefore, it seems wise to consider flexible alternative models to take into account the overdispersion or underdispersion (Hinde & Demetrio (1998)). Thus, the binomial and Poisson distributions have been generalized in several ways to handle the problem of dispersion inherent in the analysis of discrete data that may arise with the presence of aggregation of the individuals. For instance:

(i) in plant selection study the association among two plants arises when competing about the quantity of nutrients;

(ii) in biological study (Yakovlev & Tsodikov (1996) and Borges et al. (2012)), it is usually assumed that cells in a tissue are independent. However, the biological independence assumption may not be true when the dynamics of the cell population of a normal tissue is considered. It is therefore desirable to construct new models with strong biological interpretation of the dependence incorporated in the carcinogenesis process.

The binomial distribution has been generalized in various ways. Rudolfer (1990), Madsen (1993) and Luceno & Ceballos (1995) have summarized most of these generalizations. Among these extensions, there are the generalized binomial distribution introduced by Edwards (1960) and the multiplicative and the additive generalized binomial distributions which were derived by Altham (1978). The probability mass function (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 factor. On the other hand, the additive binomial distribution is a mixture of three conventional binomial models.

Altham (1978) developed the correlated Binomial model by correcting the Binomial model via the method suggested by Bahadur (1961) to encompass dependent Bernoulli variables. A three-parameter binomial distribution was derived by Paul (1985, 1987), which is a generalization of the Binomial, beta-binomial and the correlated binomial distribution proposed by Kupper & Haseman (1978). Ng (1989) developed the modified binomial distributions. In this approach, the binomial distribution is changed and the resulting distribution becomes more spread out (indicating positive correlation among the Bernoulli variables), or more peaked (indicating negative correlation among the Bernoulli variables) than the binomial distribution. A four-parameter binomial distribution was derived by Fu & Sproule (1995). This new distribution assumes values between α and β for $\alpha < \beta$, rather than the usual values 0 or 1. Lindsey (1995) and Luceno & Ceballos (1995) proposed a generalized binomial distribution which is discussed in details in Diniz et al. (2010). Chang & Zelterman (2002) generalizes the binomial distribution by considering Bernoulli variables with probability of success depending on the previous one. Tsai et al. (2003) presented a model that studies the overall error rate when testing multiple hypotheses. This model involves the distribution of the sum of dependent Bernoulli trials, and it is approximated thorough a beta-binomial structure. Instead of using the beta-binomial model, Gupta & Tao (2010) derived the exact distribution of the sum of dependent Bernoulli variables and not identically distributed. Minkova & Omey (2011) defined a new binomial distribution related to the interrupted Markov chain. Another extension of the binomial distribution is the COM-Poisson-binomial distribution (CMPB) introduced in Shmueli et al. (2005). The CMPB distribution arises as the conditional distribution of a COM-Poisson variable (Conway & Maxwell, 1962) given a sum of two COM-Poisson variables with the same dispersion parameter. It generalizes the binomial distribution and can be interpreted as the sum of dependent Bernoulli variables with a specific joint distribution (Altham et.al. 2014).

As finite Markovian models are extensively used in varies application fields, the generalized Markov-binomial model (Markov- Bernoulli model MBM, also called Markov modulated Bernoulli process (Ozekici(1997))) introduced by Edwards (1960) have been studied by many researchers from the various aspects of probability, statistics and their applications, in particular the classical problems related to the usual Bernoulli model (Anis & Gharib (1982), Arvidsson & Francke (2007), Cmey et al (2008) , Cekanavicius & Vellaisamy (2010) , Gharib & Yehia(1987), Inal(1987), Maillart et al.(2008), Minkova & Omey (2011), Ozekici (1997), Ozekici et al (2003),Pacheco et al. (2009), Yehia & Gharib(1993) and others.). Further, due to the fact that the MBM operates in a random environment depicted by a Markov chain so that the probability of success at each trial depends on the state of the environment, this model represents an interesting application of stochastic processes, and thus used by numerous authors in stochastic modelling (see for example, Switzer (1967, 1971), Pedler (1980), Xekalaki & Panaretos (2004), Arvidsson, & Francke (2007), Pires & Diniz (2012)). Moreover, from the point of view of constructing new counting distributions, the MBM introduced by Edwards(1960), represents, in fact, a new way of obtaining new counting distributions related to some Markov chain by assuming some dependency in the sequence of Bernoulli random variables that gives an additional parameter by which the Bernoulli model could be a more realistic model in practice.

The present thesis is devoted to study the probability distributions related to the Markov- Bernoulli sequence of random variables (the MBM) from some aspects such as distributional properties, characterizations, limit theorems, generalizations, and throw the light on some applications.

The thesis consists of five chapters and an introduction. The introduction is devoted to show the actuality of the subject of study and to give a historical survey about it.

Chapter one is devoted to give the basic definitions, properties and preliminary results concerning the Markov- Bernoulli sequence of random variables (MBM). Chapter one, also, throw light on some generalizations of MBM and some examples of its applications.

Chapter two is concerned with the Markov binomial and Markov negative binomial distributions, exploring their properties, characteristic functions and relations to other distributions. This chapter contains, also, a numerical study to specify the descriptive characteristics of the Markov binomial distribution.

Chapter three is devoted to investigate the properties of the Markov-Bernoulli geometric (MBG) distribution through characterizing it. Section 3.1 investigates the results of Yehia and Gharib (1993) concerning the properties of the MBG distribution analogous to those of the usual geometric distribution such as the lack of memory property and conditional moments. Section 3.2 is devoted to properties that characterize the MBG distribution using its relation to random sums. Section 3.2 is devoted to investigating some characterization results of the MBG distribution related to random sums under some moment conditions. It is worth mentioning that the results of both sections 3.2 and 3.3 are totally new and are published respectively, in Journal of Mathematics and Statistics (Vol. 10, No. 2, 186-191, 2014) and in International Journal of Statistics and Probability (Vol. 3, No. 3, 138-146, 2014).

Chapter four is devoted to investigating the limiting behavior of the sum of n- Markov-Bernoulli random variables. In section 4.1 we give a new prove for the central limit theorem using generating function. In section 4.2 we discuss in details the results of Gharib et al. (1987) concerning uniform estimates of the rate of convergence in the central limit theorem. Section 4.3 is devoted to discussing the results of Gharib et al. (1991) concerning limit theorem in the space L_{π}, $(1 \leq \pi \leq \infty)$.

In chapter five, a new method is introduced for adding two parameters to an existing distribution that extends the methods of Edwards (1960) and Marshall and Olkin (1997) for adding a parameter to a family of distributions. The method is of direct relevance to the Markov-Bernoulli geometric distribution and is applied in particular, to the one parameter Burr XII distribution to yield a three parameter extended Burr XII distribution which may serve as a competitor to such commonly used three parameters families of distributions. The results of this chapter are totally new and are submitted for publication in an international specialized journal.

TABLE OF CONTENTS

NOTATIONS AND ABBREVIATIONS

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
MBM	Markov- Bernoulli Model
MBG	Markov- Bernoulli Geomeric
mgf	Moment Generating Function
pgf	Probability Generating Function
rv (RV)	Random Variable
SF	Survival Function
$\Phi(\cdot)$	The Standard Normal Cumulative Distribution Function
pmf	The Probability Mass Function
cf	Characteristic Function
iid	Independent and Identically Distributed
MNB	Markov-Negative Binomial
E(X)	Expectation of a Random Variable X
V(X)	Variance of a Random Variable X

Cov	Covariance
Corr	Correlation
MBEB XII	Markov – Bernoulli extended Burr XII
HRF	Hazard Rate Function
MLE	Maximum Likelihood Estimation
KME	Kaplan- Meier Estimator

Chapter 1

Basic Definitions, Properties and Preliminary Results

Introduction

This chapter is devoted to introduce the basic definitions, properties and preliminary results concerning the Markov- Bernoulli sequence of random variables (Markov- Bernoulli model MBM). Some generalizations and application examples are, also, introduced. The basic references of the chapter contents are Anis & Gharib (1982), Edwards (1960), Wang (1981), Xekalaki and Panaretos (2004), Switzer (1967), and Ng et al. (1999).

1.1 The Markov-Bernoulli Sequence (The Markov-Bernoulli Model)

Edwards (1960) proposed the Markov-Bernoulli sequence of random variables as a generalization of the independent Bernoulli sequence of random variables by introducing a correlation between trials. In other words, let $X_1, X_2, ...$ be a sequence of Bernoulli random variables, each with possibly varying probabilities of success. Edwards formulated the problem as a Markov process with the following matrix of transition probabilities from one trial to the next:

$$\begin{array}{cc} & X_{i+1} \\ & \begin{array}{cc} 0 & 1 \end{array} \\ X_i \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 1-(1-\rho)p & (1-\rho)p \\ (1-\rho)(1-p) & \rho+(1-\rho)p \end{bmatrix} \end{array} = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} \qquad (1.1.1)$$

The initial probabilities are $P(X_1 = 1) = p = 1 - P(X_1 = 0) = 1 - q$, where $p \in [0,1]$ and $\rho \in [0,1]$. If $\rho = 0$, we have the usual independent Bernoulli sequence.

The resulting model (1.1) is called the Markov-Bernoulli model (MBM) or the Markov modulated Bernoulli process Ozekici (1997).

If $\rho = 1$, the Markov process remains in its initial state for ever with probability 1.