STATISTICAL MODELLING OF TIME SERIES OF COUNTS FOR A NEW CLASS OF MIXTURE DISTRIBUTIONS

KHOO WOOI CHEN

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

STATISTICAL MODELLING OF TIME SERIES OF COUNTS FOR A NEW CLASS OF MIXTURE DISTRIBUTIONS

KHOO WOOI CHEN

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2016

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: KHOO WOOI CHEN

Registration/Matric No: SHB 100009

Name of Degree: DOCTOR OF PHILOSOPHY

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

STATISTICAL MODELLING OF TIME SERIES OF COUNTS FOR A NEW CLASS OF MIXTURE DISTRIBUTIONS

Field of Study: STATISTICS

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 09 MARCH 2016

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Integer-valued correlated stochastic processes, which we often meet in the real world, are of major concern in many natural and social sciences. The classical continuous time series models which contain scalar multiplication are not able to represent count data since the integer nature of the data is not preserved. Therefore, the formulations of discrete-valued time series models for count data are apparently of significance. Much effort has been expended in the past few decades to construct discrete-valued time series models. Nevertheless, the hunt for better models is still ongoing due to the need to improve or sharpen the statistical analysis. This thesis proposes a new mixture model, the mixture of Pegram and thinning integer-valued autoregressive (MPT) processes, which is the combination of current discrete-valued time series operators. The statistical and regression properties, parameter estimation, forecasting, and graphical analysis for the new model have been examined. Model selection based upon the Akaike Information Criterion has been performed. Extensions to the moving average (MA) and autoregressive moving average (ARMA) models have also been considered. Important properties such as reversibility and regression are then discussed. The extension to the qth-order MPT process has also been investigated in the study. Previous studies have emphasized the Poisson sequence as it is an infinitely divisible distribution. In this thesis, it is shown that proposed model is able to deal with infinitely and non-infinitely divisible distributions with simpler expressions. Furthermore, the proposed MPT model is able to handle multimodality and has better performance than the current discrete-valued time series models. The available forecasting method based on the conditional expectation may not be appropriate for integer-valued time series models. Thus coherent forecasting, which is based upon the k-step ahead conditional mean, median, mode and distribution, is considered. For low count series the k-step ahead conditional distribution of the MPT model practically exhibits better performance than the other models. The

score functions and information matrix have been derived to measure the asymptotic standard errors and to analyze the variance-covariance relationship among the parameters. Parameter estimation with the maximum likelihood estimation via the Expectation-Maximization algorithm is discussed and compared with the conditional least squares method. Finally, some real life data sets from different disciplines have been applied to illustrate the analyses. The thesis is concluded with some recommendations for future work.

ABSTRAK

Nilai integer berkorelasi stokastik proses, yang kita sering bertemu di dunia sebenar, adalah perhatian utama di kawasan semulajadi dan sains sosial. Siri masa selanjar yang klasik dimana ia mengandungi skalar pendaraban adalah tidak dapat mewakili data bilang kerana sifat data bilang tidak dipeliharakan. Oleh itu, pemodelan bagi model-model nilai diskrit siri masa menjadi semakin penting. Beberapa dekad yang lalu, banyak usaha telah diperkembangkan untuk membina model-model nilai diskrit siri masa. Namun begitu, pencarian diskrit model-model yang lebih baik masih sedang dijalankan kerana bertujuan untuk memajukan atau mengasahkan analisis statistik. Tesis ini mencadangkan satu model bercampur, campuran daripada proses *Pegram* dan *Thinning* nilai integer autoregresi (MPT), di mana ia adalah hasil kombinasi daripada operator-operator nilai diskrit siri masa. Pemeriksaan model baru termasuk sifat-sifat berstatistik dan regresi, penganggaran parameter, penelahan dan grafik analisis. Pemilihan model berdasarkan *Akaike Information Criterion (AIC)* dilaksanakan. Perkembangan ke model-model purate bergerak dan autoregresi purate bergerak juga dipertimbangkan. Sifat-sifat yang penting seperti kebolehterbalikan dan regresi dibincangkan. Lanjutan yang mungkin bagi *p* tertib turut disiasatkan. Pengajian sebelum ini diutamakan dengan jujukan Poisson kerana ia diketahui sebagai taburan terbahagikan tak terhingga. Di dalam tesis ini, kami bertujuan untuk mencadang model baru yang boleh digunakan bagi taburan terbahagikan tak terhingga and terhingga dengan ungkapan yang lebih mudah. Tambahan pula, kebolehan pengendalian taburan multimod dan mempunyai prestasi yang lebih baik berbanding dengan model-model nilai diskrit semasa yang sedia ada dipaparkan. Cara penelahan yang sedia ada dimana ia mengutamakan jangkaan bersyarat mungkin tidak sesuai untuk model-model siri masa nilai integer. Jadi, penelahan koheren dimana ia berdasarkan k-langkah dahulu min bersyarat, median mod dan taburan dipertimbangkan. Bagi siri bilang yang rendah, k-langkah dahulu taburan bersyarat bagi model MPT mempamerkan prestasi yang lebih baik berbanding dengan model-model yang lain. Fungsi skor dan matrik maklumat dipertimbangkan untuk mengukur sisihan piawai asimptot dan menganalisis hubungan varians-kovarians antara parameter-parameter. Penganggaran parameter dengan penganggar kebolehjadian maksimum melalui algoritma jangkaan maximum dibincangkan dan dibuat perbandingan dengan Kaedah kuasa dua terkecil bersyarat. Akhirnya, beberapa set-set data sebenar daripada disiplin-disiplin yang berbeza digunakan untuk model ilustrasi. Tesis ini disimpulkan dengan beberapa cadangan-cadangan yang selanjutnya.

ACKNOWLEDGEMENT

Finally, it's acknowledgement. First of foremost, I would like to express my greatest gratitude to my supervisor, Prof. Ong Seng Huat, you have been a tremendous mentor for me. I would like to thank you for your aspiring guidance and invaluably constructive advices for allowing me to grow as a researcher. I would also like to thank for your continuous financial support. Your efforts to make my study completed are much appreciated. I am thankful to my co-supervisor, Prof. Atanu Biswas. Your advices in my research have been priceless. Your comments and suggestions have been helpful in my research path.

A million thanks to my family. Words cannot describe how grateful I am to my mother and my father for all of the sacrifices that you have made on behalf of me. I am indebted to you. I would like to extend my appreciation to my husband, Teh Chee Siang. Your patience and mentally support have been so significant to make today happens. I am grateful to you who have been spending innumerable sleepless nights with me when no one could answer my queries. You have always been there for me when I had difficult time in my research. My sisters have been playing a vital role during my research. You both always encourage me when I confronted frustration.

I would like to give a special thanks to Dr. Ng Choung Min. We knew each other when I was still new in research. You generously share your research skills with me to make my study possible. During my study, I had a chance to visit Indian Statistical Institute in Kolkata. I have had a great research experience with Raju Maiti, you are such a good friend who have been helping me out in every way you could.

Last but not least, I would like to thank to all those who have lent helping hand to make my thesis possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

$*$	:	Mixing operator
$\circ$	:	Binomial thinning operator / Thinning operator
X_t	:	Time series of random variable X at time t
$\tau \star_\alpha$	:	Iterated thinning operator with parameters τ and α
i.i.d.	:	Independent and identically distributed
MPT	:	Mixture of Pegram and thinning
MPT(1)	:	Mixture of Pegram and thinning first order integer-valued Autoregressive
MPT(p)	:	Mixture of Pegram and thinning pth-order integer-valued Autoregressive
MPT-MA	:	Mixture of Pegram and thinning of Moving Average
MPT-MA(1)	:	Mixture of Pegram and thinning first order integer-valued Moving Average
MPT-MA(q)	:	Mixture of Pegram and thinning qth-order integer-valued Moving Average
CLAR(1)	:	First order conditional linear Autoregressive
AR(1)	:	First order Autoregressive
AR(p)	:	pth-order Autoregressive
DAR(1)	:	Discrete-valued first order Autoregressive
MA	:	Moving-Average
MA(1)	:	First order Moving Average
INAR	:	Integer-valued Autoregressive
INAR(1)	:	First order integer-valued Autoregressive
INAR(p)	:	pth-order integer-valued Autoregressive

INAR(1)-P, PoINAR(1), Poisson INAR(1)	:	Poisson first-order integer-valued Autoregressive
INMA	:	Integer-valued Moving Average
BAR	:	Binomial Autoregressive
BAR(1)	:	First order Binomial Autoregressive
RCINAR(1)	:	Random coefficient first order integer-valued Autoregressive
IINAR(1)	:	Iterated thinning first order integer-valued Autoregressive
QINAR(1)	:	Quasi-binomial first order integer-valued Autoregressive
DARMA	:	Discrete-valued Autoregressive Moving Average
ARMA	:	Autoregressive Moving Average
ARMA(1,1)	:	First order Autoregressive Moving Average
EARMA	:	Mixed Autoregressive Moving Average Exponential sequence
EARMA(1,1)	:	First order mixed Autoregressive Moving Average Exponential sequence
pgf	:	Probability generating function
pmf	:	Probability mass function
QB	:	Quasi-Binomial
$\boldsymbol{x} = (x_0, x_1, \dots, x_n)$	:	$(n+1)$-dimensional random variable vector
$\theta = (\theta_1, \theta_2, \dots, \theta_p)$	:	p-dimensional parameter vector
$G_X, G_\varepsilon, G_Y, G_{\rho \star_\alpha X}$	:	Probability generating function for X, ε, Y and $\rho \star_\alpha X$
$G_{X,Y}$	:	Joint probability generating function of X and Y
$\mu_X, \mu_\varepsilon, \mu_\alpha$	:	Mean for X, ε and α
$\sigma_X^2, \sigma_\varepsilon^2, \sigma_\alpha^2$	:	Variance for X, ε and α
ρ_k	:	Lag-k autocorrelation function

γ_k	:	Lag-k autocovariance function
$\hat{\rho}$	:	Estimated autocorrelation value
$\hat{\gamma}$	:	Estimated autocovariance value
$P(\cdot\mid\cdot)$	:	Conditional probability function of a random variable
$E(\cdot\mid\cdot)$	:	Conditional expectation of a random variable
$\mathbb{N}_0$	:	Natural numbers
$\mathbb{C}$	:	Complex numbers
$\mathbb{Z}$	:	Integer numbers
$E(\cdot)$	:	Expectation of a random variable
$Var(\cdot)$	:	Variance of a random variable
$cov(X,Y)$	:	Covariance of X and Y
Beta (γ,β)	:	Beta distribution with shape parameters γ and β
$I[\cdot]$	:	Indicator variables
$p_x, P(X=i)$, $P_X(x)$	:	Probability distribution of X
$g(\cdot), f(\cdot)$	:	Probability density function
$Q(\boldsymbol{x};\ \vartheta)$	:	Conditional least squares of X with parameters ϑ
$X\sim D(p_0,p_1,\ldots)$	:	X is distributed as a distribution D with probability distribution $p_0,p_1,\ldots$
$t\in 0,\pm1,\pm2,\ldots$	:	Time t is the elements of set of natural numbers
$\subseteq$	:	Subset
MLE	:	Maximum likelihood estimation
EM	:	Expectation-Maximization
CMLE	:	Conditional maximum likelihood estimation
CLS	:	Conditional least squares

YW	:	Yule-Walker
GMM	:	Generalized methods of moment
OLS	:	Ordinary least squares
:	:	Likelihood function for random variable X with parameter θ
$\log L$	:	Log-likelihood function
IM	:	Information matrix
AMI	:	Auto-mutual information
PAMI	:	Partial auto-mutual information
ACVF	:	Autocovariance function
ACF	:	Autocorrelation function
PACF	:	Partial autocorrelation function
MSE	:	Mean square errors
AIC	:	Akaike information criterion
$NB(k,p)$	:	Negative Binomial distribution with parameters k and p
$Bin(N,p)$	:	Binomial distribution with index N and success probability p
$\dot{\ell}$	:	Score function
$\ddot{\ell}$	:	Observed Fisher information
$E[\ddot{\ell}]$	:	Expected Fisher information
$\xrightarrow{w}$	:	Converge weakly
$P_k(x)$	:	Conditional distribution of X_{t+k} given X_t
σ_k^2	:	Conditional variance of X_{t+k} given X_t
μ_k	:	Conditional mean of X_{t+k} given X_t
$p_k(\cdot \mid \cdot)$	:	k-step-ahead conditional probability distribution
i	:	Fisher information matrix
V	:	Inverse Fisher information matrix, variance-covariance matrix

$\boldsymbol{\theta}_{\mathbf{0}}$	:	True value of parameter vector
$\widehat{\boldsymbol{\theta}}$	:	Estimated parameter vector
$N(\mathbf{0}, \mathbf{V})$	:	Normal distribution with mean 0 and standard deviation $\sqrt{\mathbf{V}}$

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.0 Preliminary for Time Series Models

Time series modelling in the continuous context has been grabbing great attention in the past few decades. It usually appears in areas such as engineering, sociology, science and economics. A time series is a set of observations x_t, each one being recorded at a specific time t. For continuous time series, the observations are measured continuously over some time intervals, for example, $T = [0,1]$. For a discrete-time series, the observations are measured at sequential integer values over a fixed time intervals. There have been many real life examples showing discrete-time series measurement, for example, the number of Australian red wine sales and accidental deaths.

The main purpose of time series analysis is to set up a hypothetical probability model; the time series model, to represent the data. After an appropriate family of models is chosen, it is possible to estimate the parameters in the models, check for goodness-of-fit to the data, and possibly to use the proposed models to enhance our understanding of the mechanism. Once the satisfactory model has been obtained, it may be used for further study such as prediction for future observations or application in a particular field. A time series model for the observed data x_t, can be interpreted as a specification of a sequence of correlated random variables X_t of which x_t is postulated to be a realization. A general but important discussion in the study of time series models will include model stationary, autocovariance and autocorrelation functions, moment properties and the model efficiency. Like most of the study for time series models, we will consider all potential and significant properties of the proposed models in our study.

Time series of counts appears in many different contexts, usually as counts of certain events or objects in specified time intervals. This thesis will discuss mainly on the discrete correlated observations with some potential time series models. The area of discrete time series modelling involves counts which arise, for example, from traffic accidents, claim counts, number of guests staying overnight in the hotel, number of abstract reviews, and counts of price changes. In continuous time series, one conventional time series model, namely Autoregressive (AR), which was introduced by Box and Jenkins (1976) is defined by the recursion

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} \tag{1.1}$$

These models have simple interpretation and attractive properties. Although Equation (1.1) is well defined for natural numbers $\mathbb{N}_0 = \{0, 1, \ldots\}$, but the multiplication operator cannot accommodate integer values, since the multiplication of an integer with real numbers yields non-integer values. Therefore, this motivates the replacement of scalar multiplication by a different operator like the binomial thinning operator but preserves the properties of the continuous counterpart.

Binomial thinning (or thinning) operation, commonly known as Galton-Watson branching process, are probabilistic operators that can be used to handle integer values. This popular binomial thinning operator was primarily introduced by Steutel and Van Harn (1979) to define discrete self-decomposability, or more precisely of infinitely divisible integer-valued time series; a characteristic function $\psi(t)$ is said to be infinitely divisible if $\psi(t) = [\psi_n(t)]^n$ for all positive integer n, where $\psi_n(t)$ is itself a characteristic function. The fundamental interpretation and generalization of binomial thinning operation will be reviewed in detail in Chapter 2.

Pegram (1980) introduced an entirely different concept to deal with count data that specifically gives good interpretation for categorical data set. The mixture operator, abbreviated by $'*'$, has not been broadly used by researchers. Thus there is room for further development and study of the useful properties and important applications in various fields. To the best of our knowledge, Pegram's mixture operator has been investigated by Biswas and Song (2009) recently, who made a comparative study between Pegram's mixture and thinning operators. The results suggest that the mixture operator can be an alternative tool to handle count data. Therefore, it motivates us to look into the combination of both discrete operators.

1.1 Objectives

The existing thinning operator has been the sole operator to deal with count data since 1980. It is able to accommodate only infinitely divisible distributions. In non-infinitely divisible distribution such as Binomial case, the model does not have the similar form of a typical INAR(1) process (see Definition 2.1.2). Hence, many of the important properties of the typical INAR(1) model are not shared. Later, Biswas and Song (2009) studied the alternative mixing operator which was introduced by Pegram (1980). This resulted in a model that can now deal with non-infinitely divisible cases like categorical data. However, the proposed mixture model involves constant runs which may not be applicable in many real life situations. Therefore, our aim is to find a new model which is able to apply to count data, on top of that it possesses a simpler interpretation, as well as having better performance compared to the existing models.

The proposed model is defined by combining both the thinning and mixing operators. Through the combination, we aim to yield a better model by integrating the advantages of both operators. In fact, in the later chapters, one can see that our proposed model handles well the infinitely divisible distributions. For binomial case, the model provides simple structure for the transition probabilities and autocovariance and autocorrelation functions. We study the first order of AR of the new mixture process. Extensive simulation study has been carried out to observe the model behaviour. After that, we extended the order-one process to pth-order. As a step to complete the development for a family of proposed mixture model, we discussed the Moving Average (MA) of order one and its extension to qth-order. The discussion on the combination of Autoregressive Moving Average (ARMA) processes is then carried out. Likewise, the relevant properties and marginal fitting will be considered respectively in order to provide more insights into discrete-valued time series modelling.

In the literature review, there is a dearth of research work involving the area of coherent forecasting for the count data. The typical procedure for forecasting in time series models is to use conditional expectation, because this technique yields forecasts with the minimum mean squared forecast error. However, this method usually produces non-integer-valued forecasts. It has been proposed to use k-step ahead conditional probability to generate coherent forecasting for integer value forecasts. Freeland and McCabe (2004) investigated the coherent forecasting for discrete low count data. In our study, it is of interest to look into the coherent forecasting, as our model appears as a new tool to handle the count data well. Particularly, the k-step ahead conditional probability and its measurement of the central moments is considered to determine the coherency. Coherent forecasting is based on asymptotical theory. In this study, the δ-method will be applied to obtain the asymptotic distribution together with the Fisher Information matrix.

It is necessary that the superiority of a dynamic discrete time series model should be illustrated with real field applications. This dissertation will cover a few sets of count data in different discipline: internet protocol (IP) address data, criminal data, and worker compensation claim data. The data sets are readily available in current literature. See appendices for the data observed frequencies. The real applications show that the proposed model has the potential to be viable alternative model in discrete time series modelling.

1.2 Overview of Topics

The outline of this dissertation is as follows. Chapter 2 provides the literature reviews of current discrete time series operators. Details of first order of Integer-valued Autoregressive (INAR(1)) process are given. The extension of thinning operator to generalized thinning operator and the relevant discrete time series models using the generalized thinning operator are also discussed. Parameter estimation is one of the important issues to be reviewed. A literature review showed that for the INAR process and its generalized models, a majority of the discrete-valued time series models used maximum likelihood estimation (MLE), Yule-Walker (YW) estimation and conditional least square (CLS) for parameter estimation. All methods are summarized clearly in this chapter.

Chapter 3 is the main part of this thesis. This chapter starts with the model construction. We named the proposed mixture model as a mixture of thinning and Pegram's operators (MPT) process. We consider the first-order model for fundamental study. The important stationary, statistical and regression properties are discussed. Then, some specific discrete marginal distributions are fitted to the model. In particular, we show one example for the non-infinitely divisible case where the binomial marginal is well-interpreted by the model. The simple expression of the Binomial MPT process encourages its use in analyzing real life data.

We look into parameter estimation in Chapter 4 where maximum likelihood estimator via Expectation-Maximization (EM) algorithm is used to estimate the parameters for the first-order MPT (MPT(1)) process. It is well-known that the EM algorithm is suited for mixture models. Unlike in the INAR and Pegram's AR(1) model, the typical YW Equations and CLS cannot be applied to the proposed mixture model. This is because of the inseparable parameters in the autocorrelation function (ACF), see Eq. (3.12) in Chapter 3. An extensive simulation study has been carried out. The simulation study includes robustness examination. Special outlier generation schemes; the INAR(1) additive outlier and INAR(1) innovative outlier are applied to observe the ability of MLE in handling the outliers for the proposed model. Also, comparison has been carried out between the proposed model and the current existing models. Chapter 4 continues with estimating the parameters of variance-covariances matrix, which will be needed for coherent forecasting in the following chapter. Explicit expressions of score function and Fisher Information matrix are presented. Then, the expressions are carried forward to associate with the theoretical part of coherent forecasting to form a new discussion in Chapter 5. Chapter 4 concludes with numerical simulations for coherent forecasting accuracy measurements of Poisson MPT(1) model. In Chapter 5, *k*-step-ahead forecasting distribution and the relevant properties are investigated. The results present an approach for coherent forecasting in discrete time series. This is one of the main contributions of our study.

This dissertation proposed a new time series model to accommodate any discrete marginal distribution. Therefore, the formulation of other related family members for the new model is necessary, as they will be potentially useful in some real life applications. For example, the MA process is usually applied to illustrate economic data. We begin the study of the first-order MA process for the new MPT model with Poisson and Binomial marginal distributions. For the proposed model we discuss regression, fundamental properties and reversibility, followed by the investigation of qth-order MA process. The ARMA process for MPT model is then presented, together with some preliminary discussion on autocovariance and autocorrelation functions. Last but not least, a natural extension from order one to pth-order MPT processes has been investigated. It is very useful in some real life examples when the partial autocorrelation function shows more than one lag. See Section 7.4. All real life data applications for this thesis are presented in Chapter 7. The data sets include criminal counts of drug and sex offensive, internet protocol (IP) addresses, and the popular worker compensation burn claims. These data sets have been extensively discussed in the literature review. We applied the data here for the new model illustration and also for comparison purpose. The comparison is based upon Akaike Information Criterion (AIC), and it can be seen that the results are in favour of the proposed model. Chapter 8 recommends future research works. We conclude the chapter by presenting some preliminary ideas and potential research directions.

CHAPTER 2: LITERATURE REVIEWS

2.0 Introduction

This chapter acquaints the readers with the background of two important operators; the thinning and Pegram's operators. The relevant properties and significant contributions of the currently existing operators will be reviewed.

2.1 Binomial Thinning Operator

The Binomial thinning operator has been introduced to replace the scalar multiplication in order to handle count data. It is defined as follows.

Definition 2.1.1 (Binomial Thinning): Let X be the discrete random variable with range $\mathbb{N}_0$, and $\alpha \in [0,1]$. Define the random variable

$$\alpha \circ X = \sum_{i=1}^{X} Y_i \tag{2.1}$$

where Y_i is Bernoulli random variable with the success probability $P(Y_i = 1) = \alpha$ and the random variable Y_i is independent of X. Consequently, $\alpha \circ X$ arises as binomial random variable with $Bin(X, \alpha)$ such that $E[\alpha \circ X] = \alpha E[X]$ and $V[\alpha \circ X] = \alpha^2 \sigma_X^2 + \alpha(1-\alpha)\mu_X$, and the circle operator $' \circ '$ is known as binomial thinning operator. See Silva (2005, Lemma 2.1) for detailed properties of Definition 2.1. The probabilistic operator of binomial thinning is easy to interpret so that it can be applied in count modelling with several types of marginal probabilities. To adapt to different types of interpretation, the generalizations of binomial thinning are necessary. Latour (1995) proposed an immediate extension of Definition 2.1.1, so that it can work with any non-negative random variables. Brännäs and Hellström (2001) suggested that the dependency of the counting series is allowed. Another modification using signed binomial thinning is carried out by Kim and Park (2008) to handle negative values and allow for negative correlation. It is worthwhile to mention here that the generalized operators treat the thinning operator as a special case. This is useful because we could have more different interpretations for real life applications. We can now define the fundamental first order integer-valued Autoregressive (INAR(1)) process with the thinning operator as such

Definition 2.1.2 (INAR(1) Model): Let X_t be the non-negative discrete random variable, for any $\alpha \in [0, 1]$, the first order INAR (INAR(1)) process is defined by

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t \tag{2.2}$$

where ε_t is a sequence of uncorrelated non-negative integer-valued random variables having mean μ and finite variance σ^2.

McKenzie (1985) was perhaps the pioneer to study the thinning operation by introducing integer-valued Autoregressive (INAR) process. The interpretation is as such: consider there is a population at a certain time $t-1$. If we observe the population in later point of time, say t, the population maybe adjusted. This is because some individual may die off in between the time $(t-1, t]$ with probability α, and the new arrivals (ε_t) such as new born babies may be entering the population at time t. With this elegant interpretation, it can be applied in many real life situations. For examples see Freeland (1998) for claim counts and Weiβ (2007) for statistical quality control.

In many real life applications, the immigration (innovation) ε_t remains unknown. Parameter estimation and properties of ε_t can be obtained by relating it to the observations X_t. A recent model proposed by Weiβ (2012) assumed that the random variable X_t and the new immigrants (innovation) ε_t are both observable, hence the 'survivors' are proposed by the terms of $Z_{t-1} \coloneqq \alpha \circ X_{t-1} = X_t - \varepsilon_t$. With this fully observed INAR(1) process, the parameter estimation can be done directly from the observations ε_t.

The INAR(1) is a stationary process. The probability generating function (pgf) is given by

$$G_X(z) = G_X(1-\alpha+\alpha z)G_\varepsilon(z) \Leftrightarrow G_\varepsilon(z) = \frac{G_X(z)}{G_X(1-\alpha+\alpha z)} \tag{2.3}$$

where $G_X(z)$ and $G_\varepsilon(z)$ denote the pgf of X_t and ε_t, respectively. Some important basic and regression properties such as conditional expectation and joint pgf have been discussed. Also, the parameter estimation of the models have been extensively studied by several researchers; see Al-Osh and Alzaid (1987), Freeland (1998) and Brännäs (1994). In the following section, we will show several discrete marginal distributions which have been specified for innovation process in INAR(1) process.

Example 2.1.1 (Poisson INAR(1) Model): Let ε_t be an i.i.d. Poisson random variable with mean λ, then the INAR(1) process in Definition 2.1.2 follows a Poisson process $Po\left(\frac{\lambda}{1-\alpha}\right)$. Obviously X_t is a stationary Markov Chain process with Poisson marginal distribution $\frac{\lambda}{1-\alpha}$. Several notations have been given to refer to Poisson as the assumption for innovation process, i.e. INAR(1)-*P* and PoINAR(1). Table 2.1.1 simplifies some important properties of Poisson INAR(1) process. Further properties can be found in Silva (2005) and Freeland (1998).

Table 2.1.1: Fundamental properties of Poisson INAR(1) Model

Mean	$\mu_X = \dfrac{\mu_\varepsilon}{1-\alpha}$
Variance	$\sigma_X^2 = \dfrac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1-\alpha^2}$
Autocorrelation Function	$\rho_X(k) = \alpha^k$
Conditional Probability Function	$P(X_t = i \mid X_{t-1} = j) = \sum_{k=0}^{min\,(i,j)} \binom{j}{k} \alpha^k (1-\alpha)^{j-k} . P(\varepsilon_t = i-k)$
Conditional Expectation	$E(X_t \mid X_{t-1}) = \alpha X_{t-1} + \mu_\varepsilon$

Poisson INAR(1) model is a benchmark model. INAR(1) process can be fitted not only by Poisson process, but also by any discrete marginal distribution from self-decomposability family. Next, we discuss Negative Binomial marginal applied in INAR(1) model. Unlike Poisson process, Negative Binomial marginal has slightly complicated expression. Analogue with Gamma and Exponential distribution in continuous random variables, McKenzie (1987) introduced Geometric and Negative Binomial marginals. Geometric marginal is the special case of Negative Binomial marginal with $k = 1$.

Example 2.1.2 (Negative Binomial INAR(1) Model) In the case of Negative Binomial, McKenzie (1987) provided an explicit expression for distribution of ε_t, given as follows

$$P(\varepsilon_t = k) = \sum_{m=1}^{n} \binom{n}{m} \rho^{n-m} (1-\rho)^m . \binom{m+k-1}{k} (1-p)^k p^m$$

$$k > 0, n \in \mathbb{N}. \quad (2.4)$$

For more details regarding Negative Binomial and Geometric marginal distribution in INAR(1) process, one can refer to McKenzie (1986). Also see Weiβ (2008, Example 3.4) for similar expression for distribution of ε_t.

In previous discussion, we emphasized on INAR(1) model fitted with marginals of infinite range. Steutel and Van Harn (1979) and Joe (1996) stated that Eq. (2.2) is restricted to the self-decomposability family. Weiβ (2009), however, proposed Binomial AR(1) process to show that it can be fitted with the finite range marginal in a totally different expression. The model is given by

Definition 2.1.3 (Binomial AR(1) Model): Let $n \in \mathbb{N}, p \in (0,1)$ and $\omega \in [\max\left(-\frac{p}{1-p}, -\frac{1-p}{p}\right); 1]$. Define $\kappa := p \cdot (1-\omega)$ and $\alpha := \kappa + \omega$. The process $(X_t)_{\mathbb{N}_0}$, defined by the recursion

$$X_t = \alpha \circ X_{t-1} + \kappa \circ (n - X_{t-1}), \quad t \geq 1, \;\; X_0 \sim Bin(n, p), \tag{2.5}$$

where all thinning are performed independently of each other, and the thinning at time t are independent of $(X_s)_{s<t}$, is called a Binomial AR(1) process. The condition of ω guarantees that $\alpha, \kappa \in [0, 1]$. The interpretation is given here for convenience.

Suppose that a system has n mutually independent units, each of them is either in state 1 or state 0. Let X_{t-1} be the number of units being in state 1 at time $t-1$. Then $\alpha \circ X_{t-1}$ is the number of units still being in state 1 at time t, with individual transition probability α. $\kappa \circ (n - X_{t-1})$ is the number of units, which moved from state 0 to state 1 at time t, with each individual transition probability κ. Recently, the extensive research works of Binomial AR(1) model have been carried out by Prof. Weiβ Christian and his co-workers. Initially from Weiβ (2009), Weiβ and Kim (2013) studied the Binomial AR(1) model properties; moments, cumulants and estimation. Kim and Park (2010a) and Kim and Park (2010b) discussed the coherent forecasting for Binomial AR(1) process. Also see the parameter estimation for Binomial AR(1) models with applications in finance and industry by Weiβ and Kim (2013). Some important properties are tabulated in Table 2.1.2.

Table 2.1.2: Some properties of Binomial AR(1) model

Mean	$\mu_X = np$
Variance	$\sigma_X^2 = np(1-p)$
Autocorrelation Function	$\rho_X(k) = \rho^k$
Conditional Probability Function	$P(X_t = i \mid X_{t-1} = j) = \sum_{k=max\,(0,i+j-n)}^{min\,(i,j)} \binom{j}{k}\binom{n-j}{i-k} \alpha^m (1-\alpha)^{j-k} \beta^{i-k} (1-\beta)^{n-j+k-i}$
Conditional Expectation	$E(X_t \mid X_{t-1}) = \rho X_{t-1} + n\beta$

Another approach which is completely different from Definition 2.1.3 has been proposed by Al-Osh and Alzaid (1991), namely first order Binomial Autoregressive (BAR(1)) process. The main concept of the BAR is based upon the hypergeometric thinning. Suppose the random variable X with range $\{0, \dots, N\}$, then the random variable $\frac{n}{N} \circ X$ conditioned on $X = x$ is defined as

$$P\left(\frac{n}{N} \circ X \mid X = x\right) = \frac{\binom{x}{k}\binom{N-x}{n-k}}{\binom{N}{n}} \qquad (2.6)$$

Likewise, BAR model is a stationary Markov Chain process. More details and applications can be found in the paper.

2. 2 Generalized Thinning Operation

The fundamental thinning operation has been shown that it is well applicable to discrete marginal distributions. Basic concept of thinning operation is that the summation over the Bernoulli random variable with the success probability $\alpha \in [0, 1]$

resulting in the Binomial distribution with (X, α). Further extensions of thinning operation has been widely studied. Now, the success probability α becomes a random variable in the range of $(0, 1)$. The definition of this type of thinning operator, is called random coefficient thinning operator which is defined as follows

Definition 2.2.1 (Random Coefficient Thinning Operation): Let X be the random variable with the range $\mathbb{N}_0$. Let α be the random variable in the range $[0, 1]$, and let X be independent with α. Then the random variable $\alpha \circ X$ is random coefficient thinning if the operator $' \circ '$ is binomial thinning operator. By the assumption of $\mu_\alpha \coloneqq E(\alpha)$ and $\sigma_\alpha^2 \coloneqq V(\alpha)$, some important properties of random coefficient thinning are tabulated in Table 2.2.1.

Table 2.2.1: Some important properties of random coefficient thinning

$E(\alpha \circ X)$	$\mu_\alpha \cdot E(X)$
$Cov(\alpha \circ X, X)$	$\mu_\alpha \cdot V(X)$
$V(\alpha \circ X)$	$\mu_\alpha^2 \cdot V(X) + \mu_\alpha(1 - \mu_\alpha) \cdot E(X) + \sigma_\alpha^2 \cdot E[X(X-1)]$

Definition 2.2.2 (RCINAR(1) model): Let ω be the i.i.d. random variable with range $(0, 1)$, α is independent with ε_t, and each ω is independent with $(X_s)_{s<t}$. The RCINAR(1) is defined recursively,

$$X_t = \omega \circ X_{t-1} + \varepsilon_t \tag{2.7}$$

Some authors prefer to define ω be α_t to show that the ω is allowed to be random itself. The difference of the model structure is apparent by comparing Eq. (2.2) and Eq. (2.7). Zheng et al. (2007) defined the first order random coefficient autoregressive model which is abbreviated by RCINAR(1). The important properties can be found in Wei β (2008) and Zheng et al. (2007). These articles also considered parameter estimation based upon the proposed model. This case study was considered in the earlier paper by McKenzie (1985). Weiβ (2008, Lemma 3.8) defined α to be distributed according to Beta (γ, β). Then, Weiβ (2008, Example 3.9) presented the application of RCINAR(1) process with Negative Binomial marginals. Joe (1996) provided a case study where α follows beta-binomial distribution in Joe (1996, Example (d)).

Al-Osh and Aly (1992) proposed a different type of generalization. They studied the INAR(1) process with Negative Binomial marginals, and considered Geometric random variables instead of Bernoulli random variables. Weiβ (2008) named it as iterated thinning operation. The main idea of iterated thinning operation can be defined by

$$\tau \star_\alpha X = \sum_{i=1}^{(\alpha\tau)\circ X} Y_i \quad 0 < \alpha, \rho < 1 \tag{2.8}$$

where Y_i is i.i.d. random variables in range $\mathbb{N}_0$, which are independent of both X and the thinning $(\alpha\tau) \circ X$, the distribution of Y_i depends on parameter α. The operator becomes concrete by considering the pgf

$$G_{\tau \star_\alpha X}(z) = G_X(1 - \alpha\tau + \alpha\tau \cdot G_Y(z)) \tag{2.9}$$

One should notice Eq. (2.9) reduces to the fundamental thinning operator when $G_Y(z) = z$. This type of generalization leads to generalized INAR(1) process which is known as Iterated Thinning INAR(1) process, abbreviated by IINAR(1). Recently, Ristic et al. (2009) proposed a new stationary integer-valued autoregressive process of the first order with Geometric counting series. They provided a closed-form for the innovation probability mass function (pmf). Ristic et al. (2012) continued the research works on parameter estimation of an INAR(1) model with Negative Binomial. Also, the asymptotic properties have been discussed along in the paper. Last but not least, the other self-decomposability marginal like Quasi-Binomial (QB) is discussed. In the standard Poisson INAR(1) model as in Example 2.1.1, the number of retained elements is assumed to have Binomial distribution with parameters (α, x). In the development, Alzaid and Al-Osh (1993) assumed that the number of retained elements has a QB distribution with parameters (p, θ, n). Consul and Mittal (1975) studied the urn model with QB distribution, and Shenton (1986) provided some properties. Weiβ (2008) summarized the Quasi-Binomial INAR(1) (QINAR(1)) model. He stated that the corresponding innovations cannot be obtained explicitly via this model. Despite the thinning operator and the generalized version being vastly studied, in the following section, we discuss an approach which appear as an alternative tool to model the count data in discrete time series modeling.

2.3 Pegram's Mixing Operation

Pegram (1980) was the pioneer who studied the approach which appears as an alternative tool to deal with count data. The concept of transition probability matrix of an ergodic Markov Chain multinomial model is introduced, which was later extended to AR model by Pegram (1980). Biswas and Song (2009) presented a unified framework of stationary ARMA models for discrete-valued time series based upon the Pegram's mixing operator, and it is abbreviated by $'*'$. Pegram mixing operator $'*'$ is defined as follows. For two independent random variables U and V, and for a given coefficient $\phi \in (0,1)$, Pegram's mixing operator $'*'$ mixes them to produce a random variable

$$Z: Z = (U, \phi) * (V, 1 - \phi) \tag{2.10}$$

with the marginal probability function is given by

$$P(Z = j) = \phi\, P(U = j) + (1 - \phi) P(V = j), \;\; j = 0, 1, \ldots \tag{2.11}$$

It is clear that Eq. (2.11) works for any discrete distributions. Given $X \sim D(p_0, p_1, \ldots)$ means that the domain of X is $\{0, 1, \ldots\}$ and $P(X = i) = p_i, i = 0, 1, \ldots$. This operator $*$

indicates a mixture of two discrete distributions, with the respective mixing weights ϕ and $1-\phi$.

The AR(p) processes based upon Pegram's mixing operation which is given by Biswas and Song (2009) as a special case is shown as follows. Suppose that time series X_t is discrete and its marginal probability mass function is time-independent. Let $X_t \sim D(p_0, p_1, \dots)$. Suppose ε_t's are also independently and identical distributed as the same $D(p_0, p_1, \dots)$. Denote $\mu = E(X_t)$ and $\sigma^2 = var(\varepsilon_t)$.

Definition 2.3.1 (Pegram's AR(p) Model) Let X_t be a discrete-valued stochastic process such that

$$X_t = (I[X_{t-1}], \phi_1) * (I[X_{t-2}], \phi_2) * \dots * \left(I\left[X_{t-p}\right], \phi_p\right) \\ * \left(\varepsilon_t, 1-\phi_1-\phi_2-\cdots-\phi_p\right) \tag{2.12}$$

is a mixture of $(p+1)$ discrete distributions, where $I[X_{t-1}], \dots, I\left[X_{t-p}\right]$ are p point masses, $I[\cdot]$ being indicator variable; and $\varepsilon_t \sim D(p_0, p_1, \dots)$ with the respective mixing weights being $\phi_1, \phi_2, \dots, \phi_p$ and $1-\phi_1-\cdots-\phi_p, \phi_j \in (0,1), j = 1, \dots, p$ and $\sum_{j=1}^{p} \phi_j \in (0,1)$. This implies that for every $t \in 0, \pm 1, \pm 2, \dots$, the conditional probability function takes the form of

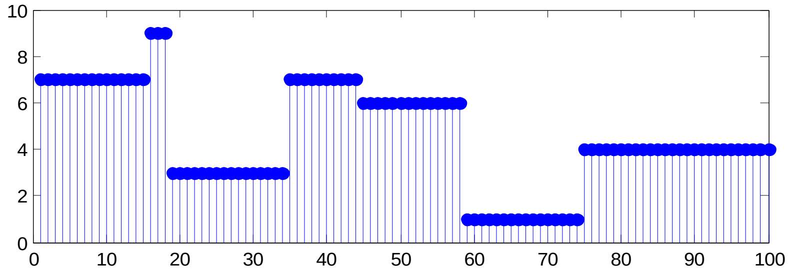
$$P\left(X_t = j | X_{t-1}, X_{t-2}, \dots, X_{t-p}\right) \\ = \left(1-\phi_1-\cdots-\phi_p\right)p_j + \phi_1 I[X_{t-1} = j] + \cdots + \phi_p I\left[X_{t-p} = j\right] \tag{2.13}$$

where $\phi_j, j = 1, \dots, p$ is chosen such that the polynomial equation $1-\phi_1 z-\cdots-\phi_p z^p = 0$ has roots lying outside of the unit disc.

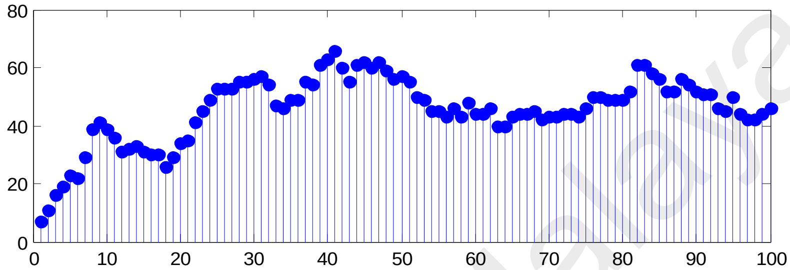
Comparatively, the Pegram's mixture operator appears to be more flexible than the thinning operator to construct Box and Jenkins' type stationary ARMA processes with arbitrary discrete marginal distributions. The flexibility yields an ARMA model for time series, particularly in categorical observations, which was unavailable with the extended thinning operators. This is simply because the categorical distribution is not infinitely divisible. Biswas and Song (2009) provided some inferences of the model such as parameter estimation. Model comparison with INAR(1) process have also been carried out in the paper. Furthermore, infant sleep data has been applied to Pegram's AR(1) process to show that the model performs well with the categorical data set. Biswas and Guha (2009) used auto-mutual information to analyze the correlation of categorical data. Recently, Song et al. (2013) studied on the statistical inferences of discrete-valued ARMA models by using categorical data.

The readers should be aware that there is another approach which is known as Jacobs and Lewis approach and it is analogue with Pegram's mixture process. Jacobs and Lewis (1978a, b) introduced the discrete-valued ARMA models (DARMA) to analyze the stationary sequences of dependent discrete random variables with specified marginal distribution. Correlation structure has been developed in the papers. However, Jacobs-Lewis approach generates the constant runs which is somehow not applicable in the real life situations. Other than these, Cui and Lund (2009) introduced a different approach of time series models for count data. They applied the renewal process to generate a correlated sequence of Bernoulli trials. The discrete marginals such as Poisson, Binomial and Geometric with the renewal models have readily been constructed. This model can be connected to the thinning process as a special case. See Cui and Lund (2009, Theorem 4) for relation of the model to Markov Chain process. Then, Cui and Lund (2010) implies the renewal process with Binomial marginal distributions. Some inferences and asymptotic properties have been studied.

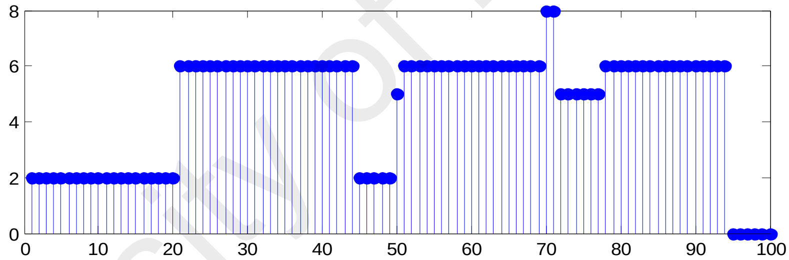
To understand more about the integer-valued Autoregressive of INAR(1), Pegram's AR(1) and Jacobs-Lewis DARMA models, we generated the sample runs for each model with Poisson marginals of parameter $\lambda = 5.0$ and $\alpha = 0.9$, respectively. Figure 2.3.1 (a)-(c) show the realizations of the models. It is clearly seen that constant runs resulting from discrete-valued first order Autoregressive (DAR(1)) by Jacobs-Lewis approach and Pegram's AR(1) may not suit most of the real life applications. Hence, our intention to seek for a better tool for discrete-valued time series models community is motivated. The models discussed here will later be compared with the new proposed model in Chapter 3.



(a)



(b)



(c)

Figure 2.3.1: The generated data by the (a) Poisson Pegram's AR(1) process; (b) Poisson INAR(1) process; (c) Poisson DAR(1) process, with parameters $\boldsymbol{\lambda = 5.0, \alpha = 0.9}$

2.4 Parameter Estimation

Parameter estimation is one of the most important aspects in statistical analysis. In this section, we intend to introduce the popular methods for parameter estimation in discrete-valued time series modelling. The well known MLE, CLS and YW Equation will be particularly discussed. The estimation problem connected to the discrete-valued time series models is more complicated than that of the conventional AR(1) process due to the complication of conditional probability functions of INAR(1) process. The conditional distribution of INAR(1) involves convolution. Comparatively, Pegram's AR(1) process has simpler conditional distribution of X_t given X_{t-1}.

2.4.1 Maximum Likelihood Estimation

To understand in details how the MLE works for the respective models, we take Definition 2.1.2 in Poisson process as an example (Al-Osh and Alzaid, 1987). Given that the likelihood function of a sample $(n+1)$ observations from the INAR(1) process is

$$L(\boldsymbol{x};\alpha,\lambda)=\left(\prod_{t=1}^{n}P(x_t)\right)\frac{[\lambda/(1-\alpha)]^{x_0}}{x_0!}\exp[-\lambda/(1-\alpha)] \tag{2.14}$$

where

$$P(\boldsymbol{x})=\exp(-\lambda)\sum_{i=0}^{\min(x_{t-1},x_t)}\frac{\lambda^{x_t-i}}{(x_t-i)!}\binom{x_{t-1}}{i}\alpha^i(1-\alpha)^{x_{t-1}-i}\quad ,t=1,2,\dots,n \tag{2.15}$$

and $\boldsymbol{x}=(x_0,x_1,\dots,x_n)$. To find MLE we take the log-likelihood function, then differentiate it and set it to be zero. Eq. (2.14) has two parameters (α,λ) to be estimated and it can be done by numerical computation. As the likelihood function contains the convolution of distribution, many researchers attempted to find a simpler way to handle the parameter estimations. Sprott (1983) suggested that one of the parameters can be estimated by the first-moment equation, and it can immediately eliminate one of the parameters. Then, the other parameter can be computed iteratively. The ML estimation can be found in most of the papers of INAR(1) models. For instance, Brannas (1994) presented the log-likelihood function explicitly for MLE. Bakouch (2009) considers the zero-truncated Poisson INAR(1) process. The log-likelihood is derived and can be solved by using **nlm** built-in function in the statistical package of R program. Jazi et al. (2012) studied the zero-inflated Poisson innovations. They compare the estimation between conditional and full maximum likelihood. Recently, Weiβ (2012) shows that it is possible to serve the innovation term ε_t as known observations. A full observation of INAR(1) process has been estimated. See Bhat and Adke (1981) for the similar estimation for branching process with known immigration. A further study by Pedeli and Karlis (2011) and Pedeli and Karlis (2013) who used the conditional ML to estimate the parameters of bivariate and multivariate INAR(1) processes respectively. Doukhan et al. (2012) investigated the relationship between weak dependence and mixing for discrete-valued processes. See 4.4 in the paper for the relevant model.

The likelihood functions of Pegram's AR(1) model is simpler and more direct than the conventional INAR(1) process. We adopted the formulation provided by Biswas and Song (2009): Given a time series data $\boldsymbol{x} = (x_0, x_1, \dots, x_n)$ from an AR(1) process where $p_j = 0, j > k$, the likelihood is

$$L(\boldsymbol{x}; \alpha, \lambda) = p_1 \left[\prod_{t=2}^{n} \{(1-\phi)p_t(x) + \phi I(X_{t-1} = x)\} \right] \tag{2.16}$$

where $p_t(x)$ has stationary solution with pmf for any discrete random variables. If $p_t(x)$ is Poisson marginal distributions, then $p_t(x) = \frac{e^{-\lambda}\lambda^x}{x!}$. Biswas and Song (2009) examined the model with categorical data. Furthermore, Biswas and Guha (2009) have done comparative study between auto-mutual information (AMI) and partial auto-mutual information (PAMI) for parameter estimation of categorical AR(1) process. More recently, Song et al. (2013) studied the MLE for categorical Pegram's AR(1) process using the sample relative frequencies. Numerical studies have been extensively presented.

2.4.2 Yule-Walker Equations

The autocovariance function (ACVF) of the discrete-valued models plays an important role in estimating the parameters by YW Equations. Here we focus on discrete-valued AR(p) models. Let the INAR(p) model be

$$X_t = \alpha_1 \circ X_{t-1} + \alpha_2 \circ X_{t-2} + \cdots + \alpha_p \circ X_{t-p} \tag{2.17}$$

Multiplying throughout Eq. (2.17) by X_{t-p} and taking the expectation

$$E(X_t X_{t-p}) = \alpha_1 E(X_{t-1} X_{t-p}) + \alpha_2 E(X_{t-2} X_{t-p}) + \cdots + \alpha_p E(X_{t-p} X_{t-p})$$

We obtain

$$\gamma_k = \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \cdots + \alpha_p \gamma_{k-p} \qquad k > 0, \qquad p = 1, 2, \ldots \tag{2.18}$$

The expectation of the innovation ε_t and X_{t-p} is zero since they are uncorrelated. Here p is the lag. If $p = 1$, Eq. (2.18) reduces to the ACVF for INAR(1) process. The ACF is easy to obtain via ACVF by dividing by γ_0 we get

$$\rho_k = \alpha_1 \rho_{k-1} + \alpha_2 \rho_{k-2} + \cdots + \alpha_p \rho_{k-p} \qquad k > 0, \qquad p = 1, 2, \ldots \tag{2.19}$$

We form the following linear equations by substituting $k = 1, 2, \ldots, p$ in Eq. (2.19). Box and Jenkins (1976) give a set of linear equations for $\phi_1, \phi_2, \ldots, \phi_p$ in terms of $\rho_1, \rho_2, \ldots, \rho_p$, that is

$$\rho_1 = \alpha_1 + \alpha_2\rho_1 + \cdots + \alpha_p\rho_{p-1}$$

$$\rho_2 = \alpha_1\rho_1 + \alpha_2 + \cdots + \alpha_p\rho_{k-2} \tag{2.20}$$

$$\vdots \qquad \vdots \qquad \vdots \quad \ldots. \quad \vdots \qquad \vdots$$

$$\rho_p = \alpha_1\rho_{p-1} + \alpha_2\rho_{p-2} + \cdots + \alpha_p$$

Then,

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} \qquad \boldsymbol{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} \qquad \boldsymbol{\mathcal{U}}_p = \begin{bmatrix} 1 & \cdots & \rho_{p-1} \\ \vdots & \ddots & \vdots \\ \rho_{p-1} & \cdots & 1 \end{bmatrix} \tag{2.21}$$

The solution for Eq. (2.21) for the parameters α in terms of autocorrelations may be written

$$\boldsymbol{\alpha} = \boldsymbol{\mathcal{U}}_p^{-1}\boldsymbol{\rho}_p \tag{2.22}$$

Otherwise, the estimators for autocovariance and autocorrelation values can be found via

$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{n-h}(X_t - \bar{X}_n)(X_{t+h} - \bar{X}_n)}{n} \qquad 0 \le h \le n-1 \tag{2.23}$$

and $\hat{\rho}(h) = \hat{\gamma}(h)/\hat{\gamma}(0)$, where $\bar{X}_n$ is the sample mean of the time series.

It is noted that the derivation of ACVF and ACF is not only limited to basic thinning operator but also the generalized thinning operation. See Zhang et al. (2012), who obtained the similar ACVF and ACF for signed generalized power series thinning operator. The proof is provided in the paper. Bakouch and Ristic (2010) consider zero truncated Poisson INAR(1) process with non-parametric estimator, which is known as YW estimation. The YW Equations approach is also applicable to Pegram's AR(1) process since the model gives similar ACVF and ACF formulations. See Biswas and Song (2009, Section 3.1). For analogue continuous case, the asymptotic distribution of YW estimator has been considered by Brockwell and Davis (1987) in section 8.7 and 8.8.

The partial autocorrelation function is a device to determine the order of AR(p) process. Denote by ϕ_{kj}, the jth coefficient in an autoregressive representation of order k, so that ϕ_{kk} is the last coefficient. From Eq. (2.19), the ϕ_{kj} satisfy the set of equations

$$\rho_j = \phi_{k1}\rho_{j-1} + \cdots + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{k(k-1)}\rho_{j-k+1} + \phi_{kk}\rho_{j-k} \quad j = 1,2,\dots,k$$

leading to the Yule-Walker equations Eq. (2.21), which may be written

$$\begin{bmatrix} 1 & \cdots & \rho_{k-1} \\ \vdots & \ddots & \vdots \\ \rho_{k-1} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_k \end{bmatrix}$$

The partial autocorrelation function can be usually found at the last coefficient of an autoregressive representation of order k, which is denoted by ϕ_{kk}. Otherwise, it can be calculated through the determinant, see Box and Jenkins (1976, page 66). For an autoregressive process of order p, the partial autocorrelation function ϕ_{kk} will be nonzero for k less than or equal to p and zero for k greater than p. This means that the partial autocorrelation function of a pth order autoregressive process has a cutoff after lag p. Hence, the researchers use it as the indicator to determine the order of an autoregressive process.

2.4.3 Conditional Least Squares Method

This section we consider the Poisson INAR(1) model for parameter estimation by the well-known CLS. The core concept of the CLS estimation is to minimize the sum of the squared distances of each observation $\boldsymbol{x} = (x_0, x_1, \dots, x_n)$ from its conditional expected value given the previous observations ζ_{t-1}, $E[X_t|\zeta_{t-1}] = \alpha X_{t-1} + \lambda$. Therefore, the CLS estimators of the parameters α and λ are obtained by minimizing the function

$$Q(\boldsymbol{x};\ \alpha, \lambda) = \sum_{t=1}^{n} \{X_t - \alpha X_{t-1} - \lambda\}^2 \tag{2.24}$$

which gives the estimated parameters explicitly

$$\begin{aligned} \hat{\alpha} &= \frac{\sum_{t=1}^{n}(X_{t-1} - \bar{X})(X_t - \bar{X})}{\sum_{t=1}^{n}(X_{t-1} - \bar{X})^2} \\ \hat{\lambda} &= \frac{1}{n}\sum_{t=1}^{n} X_t - \hat{\alpha}\frac{1}{n}\sum_{t=1}^{n} X_{t-1} \end{aligned} \tag{2.25}$$

Al-Osh and Alzaid (1987) compared the performance of CLS estimator with conditional maximum likelihood estimator (CMLE) and YW estimator. They compared the estimators in terms of bias and mean squared errors (MSE). See the discussion of Table I and Table II in the paper. Klimko and Nelson (1978) discussed the CLS estimation and inference for stochastic processes. Freeland (1998) provided a complete study and derivation of CLS estimation. Freeland and McCabe (2005) derived a corrected explicit expression for the asymptotic variance matrix of the CLS of Poisson INAR(1) process. They show that the asymptotic distribution of CLS estimators is equivalent to that of the estimators based on YW equations. Thus, both methods have similar performance in parameter estimation. Bakouch and Ristic (2010) studied the CLS estimation of zero truncated Poisson INAR(1) process. The estimated parameters are expressed explicitly and they indicated strong consistency. Asymptotical properties of the parameters have also been derived.

Other than the methods elaborated above, some authors tried generalized methods of moments (GMM) as well. Brännäs (1994) carried out a Monte Carlo simulation to compare the performance of the CLS estimator, exact ML and GMM. In fact, one important aspect in discrete-valued time series modelling is the regression models. Some estimation methods have been explored along this line. Brännäs (1995) studied ML estimation for prediction and control for time series count data model. Gourieroux et al. (1984) used ordinary least squares (OLS) approach to deal with the serially correlated data. The readers should be aware that one readily package namely **tscount** from R (Liboschik et al. 2015), which is likelihood-based methods for the framework of count time series following generalized linear models.

2.5 Concluding Remarks

Chapter 2 reviewed all potential models for discrete-valued time series models. Particularly, it emphasized the *history* of INAR(1) and Pegram's AR(1) models. The INAR(1) is defined based upon thinning operator and Pegram's AR(1) model is constructed from the idea of Pegram's mixing operation respectively. The generalized thinning INAR(1) processes have also been explored in the literature reviews. The important inferences like parameter estimation are also elaborated. The overview of Chapter 2 is important. It becomes the root to motivate us to construct a new model with better flexibility. The idea and significant inferences will be discussed in the following chapters.

CHAPTER 3: MIXED PEGRAM AND THINNING INTEGER-VALUED AUTOREGRESSIVE MODEL

3.0 Introduction

This chapter introduces a new model which is constructed from the combination of Binomial Thinning and Pegram's operators. The proposed model, known as Mixture of Pegram and Thinning Integer-Valued Autoregressive (MPT) process appears to be a new contribution to modelling time series of counts. For simplicity, the first order MPT process is considered and is abbreviated by MPT(1). It is well-known that finite mixture models provide more flexibility in empirical modelling. Furthermore, the models possess a simple interpretation and are able to cater for multimodality in the data. Therefore, it is of the interest to construct MPT(1) model which is given in Section 3.1. Section 3.2 interprets this model. Section 3.3 discusses the model stationarity and provides some important properties of the proposed model. Section 3.4 shows fittings with some marginal distributions. Several discrete marginal distributions, with finite and infinite ranges, have been considered for the MPT(1) process. Section 3.5 concludes.

3.1 Model Construction

It is important to note that the INAR(1) process which is defined in Chapter 2, Definition 2.1.2 (Eq. 2.2) comprised of two elements: the thinned part ($\alpha \circ X_{t-1}$) and the innovation term (ε_t). Pegram's mixing operation on the two independent non-negative integer-valued random variables $\alpha \circ X_{t-1}$ and ε_t with respective mixing weights of ϕ and $1 - \phi$ yields a new mixture model given below.

Definition 3.1.1 (Mixture of Pegram-INAR(1)): Consider two independent integer-valued random variables defined in Eq. (2.2), and let $\phi \in (0, 1)$. The initial value of the process X_0, has an initial distribution of $P(X_0 = i) = \pi_0$, then for every $t \in \{0, \pm 1, \pm 2, \dots\}$

$$X_t = (\phi, \alpha \circ X_{t-1}) * (1 - \phi, \varepsilon_t) \quad (3.1)$$

is a Pegram mixture of thinning and innovation processes. We denote it by MPT (mixture of Pegram and thinning). The parameter ϕ is known as mixing weight of the mixture model in Eq. (3.1), and it mixed the thinning part and the innovation term with the respective mixing weights ϕ and $1 - \phi$. The pgf is given by

$$G_{X_t}(z) = \phi G_{X_{t-1}}(1 - \alpha + \alpha z) + (1 - \phi) G_{\varepsilon_t}(z) \quad (3.2)$$

for $|z| \leq 1, z \in \mathbb{C}$.

Eq. (3.2) works for any discrete marginal distribution, including non-infinitely divisible distributions like Binomial distribution. In the following section, we will discuss in detail the MPT(1) applied in an infinite range of counts such as Poisson and Negative Binomial, as well as the finite range of counts like Binomial marginal distribution. Compared with the existing discrete-valued models in Chapter 2, the proposed model has simpler closed-form expression for the probability distribution of the innovation term ε_t.

3.2 Interpretation

We have been familiarized with the interpretation of INAR(1) process. It can be readily applied to our proposed model by integrating the idea of the mixing process. For instance, the model interpretation regarding the birth and death process, see Ross (2000, Section 6.3) for an introduction to birth and death processes. For our model interpretation, we mixed each individual at time $t - 1$ who has probability α of continuing to be alive at time t, and the number of births following certain discrete marginal distribution at each time t, with respective mixing weights of ϕ and $1-\phi$. Other interpretations in different field such as short-term disability claims and infinite server queue can be suited into the proposed model in the similar manner. The mixture model with new interpretation has been proposed here to manage different real life situations.

3.3 Properties of MPT(1) Model

First, we shall show that the MPT(1) process $\{X_t\}$ is stationary. Grunwald et al. (2000) have examined the stochastic properties and stationarity of AR(1) models with conditional linear mean (CLAR(1)). Let $(Y_t)_{t\in\mathbb{Z}}$ be a time-homogeneous first order Markov process on a sample space $S \subseteq R$. $(Y_t)_{t\in\mathbb{Z}}$ is said to have a first-order conditional linear autoregressive (CLAR(1)) structure if

$$E(Y_t \mid Y_{t-1}) = \delta Y_{t-1} + \gamma \tag{3.3}$$

where δ and γ are real numbers.

The MPT(1) process $(X_t)_{t\in\mathbb{Z}}$ of Definition 3.1.1 has a conditional linear expectation given by Eq. (3.6) and is of the form (3.3). Thus the MPT(1) process has a first-order conditional linear autoregressive (CLAR(1)) structure. It is an irreducible Feller chain and stationary with specific marginal for all *t*. Grunwald et al (2000) gave sufficient conditions for convergence to an ergodic distribution π on sample space S. For convenience the result adapted from Grunwald et al (2000) is summarized in the following theorem.

Theorem 3.3.1 (Proposition 3, Grunwald et al, 2000) Let$(X_t)_{t\in\mathbb{Z}}$be a CLAR(1) process and also an irreducible Feller chain. Furthermore if $S \subseteq [0,\infty)$ and$0 \le \delta < 1$, then $(X_t)_{t\in\mathbb{Z}}$ is ergodic and the convergence of $\parallel P_t(s,.) - \pi(.) \parallel \to 0, t \to \infty$ is geometrically fast, where $P_t(s,A) = P(X_t \in A | X_0 = s)$ and $\parallel \cdot \parallel$ is the total variation norm. With specified marginals for all t, $(X_t)_{t\in\mathbb{Z}}$ is a stationary process. As a consequence of $E[X_t] = \lambda$, $var(X_t) = \lambda$ (see Theorem 5.1.3 and 5.1.4), we get the following results regarding the stationary process mean and variance for MPT(1).

Proposition 3.3.1 (Stationary process mean): Let $\{X_t\}$ be a stationary process with Poisson (λ). For $|E(X_0)| < \infty$, $\alpha \in [0,1], \phi \in (0,1)$. If $E(X_0) = \lambda$, then $E(X_t) = \lambda$ for $t \geq 0$.

Proposition 3.3.2 (Stationary process variance): Let $\{X_t\}$ be a stationary process with Poisson (λ) with $\alpha \in [0,1], \phi \in (0,1)$, then $var(X_t) = \lambda$ for $t \geq 0$.

Particularly, the stationarity of Poisson MPT(1) will be discussed in the next section. Now, we present the regression properties of MPT(1) process.

Theorem 3.3.2 (Conditional pgf of MPT(1) Process): Let X_t be a stationary process following Definition 3.1.1. By taking $G_{X_t|X_{t-1}}(z) = E(z^{X_t}|X_{t-1})$, it is easy to obtain the conditional pgf of the MPT(1) process which is given by

$$G_{X_t|X_{t-1}}(z) = \phi(1-\alpha+\alpha z)^{X_{t-1}} + (1-\phi)G_{\varepsilon_t}(z) \tag{3.4}$$

Proof:

Conditional pgf of $X_t|X_{t-1}$ is given by

$$\begin{aligned} G_{X_t|X_{t-1}}(z) &= E(z^{X_t}|X_{t-1}) \\ &= \phi E(z^{\alpha\circ X_{t-1}}|X_{t-1}) + (1-\phi)E(z^{\varepsilon_t}|X_{t-1}) \\ &= \phi(1-\alpha+\alpha z)^{X_{t-1}} + (1-\phi)G_{\varepsilon_t}(z) \end{aligned}$$

The corresponding conditional probability function takes the form

$$P(X_t = i|X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + (1-\phi)P(\varepsilon_t = i) \tag{3.5}$$

for all $t \in 0, 1, 2, \ldots$.

Corollary 3.3.1 (Conditional Moments of MPT(1) Process): The conditional expectation of X_t given X_{t-1} is

$$E(X_t|X_{t-1}) = \phi\alpha\, X_{t-1} + (1-\phi)E(\varepsilon_t) \tag{3.6}$$

and the conditional variance is defined by

$$Var(X_t|X_{t-1}) = E\left({X_t}^2|X_{t-1}\right) - E(X_t|X_{t-1})^2 \tag{3.7}$$

where $E\left({X_t}^2|X_{t-1}\right) = \sum_{x_t=0}^{\infty} {x_t}^2\, P(X_t|X_{t-1})$; $P(X_t|X_{t-1})$ is the conditional probability in Eq. (3.5). The formulation of conditional mean and variance is determined based on the marginal assumption of random variable X_t. See the following examples. The autocorrelation is important in measuring the serial dependence. It can be easily obtained using Eq. (3.7)**.** Note that

$$E(X_t X_{t-1}|X_{t-1}) = \phi\alpha X_{t-1}^2 + (1-\phi)\mu_\varepsilon X_{t-1} \tag{3.8}$$

where $\mu_\varepsilon = E(\varepsilon_t)$ and $E(X_t) = \frac{(1-\phi)\mu_\varepsilon}{1-\phi\alpha}$, and hence

$$E(X_t X_{t-1}) = \phi\alpha E(X_{t-1}^2) + (1-\phi)\mu_\varepsilon \frac{(1-\phi)\mu_\varepsilon}{1-\phi\alpha} \tag{3.9}$$

Lemma 3.3.1 (Lag-one Autocovariance Function): Let X_t be a process follows Definition 3.1.1, the lag-one ACVF is given by

$$Cov(X_t, X_{t-1}) = \phi\alpha Var(X_{t-1}) \tag{3.11}$$

Lemma 3.3.2 (Lag-one Autocorrelation Function): Let X_t be a process follows Definition 3.1.1, the lag-one ACF is given by

$$corr(X_t, X_{t-1}) = \rho_x(1) = \phi\alpha \tag{3.12}$$

It can be observed that the ACVF and ACF have similar expression as in the mixture model in Eq. (2.13). By induction, we obtained the lag-h ACF of the MPT(1) process to be $\rho_x(h) = corr(X_t, X_{t-h}) = (\phi\alpha)^{|h|}$. Since $\sum_{h=-\infty}^{\infty}(\phi\alpha)^{|h|} < \infty$, the spectral density function of the process is given by

$$f_{XX}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma(h)e^{-ih\omega} = \frac{\gamma_0}{2\pi}\frac{1-(\phi\alpha)^2}{1+(\phi\alpha)^2 - 2\phi\alpha cos\omega} \tag{3.13}$$

where $\omega \in (-\pi, \pi]$ and $\gamma(h) = Cov(X_t, X_{t-h})$. The joint pgf of the process is significant to define the time-reversibility of the process.

Theorem 3.3.3 (Joint pgf of MPT(1) Process): Let X_t be a process following Definition 3.1.1, the joint pgf is given by

$$G_{X_t,X_{t-1}}(z_1, z_2) = \phi G_X\big(z_2(1-\alpha+\alpha z_1)\big) + (1-\phi)G_\varepsilon(z_1)G_X(z_2) \tag{3.14}$$

Proof:

$$\begin{aligned}
G_{X_t,X_{t-1}}(z_1, z_2) &= E({z_1}^{X_t}\,{z_2}^{X_{t-1}}) \\
&= \phi E(z^{\alpha\circ X_{t-1}}\,{z_2}^{X_{t-1}}) + (1-\phi)E(z^{\varepsilon_t}\,{z_2}^{X_{t-1}}) \\
&= \phi E\big({z_2}^{X_{t-1}}E({z_1}^{\alpha\circ X_{t-1}}|X_{t-1})\big) + (1-\phi)E({z_1}^{\varepsilon_t})E({z_2}^{X_{t-1}}) \\
&= \phi E[z_2(1-\alpha+\alpha z_1)^{X_{t-1}}] + (1-\phi)G_{\varepsilon_t}(z_1)G_{X_{t-1}}(z_2) \\
&= \phi G_{X_{t-1}}\big(z_2(1-\alpha+\alpha z_1)\big) + (1-\phi)G_{\varepsilon_t}(z_1)G_{X_{t-1}}(z_2)
\end{aligned}$$

The innovation term still remains unknown thus far. In the next section, we show some examples by fitting the power series distributions like Poisson, Negative Binomial and Binomial in MPT(1) model. μ_ε and ${\sigma_\varepsilon}^2$ together with their respective discrete marginal distribution will be defined accordingly. Also, the corresponding simulated sample paths will be generated.

3.4 MPT(1) Model with Discrete Marginal Distributions

The preliminary study begins with simple marginal distributions. Infinite range of discrete marginal distributions have been applied to MPT(1) model. Important fundamental and regression properties will be discussed. Particularly, Poisson, Negative Binomial and new Geometric (Ristic et al., 2009) will be involved in the study. The discussion on finite range of Binomial marginal will also be presented to show that the proposed model is compatible. Next, the expression of several discrete marginals fitted in the MPT(1) model have been performed.

3.4.1 Poisson Process

Theorem 3.4.1 (Stationary Poisson MPT(1) Process) Let X_t be a process following MPT(1) model of Definition 2.1.1. Then X_t is a stationary process and possesses a unique stationary Poisson marginal with mean (λ). The conditional pgf of $X_t|X_{t-1}$ is given by

$$G_{X_t|X_{t-1}}(z) = \phi(1-\alpha+\alpha z)^{X_{t-1}} + e^{\lambda(z-1)} - \phi e^{\lambda\alpha(z-1)}$$

with innovation process ε_t has pgf given by

$$G_{\varepsilon_t}(z) = \frac{1}{1-\phi}\left\{e^{\lambda(z-1)} - \phi e^{\lambda\alpha(z-1)}\right\} \tag{3.15}$$

Example 3.4.1 (Poisson MPT(1) Model): Let X_t be a stationary process with Poisson marginals $Poi(\lambda)$. The transition probability is given by

$$P(X_t = i|X_{t-1} = j) = \phi\binom{j}{i}\alpha^i(1-\alpha)^{j-i} + (1-\phi)P(\varepsilon_t = i)$$

where the pmf of innovation process ε_t is

$$P(\varepsilon_t = i) = \frac{1}{1-\phi}\left\{\frac{e^{-\lambda}\lambda^i}{i!} - \phi\frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right\} \quad i = 0,1,\dots \tag{3.16}$$

and the condition $e^{\lambda(1-\alpha)} < \frac{1}{\phi}$ is necessary for (3.16) to be defined as positive. The moments of innovation process ε_t are

$$E(\varepsilon_t) = \left(\frac{1-\alpha\phi}{1-\phi}\right)\lambda = \mu_\varepsilon$$

$$Var(\varepsilon_t) = \frac{1}{1-\phi}(\lambda^2+\lambda) - \frac{\phi}{1-\phi}(\lambda\alpha + \lambda^2\alpha^2) - \left(\frac{1-\alpha\phi}{1-\phi}\right)^2 \lambda^2$$

Corollary 3.4.1 (Conditional Moments of Poisson MPT(1) Process): Let X_t be a process following Theorem 3.4.1, the conditional moments are

(a) $E(X_t|X_{t-1}) = \phi\alpha X_{t-1} + (1-\alpha\phi)\lambda$

(b) $var(X_t|X_{t-1}) = \phi\alpha^2(1-\phi)X_{t-1}^2 + \phi\alpha[(1-\alpha) - 2(1-\phi\alpha)\lambda]X_{t-1} + (1-\phi\alpha)\lambda$
$+\phi\alpha(2-\alpha(1-\phi))\lambda^2$

(3.17)

Poisson process applied to time series models are often taken as benchmark in real life applications due to its simplicity for possessing single parameter. We generate the realizations of Example 3.4.1 with **stem** built-in function in Matlab 2013a to compare the paths with Figure 2.3.1 in Chapter 2. The sample path of Figure 3.4.1 gives better and reasonable sequences for time series modelling.

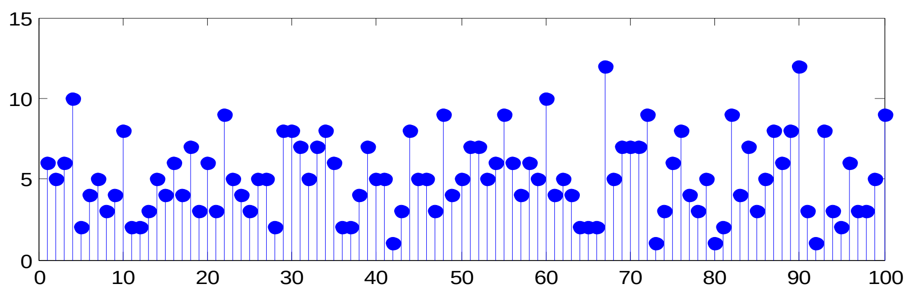


Figure 3.4.1: The realizations by Poisson MPT(1) process with parameters $\boldsymbol{\phi = 0.5,\ \lambda = 5.0, \alpha = 0.9}$

Figure 3.4.2 shows some sample paths generated by Example 3.4.1 with the aids of inverse transform method. Samples of size 100 have been generated with some arbitrary parameter values: $\lambda = 1.0, 2.0, 3.0$; $\alpha = 0.3, 0.5$; $\phi = 0.1, 0.2$. The sample paths of Poisson MPT(1) process are generated through inverse transform method using the cumulative distribution of Example 3.4.1. Histogram in Figure 3.4.2 shows that for a larger λ, larger count values are involved, while for small λ we tend to get high frequencies of small counts. Figure 3.4.3 presents the pmf of Poisson MPT(1) process with several combinations of parameter values. It can be seen that the varying of parameter values determines different shapes of the distribution. The distribution is long-tailed with zero probability when x is more than 8.

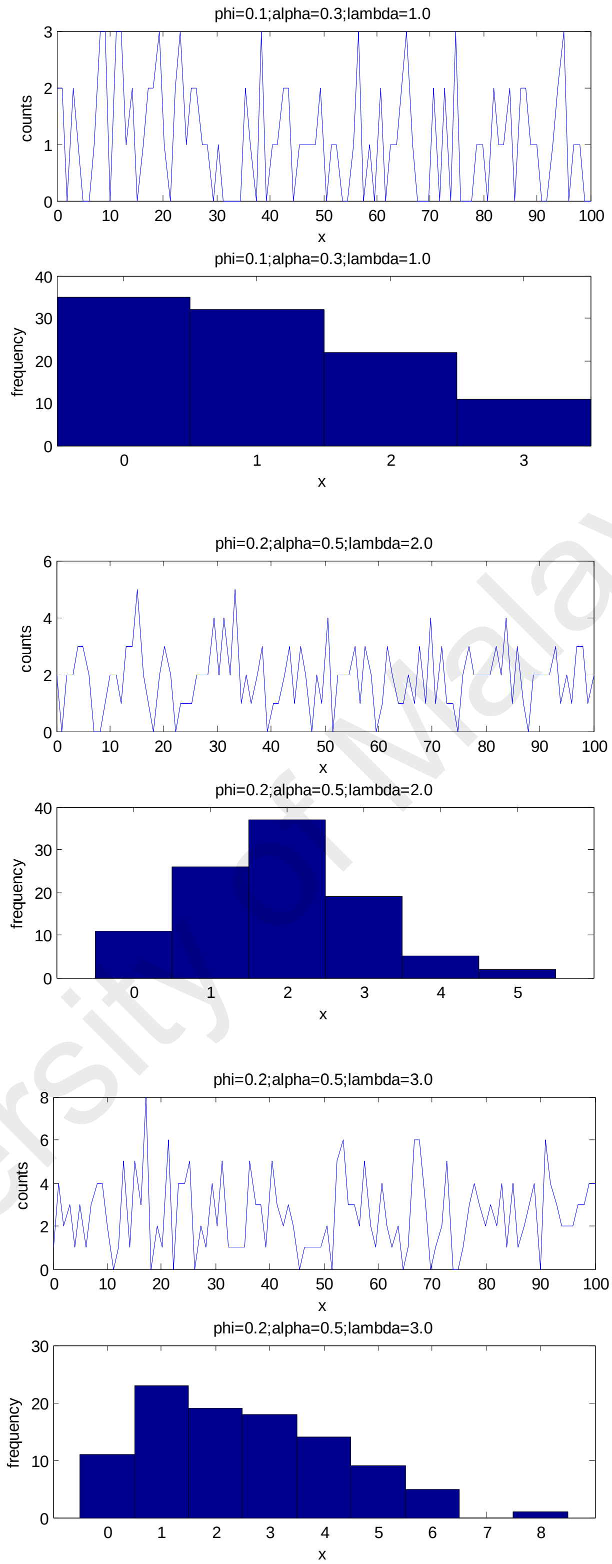


Figure 3.4.2: Simulated sample paths and histogram of Poisson MPT(1) process for $\boldsymbol{\lambda = 1.0, 2.0, 3.0;\ \alpha = 0.3, 0.5;\ \phi = 0.1, 0.2}$

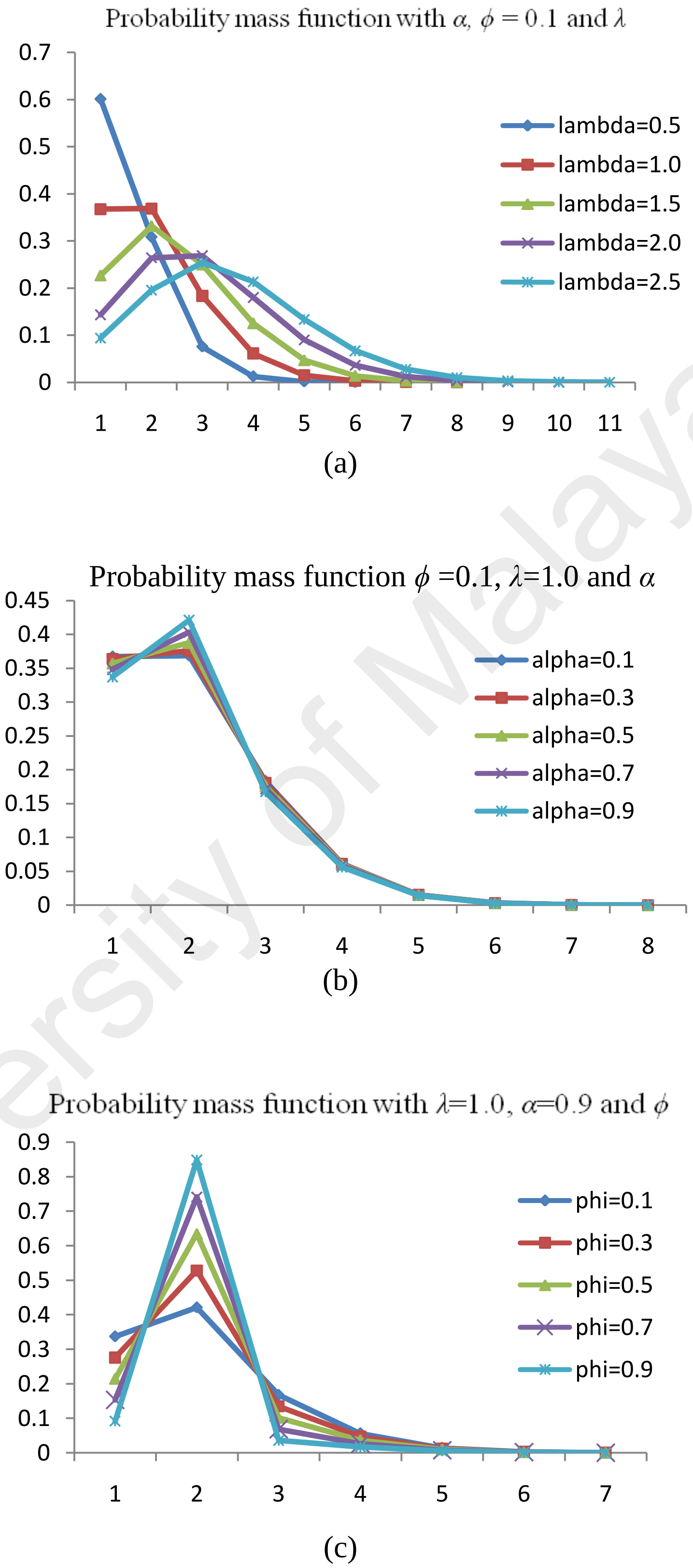


Figure 3.4.3: Simulated probability mass function with various combination of parameters. x-axis represents count data (x) and y-axis represents pmf.

3.4.2 Negative Binomial Process

Example 3.4.2 (Negative Binomial MPT(1) Model): Let $\{X_t\}$ be a process with Negative Binomial marginals, $NB(k,P), k > 0, P > 0$ and $Q = 1 + P$. Then the innovation process ε_t has pgf given by

$$G_{\varepsilon_t}(z) = \frac{1}{1-\phi}\{(1+P-Pz)^{-k} - \phi(1+\alpha P-\alpha Pz)^{-k}\} \tag{3.18}$$

To ensure positive innovation term, the marginal distribution must fulfil the condition of $\phi < \left(\frac{1+P}{1+\alpha P}\right)^{-k}$. Consequently the mean for innovation term is

$$E(\varepsilon_t) = \left(\frac{1-\alpha\phi}{1-\phi}\right)kP \tag{3.19}$$

and the variance is

$$\begin{aligned} Var(\varepsilon_t) = \frac{1}{1-\phi}[(kP)^2 + kP(1+P)] - \frac{\phi}{1-\phi}[\alpha kP + \alpha^2(kP)^2] \\ - \left(\frac{1-\alpha\phi}{1-\phi}\right)^2 (kP)^2 \end{aligned} \tag{3.20}$$

The conditional expectation is simply

$$E(X_t|X_{t-1}) = \phi\alpha X_{t-1} + (1-\alpha\phi)kP$$

and $E(X_t^2|X_{t-1}) = \phi\alpha(1-\alpha)X_{t-1} + \phi\alpha^2 {X_{t-1}}^2 + kP[1+P(1+k)] - \phi k\alpha P[1 + \alpha P(1+k)]$ can be used to form conditional variance. Figure 3.4.4 and Figure 3.4.5 present the simulated sample paths and frequency histogram of Negative Binomial MPT(1) process, as well as the probability mass functions with arbitrary parameter values. When r is increased, the mean of the simulated data shifted to the right, resulting large sample mean. Also, we tend to obtain larger count values. Second and third histogram in Figure 3.4.4 show that the NB MPT(1) model seems to detect multimodality in the simulated data. For Geometric MPT(1) process, we simply assign $k = 1$. Figure 3.4.5 presents the simulated path for different value of P. For large value of P, the model gives either 0 or 1 for counts.

Comparatively with Example 2.1.2, the form of innovation process in Example 3.4.2 gives simpler expression for the fitting of Negative Binomial marginal distribution. For this point of view, the proposed NB MPT(1) process can be very useful in real application. We are always reserved to seek for simpler and better models as our first aim.

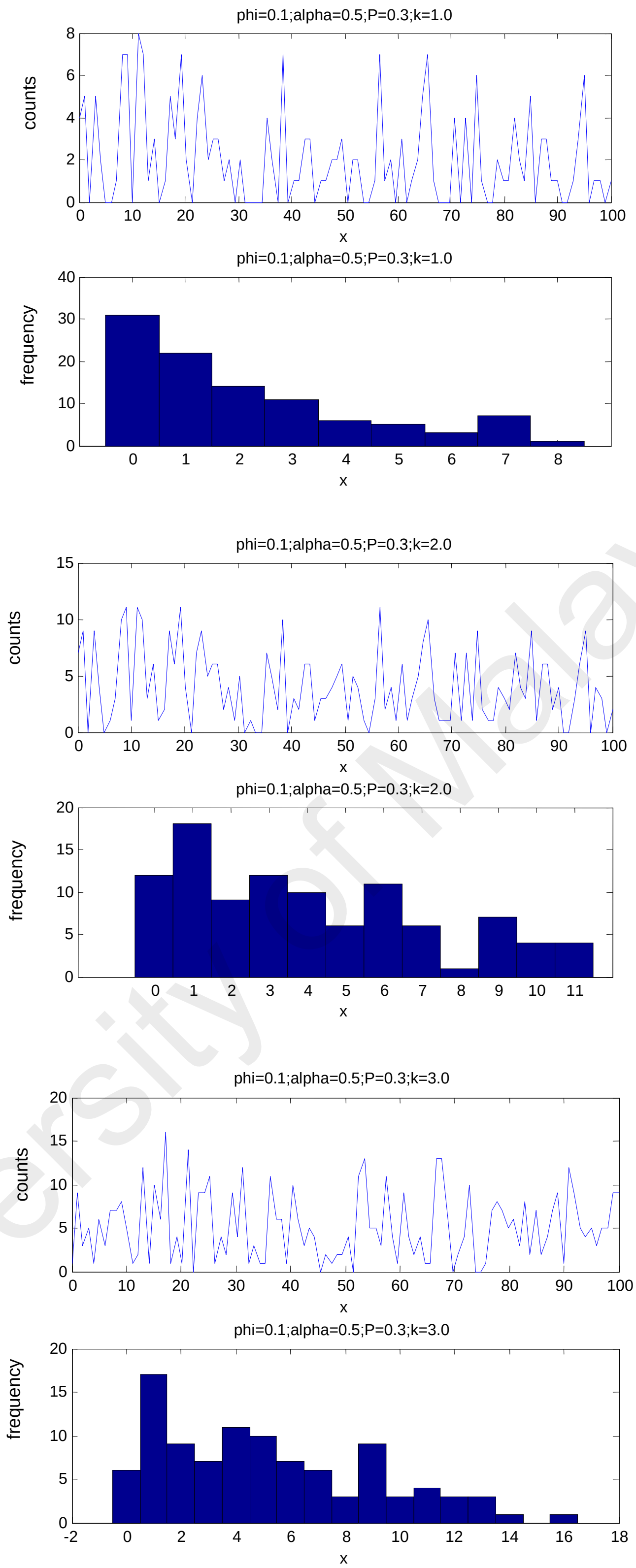


Figure 3.4.4: Simulated sample paths and frequency histogram of Negative Binomial MPT(1) process for $\phi = 0.1;\ \alpha = 0.5; P = 0.3;$ $k = 1.0, 2.0, 3.0.$

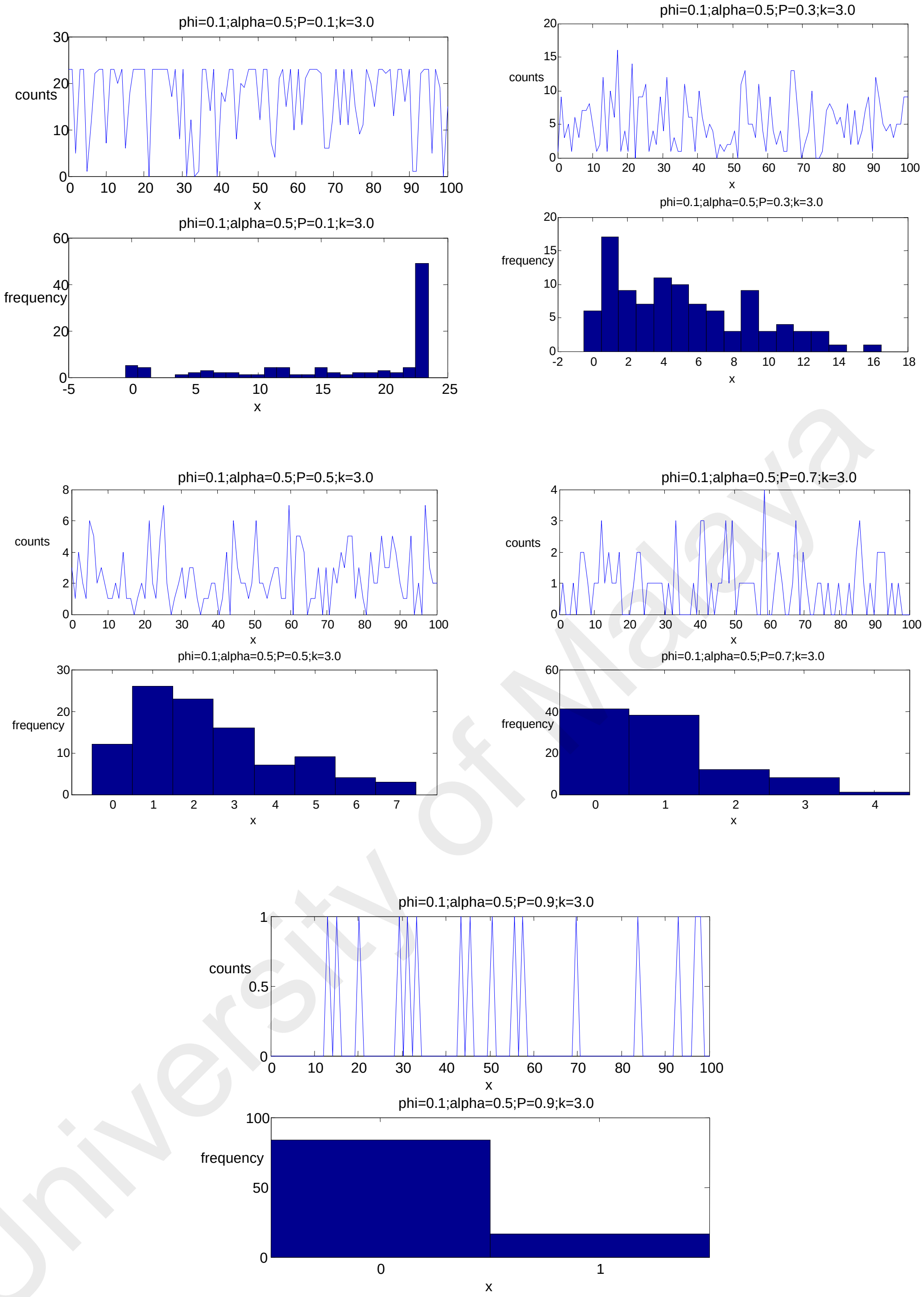


Figure 3.4.5: Simulated sample paths and frequency histogram of Negative Binomial MPT(1) process for $\phi = 0.1$; $\alpha = 0.5$; $P = 0.1, 0.3, 0.5, 0.7, 0.9$; $k = 3.0$

3.4.3 New Geometric Process

Example 3.4.3 (New Geometric MPT(1) Process): Let $\{X_t\}$ be a process with new Geometric marginal $Geo(1, p)$. Then the pgf of the innovation term is given by

$$G_{\varepsilon_t}(z) = \frac{1}{1-\phi}\left\{\frac{1}{1+p-pz} - \phi\left[\frac{1+\alpha-\alpha z}{1+\alpha(1+p)-\alpha(1+p)z}\right]\right\} \tag{3.21}$$

It is easy to derive the conditional pmf of (3.21) by multiple derivatives which is given by

$$P(\varepsilon_t = 0) = \frac{1}{1-\phi}\left\{\frac{1}{1+p} - \phi\frac{1+\alpha}{1+\alpha(1+p)}\right\}$$

$$P(\varepsilon_t = l) = \frac{1}{1-\phi}\left\{\frac{p^l}{(1+p)^{l+1}} - \alpha^l\phi p\left[\frac{(1+p)^{l-1}}{\{1+\alpha(1+p)\}^{l+1}}\right]\right\} \quad l = 1,2,3,\ldots \tag{3.22}$$

with the mean and variance of the innovation process are

$$E(\varepsilon_t) = \left(\frac{1-\alpha\phi}{1-\phi}\right)p \tag{3.23}$$

and

$$Var(\varepsilon_t) = \frac{2p^2}{1-\phi} - \frac{\phi}{1-\phi}\alpha^2 p(1+p) - \frac{(1-\phi\alpha)p}{1-\phi}\left(1 - \frac{(1-\phi\alpha)p}{1-\phi}\right) \tag{3.24}$$

One suggestion to obtain the conditional mean and variance of new Geometric is via k-step-ahead conditional statistical measures with $k = 1$. The form is relatively complex and it is omitted here and could be considered for further research work. The discrete marginals expressed in Examples 3.4.1-3.4.3 are appropriate to analyze overdispersed data. From Figure 3.4.5 we observe that the lower p value in new Geometric generates low count data, for an example $p = 0.3$ generates $x = 0, 1, 2$. For high value of $p = 0.9$ it generates the count data $x = 0, 1, \ldots, 7$. The simulated pmf provides similar pattern for fixed ϕ and α. The patterns may differ for different values of the parameters. Likewise, the new Geometric MPT(1) process has a long-tailed distribution.

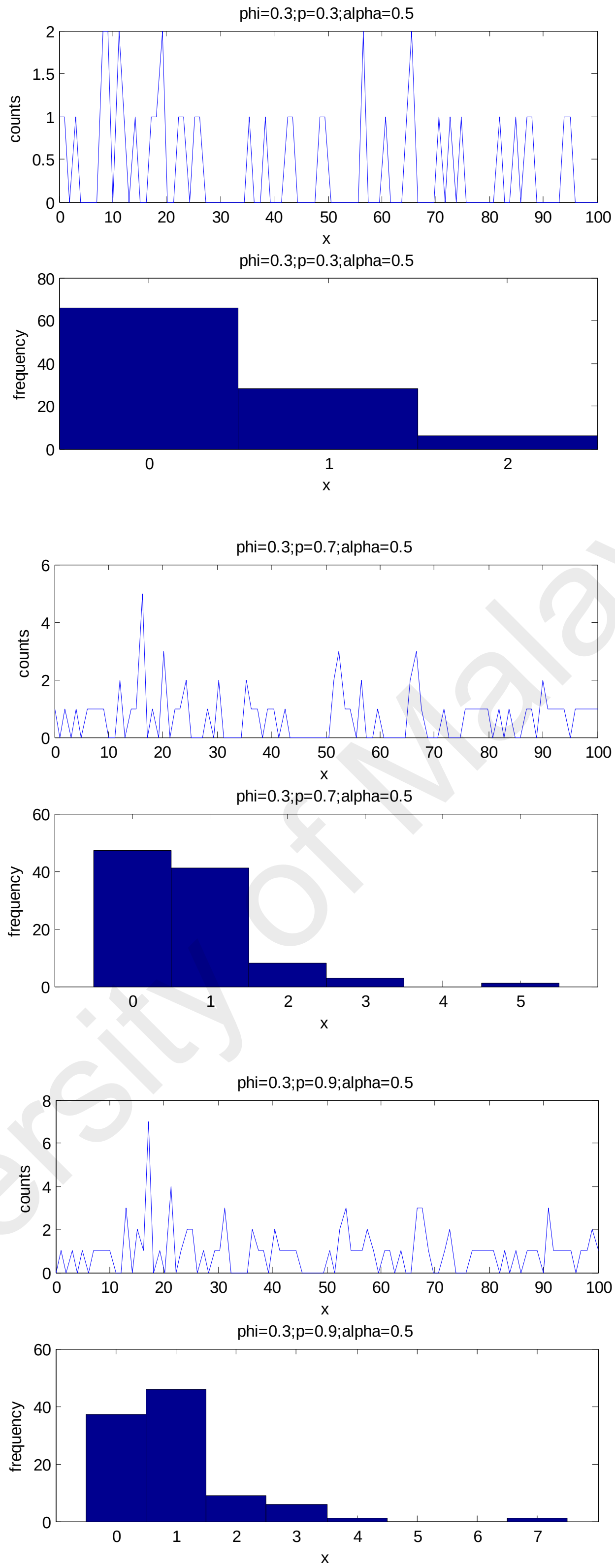


Figure 3.4.6: Simulated sample paths and frequency histogram of New Geometric MPT(1) process for $\phi = 0.3;\ \alpha = 0.5; p = 0.3, 0.7, 0.9$

3.4.4 Binomial Process

Example 3.4.4 (Binomial MPT(1) Process): Let $\{X_t\}$ be a process with Binomial marginals $Bin(N,p)$, N is a positive integer and $0 < p < 1$. Then the innovation process ε_t has pgf given by

$$G_{\varepsilon_t}(z) = \frac{1}{1-\phi}\left\{\binom{N}{i} p^i (1-p)^{n-i} - \phi \binom{N}{i} (\alpha p)^i (1-\alpha p)^{n-i}\right\} \tag{3.25}$$

Similar to Negative Binomial MPT(1) process, the positive innovation term of Binomial MPT(1) must fulfil the condition of $\phi < \left(\frac{1-p}{1-\alpha p}\right)^N$. The mean and variance of innovation process can be obtained from Eq. (3.25) to arrive at

$$E(\varepsilon_t) = \left(\frac{1-\alpha\phi}{1-\phi}\right) Np \tag{3.26}$$

$$\begin{aligned} Var(\varepsilon_t) = {} & \frac{1}{1-\phi}[(Np)^2 + Npq] - \frac{\phi}{1-\phi}[(N\alpha p)^2 + N\alpha p(1-\alpha p)] \\ & - \left(\frac{1-\alpha\phi}{1-\phi}\right)^2 (Np)^2 \end{aligned} \tag{3.27}$$

The conditional mean is given by

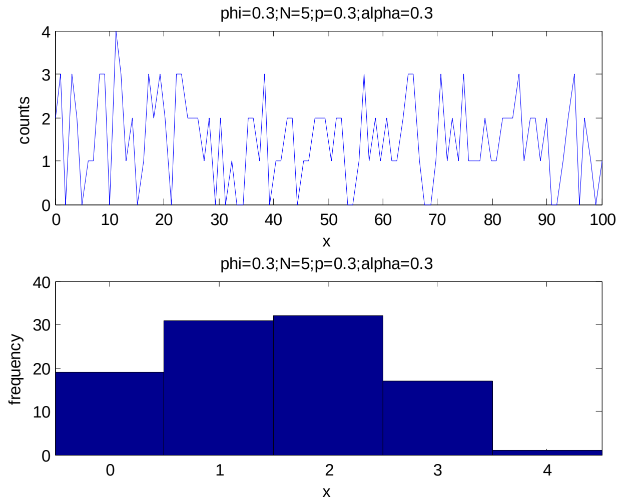
$$E(X_t|X_{t-1}) = \phi\alpha X_{t-1} + (1-\alpha\phi)Np$$

and the conditional variance can be obtained via

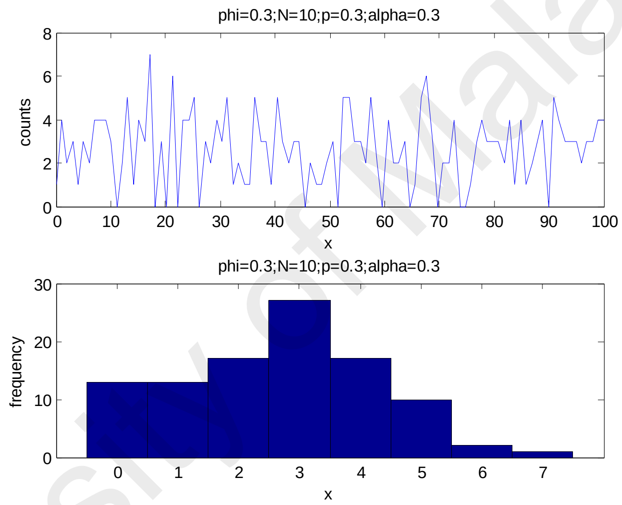
$$\begin{aligned} E(X_t^2|X_{t-1}) = {} & \phi\alpha X_{t-1}[(1-\alpha) + \alpha X_{t-1}] + Np[(1-p) + (Np)] \\ & - \phi N\alpha p[(1-\alpha p) + N\alpha p] \end{aligned}$$

and conditional expectation is as shown. In similar way we generate the realizations by Binomial MPT(1) model which is shown in Figure 3.4.6 and Figure 3.4.7. Similarly, Figure 3.4.6 shows that the larger sample sizes produce larger counts. One can also see the mixture of two binomial distributions presents bell shaped distribution when N, the sample size, is increased.

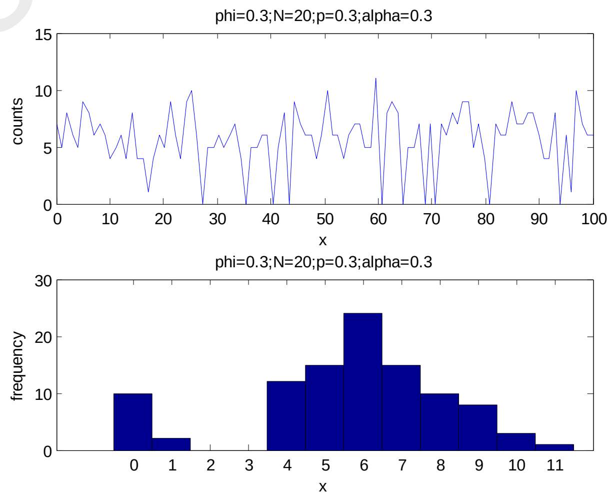
Unlike the Binomial AR(1) model as defined in Definition 2.1.3, the proposed Binomial MPT(1) model holds the similar expression of Definition 3.1.1. Majority properties are easily shared for any marginal fitting. The only concern is that for Binomial case, we consider the domain of x takes $x \in [0, N]$. The real data applications of discrete marginal distributions in MPT(1) process will be shown in the later chapter. See Chapter 7 for the comparison with the currently existing discrete time series models.



(a)



(b)



(c)

Figure 3.4.7: Simulated sample paths and frequency histogram of Binomial MPT(1) process for $\alpha = 0.3$; $\phi = 0.3$; $p = 0.3$; (a) $N = 5$, (b) $N = 10$, (c) $N = 20$

3.5 Concluding Remarks

Chapter 3 proposes a discrete time series model arising from the combination of the existing binomial thinning (or thinning) and Pegram's operators. The mixture of thinning and Pegram's operators is introduced with the first order AR process, abbreviated by MPT(1), as for the preliminary study. Model stationarity has been discussed. The proposed model has also simpler interpretation (compare Example 3.4.2 and Example 2.1.2) and able to accommodate Binomial distribution by holding the similar properties as the infinitely divisible cases. Furthermore, mixture model is popular with catering to the multimodality in the data. It appears as a new and an important contribution in the community of discrete time series models.

CHAPTER 4: PARAMETER ESTIMATION

4.0 Outline

Parameter estimation is one important aspect in statistical analysis. This chapter discusses the likelihood theory and estimating functions. The well-known MLE via EM algorithm and CLS are particularly discussed in Section 4.1. Design of simulation study is explained in Section 4.2. Furthermore, we study the robustness of MLE to outliers according to the contamination schemes. The breakdown point of the estimator will also be discussed. The results have been tabulated in Section 4.3. In Section 4.4, we derive an explicit expression of the score function and the expected Fisher information matrix for Poisson MPT(1) process. These expressions are very useful for coherent forecasting which will be elaborated in the next chapter. Section 4.5 presents the simulation study descriptive measures of forecasting accuracy. Section 4.6 concludes.

4.1 Likelihood Theory and Estimating Functions

A unified approach is introduced in this section to perform estimation through applying the estimating functions. The standard estimating techniques have been provided by Newey and McFadden (1994), such as the popular maximum likelihood, generalized method of moments, and minimum distance. Klimko and Nelson (1978) studied the CLS method in stochastic processes. The conditions for a valid and regular estimating function are clear. It can be found elsewhere but given here as the initial introduction of the section for parameter estimation. Given a valid and regular estimating function for all the estimators $\theta = (\theta_1, \theta_2, \dots, \theta_n)$, it must satisfy (i) the covariance matrix is positive definite, (ii) the function is almost surely differentiable with respect to the components of θ and, (iii) the function is non-singular. One should notice that the proposed model in Chapter 3 fulfil the conditions. Throughout this thesis, MLE via EM algorithm have been studied for parameter estimation. The CLS method is used for comparison purpose. An extensive simulation results have been tabulated here.

4.1.1 Expectation-Maximization (EM) Algorithm

To obtain the parameter estimation by MLE via EM algorithm, we first define the log-likelihood function. Let the marginal mass function $p_j, j = 0, 1, \dots,$ and let $\phi \in (0, 1)$, denote the log-likelihood function L as follows.

$$\log L = \log P(X_1) + \sum_{t=2}^{T} \log P(X_t | X_{t-1}) \tag{4.1}$$

where the conditional probability function is defined in Chapter 3, Eq. (3.5). Differentiate Eq. (4.1) and letting it equal to zero to solve the equation. The MLE can be easily solved numerically using **mle** in Matlab 2013a built-in package, by defining the objective function according to one's preference. Here, Eq. (4.1) is the objective function for our model. The initial value can be obtained via method of moments to reduce the number of iterations.

MLE works for majority likelihood functions. However, solving Eq. (4.1) with the built-in function may not work for all marginals considered in MPT(1) model, especially estimating the parameter of mixing weight. To estimate the parameters for MPT(1) model which is a finite mixture distribution, we consider using MLE via EM algorithm because it is known to work well for finite mixture distributions. EM algorithm comprises two steps; the expectation step (E-step) and the maximization step (M-step). The EM algorithm implemented here follows Karlis and Xekalaki (1999, Section 2) which is described as such:

Suppose a k-finite mixture of a distribution with probability density function $g(x)$ and components $f(x|\theta)$ is defined by

$$g(x) = \sum_{j=1}^{k} \phi_j f(x|\theta_j) \tag{4.2}$$

where $\phi_j > 0$ for $j = 1, \ldots, k$; $\sum_{j=1}^{k} \phi_j = 1$ are the mixing proportions and θ_j are the parameters for each subpopulation. The mixing proportion ϕ_j can be regarded as the probability that a randomly selected observation belongs to the j-th subpopulation. The algorithm is

E-step: With the current estimates ϕ_j^{old} and $\mu(\theta_j^{old})$ calculate

$$w_{ij} = \frac{p_j^{old} f\left(x_i \middle| \theta_j^{old}\right)}{g(x_i)}, \quad i = 1, \ldots, k, j = 1, \ldots, k \tag{4.3}$$

M-step: Obtain the new estimates of the parameters $\mu\left(\theta_j\right)$ and ϕ_j from

$$\mu\left(\theta_j^{new}\right) = \frac{\sum_{i=1}^{n} w_{ij} x_i}{\sum_{i=1}^{n} w_{ij}} \quad \text{and} \quad \phi_j^{new} = \frac{\sum_{i=1}^{n} w_{ij}}{n} \qquad j = 1, \ldots, k \tag{4.4}$$

It is well-known that the mean of a mixture is the weighted mean of the means of all components weighted by the mixing proportion. We look into both equations particularly. Eq. (4.4) consists two parts of estimation; the parameters of the distributions and the mixing proportions. Karlis and Xekalaki (1999) suggested that the mean parameters of distributions can be avoided to reduce the calculation by almost $100/(2k-1)\%$. For an example, the mean parameter of Poisson distribution is just simply λ. For MPT(1) model, it is a finite mixture model formulated by a thinning process $(\alpha \circ X_t)$ and an innovation term (ε_t). The thinning process is a compound distribution and so the mean parameter is formed by the convolution of two distributions. Sprott (1983) discussed the convolution of Poisson and Binomial distribution. In the paper, he explained clearly that one of the parameter can be estimated from the first-moment equation. This can be used to instantly eliminate one of the parameters, and then another parameter can be estimated iteratively. Al-Osh and Alzaid (1987) adopted the idea in estimating the parameters for INAR(1) model. It is a typical representation for the convolution of Poisson-Binomial distribution which contains two parameters to be estimated. Similar approach has been considered to solve the convolution of mean parameter $\lambda\alpha$ (See Example 3.4.1) in Poisson MPT(1) process. After that, the mixing weight can be found iteratively using the right-hand-side formula in Eq. (4.4), until convergence is reached with a margin of error of 0.001. An effective initial value is suggested to reduce the computation time. The use of method of moments (Fruman and Lindsay, 1994) appears as a good method to obtain initial value for MLE. See also McLachlan and Peel (2000) for an in-depth discussion of EM algorithm.

4.1.2 Conditional Least Squares Method

CLS method is another popular method for estimation. The CLS estimates are obtained by minimizing

$$Q(\boldsymbol{x};\alpha,\phi,\lambda) = \sum_{t=2}^{n}(X_t - E(X_t|X_{t-1}))^2 \tag{4.5}$$

The conditional expectation can be obtained as in Eq. (3.6) by setting $E(\varepsilon_t) = \mu_\varepsilon$. Parameter estimation can be done by minimizing Eq. (4.5) as an optimization problem. However, the simulation results indicated that the CLS method fails to estimate the parameter of mixing proportion (ϕ) because of the inseparable nature of parameters ϕ and α in ACVF of Eq. (3.11).

Remark:

The discrete time series models such as INAR and Pegram's AR(1) models hold some similar characteristics with continuous AR(1) model. This is due to the fact that they share common estimation methods. Besides MLE and CLS methods, YW Equation has been used to estimate the parameters for existing discrete-valued time series models. See Brockwell and Davis (1987) and Chapter 2, Section 2.4.2 for more information. However, YW Equation cannot be applied for Poisson MPT(1) model for similar reason as encountered by CLS methods.

4.2 Design of Numerical Study on Parameter Estimation

Previous section discusses the theory and the estimators for estimating Poisson MPT(1) process. To get some relative merits of MLE via EM algorithm and CLS, we have done a simulation study for both estimation methods, which are programmed in MATLAB on computers (8GB RAM) running Windows Vista. Two types of generated data sets are considered as follow.

4.2.1 Without Contamination

For the data set without contamination, the data used in the study is generated from MPT(1) models. As many as 1000 Monte Carlo samples were generated for each sample size $n \in \{100, 500, 1000, 5000, 10000\}$ with different combination of the parameters. The standard errors of the estimated parameters are calculated. Also, we considered parallel computing with 4 workers for CLS estimation as it required long running time.

4.2.2 With Contamination

For the data sets with outliers, we repeated the data generation as mentioned in subsection 4.2.1, and adding in the outliers according to the contamination scheme: A small amount of outliers (1% with respect to the sample size *n*) are placed further away from the rest of the data- (i) 3 counts and (ii) 6 counts away from the maximum of X, that is at the positions of $\max(X) + 3$ and $\max(X) + 6$.

Then, we examined the breakdown point of MLE for Poisson MPT(1) model. The breakdown point here is defined as the MLE estimation for λ deviates by a magnitude exceeding 1. We focused only on the breakdown point of MLE here, since the simulation results without the presence outliers indicate the incompetency of CLS. See the results in Table 4.3.1. The percentage of outliers is increased gradually starting from 1% until MLE breaks down. For a more comprehensive study, two types of outlier have been considered: Additive outliers (AO) and Innovative outliers (IO). For both outliers we adopted the definitions from Barczy et al. (2010) which are provided for readers' convenience.

Definition 4.2.1 (Additive Outliers Poisson INAR(1)): A stochastic process $(Y_k)_{k\in\mathbb{Z}_+}$ is called an INAR(1) model contaminated with finitely many additive outliers if

$$Y_k = X_k + \sum_{i=1}^{I} \delta_{k,\,s_i}\theta_i\,, \quad k \in \mathbb{Z}_+ \tag{4.6}$$

where $(X_k)_{k\in\mathbb{Z}_+}$ is an INAR(1) process given by Eq. (2.2), $EX_0^2 < \infty, E\varepsilon_1^2 < \infty, P(\varepsilon_1 \neq 0) > 0$, and $I \in \mathbb{N}, s_i, \theta_i \in \mathbb{N}, i = 1, \dots, I$ such that $s_i \neq s_j$ if $i \neq j, i, j = 1, \dots, I$. In (4.6), $\theta_i, i = 1, \dots, I$, represents the ith additive outlier's size and $\delta_{k,\,s_i}$ is an impulse taking the value 1 if $k = s_i$ and 0 otherwise.

Definition 4.2.2 (Innovational Outliers Poisson INAR(1)): Let $(\varepsilon_l)_{l\in\mathbb{N}}$ be an i.i.d. sequence of non-negative integer-valued random variables. A stochastic process $(Y_k)_{k\in\mathbb{Z}_+}$ is called an INAR(1) model with finitely many innovational outliers if

$$Y_k = \sum_{j=1}^{Y_{k-1}} \xi_{k,j} + \eta_k\,, \quad k \in \mathbb{N} \tag{4.7}$$

where for all $k \in \mathbb{N}, \left(\xi_{k,j}\right)_{j\in\mathbb{N}}$ is a sequence of i.i.d. Bernoulli random variable with mean $\alpha \in (0,1)$ such that these sequences are mutually independent and independent of the sequence $(\varepsilon_l)_{l\in\mathbb{N}}$, and Y_0 is a non-negative integer-valued random variable independent of the sequence $\left(\xi_{k,j}\right)_{j\in\mathbb{N}}$, $k \in \mathbb{N}$ and $(\varepsilon_l)_{l\in\mathbb{N}}$, and

$$\eta_k := \varepsilon_k + \sum_{i=1}^{I} \delta_{k,s_i}\theta_i\,, \quad k \in \mathbb{Z}_+$$

where $I \in \mathbb{N}$ and $s_i, \theta_i \in \mathbb{N}, i = 1, \dots, I$, and $EY_0^2 < \infty, E\varepsilon_1^2 < \infty, P(\varepsilon_1 \neq 0) > 0$.

In this contamination study, we generate 1000 Monte Carlo samples of each sample size $n = 10000$ for each contamination percentage. The contamination data is generated from Poisson INAR(1) additive and innovative outliers respectively, in varying percentage (δ), starting from 1%. We use Eq. (4.6) to generate additive outlier Poisson INAR(1) (Y_k), where $k = 1,2,\dots,\delta n$. By looking into Eq. (4.6), we generate the Poisson INAR(1) process with parameter $\lambda = 2.0, \alpha = 0.7$, and we consider that there is only one outlier $(I = 1)$ occurs in the Poisson INAR(1) process, at the known position of $s = \frac{\delta n}{2}$ with size Poisson $(\lambda = 4)$. For an example, if we consider $\delta = 1\%$, then $k = 1,2,\dots,100$; which means the length of the Poisson INAR(1) additive outlier is 100, and the outlier occurs at the position of $s = \frac{100}{2} = 50$. When $k = s = 50$, the impulse taking 1 and the outlier occurs. The rest of 9900 data which are generated from MPT(1) process is then combined with 100 data generated by Poisson INAR(1) additive outlier to form the contamination data. We do the similar simulation study by increasing the contamination percentage. Also, we generate Poisson INAR(1) innovative outlier with the similar technique using Eq. (4.7).

4.3 Simulation Results of Parameter Estimation

We will present the simulation results by Poisson and Binomial MPT(1) model, respectively. The design for simulation study which has been expressed in Section 4.2 is applicable to Poisson MPT(1) process. We considered only the design in subsection

4.2.1 for Binomial MPT(1) process. Upon each tabulated result there will be a conclusive discussion of the numerical study.

4.3.1 Poisson MPT(1) Model

The data is generated by Poisson MPT(1) model without contamination. Four different combinations of parameters are considered and standard errors are calculated along. Besides, the computation time are also compared. There are three conclusions that can be highlighted from Table 4.3.1;

- It is obviously seen that the parameter estimation by EM algorithm is better than CLS,
- As expected, all parameters estimated by EM algorithm converged with the smaller standard errors when the sample size increased,
- Consistency is apparent for EM algorithm,
- The computation time by EM algorithm is much lower compared with CLS, and
- CLS method is not able to handle the estimation; the estimation is diverged for some parameters.

Next, the numerical study is carried out with the data generated by Poisson MPT(1) model. The outliers size of 3 and 6 are placed at the position $\max(x) + 3$ and $\max(x) + 6$ respectively. The sample size of outliers is 1% of each sample size. For comparison purpose, the combinations of parameters remain the same. We obtained similar conclusions from Table 4.3.2 and Table 4.3.3 as bulleted in the following;

- The parameters estimated by EM algorithm are not greatly affected by the data with outliers, the size of 3 and 6, with respect to 1% of the sample size, added in,
- Similarly, all parameters estimated by EM algorithm converged with small standard errors,

- EM algorithm has short computation time compare with CLS method, and
- CLS method fails to estimate the parameter $\hat{\phi}$.

Remark:

From the tables we know that CLS did not perform well in estimating the parameters. The estimation for mixing proportion (ϕ) is even far away from the true values. Also, the consistency is not preserved for CLS method. As we mentioned earlier this is due to the inseparable parameters ϕ and α existing in MPT(1) model. We have done the CLS method here for comparison purpose. A distinct difference can be observed by comparing the current results with the well-known thinning based INAR(1) model and Pegram's AR(1); CLS performs well for INAR(1) and Pegram's AR(1) but *not* for MPT(1) model. See Al-Osh and Alzaid (1987) and Biswas and Song (2009) for the simulation results of CLS for INAR(1) and Pegram's AR(1) models.

Table 4.3.4 tabulates the outlier handled by Poisson INAR(1) and Pegram's AR (1) models. It can be noticed that it shows poor performance for the estimation when the outliers are added in.

Since the MLE via EM algorithm has satisfactory performance in estimating the parameters for mixture model, it is therefore of great interest to be acquainted with the breakdown point of the MLE by EM algorithm. Both types of outliers presented similar pattern and effect in this contamination study as displayed in Figure 4.3.1. The parameter deviation involves a slight difference at 10^{-3} between two types of outliers. Figure 4.3.2 shows the deviation percentage of the respective outliers. As expected, the deviation increased gradually for all parameters when the outliers are added; it can be seen that λ deviates the most and ϕ deviates the least. It is observed that MLE for λ deviates by a magnitude exceeding 1 when more than 18% of AOs or 19% of IOs are added. As expected, the MLE becomes inefficient when the contamination is increased.

Table 4.3.1: Parameter estimates, standard errors (in brackets) by MLE (EM Algorithm) and CLS for simulated Poisson MPT(1) samples

True values: (α, λ, ϕ)	EM Algorithm				CLS			
	Duration (sec)	$\hat{\alpha}_{ML}$	$\hat{\lambda}_{ML}$	$\hat{\phi}_{ML}$	Duration (sec)	$\hat{\alpha}_{CLS}$	$\hat{\lambda}_{CLS}$	$\hat{\phi}_{CLS}$
(0.3, 1.0, 0.4) n=100	97.32	0.30108 (0.00280)	1.00510 (0.00943)	0.40030 (0.00442)	17213.85	0.49317 (0.31235)	0.96570 (0.11902)	0.21409 (0.22435)
n=500	300.52	0.30211 (0.00056)	0.99464 (0.00185)	0.40119 (0.00087)	15125.25	0.39031 (0.24149)	0.99639 (0.07320)	0.175513 (0.24898)
n=1000	390.25	0.29977 (0.00029)	1.00171 (0.00098)	0.40034 (0.00040)	17191.15	0.35965 (0.24075)	0.99845 (0.06147)	0.16240 (0.25933)
n=5000	1530.07	0.30011 (0.00015)	1.00101 (0.00068)	0.40020 (0.00032)	16946.62	0.36479 (0.23597)	1.00533 (0.06247)	0.16384 (0.25859)
n=10000	2897.63	0.30001 (0.00009)	1.00011 (0.00032)	0.40011 (0.00029)	28499.01	0.26602 (0.21047)	1.00492 (0.03444)	0.13428 (0.28188)
(0.7, 1.0, 0.1) n=100	94.57	0.71012 (0.00624)	0.99300 (0.00844)	0.09788 (0.00314)	16490.24	0.48534 (0.32530)	0.96557 (0.12039)	0.21040 (0.16319)
n=500	219.76	0.70215 (0.00142)	0.99896 (0.00201)	0.09993 (0.00048)	17033.84	0.38395 (0.39117)	1.00082 (0.07777)	0.17967 (0.13444)
n=1000	377.807	0.70260 (0.00065)	0.99715 (0.00093)	0.10627 (0.00040)	17356.61	0.35657 (0.41223)	1.00055 (0.06272)	0.16751 (0.12226)
n=5000	1639.40	0.69960 (0.00980)	1.0008 (0.01417)	0.0838 (0.01653)	21500.82	0.29477 (0.46131)	1.00432 (0.04118)	0.14211 (0.10602)
n=10000	3010.70	0.70020 (0.00690)	0.9999 (0.00994)	0.0840 (0.01632)	27182.26	0.26580 (0.48038)	1.00697 (0.03643)	0.13583 (0.09940)
(0.3, 2.0, 0.2) n=100	93.87	0.31179 (0.00263)	1.93560 (0.01630)	0.19748 (0.00448)	15388.35	0.62023 (0.39166)	1.89417 (0.19194)	0.15174 (0.09888)
n=500	211.91	0.30817 (0.00060)	1.94938 (0.00380)	0.19196 (0.00094)	15945.23	0.51679 (0.31954)	1.93957 (0.11821)	0.12088 (0.10651)
n=1000	350.88	0.30901 (0.00034)	1.94241 (0.00218)	0.19137 (0.00057)	16789.25	0.48299 (0.28991)	1.94673 (0.10607)	0.10620 (0.11173)
n=5000	1415.73	0.30012 (0.00301)	1.99964 (0.01980)	0.18573 (0.01547)	20506.58	0.39616 (0.24137)	1.94661 (0.09018)	0.09018 (0.12262)
n=10000	2631.09	0.29995 (0.00201)	2.00011 (0.01374)	0.18625 (0.01476)	25731.42	0.36790 (0.22824)	1.94707 (0.06976)	0.08241 (0.12769)
(0.5, 3.0, 0.1) n=100	96.73	0.52078 (0.00402)	2.89290 (0.02190)	0.09567 (0.00258)	15490.38	0.72025 (0.29239)	2.86644 (0.23234)	0.12346 (0.08022)
n=500	219.04	0.51880 (0.00104)	2.89328 (0.00585)	0.09508 (0.00054)	16008.55	0.65311 (0.24833)	2.90443 (0.15644)	0.09521 (0.05530)
n=1000	371.67	0.51665 (0.00061)	2.90431 (0.00347)	0.09460 (0.00028)	16472.98	0.62293 (0.24483)	2.90409 (0.14280)	0.08336 (0.05111)
n=5000	1508.90	0.50032 (0.00411)	2.99821 (0.02454)	0.08654 (0.01401)	21708.08	0.54025 (0.22151)	2.90893 (0.11357)	0.06530 (0.05195)
n=10000	2859.74	0.50021 (0.00282)	2.99884 (0.01661)	0.08715 (0.01343)	26481.25	0.50393 (0.22277)	2.91778 (0.09855)	0.05631 (0.05508)

Table 4.3.2: Parameter estimates, standard errors (in brackets) by MLE (EM algorithm) and CLS for Poisson MPT(1) with outlier size of 3

True values: (α, λ, ϕ)	EM Algorithm				CLS			
	Duration (sec)	$\hat{\alpha}_{ML}$	$\hat{\lambda}_{ML}$	$\hat{\phi}_{ML}$	Duration (sec)	$\hat{\alpha}_{CLS}$	$\hat{\lambda}_{CLS}$	$\hat{\phi}_{CLS}$
(0.3, 1.0, 0.4) n=100	100.27	0.28503 (0.03165)	1.06242 (0.11957)	0.36817 (0.05012)	15553.61	0.43887 (0.43811)	0.96282 (0.11809)	0.19115 (0.23918)
n=500	241.83	0.28037 (0.02316)	1.07207 (0.08610)	0.38204 (0.02567)	15193.39	0.74060 (0.46548)	1.05936 (0.13188)	0.40607 (0.10237)
n=1000	410.06	0.27903 (0.02267)	1.07619 (0.08307)	0.38522 (0.02023)	16795.51	0.77690 (0.49325)	1.07020 (0.12351)	0.45063 (0.10363)
n=5000	1846.14	0.27673 (0.02364)	1.08433 (0.08588)	0.39172 (0.01086)	22581.36	0.78954 (0.50928)	1.08462 (0.12349)	0.49853 (0.12471)
n=10000	3567.28	0.27602 (0.02418)	1.08701 (0.08787)	0.39358 (0.00832)	26421.84	0.80702 (0.51724)	1.09617 (0.12332)	0.52310 (0.14234)
(0.7, 1.0, 0.1) n=100	98.42	0.66401 (0.07621)	1.06494 (0.12440)	0.08344 (0.02231)	15473.90	0.45064 (0.35088)	0.96274 (0.11774)	0.19466 (0.14907)
n=500	305.54	0.65531 (0.05323)	1.07044 (0.08464)	0.09882 (0.00667)	16002.37	0.75994 (0.18759)	1.07657 (0.14596)	0.38838 (0.30542)
n=1000	380.36	0.65211 (0.05162)	1.07431 (0.08075)	0.10064 (0.00501)	17765.35	0.77309 (0.14468)	1.07040 (0.12713)	0.44848 (0.35968)
n=5000	1574.45	0.64591 (0.05496)	1.08392 (0.08542)	0.10147 (0.00303)	21921.76	0.80338 (0.14580)	1.09627 (0.13274)	0.51430 (0.42110)
n=10000	2964.63	0.64402 (0.05652)	1.08724 (0.08808)	0.10118 (0.00251)	27558.60	0.80869 (0.14621)	1.09709 (0.12716)	0.52476 (0.43046)
(0.3, 2.0, 0.2) n=100	95.23	0.28984 (0.02296)	2.0804 (0.16681)	0.18791 (0.03657)	15606.21	0.60658 (0.37853)	1.89170 (0.18454)	0.13971 (0.10177)
n=500	304.24	0.28773 (0.01536)	2.08772 (0.10943)	0.19055 (0.01905)	16808.02	0.78412 (0.50474)	2.01109 (0.13873)	0.28057 (0.11404)
n=1000	370.48	0.28740 (0.01425)	2.08876 (0.10041)	0.19133 (0.01437)	16904.27	0.79826 (0.51201)	2.02302 (0.13271)	0.32231 (0.14359)
n=5000	1604.23	0.28591 (0.01444)	2.09885 (0.10121)	0.19349 (0.00862)	20513.52	0.82262 (0.53233)	2.05365 (0.11908)	0.37666 (0.18682)
n=10000	2796.21	0.28534 (0.01496)	2.10324 (0.10450)	0.19477 (0.00681)	27001.59	0.82854 (0.53635)	2.05907 (0.11725)	0.39616 (0.20515)
(0.5, 3.0, 0.1) n=100	89.45	0.48704 (0.03132)	3.09067 (0.20045)	0.08645 (0.02143)	16819.85	0.71064 (0.28448)	2.845844 (0.23378)	0.12228 (0.07966)
n=500	300.75	0.48576 (0.01895)	3.09051 (0.11970)	0.09200 (0.01132)	17052.51	0.82990 (0.31256)	2.98836 (0.15461)	0.23274 (0.13256)
n=1000	390.45	0.48431 (0.01816)	3.09857 (0.11376)	0.09452 (0.00822)	18325.65	0.83910 (0.32541)	3.00775 (0.14958)	0.27033 (0.14285)
n=5000	1499.47	0.48225 (0.01835)	3.11127 (0.11438)	0.09833 (0.00382)	21499.41	0.85877 (0.36901)	3.02777 (0.11811)	0.31093 (0.21777)
n=10000	2856.43	0.48127 (0.01901)	3.11717 (0.09975)	0.09977 (0.00288)	23643.41	0.85445 (0.36471)	3.02381 (0.11455)	0.33666 (0.24216)

Table 4.3.3: Parameter estimates, standard errors (in brackets) by MLE (EM algorithm) and CLS for Poisson MPT(1) with outlier size of 6

True values: (α, λ, ϕ)	EM Algorithm				CLS			
	Duration (sec)	$\hat{\alpha}_{ML}$	$\hat{\lambda}_{ML}$	$\hat{\phi}_{ML}$	Duration (sec)	$\hat{\alpha}_{CLS}$	$\hat{\lambda}_{CLS}$	$\hat{\phi}_{CLS}$
(0.3, 1.0, 0.4) n=100	93.00	0.27660 (0.03491)	1.09389 (0.13720)	0.37469 (0.04814)	15021.44	0.41283 (0.26443)	0.95978 (0.11643)	0.18476 (0.24300)
n=500	244.27	0.27270 (0.02970)	1.10212 (0.11247)	0.39362 (0.02019)	14746.51	0.82893 (0.02465)	1.07664 (0.16834)	0.53108 (0.15632)
n=1000	362.93	0.27200 (0.02918)	1.10405 (0.10912)	0.39847 (0.01309)	15562.05	0.84942 (0.55625)	1.09342 (0.17564)	0.58707 (0.20020)
n=5000	1591.60	0.26947 (0.03078)	1.11355 (0.11467)	0.40394 (0.00763)	19826.49	0.87964 (0.58346)	1.13116 (0.17649)	0.63977 (0.24656)
n=10000	3432.56	0.26885 (0.03129)	1.11601 (0.11666)	0.40531 (0.00710)	24272.17	0.88196 (0.58549)	1.13003 (0.16863)	0.64964 (0.25514)
(0.7, 1.0, 0.1) n=100	94.32	0.64644 (0.08112)	1.09247 (0.13671)	0.08432 (0.02152)	16127.20	0.41026 (0.37136)	0.96459 (0.11222)	0.18747 (0.14196)
n=500	315.64	0.63704 (0.06862)	1.10081 (0.11121)	0.10004 (0.00651)	14997.14	0.83413 (0.16903)	1.06906 (0.16287)	0.52997 (0.43843)
n=1000	400.16	0.63471 (0.06801)	1.10380 (0.10907)	0.10207 (0.00541)	15137.46	0.85039 (0.17461)	1.09530 (0.17291)	0.58869 (0.49383)
n=5000	1595.45	0.62890 (0.07172)	1.11331 (0.11440)	0.10222 (0.00390)	19329.90	0.87762 (0.19059)	1.12508 (0.18126)	0.64578 (0.54880)
n=10000	2840.75	0.62700 (0.07335)	1.11667 (0.1173)	0.10154 (0.00312)	24502.06	0.88173 (0.19270)	1.13516 (0.17207)	0.65440 (0.55712)
(0.3, 2.0, 0.2) n=100	90.23	0.28665 (0.0235)	2.10333 (1.1123)	0.19274 (0.0366)	15743.62	0.56868 (0.35220)	1.89503 (0.18500)	0.13316 (0.10310)
n=500	311.74	0.2842 (0.01822)	2.1133 (1.11551)	0.19558 (0.01711)	14435.57	0.83756 (0.54682)	2.03249 (0.16514)	0.39135 (0.20743)
n=1000	390.58	0.28318 (0.01801)	2.12014 (1.12115)	0.19759 (0.01282)	14990.24	0.86657 (0.57287)	2.04739 (0.15586)	0.43808 (0.24711)
n=5000	1661.33	0.28181 (0.01840)	2.12952 (1.12975)	0.20167 (0.00611)	19275.59	0.87770 (0.58238)	2.08366 (0.15164)	0.49356 (0.29803)
n=10000	2800.71	0.28127 (0.01898)	2.13382 (1.13391)	0.20361 (0.00611)	25481.81	0.88078 (0.58471)	2.08288 (0.14692)	0.51259 (0.31639)
(0.5, 3.0, 0.1) n=100	100.45	0.48431 (0.0318)	3.10727 (0.2075)	0.08791 (0.0204)	16458.43	0.70093 (0.27795)	2.86616 (0.23645)	0.11444 (0.07312)
n=500	320.75	0.48008 (0.02340)	3.14966 (0.14960)	0.09972 (0.00784)	17582.36	0.85169 (0.01622)	3.00236 (0.00773)	0.32511 (0.01065)
n=1000	395.15	0.47951 (0.02240)	3.12944 (0.14238)	0.10212 (0.00630)	19214.25	0.87376 (0.01206)	3.02940 (0.00510)	0.372904 (0.00881)
n=5000	1593.21	0.47756 (0.02298)	3.14174 (0.14411)	0.10400 (0.00601)	21173.76	0.89555 (0.40073)	3.04691 (0.14561)	0.41834 (0.32179)
n=10000	2805.43	0.47691 (0.02333)	3.14567 (0.1470)	0.10444 (0.00643)	25753.34	0.89853 (0.40310)	3.06180 (0.14149)	0.43373 (0.33679)

Table 4.3.4: ML estimates and standard errors (in bracket) for INAR(1) and Pegram's AR(1) with outliers

Method	Thinning INAR(1)				Pegram's AR(1)			
Outlier	max(X)+3		max(X)+6		max(x)+3		max(X)+6	
True Values: (α, λ)	$\hat{\alpha}_{ML}$	$\hat{\lambda}_{ML}$	$\hat{\alpha}_{ML}$	$\hat{\lambda}_{ML}$	$\hat{\alpha}_{ML}$	$\hat{\lambda}_{ML}$	$\hat{\alpha}_{ML}$	$\hat{\lambda}_{ML}$
(0.3, 1.0) 100	0.2121 (0.1274)	1.1795 (0.2490)	0.1671 (0.1550)	1.2663 (0.3138)	0.2025 (0.1388)	1.0603 (0.1480)	0.1533 (0.1687)	1.0907 (0.1650)
500	0.4437 (0.1488)	0.8378 (0.1758)	0.5217 (0.2239)	0.7379 (0.2686)	0.5108 (0.2135)	1.0750 (0.0876)	0.5497 (0.2515)	1.0995 (0.1163)
1000	0.4891 (0.1920)	0.770 (0.2349)	0.5800 (0.2812)	0.6465 (0.3563)	0.5091 (0.2117)	1.0731 (0.0851)	0.6083 (0.3093)	1.1045 (0.1133)
5000	0.5490 (0.2503)	0.6853 (0.3170)	0.6433 (0.3440)	0.5524 (0.4487)	0.5719 (0.2732)	1.0813 (0.0839)	0.6766 (0.3772)	1.1128 (0.1147)
10000	0.5648 (0.2660)	0.6628 (0.3392)	0.6591 (0.3597)	0.5395 (0.4715)	0.5887 (0.1741)	1.0846 (0.1081)	0.6924 (0.3929)	1.1148 (0.1159)
(0.3, 2.0) 100	0.2257 (0.1156)	2.2802 (0.4096)	0.1983 (0.1311)	2.3707 (0.4713)	0.2243 (0.1267)	2.0776 (0.2003)	0.1895 (0.1462)	2.1001 (0.2171)
500	0.4135 (0.1205)	1.7292 (0.3000)	0.4728 (0.1763)	1.5727 (0.4422)	0.4232 (0.1302)	2.0822 (0.1203)	0.4907 (0.1940)	2.1134 (0.1410)
1000	0.4508 (0.1540)	1.6221 (0.3895)	0.5167 (0.2185)	1.4408 (0.5659)	0.4647 (0.1677)	2.0877 (0.1078)	0.5425 (0.2442)	2.1172 (0.1318)
5000	0.5024 (0.2040)	1.4776 (0.5272)	0.5753 (0.2761)	1.2730 (0.7297)	0.5184 (0.2201)	2.0973 (0.1010)	0.6053 (0.3062)	2.1282 (0.1313)
10000	0.5170 (0.2184)	1.4349 (0.5694)	0.5917 (0.2925)	1.2259 (0.7766)	0.5341 (0.1576)	2.1016 (0.1000)	0.6194 (0.3201)	2.1326 (0.1343)
(0.5, 2.0) 100	0.3970 (0.1337)	2.4610 (0.5977)	0.3471 (0.1754)	2.6956 (0.7983)	0.3836 (0.1587)	2.0825 (0.2671)	0.3166 (0.2099)	2.1015 (0.2705)
500	0.5599 (0.0689)	1.8038 (0.2444)	0.5888 (0.0941)	1.6986 (0.3313)	0.5716 (0.0818)	2.0805 (0.1377)	0.6091 (0.1138)	2.1151 (0.1637)
1000	0.5878 (0.0915)	1.6916 (0.3279)	0.6258 (0.1279)	1.5469 (0.4635)	0.6043 (0.1085)	2.0883 (0.1179)	0.6577 (0.1594)	2.1142 (0.1384)
5000	0.6296 (0.1309)	1.5276 (0.4782)	0.6743 (0.1752)	1.3520 (0.6516)	0.6500 (0.1513)	2.0948 (0.1009)	0.7103 (0.2110)	2.1256 (0.1305)
10000	0.6412 (0.1421)	1.4808 (0.5233)	0.6862 (0.1869)	1.3038 (0.6989)	0.6617 (0.1629)	2.1012 (0.1042)	0.7222 (0.2228)	2.1278 (0.1304)

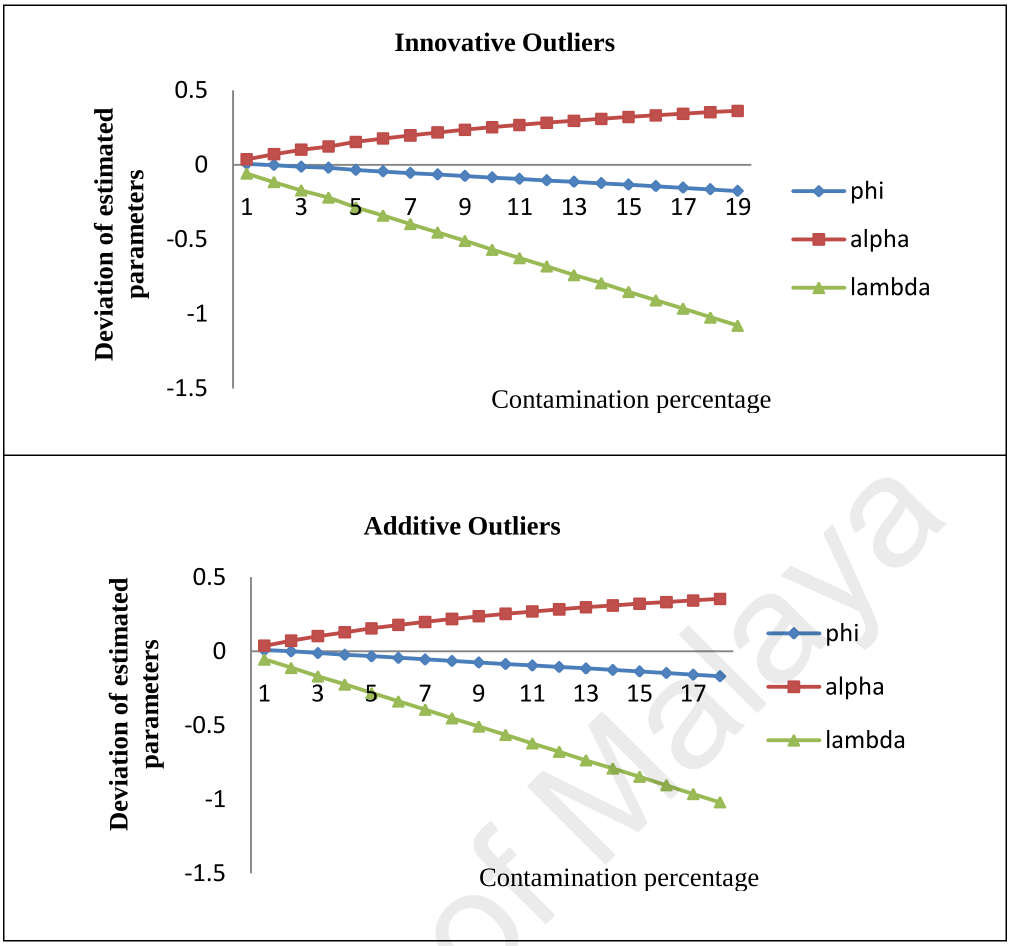


Figure 4.3.1: Breakdown point of MLE via EM algorithm with AO and IO outliers for MPT(1) process for the true parameters of $(\phi, \alpha, \lambda) = (0.1, 0.7, 1.0)$

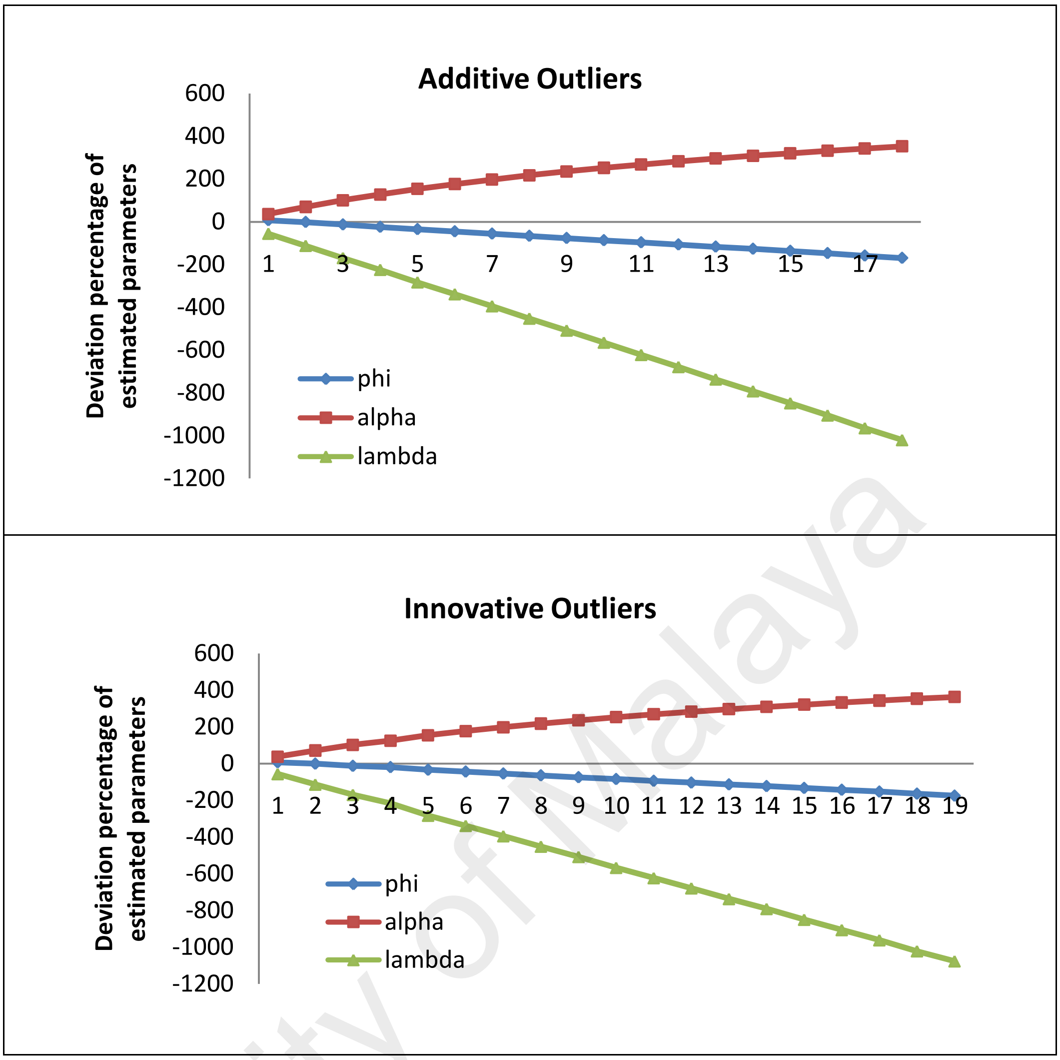


Figure 4.3.2: Deviation percentage of estimated parameters for AO and IO outliers for true parameters $(\boldsymbol{\phi}, \boldsymbol{\alpha}, \boldsymbol{\lambda}) = (\mathbf{0.1}, \mathbf{0.7}, \mathbf{1.0})$

4.3.2 Binomial MPT(1) Model

We present the Binomial MPT(1) model, which has entirely different interpretation for real life application. The simulation design in Section 4.2.1 is still applicable to Binomial MPT(1) Model. To run the simulation study we fixed the index N in the study of the behaviour of Binomial MPT(1) process by numerical computation, with $N = 5$ for different combinations of the parameters (ϕ, p, α). Again, the CLS method has been presented for comparison. The simulation results have been tabulated in Table 4.3.5. Similarly, as with the Poisson MPT(1) process, overall the EM algorithm performed much better compared with CLS. The estimates converged to the true values when the sample size is increased. Therefore empirically, consistency is achieved. Likewise, CLS gives poor estimates for the parameters, especially the mixing proportion. Also, the computation time of CLS is much longer than the EM algorithm.

Table 4.3.5: Parameter estimates, standard errors (in brackets) by MLE (EM Algorithm) and CLS for Binomial MPT(1), $N = 5$

Parameter: (ϕ, p, α)	EM Algorithm				CLS			
$N = 5$	Duration (sec)	$\hat{\phi}_{ML}$	$\hat{p}_{ML}$	$\hat{\alpha}_{ML}$	Duration (sec)	$\hat{\phi}_{CLS}$	$\hat{p}_{CLS}$	$\hat{\alpha}_{CLS}$
(0.2,0.3,0.5) n=100	17.58	0.19679 (0.02290)	0.28999 (0.02197)	0.48332 (0.03662)	4630.84	0.11061 (0.10432)	0.28791 (0.02422)	0.54287 (0.27623)
n=500	64.71	0.19937 (0.00856)	0.28998 (0.01354)	0.48330 (0.02257)	4965.51	0.09717 (0.11427)	0.29195 (0.01568)	0.44868 (0.25820)
n=1000	120.56	0.19970 (0.00616)	0.28999 (0.01195)	0.48332 (0.01992)	5490.92	0.09234 (0.11848)	0.29249 (0.01307)	0.43172 (0.25980)
n=5000	627.89	0.20013 (0.00290)	0.28989 (0.01053)	0.48315 (0.01755)	9196.58	0.08253 (0.12653)	0.29210 (0.01058)	0.36473 (0.27801)
n=10000	1169.17	0.20000 (0.00199)	0.29000 (0.01019)	0.48334 (0.01698)	14568.31	0.07826 (0.13088)	0.29254 (0.00940)	0.33745 (0.28501)
(0.2,0.5,0.7) n=100	41.28	0.16040 (0.04653)	0.50021 (0.02158)	0.70029 (0.03021)	2009.21	0.03714 (0.16393)	0.45802 (0.04709)	0.63930 (0.26916)
n=500	100.83	0.17002 (0.03241)	0.50002 (0.00981)	0.70003 (0.01373)	2073.94	0.03519 (0.16574)	0.45874 (0.04368)	0.60249 (0.29243)
n=1000	165.19	0.17256 (0.02898)	0.50014 (0.00724)	0.70019 (0.01014)	2295.93	0.03592 (0.16495)	0.46041 (0.04140)	0.60948 (0.27848)
n=5000	640.40	0.17603 (0.02409)	0.49982 (0.00323)	0.69975 (0.00452)	3886.23	0.03406 (0.16665)	0.46097 (0.03986)	0.55902 (0.29808)
n=10000	1082.21	0.18434 (0.01628)	0.49996 (0.00228)	0.69994 (0.00319)	5886.57	0.03219 (0.16851)	0.45965 (0.04099)	0.53036 (0.30838)
(0.3,0.1,0.3) n=100	37.38	0.28332 (0.02813)	0.09999 (0.01318)	0.29998 (0.03953)	5060.31	0.22951 (0.15820)	0.10721 (0.01527)	0.38194 (0.25752)
n=500	78.39	0.28916 (0.01437)	0.09988 (0.00599)	0.29965 (0.01798)	5277.32	0.19576 (0.16988)	0.11017 (0.01351)	0.31415 (0.22589)
n=1000	130.67	0.28941 (0.01218)	0.10030 (0.00423)	0.30089 (0.01270)	5867.84	0.18116 (0.17525)	0.11028 (0.01289)	0.30193 (0.22655)
n=5000	556.75	0.29003 (0.01027)	0.10006 (0.00183)	0.30017 (0.00549)	10588.78	0.16254 (0.18506)	0.11453 (0.01153)	0.24984 (0.22309)
n=10000	1097.92	0.29002 (0.01014)	0.10006 (0.00133)	0.30017 (0.00398)	15808.54	0.14583 (0.19510)	0.11029 (0.01116)	0.24518 (0.22070)
(0.5,0.1,0.3) n=100	35.81	0.46378 (0.05257)	0.10003 (0.01312)	0.30010 (0.03934)	5166.66	0.21623 (0.31448)	0.11288 (0.01839)	0.39660 (0.27387)
n=500	78.14	0.47568 (0.02891)	0.10004 (0.00610)	0.30011 (0.01831)	4862.87	0.19350 (0.33237)	0.11538 (0.01800)	0.32736 (0.23263)
n=1000	130.38	0.47789 (0.02411)	0.09990 (0.00427)	0.29971 (0.01281)	5737.09	0.18182 (0.34247)	0.11626 (0.01762)	0.29260 (0.22614)
n=5000	570.47	0.47841 (0.02193)	0.09998 (0.00187)	0.29993 (0.00560)	10141.69	0.15134 (0.36891)	0.11647 (0.01716)	0.26175 (0.22392)
n=10000	1091.60	0.47843 (0.02174)	0.10001 (0.00135)	0.30004 (0.00405)	15643.30	0.14175 (0.37728)	0.11645 (0.01695)	0.25631 (0.22226)

4.4 Score Function and Fisher Information for Poisson MPT(1) Model

We derive the expressions of score function and observed Fisher information for Poisson MPT(1) model. Note that the conditional likelihood can be written as $L(\alpha,\lambda,\phi) = \prod_{t=1}^{n} P(X_t|X_{t-1})$, where $P(X_t|X_{t-1})$ is defined in (3.5) with the innovation pmf form of (3.16).

The first derivatives of the log-likelihood define the score functions. Let $\dot{\ell}_\alpha, \dot{\ell}_\lambda, \dot{\ell}_\phi$ be the score functions with respect to α, λ, ϕ, while the respective score functions are defined by

$$\dot{\ell}_\alpha = \frac{\partial}{\partial\alpha}\ell(\alpha,\lambda,\phi;\ X_0, X_1, \dots, X_n) = \sum_{t=1}^{n} \frac{\frac{\partial}{\partial\alpha} P(X_t|X_{t-1})}{P(X_t|X_{t-1})} \tag{4.7a}$$

$$\dot{\ell}_\lambda = \frac{\partial}{\partial\lambda}\ell(\alpha,\lambda,\phi;\ X_0, X_1, \dots, X_n) = \sum_{t=1}^{n} \frac{\frac{\partial}{\partial\lambda} P(X_t|X_{t-1})}{P(X_t|X_{t-1})} \tag{4.7b}$$

$$\dot{\ell}_\phi = \frac{\partial}{\partial\phi}\ell(\alpha,\lambda,\phi;\ X_0, X_1, \dots, X_n) = \sum_{t=1}^{n} \frac{\frac{\partial}{\partial\phi} P(X_t|X_{t-1})}{P(X_t|X_{t-1})} \tag{4.7c}$$

where $\ell(\cdot)$ is the log-likelihood. The partial derivative of the conditional probability is found by making use of the following derivative

$$\frac{\partial}{\partial\alpha}\left[\alpha^i(1-\alpha)^{j-i}\right] = \left[\frac{i}{\alpha(1-\alpha)} - \frac{j}{1-\alpha}\right]\alpha^i(1-\alpha)^{j-i}$$

and the relation

$$i\binom{j}{i} = j\binom{j-1}{i-1}$$

Also,

$$\frac{\partial}{\partial\alpha}\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right] = \lambda\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right]$$

and

$$\frac{\partial}{\partial \lambda}\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right] = \alpha\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right]$$

Besides that,

$$\frac{\partial}{\partial \lambda}\left[\frac{e^{-\lambda}(\lambda)^i}{i!}\right] = \frac{e^{-\lambda}(\lambda)^{i-1}}{(i-1)!} - \frac{e^{-\lambda}(\lambda)^i}{i!}$$

Now we are ready to derive the partial derivatives of the conditional probability with respect to α, λ and ϕ.

Proposition 4.4.1 The derivatives of $P(X_t|X_{t-1})$ with respect to α, ϕ and λ are given by

$$\begin{aligned}
&\frac{\partial}{\partial \alpha} P(X_t = i|X_{t-1} = j) \\
&= \phi\binom{j}{i}\left[\frac{i}{\alpha(1-\alpha)} - \frac{j}{1-\alpha}\right]\alpha^i(1-\alpha)^{j-i} - \phi\lambda\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right] \\
&= \phi\binom{j}{i}\frac{i}{\alpha(1-\alpha)}\alpha^i(1-\alpha)^{j-i} - \phi\binom{j}{i}\frac{j}{1-\alpha}\alpha^i(1-\alpha)^{j-i} \\
&\qquad\qquad - \phi\lambda\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right] \\
&= \frac{\phi j}{1-\alpha}\left[\binom{j-1}{i-1}\alpha^{i-1}(1-\alpha)^{j-i} - \binom{j}{i}\alpha^i(1-\alpha)^{j-i}\right] \\
&\qquad\qquad - \phi\lambda\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right]
\end{aligned}$$

$$\frac{\partial}{\partial \lambda} P(X_t = i|X_{t-1} = j) = \frac{e^{-\lambda}(\lambda)^{i-1}}{(i-1)!} - \frac{e^{-\lambda}(\lambda)^i}{i!} - \phi\alpha\left[\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right]$$

$$\frac{\partial}{\partial \phi} P(X_t = i|X_{t-1} = j) = \binom{j}{i}\alpha^i(1-\alpha)^{j-i} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}$$

It should be noticed here that $i \le j$, since X_t and X_{t-1} may take any integer value and so we adopt the convention that for $i > j$ and $j = 0$, the binomial coefficient goes zero.

The second derivatives with respect to α, λ and ϕ are not hard to be computed as follows:

Proposition 4.4.2 The score function with respect to α, ϕ and λ are given by

$$\dot{\ell}_\alpha = \sum_{t=1}^{n} \frac{\frac{\phi j}{1-\alpha}\left(\binom{j-1}{i-1}\alpha^{i-1}(1-\alpha)^{j-i} - \binom{j}{i}\alpha^{i}(1-\alpha)^{j-i}\right) - \lambda\phi\left(\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{(i)!}\right)}{P(X_t|X_{t-1})} \tag{4.8a}$$

$$\dot{\ell}_\phi = \sum_{t=1}^{n} \frac{\binom{j}{i}\alpha^{i}(1-\alpha)^{j-i} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{i!}}{P(X_t|X_{t-1})} \tag{4.8b}$$

$$\dot{\ell}_\lambda = \sum_{t=1}^{n} \frac{\frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} - \frac{e^{-\lambda}\lambda^{i}}{i!} - \alpha\phi\left(\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} - \frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{i!}\right)}{P(X_t|X_{t-1})} \tag{4.8c}$$

Proposition 4.4.3 The second derivatives of the conditional probability are

$$\frac{\partial^2}{\partial\alpha^2}P(X_t = i|X_{t-1} = j) = \frac{\partial}{\partial\alpha}\left[\frac{\partial}{\partial\alpha}P(X_t = i|X_{t-1} = j)\right]$$

$$= \frac{\phi j}{(1-\alpha)^2}\left\{2(1-j)\binom{j-1}{i-1}\alpha^{i-1}(1-\alpha)^{j-i}\right.$$

$$\left. + (j-1)\left(\binom{j-2}{i-2}\alpha^{i-2}(1-\alpha)^{j-i-1} + \binom{j}{i}\alpha^{i}(1-\alpha)^{j-i}\right)\right\}$$

$$-\lambda^2\phi\left\{\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-2}}{(i-2)!} - 2\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} + \frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{(i)!}\right\}$$

$$\frac{\partial^2}{\partial\lambda^2}P(X_t = i|X_{t-1} = j) = \frac{\partial}{\partial\lambda}\left[\frac{\partial}{\partial\lambda}P(X_t = i|X_{t-1} = j)\right]$$

$$= \frac{e^{-\lambda}\lambda^{i-2}}{(i-2)!} - 2\frac{e^{-\lambda}\lambda^{i-1}}{(i-1)!} + \frac{e^{-\lambda}\lambda^{i}}{(i)!} - \alpha^2\phi\left\{\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-2}}{(i-2)!} - 2\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!} + \frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{(i)!}\right\}$$

$$\frac{\partial^2}{\partial\phi^2}P(X_t|X_{t-1}) = \frac{\partial}{\partial\phi}\left[\frac{\partial}{\partial\phi}P(X_t = i|X_{t-1} = j)\right] = 0$$

The variance covariance matrix off-diagonal entries are also derived:

$$\frac{\partial^2}{\partial\alpha\,\partial\lambda}P(X_t=i|X_{t-1}=j)=\frac{\partial}{\partial\alpha}\left[\frac{\partial}{\partial\lambda}P(X_t=i|X_{t-1}=j)\right]$$
$$=-\phi\left\{\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!}-\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{(i)!}\right\}$$
$$-\alpha\lambda\phi\left\{\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-2}}{(i-2)!}-2\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!}+\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{(i)!}\right\}$$

$$\frac{\partial^2}{\partial\alpha\,\partial\phi}P(X_t=i|X_{t-1}=j)=\frac{\partial}{\partial\alpha}\left[\frac{\partial}{\partial\phi}P(X_t=i|X_{t-1}=j)\right]$$
$$=\frac{j}{1-\alpha}\left\{\binom{j-1}{i-1}\alpha^{i-1}(1-\alpha)^{j-i}-\binom{j}{i}\alpha^{i}(1-\alpha)^{j-i}\right\}$$
$$-\lambda\left\{\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!}-\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{(i)!}\right\}$$

$$\frac{\partial^2}{\partial\phi\,\partial\lambda}P(X_t=i|X_{t-1}=j)=\frac{\partial}{\partial\phi}\left[\frac{\partial}{\partial\lambda}P(X_t=i|X_{t-1}=j)\right]$$
$$=-\alpha\left\{\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i-1}}{(i-1)!}-\frac{e^{-\lambda\alpha}(\lambda\alpha)^{i}}{(i)!}\right\}$$

The off-diagonal entries of $\frac{\partial^2}{\partial\alpha\,\partial\lambda},\frac{\partial^2}{\partial\alpha\,\partial\phi}$ and $\frac{\partial^2}{\partial\phi\,\partial\lambda}$ is equal to $\frac{\partial^2}{\partial\lambda\,\partial\alpha}$,$\frac{\partial^2}{\partial\phi\,\partial\alpha}$ and $\frac{\partial^2}{\partial\lambda\,\partial\phi}$ respectively to produce the symmetric variance-covariance matrix. The second derivatives of the log-likelihood determine the observed Fisher information. We denote $\ddot{\ell}_{\alpha\alpha},\ddot{\ell}_{\phi\phi},\ddot{\ell}_{\lambda\lambda},\ddot{\ell}_{\alpha\lambda},\ddot{\ell}_{\alpha\phi},\ddot{\ell}_{\phi\lambda}$ as the second derivatives of the log-likelihood with respect to α,ϕ and λ, and so the observed Fisher Information for Poisson MPT(1) is presented as follow,

Proposition 4.4.4 Let $\ddot{\ell}_{\alpha\alpha},\ddot{\ell}_{\phi\phi},\ddot{\ell}_{\lambda\lambda},\ddot{\ell}_{\alpha\lambda},\ddot{\ell}_{\alpha\phi},\ddot{\ell}_{\phi\lambda}$ denote the second derivatives of the log-likelihood with respect to $\alpha,\phi$and λ. The observed Fisher Information for MPT(1) is presented as follows.

$$\ddot{\ell}_{\alpha\alpha} = \sum_{t=1}^{n} \frac{P(X_t|X_{t-1})\frac{\partial^2}{\partial\alpha^2}P(X_t|X_{t-1}) - \left(\frac{\partial}{\partial\alpha}P(X_t|X_{t-1})\right)^2}{P(X_t|X_{t-1})^2} \tag{4.9a}$$

$$\ddot{\ell}_{\phi\phi} = \sum_{t=1}^{n} \frac{P(X_t|X_{t-1})\frac{\partial^2}{\partial\phi^2}P(X_t|X_{t-1}) - \left(\frac{\partial}{\partial\phi}P(X_t|X_{t-1})\right)^2}{P(X_t|X_{t-1})^2} \tag{4.9b}$$

$$\ddot{\ell}_{\lambda\lambda} = \sum_{t=1}^{n} \frac{P(X_t|X_{t-1})\frac{\partial^2}{\partial\lambda^2}P(X_t|X_{t-1}) - \left(\frac{\partial}{\partial\lambda}P(X_t|X_{t-1})\right)^2}{P(X_t|X_{t-1})^2} \tag{4.9c}$$

$$\ddot{\ell}_{\alpha\lambda} = \sum_{t=1}^{n} \frac{P(X_t|X_{t-1})\frac{\partial^2}{\partial\lambda\partial\alpha}P(X_t|X_{t-1}) - \frac{\partial}{\partial\lambda}P(X_t|X_{t-1})\frac{\partial}{\partial\alpha}P(X_t|X_{t-1})}{P(X_t|X_{t-1})^2} \tag{4.9d}$$

$$\ddot{\ell}_{\phi\alpha} = \sum_{t=1}^{n} \frac{P(X_t|X_{t-1})\frac{\partial^2}{\partial\phi\partial\alpha}P(X_t|X_{t-1}) - \frac{\partial}{\partial\phi}P(X_t|X_{t-1})\frac{\partial}{\partial\alpha}P(X_t|X_{t-1})}{P(X_t|X_{t-1})^2} \tag{4.9e}$$

$$\ddot{\ell}_{\phi\lambda} = \sum_{t=1}^{n} \frac{P(X_t|X_{t-1})\frac{\partial^2}{\partial\phi\partial\lambda}P(X_t|X_{t-1}) - \frac{\partial}{\partial\phi}P(X_t|X_{t-1})\frac{\partial}{\partial\lambda}P(X_t|X_{t-1})}{P(X_t|X_{t-1})^2} \tag{4.9f}$$

where the first and second derivatives are derived above. To find the elements for Fisher information, the expected Fisher information can be calculated numerically using the results of second derivatives of log-likelihood function, for which the time series data comprised of low counts. Otherwise, one can consider applying observed Fisher information to obtain standard errors of the estimates. Noting that the expected Fisher information is a function of (X_t, X_{t-1}), we gain

Proposition 4.4.5 The elements of Fisher information matrix are given by

$$\begin{aligned} &E[\ddot{\ell}_{\alpha\alpha}] \\ &= \sum_{t=2}^{n} E[h(X_t, X_{t-1})] \\ &= \sum_{t=2}^{n} \sum_{all\ \{x_t, x_{t-1}\}} h(x_t, x_{t-1}) P(X_t = x_t, X_{t-1} = x_{t-1}) \end{aligned} \tag{4.10a}$$

$$= (n-1) \sum_{all\ \{x_t, x_{t-1}\}} P(X_{t-1} = x_{t-1}) \left(\frac{\partial^2}{\partial \alpha^2} P(X_t|X_{t-1}) - \frac{\left(\frac{\partial}{\partial \alpha} P(X_t|X_{t-1}) \right)^2}{P(X_t|X_{t-1})} \right)$$

$$E[\ddot{\ell}_{\phi\phi}] = (n-1) \sum_{all\ \{x_t, x_{t-1}\}} P(X_{t-1} = x_{t-1}) \left(\frac{\partial^2}{\partial \phi^2} P(X_t|X_{t-1}) - \frac{\left(\frac{\partial}{\partial \phi} P(X_t|X_{t-1}) \right)^2}{P(X_t|X_{t-1})} \right) \quad (4.10b)$$

$$E[\ddot{\ell}_{\lambda\lambda}] = (n-1) \sum_{all\ \{x_t, x_{t-1}\}} P(X_{t-1} = x_{t-1}) \left(\frac{\partial^2}{\partial \lambda^2} P(X_t|X_{t-1}) - \frac{\left(\frac{\partial}{\partial \lambda} P(X_t|X_{t-1}) \right)^2}{P(X_t|X_{t-1})} \right) \quad (4.10c)$$

$$E[\ddot{\ell}_{\alpha\lambda}]$$
$$= (n-1) \sum_{all\ \{x_t, x_{t-1}\}} P(X_{t-1} = x_{t-1}) \left(\frac{\partial^2}{\partial \lambda \partial \alpha} P(X_t|X_{t-1}) - \frac{\frac{\partial}{\partial \lambda} P(X_t|X_{t-1}) \frac{\partial}{\partial \alpha} P(X_t|X_{t-1})}{P(X_t|X_{t-1})} \right) \quad (4.10d)$$

$$E[\ddot{\ell}_{\alpha\phi}]$$
$$= (n-1) \sum_{all\ \{x_t, x_{t-1}\}} P(X_{t-1} = x_{t-1}) \left(\frac{\partial^2}{\partial \phi \partial \alpha} P(X_t|X_{t-1}) - \frac{\frac{\partial}{\partial \phi} P(X_t|X_{t-1}) \frac{\partial}{\partial \alpha} P(X_t|X_{t-1})}{P(X_t|X_{t-1})} \right) \quad (4.10e)$$

$$E[\ddot{\ell}_{\phi\lambda}]$$
$$= (n-1) \sum_{all\ \{x_t, x_{t-1}\}} P(X_{t-1} = x_{t-1}) \left(\frac{\partial^2}{\partial \phi \partial \lambda} P(X_t|X_{t-1}) - \frac{\frac{\partial}{\partial \phi} P(X_t|X_{t-1}) \frac{\partial}{\partial \lambda} P(X_t|X_{t-1})}{P(X_t|X_{t-1})} \right) \quad (4.10f)$$

The expected Fisher information is used to compute the asymptotic distribution of the parameter estimates. Bu et al. (2008) derived expressions for the score function and the Fisher information matrix, which form the basis for ML estimation and inference. In the paper, they showed that the score function and Fisher information matrix can be neatly represented as conditional expectations. Freeland and McCabe (2004a) considered a similar approach in the information matrix (IM) test. See Silva et al. (2005) for expected IM derivation for replicated time series sequences.

4.5 Simulation Study of Forecasting Accuracy Measurements

Based upon the forecasting accuracy measurement methods expressed in Chapter 5, Section 5.4, we run some simulations here to obtain better understanding of coherent forecasting. The data set is generated from Geometric Pegram's AR(1) process with known parameters. We generate three sets of data based on parameters (α, p), which are $(0.3,0.5)$, $(0.5,0.5)$ and $(0.8,0.2)$. Simulations are done with data size 1000 for 100 trials, for each set of parameter combination. Then, the data is divided into half; first 500 observations are used for parameter estimation and the rest 500 observations are used to fit all descriptive measures. Table 4.5.1 displays the parameter estimation for all three models. The simulation results show that the parameters are close to the true parameters. We report the results for descriptive measures in Table 4.5.2, Table 4.5.3 and Table 4.5.4. Table 4.5.2 shows the estimated PRMSE, PMAE and PTP of Poisson MPT(1) process for $h = 1, \dots, 4$. One can see that the both PRMSE and PMAD increase in h and PTP decreases in h. The results show that the MPT(1) is able to handle the data which has the nature of Geometric margins. Also, the PTP values show the advantage of using median predictor and mode predictor over mean predictor. For an example, for the parameter of $\hat{\phi}$ =0.1649, $\hat{\alpha}$ =0.3004, $\hat{\lambda}$ =0.9994, the mean predictor gives about 25% to be the same with the true values, but for median and mode predictors, they both give about 47%. Besides, the measurements are consistent when h is increased. On the other hand, INAR(1) and Pegram's AR(1) presents uncertain trends in the descriptive measures. This shows that the INAR(1) and Pegram's AR(1) are not the proper models to handle the generated data.

Table 4.5.1: Parameter estimation and standard errors (in bracket), for MPT(1), INAR(1) and Pegram's AR(1) with Poisson marginal

True Parameters (ϕ, α, λ)	Estimated parameters		
	MPT(1)	INAR(1)	AR(1)
(0.2,0.3,1.0)	$\hat{\phi}$ =0.2422(0.0434) $\hat{\alpha}$ =0.3004(0.0081) $\hat{\lambda}$ =0.9994(0.0268)	$\hat{\alpha}$ =0.2975(0.0510) $\hat{\lambda}$ =1.0012(0.0154)	$\hat{\alpha}$ =0.2979(0.0329) $\hat{\lambda}$ =0.9976(0.0096)
(0.2,0.5,1.5)	$\hat{\phi}$ =0.2541(0.0543) $\hat{\alpha}$ =0.5004(0.0111) $\hat{\lambda}$ =1.4995(0.0333)	$\hat{\alpha}$ =0.4979(0.0240) $\hat{\lambda}$ =1.5556(0.0704)	$\hat{\alpha}$ =0.4983(0.0317) $\hat{\lambda}$ =1.5044(0.0226)
(0.2,0.8,4.0)	$\hat{\phi}$ =0.2675(0.0679) $\hat{\alpha}$ =0.8424(0.0475) $\hat{\lambda}$ =3.8011(0.2211)	$\hat{\alpha}$ =0.7991(0.0203) $\hat{\lambda}$ =3.7824(0.2139)	$\hat{\alpha}$ =0.7988(0.0198) $\hat{\lambda}$ =3.7916(0.2138)

Table 4.5.2: Estimated PRMSE, PMAD and PTP values through mean, median and mode for Poisson MPT(1) process

Parameters $(\hat{\phi}, \hat{\alpha}, \hat{\lambda})$	h-step	PTP			PRMSE	PMAD
		mean	median	mode		
$\hat{\phi}$ =0.1649 $\hat{\alpha}$ =0.3004 $\hat{\lambda}$ =0.9994	1	25.55	46.91	46.91	1.30	1.04
	2	25.35	46.91	46.91	1.41	1.05
	3	25.35	46.71	46.71	1.41	1.07
	4	25.35	46.71	46.71	1.41	1.05
$\hat{\phi}$ =0.2541 $\hat{\alpha}$ =0.5004 $\hat{\lambda}$ =1.4995	1	39.72	43.11	43.11	1.56	1.27
	2	15.57	42.91	42.91	1.76	1.29
	3	15.57	42.71	42.71	1.78	1.29
	4	15.57	42.51	42.51	1.78	1.29
$\hat{\phi}$ =0.2675 $\hat{\alpha}$ =0.8424 $\hat{\lambda}$ =3.8011	1	16.17	19.96	19.96	2.93	3.68
	2	7.78	19.96	19.96	3.66	3.65
	3	7.78	19.76	19.76	3.85	3.63
	4	7.78	19.76	19.76	3.89	3.62

Table 4.5.3: Estimated PRMSE, PMAD and PTP values through mean, median and mode for Poisson INAR(1) process

Parameters	h-step	PTP			PRMSE	PMAD
$(\hat{\alpha}, \hat{\lambda})$		mean	median	mode		
$\hat{\alpha}$ =0.2975	1	3.60	67.10	67.10	2.09	1.28
$\hat{\lambda}$ =1.0012	2	2.30	67.17	67.17	3.25	1.29
	3	2.30	67.23	67.23	3.90	1.29
	4	2.31	67.30	67.30	4.26	1.22
$\hat{\alpha}$ =0.4983	1	20.40	73.40	73.40	0.94	0.58
$\hat{\lambda}$ =1.4990	2	7.71	73.47	73.47	1.38	0.62
	3	7.72	73.55	73.55	1.58	0.56
	4	7.72	73.62	73.62	1.68	0.59
$\hat{\alpha}$ =0.7991	1	2.4	8.20	8.20	3.39	2.16
$\hat{\lambda}$ =3.7824	2	1.8	56.56	56.56	6.10	2.15
	3	1.80	56.61	56.61	8.26	2.10
	4	1.81	56.67	56.67	9.98	2.14

Table 4.5.4: Estimated PRMSE, PMAD and PTP values through mean, median and mode for Poisson Pegram's AR(1) process

Parameters	h-step	PTP			PRMSE	PMAD
$(\hat{\alpha}, \hat{\lambda})$		mean	median	mode		
$\hat{\alpha}$ =0.2979	1	12.40	75.50	75.50	0.87	0.63
$\hat{\lambda}$ =0.9976	2	12.31	75.58	75.58	1.12	0.71
	3	12.32	75.55	75.55	1.19	0.68
	4	12.34	75.53	75.53	1.22	0.50
$\hat{\alpha}$ =0.4983	1	17.60	11.1	11.1	0.92	1.04
$\hat{\lambda}$ =1.5044	2	17.62	71.47	71.47	1.38	0.95
	3	17.64	71.44	71.44	1.60	0.91
	4	17.65	71.40	71.40	1.71	0.89
$\hat{\alpha}$ =0.7988	1	25.9	8.50	8.50	0.77	2.07
$\hat{\lambda}$ =3.7916	2	14.81	8.51	8.51	1.37	2.05
	3	11.02	57.31	57.31	1.85	1.96
	4	11.03	57.37	57.37	2.33	2.02

4.6 Concluding Remarks

This chapter plays an important role for data analysis. For any proposed model, the parameters have to be estimated first before moving forward for real application. We have shown that the EM algorithm has good performance against CLS method. It provides good estimation for mixture model. We have done some simulation study for both Poisson and Binomial MPT(1) model. Also, it is noticeable that Poisson INAR(1) and Pegram's AR(1) processes are not able to handle the data with outliers. On the contrary we see that Poisson MPT(1) process was not greatly affected by small amount of outliers. The breakdown point has been examined.

To further investigate the coherent forecasting in next chapter, we first require some fundamental theories. Score functions and expected Information Matrix have been developed to study the relationship of variance-covariance matrix. See Chapter 5 for more in-depth discussion.

CHAPTER 5: COHERENT FORECASTING

5.0 Background

Previous chapter discussed on parameter estimation of MPT(1) model, specifically in Poisson marginal. Once the parameters have been identified, the study on coherent forecasting of Poisson MPT(1) process can be further carried out in this chapter. One main objective of modelling the time series data is to forecast the future values as a given threshold of the interested variables. Traditional forecasting in continuous time series modelling is based upon conditional expectation, in which the technique will yield forecasts with minimum mean squared error. However, this method commonly results in non-integer-valued forecasts, which are thus not a reality in the context of count data models. Therefore, it is suggested here that the k-step-ahead conditional distribution can be used instead to forecast the integer-valued time series models. Freeland and McCabe (2004b) suggested using the median (and mode) of probability distribution because the median always lies in the support and is thus coherent. However, they give a conclusive result to say that it would be more informative to consider probability distribution in the support.

In the literature, there are some researchers who have been concerned with the coherency in the forecasting of INAR process. The supportive elements for computing the asymptotic distributions such as score functions and inverse Fisher IM have been expressed in terms of conditional expectation by Freeland and McCabe (2004a). This technique created a new page in coherent forecasting study. Simultaneously but in a separate work, Freeland and McCabe (2004b) exemplified the idea with real data analysis of wage loss claim counts. Silva et al. (2005) considered expected Fisher information matrix for replicated INAR(1) process. Later, Bu and McCabe (2008) continued to conduct forecasting in the INAR(p) models. Silva et al. (2009) considered coherent forecasts using Bayesian methodology for the numerical computing. Several researchers have also considered the Bayesian method in obtaining the confidence intervals. For instances, McCabe and Martin (2005) considered the Bayesian predictions for low count time series discrete valued data. Other than that, Jung and Tremayne (2006) used block-bootstrap techniques to estimate the asymptotic standard errors for coherent forecasting.

Having reviewed all these works, it is noticed that there is not much research involving coherent forecasting for discrete-valued models. Therefore, we are motivated to consider coherent forecasting for Poisson MPT(1) model in this chapter. Section 5.1 gave the forecasting properties of k-step-ahead forecasting distributions. The expressions for score functions and Fisher information matrix as previously discussed in subsection 4.4 will be further investigated in Section 5.2. The readily estimated parameters have also been applied to compute the conditional distribution in forecasting. Section 5.3 examines the prediction intervals. Section 5.4 elaborate some descriptive measures of forecasting accuracy. Section 5.5 concludes.

5.1 *k*-step-ahead Forecasting Distribution of Poisson MPT(1) Model

This section constructed k-step ahead conditional distributions for Poisson MPT(1) model. Example 3.4.1 can be viewed as a special case when $k = 1$. Let X_t be a Poisson process with mean λ, the following results have been developed for the forecasting distribution of Poisson MPT(1) process.

Theorem 5.1.1 (*k*-step ahead Conditional Distribution): In the Poisson MPT(1) process the conditional pgf of X_{t+k} given X_t is given by

$$G_{X_{t+k}|X_t}(z) = \phi^k (1 - \alpha^k + \alpha^k z)^{X_t} + e^{\lambda(z-1)} - \phi^k e^{\lambda\alpha^k (z-1)} \tag{5.1}$$

Proof:

We prove the results by mathematical induction.

The 1-step-ahead conditional pgf is given by

$$G_{X_{t+1}|X_t}(z) = E(z^{X_{t+1}}|X_t) = \phi(1-\alpha+\alpha z)^{X_t} + e^{\lambda(z-1)} - \phi e^{\lambda\alpha(z-1)}$$

The 2-step-ahead conditional pgf is given by

$$G_{X_{t+2}|X_t}(z) = E(z^{X_{t+2}}|X_t) = E_{X_{t+1}|X_t}E(z^{X_{t+2}}|X_{t+1})$$

$$= E_{X_{t+1}|X_t}(\phi(1-\alpha+\alpha z)^{X_{t+1}}|X_t) + e^{\lambda(z-1)} - \phi e^{\lambda\alpha(z-1)}$$

$$= \phi^2(1-\alpha^2+\alpha^2 z)^{X_t} + e^{\lambda(z-1)} - \phi^2 e^{\lambda\alpha^2(z-1)}$$

By induction, assuming that k-step-ahead conditional pgf (5.1) holds, then

$$G_{X_{t+k+1}|X_t}(z) = E(z^{X_{t+k+1}}|X_t) = E_{X_{t+1}|X_t}E(z^{X_{t+k+1}}|X_{t+1})$$

$$= \phi^{k+1}(1-\alpha^{k+1}+\alpha^{k+1}z)^{X_t} + e^{\lambda(z-1)} - \phi^{k+1}e^{\lambda\alpha^{k+1}(z-1)}$$

Hence, Theorem 5.1.1 is true for all k.

Corollary 5.1.1 (k-step-ahead Conditional Expectation): Let $E(X_t) = \mu_x$, the k-step-ahead conditional pgf results in the conditional expectation

$$E[X_{t+k}|X_t] = (\phi\alpha)^k X_t + (1-(\phi\alpha)^k)\mu_x \tag{5.2}$$

Proof:

Taking first derivative of conditional pgf and letting $z = 1$, we get

$$G'_{X_{t+k}|X_t}(z) = \phi^k\alpha^k X_t(1-\alpha^k+\alpha^k z)^{X_t-1} + \lambda e^{\lambda(z-1)} - \lambda\phi^k\alpha^k e^{\lambda\alpha^k(z-1)}$$

$$G'_{X_{t+k}|X_t}(z)|_{z=1} = \phi^k\alpha^k X_t + \lambda - \lambda\phi^k\alpha^k$$

$$= \phi^k\alpha^k X_t + \lambda(1-\phi^k\alpha^k)$$

It is slightly complicated to obtain k-step-ahead conditional variance, but it can be found as follow.

Corollary 5.1.2 (k-step-ahead Conditional Variance): The k-step-ahead conditional variance is

$$Var[X_{t+k}|X_t] = \phi^k\alpha^{2k}({X_t}^2 - X_t) + (1-\phi^k\alpha^{2k})\lambda^2 + \phi^k\alpha^k X_t + \lambda(1-\phi^k\alpha^k) - (\phi^k\alpha^k X_t + \lambda(1-\phi^k\alpha^k))^2 \tag{5.3}$$

Proof:

We apply the formula $Var[X_{t+k}|X_t] = G''_{X_{t+k}|X_t}(1) + G'_{X_{t+k}|X_t}(1) - \left[G'_{X_{t+k}|X_t}(1)\right]^2$, and the second derivatives are

$$G''_{X_{t+k}|X_t}(z) = \phi^k\alpha^{2k}X_t(X_t-1)(1-\alpha^k+\alpha^k z)^{X_t-2} + \lambda^2 e^{\lambda(z-1)} - \lambda^2\phi^k\alpha^{2k}e^{\lambda\alpha^k(z-1)}$$

$$G''_{X_{t+k}|X_t}(z)|_{z=1} = \phi^k\alpha^{2k}\left({X_t}^2 - X_t\right) + \lambda^2(1-\phi^k\alpha^{2k})$$

By letting $k \to \infty$, we obtain the unconditional expectation and variance of the MPT(1) process: $E[X_t] = \lambda$, $var(X_t) = \lambda$.

Theorem 5.1.2 (lag-k Autocovariance Function): The lag-k ACVF is given by

$$\gamma(k) = (\phi\alpha)^k Var(X_{t-k}) \tag{5.4}$$

Proof:

We prove the result by using the equation

$cov(X_t\ X_{t-k})$

$= E(X_{t-k}\{(\phi\alpha)X_{t-1} + (1-\phi)\mu_\varepsilon\}) - E(X_t)E(X_{t-k})$

$= (\phi\alpha)E(X_{t-k}X_{t-1}) + (1-\phi)\mu_\varepsilon\mu_x - {\mu_x}^2$

$= (\phi\alpha)E(X_{t-k}\{(\phi\alpha)X_{t-2} + (1-\phi)\mu_\varepsilon\}) + (1-\phi)\mu_\varepsilon\mu_x - {\mu_x}^2$

$= (\phi\alpha)^2E(X_{t-k}X_{t-2}) + (\phi\alpha)(1-\phi)\mu_\varepsilon\mu_x + (1-\phi)\mu_\varepsilon\mu_x - {\mu_x}^2$

$= \qquad \vdots$

$$= (\phi\alpha)^k E\left({X_{t-k}}^2\right) + \sum_{j=0}^{k-1}(\phi\alpha)^j(1-\phi)\mu_\varepsilon\mu_x - \mu_x^2$$

$$= (\phi\alpha)^k Var(X_{t-k}) + (\phi\alpha)^k\mu_x^2 + \sum_{j=0}^{k-1}(\phi\alpha)^j(1-\alpha\phi){\mu_x}^2 - {\mu_x}^2$$

$$= (\phi\alpha)^k Var(X_{t-k}) + (\phi\alpha)^k\mu_x^2 + \sum_{j=0}^{k-1}(\phi\alpha)^j{\mu_x}^2 - \sum_{j=0}^{k-1}(\phi\alpha)^{j+1}{\mu_x}^2 - {\mu_x}^2$$

$= (\phi\alpha)^k Var(X_{t-k}) + \mu_x^2 - \mu_x^2$

$= (\phi\alpha)^k Var(X_{t-k})$

Corollary 5.1.3 (lag-k Autocorrelation Functions): The lag-k autocorrelation function is given by

$$\rho(k) = (\phi\alpha)^k \tag{5.5}$$

It is easy to obtain Eq. (5.5) by just taking $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$, where $\gamma(0)$ is the variance of X_t. Furthermore, we found some immediate results from the propositions above, which are given as follow.

Theorem 5.1.3 (Mean of X_t): Let μ_k denote the conditional mean of $X_{t+k}|X_t$ and let μ_x be the mean of Poisson marginal distribution with $\frac{(1-\phi)\mu_\varepsilon}{1-\phi\alpha}$. Then, $\mu_k \xrightarrow{w} \mu_x$. That is μ_k

converges weakly to μ_x or $X_{t+k}|X_t$ has a Poisson limiting distribution with mean $\frac{(1-\phi)\mu_\varepsilon}{1-\phi\alpha}$.

Proof:

From (5.2), we have

$$\lim_{k\to\infty} E[X_{t+k}|X_t] = \lim_{k\to\infty} (\phi\alpha)^k X_t + (1-(\phi\alpha)^k)\mu_x$$
$$= \mu_x = \frac{(1-\phi)\mu_\varepsilon}{1-\phi\alpha}$$

Theorem 5.1.4 (Variance of X_t): Let σ_k^2 denote the conditional variance of $X_{t+k}|X_t$ and let σ_X^2 be the variance of Poisson marginal distribution. Then, $\sigma_k^2 \xrightarrow{w} \sigma_X^2$. That is σ_k^2 converges weakly to σ_X^2.

Proof:

From (5.3), we have

$$\lim_{k\to\infty} Var[X_{t+k}|X_t] = \lim_{k\to\infty} \phi^k\alpha^{2k}\left({X_t}^2 - X_t\right) + (1-\phi^k\alpha^{2k})\lambda^2 + \phi^k\alpha^k X_t$$
$$+\lambda(1-\phi^k\alpha^k) - \left(\phi^k\alpha^k X_t + \lambda(1-\phi^k\alpha^k)\right)^2$$
$$= \lambda = Var(X_t)$$

Theorem 5.1.5 (Marginal Distribution of X_t): Let $P_k(x)$ denote the distribution of $X_{t+k}|X_t$ and let $P_X(x)$ be the distribution of a Poisson random variable which is $P_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,\ldots$. Then, $P_k(x) \xrightarrow{w} P_X(x)$. That is $P_k(x)$ converges weakly to $P_X(x)$. The proof is given as follow.

Proof:

Taking multiple derivatives of (5.1) we have

$$\lim_{k\to\infty} P(X_{t+k} = x|X_t = x_t)$$

$$= \lim_{k\to\infty} \phi^k \binom{x_t}{x} (\alpha^k)^x (1-\alpha^k)^{X_t - x} + \lim_{k\to\infty} \sum_{i=0}^{k-1} \phi^i \left\{ \frac{e^{-\lambda\alpha^i}\left(\lambda\alpha^i\right)^x}{x!} - \phi^k \frac{e^{-\lambda\alpha^{i+1}}\left(\lambda\alpha^{i+1}\right)^x}{x!} \right\}$$

$$= \lim_{k\to\infty} \left\{ \frac{e^{-\lambda}\lambda^x}{x!} + \sum_{i=1}^{k-1} \phi^i \frac{e^{-\lambda\alpha^i}(\lambda\alpha^i)^x}{x!} - \sum_{j=1}^{k} \phi^j \frac{e^{-\lambda\alpha^j}(\lambda\alpha^j)^x}{x!} \right\}$$

$$= \lim_{k\to\infty} \left\{ \frac{e^{-\lambda}\lambda^x}{x!} - \phi^k \frac{e^{-\lambda\alpha^k}(\lambda\alpha^k)^x}{x!} \right\}$$

$$= \frac{e^{-\lambda}\lambda^x}{x!} = P_X(x)$$

5.2 Point Mass Forecasts

Point mass forecasts and k-step-ahead conditional distribution have been developed in this section. Point masses have been applied to forecast by simply referring to the median, and cumulative distributions may determine the median forecast. However, the median may not be very informative for low counts. Therefore, using the probability distribution considered here would be more appropriate to forecast the outcomes.

Theorem 5.1.1 gives the distribution of X_{t+k} given X_t is a mixture distribution. The conditional pmf of Poisson MPT(1) process of X_{t+k} given X_t is,

$$P(X_{t+k} = x|X_t = x_t) = \phi^k \binom{x_t}{x} (\alpha^k)^x (1-\alpha^k)^{X_t - x} + \frac{e^{-\lambda}\lambda^x}{x!} - \phi^k \frac{e^{-\lambda\alpha^k}(\lambda\alpha^k)^x}{x!} \quad (5.6)$$

The expression shown is relatively simpler than the k-step-ahead conditional distribution of Poisson INAR(1) model. See Freeland (1998, Eq. (3.3.1)). In the following section, we illustrate the conditional mean, median, mode and point mass forecasts with real data set. The conditional mode is easy to find since for the non-negative integer point, x, the k-step-ahead conditional probability mass, $p_k(x|X_t)$, is the largest.

5.3 Prediction Intervals

Knowledge of the parameters $\boldsymbol{\theta} = (\alpha, \lambda, \phi)'$ is required to implement the forecasting ideas. The parameters have been prior estimated to compute $p_k(x|X_n)$. Suppose we have a sample of size n and denote the ML estimates for this sample by $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}$. Under some mild regularities conditions as in Freeland (1998, Section 4.1), $\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}}{}'$ is asymptotically normal with mean $\boldsymbol{\theta}_{\mathbf{0}}{}'$ and variance $n^{-1}i^{-1}$, where i is denoted as Fisher information matrix, and $\boldsymbol{\theta}_{\mathbf{0}}{}' = (\alpha_0, \lambda_0, \phi_0)'$ are the true parameters; which have been estimated by MLE via EM algorithm. The inverse Fisher information matrix can be obtained via the following theorem,

Theorem 5.3.1 (Inverse Fisher Information Matrix): Denote the parameters as $\boldsymbol{\theta} = (\alpha, \lambda, \phi)'$. We are interested to compute $\widehat{\boldsymbol{\theta}}$, where $\widehat{\boldsymbol{\theta}}$ is asymptotically normally distributed around the true parameter values for Eq. (5.6), i.e. $\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{0}}\right) \sim N(\mathbf{0}, \mathbf{V})$ for some variance covariance matrix $\mathbf{V}$, where $\mathbf{V}$ is the inverse Fisher information matrix. The inverse matrix can be written as

$$\mathbf{V} = \left(-E\left[\frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial\theta\partial\theta'}\right]\right)^{-1}$$

where

$$\frac{\partial^2 \ln P(X_t|X_{t-1})}{\partial\theta\partial\theta'} = \begin{bmatrix} \ddot{\ell}_{\alpha\alpha} & \ddot{\ell}_{\alpha\lambda} & \ddot{\ell}_{\phi\alpha} \\ \ddot{\ell}_{\alpha\lambda} & \ddot{\ell}_{\lambda\lambda} & \ddot{\ell}_{\phi\lambda} \\ \ddot{\ell}_{\phi\alpha} & \ddot{\ell}_{\phi\lambda} & \ddot{\ell}_{\phi\phi} \end{bmatrix}$$

The elements of Fisher information matrix are calculated by truncating the infinite sums to m, which corresponds to substituting the sample space of $\{0,1,2,\ldots\}$ of X_t by the sample space of $\{0,1,2,\ldots,m\}$. The value for m is selected so that $P(X_t > m) < 10^{-15}$ (Silva et al., 2005). These elements can be computed numerically using the derivatives. The expectation of second derivatives can be computed numerically by using the formulas in Chapter 4, (4.10a) - (4.10f). Now, we applied the δ- method for finding the asymptotic distribution of function $p_k(x|X_n;\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}})$. By referring to Theorem 3.4.1 (Freeland, 1998), we obtain the following asymptotical property,

Theorem 5.3.2 (Asymptotic Normal Distribution): We consider the k-step-ahead conditional probability here as $p_k(x|X_n;\widehat{\boldsymbol{\theta}}_{\boldsymbol{n}})$, which for a fixed x and sample size n, it has an asymptotically normal distribution with $p_k(x|X_n;\boldsymbol{\theta}_{\boldsymbol{0}})$ and variance

$$\begin{aligned}
&{\sigma_k}^2(x;\ \alpha_0,\lambda_0,\phi_0)\\
&= n^{-1}\left[\nu_\alpha\left(\left.\frac{\partial p_k}{\partial\alpha}\right|_{\alpha=\alpha_0,\lambda=\lambda_0,\phi=\phi_0}\right)^2 + \nu_\lambda\left(\left.\frac{\partial p_k}{\partial\lambda}\right|_{\alpha=\alpha_0,\lambda=\lambda_0,\phi=\phi_0}\right)^2\right.\\
&+ \nu_\phi\left(\left.\frac{\partial p_k}{\partial\phi}\right|_{\alpha=\alpha_0,\lambda=\lambda_0,\phi=\phi_0}\right)^2 + 2\nu_{\alpha\lambda}\left.\frac{\partial p_k}{\partial\alpha}\frac{\partial p_k}{\partial\lambda}\right|_{\alpha=\alpha_0,\lambda=\lambda_0,\phi=\phi_0}\\
&\left.+ 2\nu_{\alpha\phi}\left.\frac{\partial p_k}{\partial\alpha}\frac{\partial p_k}{\partial\phi}\right|_{\alpha=\alpha_0,\lambda=\lambda_0,\phi=\phi_0} + 2\nu_{\lambda\phi}\left.\frac{\partial p_k}{\partial\lambda}\frac{\partial p_k}{\partial\phi}\right|_{\alpha=\alpha_0,\lambda=\lambda_0,\phi=\phi_0}\right]
\end{aligned}\tag{5.7}$$

where ν_α, ν_λ and ν_ϕ are the diagonal elements and $\nu_{\alpha\lambda}$, $\nu_{\alpha\phi}$ and $\nu_{\lambda\phi}$ are the off-diagonal elements of **V** in Theorem 5.3.1, respectively. Here

$$\begin{aligned}
\frac{\partial}{\partial\alpha}p_k(x|X_n) = \frac{\phi^k k\alpha^{k-1}X_n}{1-\alpha^k}&\left(\binom{X_n-1}{x-1}(\alpha^k)^{x-1}(1-\alpha^k)^{X_n-x}\right.\\
&\left.-\binom{X_n}{x}(\alpha^k)^x(1-\alpha^k)^{X_n-x}\right)\\
&-\lambda\phi^k k\alpha^{k-1}\left(\frac{e^{-\lambda\alpha^k}(\lambda\alpha^k)^{x-1}}{(x-1)!}-\frac{e^{-\lambda\alpha^k}(\lambda\alpha^k)^x}{x!}\right)
\end{aligned}$$

$$\frac{\partial}{\partial\lambda}p_k(x|X_n) = \frac{e^{-\lambda}(\lambda)^{x-1}}{(x-1)!} - \frac{e^{-\lambda}(\lambda)^{x}}{(x)!} - (\alpha\phi)^k\left(\frac{e^{-\lambda\alpha^k}(\lambda\alpha^k)^{x-1}}{(x-1)!} - \frac{e^{-\lambda\alpha^k}(\lambda\alpha^k)^{x}}{x!}\right)$$

$$\frac{\partial}{\partial\phi}p_k(x|X_n) = k\phi^{k-1}\left[\binom{X_n}{x}(\alpha^k)^x(1-\alpha^k)^{X_n-x} - \frac{e^{-\lambda\alpha^k}(\lambda\alpha^k)^{x}}{x!}\right]$$

Thus, we can compute a 95% confidence interval for $p_k(x|X_n;\ \alpha_0,\lambda_0,\phi_0)$, based on its asymptotic distribution, by means of

$$p_k\left(x\middle|X_n;\ \hat{\alpha}_n,\hat{\lambda}_n,\hat{\phi}_n\right) \pm 1.96\sigma_k(x;\ \alpha_0,\lambda_0,\phi_0) \tag{5.8}$$

5.4 Descriptive Measures of Forecasting Accuracy

We describe some methods in descriptive measures of forecasting accuracy. Given a set of observed data $\{Y_1, Y_2, \ldots, Y_n, Y_{n+1} \ldots, Y_{n+m}\}$ of size $(n+m)$, the data is partitioned into two sets; $\{Y_1, Y_2, \ldots, Y_n\}$ is called training set and $\{Y_{n+1} \ldots, Y_{n+m}\}$ is called testing set. The training data set containing first n observations is used for parameter estimation and the rest of the data which containing m observations is used for forecasting accuracy measurements. There are three measurements for forecasting accuracy which are described as follow,

i) Prediction root mean squared error (denoted by PRMSE), which is defined as

$$PRMSE = \sqrt{\frac{1}{m}\sum_{h=1}^{m}\left(Y_{t+h} - \hat{Y}_{t+h}{}^{me}\right)^2}$$

where $\hat{Y}_{t+h}{}^{me}$ is the mean of the estimated h-step ahead forecasting distribution Y_{t+h} given Y_t given in Eq. (5.6).

ii) Prediction mean absolute deviation (denoted by PMAD), which is defined as

$$PMAE = \frac{1}{m}\sum_{h=1}^{m}\left|Y_{t+h} - \hat{Y}_{t+h}{}^{med}\right|$$

where $\hat{Y}_{t+h}{}^{med}$ is the median of the estimated h-step ahead forecasting distribution Y_{t+h} given Y_t given in Eq. (5.6).

iii) Percentage of true predictions (PTP), which is defined as

$$PTP = \frac{1}{m}\sum_{h=1}^{m} I\left(Y_{t+h} - \hat{Y}_{t+h}\right) \times 100\%$$

where $I(\cdot)$ is an indicator function. Here, $\hat{Y}_{t+h}$ can be the predictive mean, median and mode (round to the nearest integer) in Eq. (5.6).

Remark: The PRMSE(h) and PMAD(h) should intuitively increase in h whereas the PTP(h) should intuitively decrease in h.

All the methods above are used to measure the performance of forecasting models, Poisson MPT(1), Poisson INAR(1) and Poisson Pegram's AR(1) processes. Some simulation has been done based on these forecasting measurements. See the simulation results in Section 4.5, Chapter 4 for more information.

5.5 Concluding Remarks

This chapter discussed the significance of coherent forecasting in discrete-valued time series modelling and subsequently developed coherent forecasting for a new Poisson MPT(1) process. Important quantities of the new model, such as score functions and Fisher information matrix, have also been derived. Prediction intervals together with asymptotical normal distribution have been studied. Numerical simulation has been carried out to compare it with existing models, and it is noticed that the results are in favour of the proposed model. Real data has been fitted to the forecasting distribution for illustration. See Chapter 7 for real data illustration.

CHAPTER 6: MIXED AUTOREGRESSIVE MOVING AVERAGE MODELS

6.0 Introduction

The first-order mixed Autoregressive (AR) model, a mixture of Pegram and Thinning (MPT) model, its relevant properties, parameter estimation with simulation studies have been investigated in detail. For a more complete study of time series models for the MPT family, Moving Average (MA) and higher order of AR (AR(p)) processes, as well as the mixed Autoregressive Moving Average (ARMA) of MPT processes will be discussed in this Chapter together with fitting for some discrete marginal distributions. Chapter 6 comprises several sections. Section 6.1 gives the literature reviews of existing ARMA models. The first-order integer-valued MPT Moving Average (MPT-MA(1)) model is constructed in Section 6.2. The sample paths of Poisson and Binomial marginals are displayed to illustrate realizations of the process. The important property of reversibility has been considered. Section 6.3 extends the order-one MPT-MA process to higher order of MPT-MA processes; it is named as MPT of q-th order integer-valued MA processes which is abbreviated by MPT-MA(q). Section 6.4 is an immediate extension from Chapter 3. The higher order of MPT processes have been considered. The first-order integer-valued MPT Autoregressive Moving Average (MPT-ARMA(1,1)) model has been developed in Section 6.5. All relevant statistical and regression properties have been studied for each of the considered models. Section 6.6 concludes.

6.1 Background of ARMA Processes

There have been much research studies on ARMA processes in continuous time series modelling for past few decades. Back in 1970s, Jacobs and Lewis (1977) were perhaps the first to develop a mixed ARMA(1,1) with exponential sequence (EARMA(1,1)). Later, this model was then extended to a general mixed EARMA(p,q)

(Lawrence and Lewis, 1980). McKenzie (1981) continued the development of the EARMA model by extending the range of the correlation to negative values. The vast exploration of ARMA models motivates modelling for the discrete case in a similar direction. Discrete-valued ARMA models are rare. McKenzie (1988) developed a family of models for discrete-time processes with Poisson marginal distributions. Brannas and Hall (2001) presented a new characterization of an integer-valued MA models. They derived the moments and probability generating functions for four MA model variants. Al-Osh and Alzaid (1988) provided various properties of an integer-valued Moving Average (INMA) model, such as joint distribution, regression, time reversibility, partial correlations and conditional regression properties. The non-Markovian INMA model has been investigated for the Poisson sequences only. Aly and Bouzar (1994) provided a detailed and complete discussion on the ARMA models and the generalizations in terms of probability generating functions. On the other hand, Jacobs and Lewis (1978a, b) introduced an entirely different discrete-valued time series models, the discrete-valued Autoregressive Moving Average models, in which the sequence X_t is taken as a probabilistic linear combination of i.i.d. integer-valued random variables. Also, Biswas and Song (2009) introduced DARMA models, analogue to Jacobs-Lewis approach but with simpler interpretation, to accommodate categorical data. The literature review suggested the construction and extension of the MPT processes. The proposed models have been based upon the structure of the well-known MA and ARMA models. The study with regards to autocovariance and autocorrelation structures which are analogue of continuous ARMA models will also be considered for the discrete-valued models.

6.2 Mixture of Pegram and Thinning First Order Integer-Valued Moving Average Process

Here, we focus on a correlated sequence of $\{X_t\}$. A stationary sequence $\{\varepsilon_t\}$ of i.i.d non-negative integer-valued random variable is introduced. The sequence of $\{X_t\}$ and $\{\varepsilon_t\}$ are uncorrelated. We have the following definition of a mixed Pegram and thinning MA process.

Definition 6.2.1 (Mixture of Pegram-INMA(1) process): Let $\{X_t\}_{t\in\mathbb{N}}$ be a sequence independent of time. Suppose the sequence $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is an i.i.d. non-negative integer-valued random variables having mean μ_ε and variance σ_ε^2. Let $\beta \in [0,1]$ with $P(W_i = 1=1-P(Wi=0)=\beta$. Define the random variable $\beta\circ\varepsilon t$ by

$$\beta \circ \varepsilon_t = \sum_{i=1}^{\varepsilon_t} W_i$$

with the sequence of $\{W_i\}_{i\in\mathbb{N}}$ is i.i.d. random variables independent of $\{X_t\}_{t\in\mathbb{N}}$ and $\{\varepsilon_t\}_{t\in\mathbb{N}}$. Then, we say that X_t is a mixture of Pegram and thinning first order integer-valued Moving Average, abbreviated by MPT-MA(1) process if it admits the form

$$X_t = (\theta_0, \beta_0 \circ \varepsilon_t) * (\theta_1, \beta_1 \circ \varepsilon_{t-1}) \tag{6.1}$$

where $\theta_0, \theta_1 \in (0,1)$, $\beta_0, \beta_1 \in [0,1]$ are the mixing weights, $\sum_{i=0}^{q} \theta_i = 1$ and $\sum_{i=0}^{q} \beta_i = 1$. Here, $q = 1$. The pgf is given by

$$G_X(z) = \theta_0 G_\varepsilon(1 - \beta_0 + \beta_0 z) + \theta_1 G_\varepsilon(1 - \beta_1 + \beta_1 z) \tag{6.2}$$

where $G_\varepsilon(z)$ is the pgf of ε_t.

For simplicity we assign the value of $\beta_0 = 1$ here. It is reminded that $1 \circ X = X$ (Al-Osh and Alzaid, 1987). However it is possible for β_0 to take any real numbers between

0 and 1 inclusive. See Brannas and Hall (2001) discussed four models for potential real life applications. They show that β_0 could be assigned differently to suit the respective models.

6.2.1 Model Interpretation

It is obvious that model (6.1) is a mixture model consisting of two independent components ε_t and ε_{t-1}($\beta \circ \varepsilon_{t-i}$ and $\beta \circ \varepsilon_{t-j}$ are independent for $i \neq j$), mixed by Pegram's operator (Pegram, 1980) with respect to the mixing parameter, as in the MPT(1) model. Note that $\beta \circ \varepsilon_{t-1}$ is a binomial random variable having the parameters of $(\varepsilon_{t-1}, \beta)$; here we assume $U_{t-1} = \beta \circ \varepsilon_{t-1}$, and ε_t is i.i.d. discrete-valued random variables having mean μ and variance σ_ε^2. $\{X_t\}$ is dependent random variable. The interpretation of the MPT-MA(1) process is now apparent. U_{t-1} represents the survivals at time $t-1$, and ε_t represents arrivals at time t, where both counting processes are mixed with the mixing proportions of θ_0 and θ_1 respectively. Realizations of many counting processes such as the number of patients staying in a hospital, the number of customers in the department store and the stock transaction could be described by MPT-MA(1) model.

6.2.2 Properties of MPT-MA(1) Process

Similar to the MA and INMA(1) processes, MPT-MA(1) process is a non-Markovian time series model. This section develops some structure properties as follow,

Theorem 6.2.1 (Conditional pgf of MPT-MA(1) process): Let $\{\varepsilon_t\}$ be any discrete random variable, then the conditional pgf corresponding to Definition 6.2.1 is

$$G_{X_t|\varepsilon_t,\varepsilon_{t-1}}(z) = \theta_0(z)^{\varepsilon_t} + \theta_1(1-\beta_1+\beta_1 z)^{\varepsilon_{t-1}} \tag{6.4}$$

Proof: We have

$$\begin{aligned}
G_{X_t|\varepsilon_t,\varepsilon_{t-1}}(z) &= E(z^{X_t}|\varepsilon_t,\varepsilon_{t-1}) \\
&= \theta_0 E(z^{\varepsilon_t}|\varepsilon_t,\varepsilon_{t-1}) + \theta_1 E\left(z^{\beta_1\circ\varepsilon_{t-1}}|\varepsilon_t,\varepsilon_{t-1}\right) \\
&= \theta_0 E(z^{\varepsilon_t}|\varepsilon_t) + \theta_1 E\left(z^{\beta_1\circ\varepsilon_{t-1}}|\varepsilon_{t-1}\right) \\
&= \theta_0 (z)^{\varepsilon_t} + \theta_1(1-\beta_1+\beta_1 z)^{\varepsilon_{t-1}}
\end{aligned}$$

Corollary 6.2.1 (Conditional Probability Function): We obtain the conditional probability function as

$$P(X_t = i|\varepsilon_t,\varepsilon_{t-1}) = \theta_0 P(\varepsilon_t = i) + \theta_1 \binom{\varepsilon_{t-1}}{i}(\beta_1)^i(1-\beta_1)^{\varepsilon_{t-1}-i}$$

Finally, by assuming the value of $\beta_0 = 1$ we show that the ACRF of X_t may be derived analogue to Box-Jenkins (1970):

$$\rho_X(k) = \begin{cases} \theta_0\theta_1\beta_1, & k = 1 \\ 0 & k > 1 \end{cases} \tag{6.5}$$

6.2.3 Fitting of Discrete Marginal Distributions

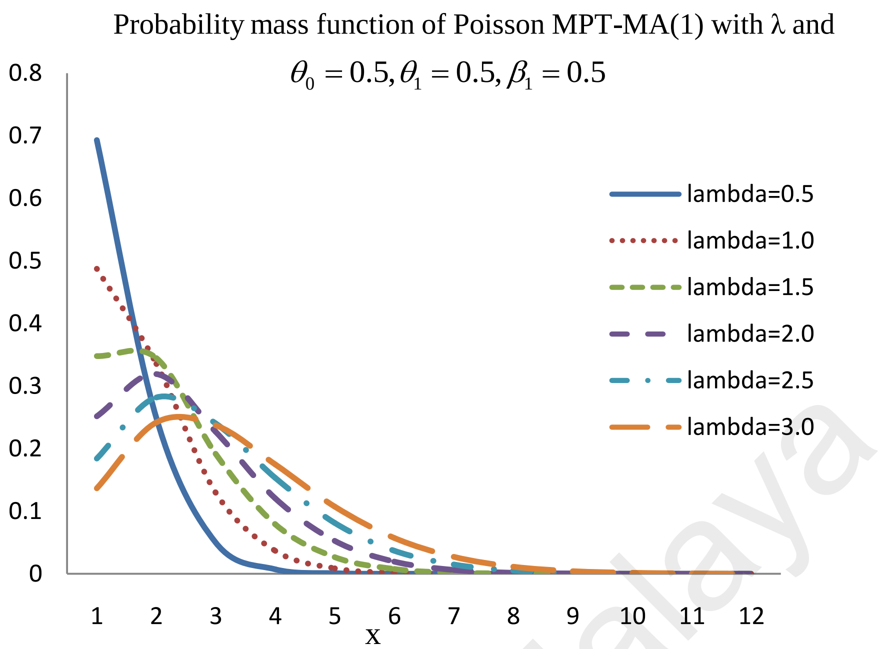
We show that MPT-MA(1) can be fitted by any discrete marginal distribution, including non infinitely divisible distributions like the binomial distribution.

Example 6.2.1 (Poisson MPT-MA(1) Process): Let $\{\varepsilon_t\}$ be a Poisson process with mean λ and $\beta_0 = 1$, then the pgf of Eq. (6.2) is given by

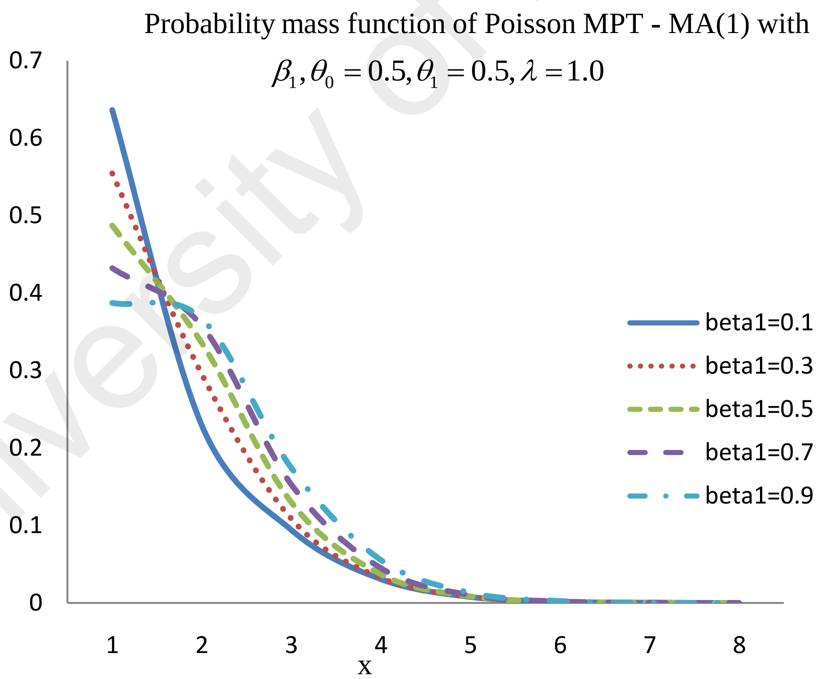
$$G_X(z) = \theta_0 e^{\lambda(z-1)} + \theta_1 e^{\lambda\beta_1(z-1)} \tag{6.6}$$

A compound distribution has pgf form of $G_1(G_2(z))$, Eq. (6.6) shows that X_t has compound Poisson mixture process with the mean $(\theta_0 + \beta_1\theta_1)\lambda$ and variance $\big(\theta_0 +$

${\beta_1}^2\theta_1\big)\lambda^2$. The condition $\frac{\theta_0}{\theta_1} < e^{\lambda(1-\beta_1)}$ ensures that the pmf of X_t is strictly positive. The simulated pmf, sample paths and respective histograms of Poisson MPT-MA(1) process are displayed in the following graphs.



(a)



(b)

Figure 6.2.1: Simulated probability mass function with different combination of parameters

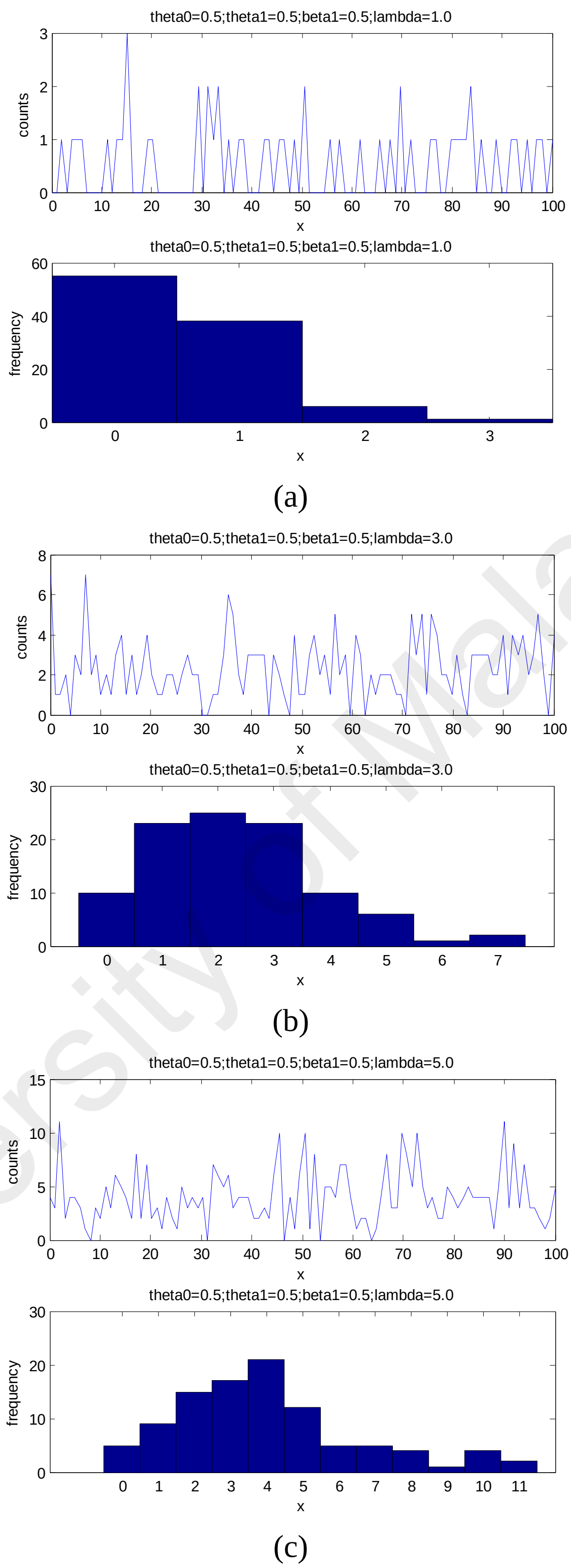


Figure 6.2.2: Simulated sample paths and histogram of Poisson MPT-MA(1) process for $\lambda = 1.0, 2.0, 3.0$; $\theta_0, \theta_1 = 0.5$; $\beta_1 = 0.5$

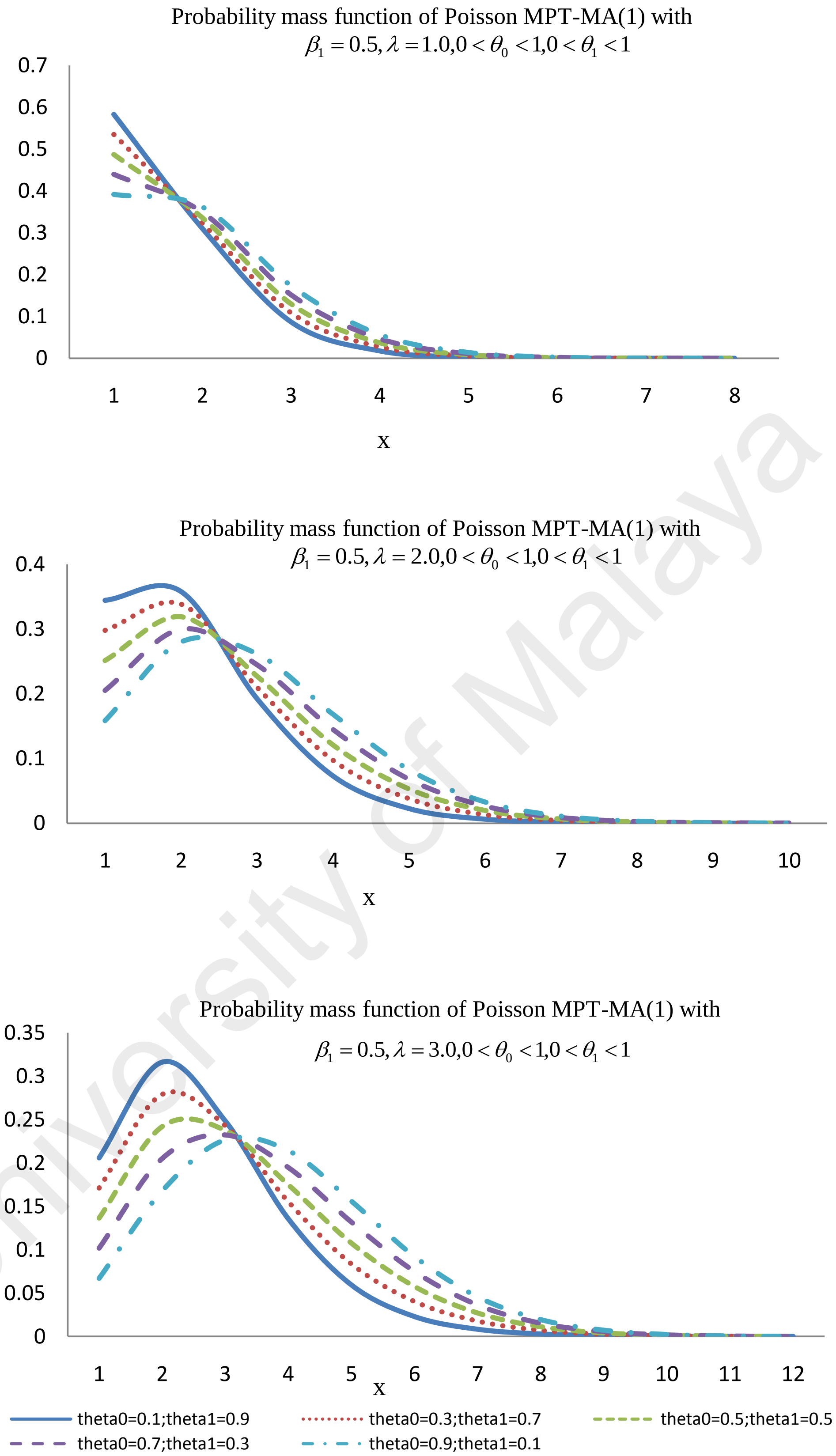


Figure 6.2.3: Probability mass function of Poisson MPT-MA(1) Process with different combination of parameters

In Figure 6.2.1, we varied the parameters of λ and β_1 to observe the pattern of pmf of Poisson MA(1) model. It is seen that the distribution is long tailed. The skewness to the right is getting more pronounced when parameter λ increases. Figure 6.2.2 shows the simulated sample paths and histogram of Poisson MA(1) process. When λ is increased, we tend to obtain higher counts in the sample paths. It can be seen in Figure 6.2.3, the parameter λ determines the shape of the distribution. The bell-shaped and long tailed distribution is more apparent when λ is large. Comparatively, both Figure 3.4.3(a) and 6.2.1(a) gives similar pattern for the pmf. This is the similarity between Poisson AR(1) and Poisson MA(1) processes. It is of interest to compare both models in real life applications. It is noted that our model is practically applicable for any discrete marginal distribution. This is the main different with INMA(1) process, which accommodates only Poisson marginal (Al-Osh and Alzaid, 1988).

Example 6.2.2 (Binomial MPT-MA(1) process): Let the random variables $\{\varepsilon_t\}$ be the Binomial process with the parameters (N, p), where N is the number of trials and p is the probability of success. Let $\beta_0 = 1$ and the pgf of $\{X_t\}$ is given by

$$G_X(z) = \theta_0(1 - p + pz)^N + \theta_1(1 - \beta_1 p + \beta_1 pz)^N \tag{6.8}$$

where $p, \beta_1 \in [0,1]$.

We simulated the pmf of binomial MA(1) process with arbitrary combination of parameters in Figure 6.2.4. It is clearly seen that binomial MA(1) process appears to have two peaks for some parameter combinations, especially when we fixed the parameter values for of θ_0, θ_1 and β_1 and vary the parameter p. In Figure 6.2.4(b), there are two modes for smaller values of β_1, and this evolves to a nice bell shape when β_1 gets larger. We fixed the parameters p and β_1, and vary the parameters θ_0, θ_1 in Figure 6.2.4 (c). Initially with small value of θ_0 and large value of θ_1, the combination results in a bell-shaped curve which is skewed to the right. When both parameters are nearly equal, two modes appear. Finally, large θ_0 and small θ_1, results in a bell-shaped curve with skewing to the left. These phenomena shows that the binomial MA(1) process holds important characteristic of a mixture distribution. It may be the model to cater for data with multimodality in real life application. Additional information for the binomial MA(1) process are obtained from the simulated sample paths and frequency histogram shown in Figure 6.2.5. Figure 6.2.5 (a) shows a distribution skewed to the right while Figure 6.2.5 (b) is almost bell-shaped. We conclude that the larger sample mean will provide higher counts in the sample paths. Besides that, left-skewed bimodal distribution appears in the simulated sample path as shown in Figure 6.2.5 (c). This indicates that the binomial MA(1) process is a potential candidate to handle different shapes in count data and is thus flexible model for real life application.

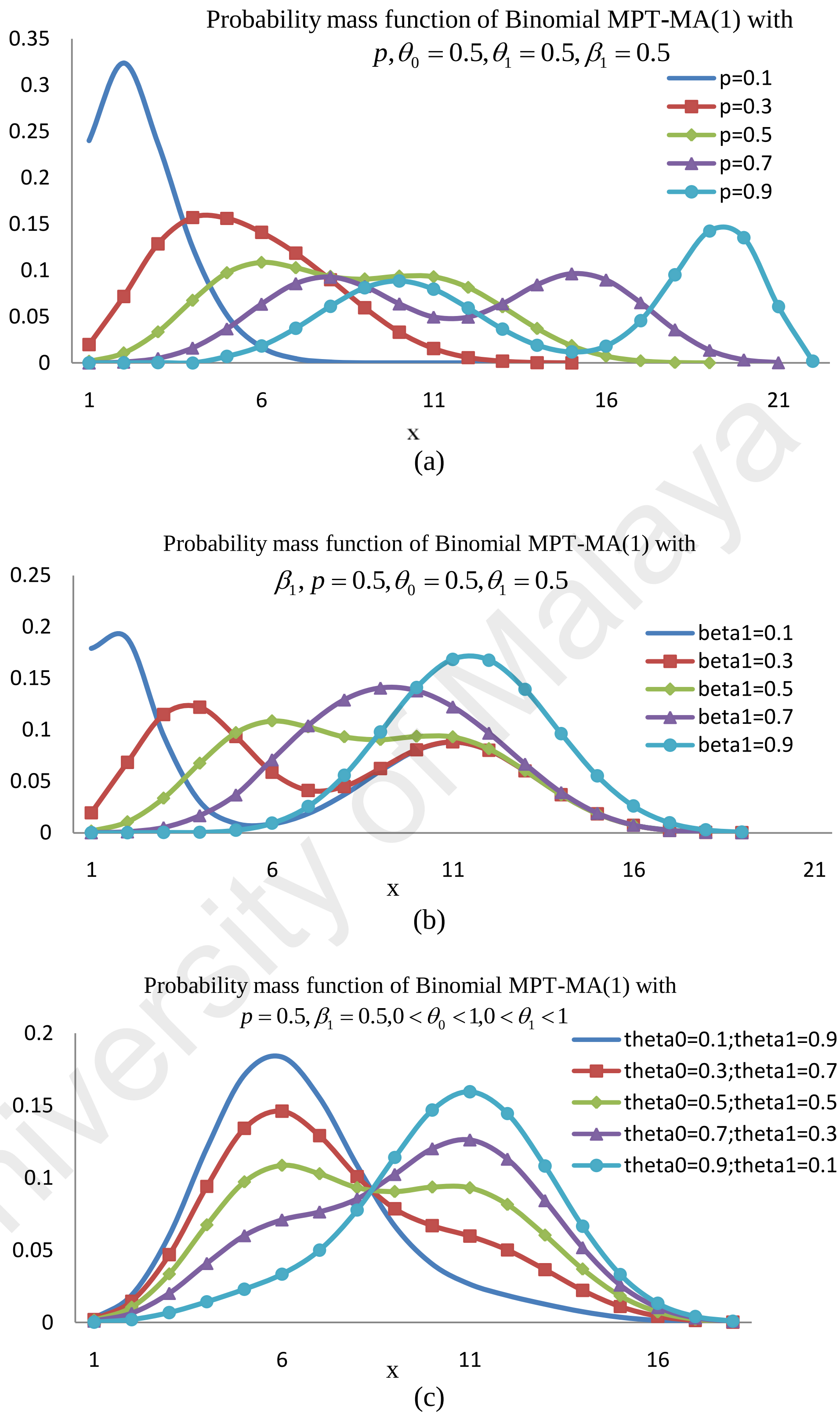
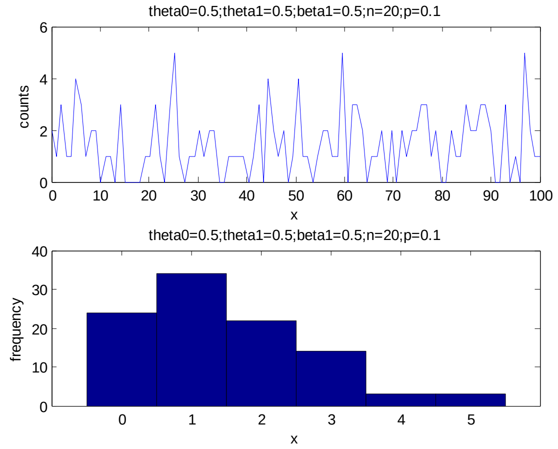
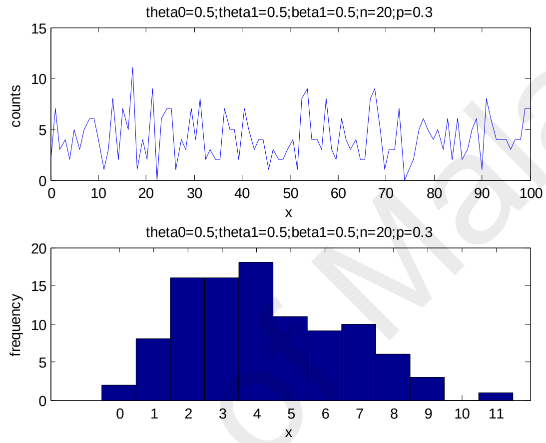


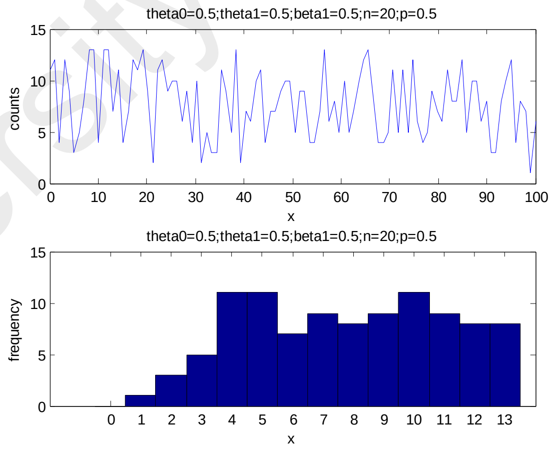
Figure 6.2.4: Simulated probability mass function of binomial MPT-MA(1) process with different combination of parameters



(a)



(b)



(c)

Figure 6.2.5: Simulated sample paths and frequency histogram of binomial MPT-MA(1) process for $N = 20$; $\theta_0, \theta_1 = 0.5$; $\beta_1 = 0.5$; $(a) p = 0.1$; $(b) p = 0.3$; $(c) p = 0.5$

6.3 Mixture of Pegram and Thinning of qth-Order Moving Average Process

In order to attain flexibility in the discrete time series modelling, it is natural to extend MPT-MA(1) to higher order of MA process, which is abbreviated here with MPT-MA(q). At the first sight, it seems reasonable to define the MPT-MA(q) process based upon Definition 6.2.1 extension as defined by

Definition 6.3.1 (Mixture of Pegram-INMA(q) process): Let X_t be a discrete-valued stochastic process such that

$$X_t = (\theta_0, \beta_0 \circ \varepsilon_t) * (\theta_1, \beta_1 \circ \varepsilon_{t-1}) * (\theta_2, \beta_2 \circ \varepsilon_{t-2}) * \dots * (\theta_q, \beta_q \circ \varepsilon_{t-q}) \qquad (6.10)$$

where $\sum_{i=0}^{q} \theta_i = 1$ and $\theta_i \geq 0$, and $\{\varepsilon_t\}$ is an i.i.d. random variable having mean μ_ε and variance σ_ε^2. The pgf is

$$G_X(z) = \sum_{i=0}^{q} \theta_i G_{\varepsilon_{t-i}}(1 - \beta_i + \beta_i z) \qquad (6.11)$$

Example 6.3.1 (Poisson MPT-MA(q) Process): Let ε_t be Poisson(λ), then the pgf is given by

$$G_X(z) = \sum_{i=0}^{q} \theta_i e^{\lambda \beta_i (z-1)} \qquad (6.12)$$

Since $\mu_\varepsilon = \sigma_\varepsilon^2 = \lambda$, it is easy to see that the expectation of X_t is $E(X_t) = \sum_{i=0}^{q} \theta_i \beta_i \lambda$ and variance $Var(X_t) = \sum_{i=0}^{q} \theta_i {\beta_i}^2 \lambda^2$. Therefore, the random variable X_t is a mixture of compound Poisson marginal distributions.

Theorem 6.3.1 (Conditional pgf of MPT-MA(q) Process): Let $\{\varepsilon_t\}$ be any discrete random variables, then the conditional pgf that corresponds to Definition 6.3.1 is generally given by

$$G_{X_t|\varepsilon_t,\varepsilon_{t-1},\ldots,\varepsilon_{t-q}}(z) = \sum_{i=0}^{q} \theta_i (1 - \beta_i + \beta_i z)^{\varepsilon_{t-i}}$$

Corollary 6.3.1 (Conditional pmf of MPT-MA(q) Process): The conditional probability function for MPT-MA(q) process is

$$P\left(X_t = i \middle| \varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}\right) = \theta_0 P(\varepsilon_t = i) + \sum_{j=1}^{q} \theta_j \binom{\varepsilon_{t-j}}{i} \left(\beta_j\right)^i \left(1 - \beta_j\right)^{\varepsilon_{t-j}-i} \quad (6.13)$$

Corollary 6.3.2 (Conditional expectation of MPT-MA(q) Process): By taking the first derivative of Theorem 6.3.1 and let $z = 1$, the conditional expectation is given by

$$E\left(X_t \middle| \varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}\right) = \sum_{i=0}^{q} \theta_i \beta_i \varepsilon_{t-i} \quad (6.14)$$

Corollary 6.3.3 (Conditional variance of MPT-MA(q) Process): By the similar expression we obtain the conditional variance as follow

$$Var\left(X_t \middle| \varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}\right) = \sum_{i=0}^{q} \theta_i {\beta_i}^2 \varepsilon_{t-i}(\varepsilon_{t-i} - 1) - \sum_{i=0}^{q} \theta_i \beta_i \varepsilon_{t-i} + \left(\sum_{i=0}^{q} \theta_i \beta_i \varepsilon_{t-i}\right)^2 \quad (6.15)$$

Proposition 6.3.1 (lag-k Autocovariance function of MPT-MA(q) process): Let ε_t be i.i.d. random variable as stated in Definition 6.3.1, the lag-k autocovariance function of MPT-MA(q) process is given by

$$\gamma(k) = \begin{cases} {\theta_0}^2 + (\theta_1\beta_1)^2 + (\theta_2\beta_2)^2 + \cdots + \left(\theta_q\beta_q\right)^2 {\sigma_\varepsilon}^2 \ ; k = 0 \\ \theta_0(\theta_k\beta_k) + (\theta_1\beta_1)(\theta_{k+1}\beta_{k+1}) + \cdots + \left(\theta_q\beta_q\right)\left(\theta_{q-k}\beta_{q-k}\right){\sigma_\varepsilon}^2 \ ; k = 1,2,\ldots,q \\ 0 \ ; k > q \end{cases} \quad (6.16)$$

Proof:

Let $cov(X_t, X_{t-k})$ be the autovariance of X_t and X_{t-k}. Then,

$$\begin{aligned} cov(X_t, X_{t-k}) &= E(X_t\, X_{t-k}) \\ &= E(\{\theta_0\beta_0\varepsilon_t + \theta_1\beta_1\varepsilon_{t-1} + \cdots + \theta_q\beta_q\varepsilon_{t-q}\}\{\theta_0\beta_0\varepsilon_{t-k} + \theta_1\beta_1\varepsilon_{t-k-1} \\ &\qquad + \cdots + \theta_q\beta_q\varepsilon_{t-q-k}\}) \\ &= E\left((\theta_0\beta_0)^2\varepsilon_t\varepsilon_{t-k} + (\theta_0\beta_0)(\theta_1\beta_1)\varepsilon_t\varepsilon_{t-k-1} + \cdots \right. \\ &\qquad \left. + \left(\theta_q\beta_q\right)^2\varepsilon_{t-q}\varepsilon_{t-k-q}\right) \end{aligned}$$

Substitute $k = 0,1, \dots, q$ to obtain Proposition 6.3.1. Subsequently, Proposition 6.3.2 is obtained.

Proposition 6.3.2 (lag-k Autocorrelation function of MPT-MA(q) process) The lag-k autocorrelation function is given by

$$\rho(k) = \begin{cases} 1 \;\; ; k = 0 \\ \dfrac{\theta_0(\theta_k\beta_k) + (\theta_1\beta_1)(\theta_{k+1}\beta_{k+1}) + \cdots + \left(\theta_q\beta_q\right)\left(\theta_{q-k}\beta_{q-k}\right)}{{\theta_0}^2 + (\theta_1\beta_1)^2 + (\theta_2\beta_2)^2 + \cdots + \left(\theta_q\beta_q\right)^2} ; k = 1,2, \dots, q \\ 0 \, ; k > q \end{cases} \tag{6.17}$$

6.4 Mixture of Pegram and Thinning of pth-Order Autoregressive Processes

This section is an immediate extension of Chapter 3. It is natural to extend the order one MPT model to pth-order. Here, we abbreviate the mixture of Pegram and thinning p th-order integer-valued Autoregressive process as MPT(p) to avoid confusion. Extending from Definition 3.1.1, we define the MPT p th-order integer-valued AR process as follow,

Definition 6.4.1 (Mixture of Pegram-INAR(p)): Let ε_t be an i.i.d process with range $\mathbb{N}_0$ which having mean $E(\varepsilon_t) = \mu_\varepsilon$ and variance $var(\varepsilon_t) = {\sigma_\varepsilon}^2$. Let $\alpha \in (0,1)$ and ϕ_j be the mixing weights where $\phi_j \in (0,1), j = 1, \dots, p$ and $\sum_{j=1}^{p}\phi_j = 1$. For every $t = 0, \pm 1, \pm 2, \dots$ a discrete-valued stochastic process $(X_t)_{\mathbb{Z}}$ is defined by

$$X_t = (\phi_1, \alpha \circ_t X_{t-1}) * \ldots * (\phi_p, \alpha \circ_t X_{t-p}) * (1 - \phi_1 - \phi_2 - \cdots - \phi_p, \varepsilon_t) \qquad (6.18)$$

The time index t below the thinning operation indicates that the corresponding thinning is used to define the discrete-valued stochastic process X_t, which is uncorrelated with ε_t, and the notation $*$ indicates Pegram's mixing operator. Equation (6.18) is called MPT(p) process if it follows the assumptions (i) ε_t is independent of thinning at time t $(\circ_t)$, $\alpha \circ_t X_t$ is independent of $\alpha \circ_t (X_s)_{s<t}$, (ii) the mixing weights $\phi_j, j = 1, \ldots, p$ is independent of ε_t and the thinning at time t, and (iii) the conditional distribution on X_t $P(\alpha \circ_{t+1} X_t, \ldots, \alpha \circ_{t+p} X_t | X_t = x_t, \mathcal{H}_{t-1})$ is equal to $P(\alpha \circ_{t+1} X_t, \ldots, \alpha \circ_{t+p} X_t | X_t = xt$, where for $s \le t-1$ and $j=1,\ldots,p$, $\mathcal{H}t-1$ abbreviates the process history of all Xs and $\alpha \circ_{s+j} X_s$.

The MPT(p) process is defined with its pgf which has to fulfil

$$\begin{aligned} G_{X_t}(z) &= \sum_{j=1}^{p} \phi_j E(z^{\alpha \circ_t X_{t-j}}) + \left(1 - \sum_{j=1}^{p} \phi_j\right) E(z^{\varepsilon_t}) \\ &= \sum_{j=1}^{p} \phi_j G_{X_{t-j}}(1 - \alpha + \alpha z) + \left(1 - \sum_{j=1}^{p} \phi_j\right) G_{\varepsilon_t}(z) \end{aligned}$$

Theorem 6.4.1 (Conditional pgf of MPT(p) Process): The conditional pgf takes the form

$$G_{X_t|X_{t-1},\ldots,X_{t-p}}(z) = \sum_{j=1}^{p} \phi_j \,(1 - \alpha + \alpha z)^{X_{t-j}} + \left(1 - \sum_{j=1}^{p} \phi_j\right) G_{\varepsilon_t}(z) \qquad (6.19)$$

Corollary 6.4.1 (Conditional pmf of MPT(p) Process): The probability mass function is given by

$$P(X_t = x_t | X_{t-1}, \dots) = P(X_t = x_t | X_{t-1}, \dots, X_{t-p})$$
$$= \sum_{j=1}^{p} \phi_j \binom{x_{t-j}}{x_t} \alpha^{x_t} (1-\alpha)^{x_{t-j}-x_t} + \left(1 - \sum_{j=1}^{p} \phi_j\right) P(\varepsilon_t = x_t) \tag{6.20}$$

Corollary 6.4.2 (Conditional Expectation): The conditional expectation is given by

$$E(X_t | X_{t-1}, \dots, X_{t-p}) = \alpha \sum_{j=1}^{p} \phi_j X_{t-j} + \left(1 - \sum_{j=1}^{p} \phi_j\right) \mu_\varepsilon \tag{6.21}$$

Corollary 6.4.3 (Conditional Variance): The conditional variance in quadratic form is given by

$$Var(X_t | X_{t-1}, \dots, X_{t-p})$$
$$= \alpha \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{p} \alpha^2 \phi_j X_{t-j} (X_{t-j} - 1) + \left(1 - \sum_{j=1}^{p} \phi_j\right) (\sigma_\varepsilon^2 + \mu_\varepsilon^2)$$
$$-\alpha^2 \sum_{j=1}^{p} {\phi_j}^2 {X_{t-j}}^2 - \left(1 - \sum_{j=1}^{p} \phi_j\right)^2 \mu_\varepsilon^2 - 2\alpha \sum_{j=1}^{p} \phi_j X_{t-j} \left(1 - \sum_{j=1}^{p} \phi_j\right) \mu_\varepsilon \tag{6.22}$$

Next, we would like to study the autocovariance and autocorrelation functions. The assumption that all thinnings are identical and independent holds true, and they are derived as follows.

Theorem 6.4.2 (Autocovariance Function of MPT(p) Models): Let $(X_t)_{\mathbb{Z}}$ be a stationary MPT(p) process according to Definition 6.4.1. Let $\gamma(k) = Cov(X_t, X_{t-k})$ denote the autocovariance function and

$$\mu(j,k) = E[(\alpha \circ_t X_{t-j}) \cdot X_{t-k}] - \alpha \cdot E[X_{t-j} \cdot X_{t-k}], \quad k \geq 1$$

The autocovariance function is

$$\gamma(k) = \alpha \cdot \sum_{j=1}^{p} \phi_j \cdot \gamma(|k-j|) + \sum_{j=k+1}^{p} \phi_j \cdot \mu(j,k) \tag{6.23}$$

where for $j > k$,

$$\mu(j,k) = \alpha \sum_{m=1}^{j-k-1} \phi_j \, \mu(j, k+m) + \phi_{j-k}\left[cov\left(\alpha \circ_t X_{t-j}, \alpha \circ_{t-k} X_{t-j}\right) - \alpha^2 \sigma_X^2\right]$$

otherwise $j \leq k, \mu(j,k) = 0$.

Proof:

$$cov(X_t, X_{t-k}) = E(X_t X_{t-k}) - E(X_t)E(X_{t-k})$$

$$\begin{aligned}
E(X_t X_{t-k}) \quad &= E\left[\left\{\sum_{j=1}^{p} \phi_j \left(\alpha \circ X_{t-j}\right) + \left(1 - \sum_{j=1}^{p} \phi_j\right) \varepsilon_t \right\} X_{t-k}\right] \\
&= E\left[\sum_{j=1}^{p} \phi_j \left(\alpha \circ X_{t-j}\right) \cdot X_{t-k} + \left(1 - \sum_{j=1}^{p} \phi_j\right) \varepsilon_t \cdot X_{t-k}\right] \\
&= \sum_{j=1}^{p} \phi_j \, E\left[\left(\alpha \circ X_{t-j}\right) \cdot X_{t-k}\right] + \left(1 - \sum_{j=1}^{p} \phi_j\right) E(\varepsilon_t) E(X_{t-k}) \\
&= \sum_{j=1}^{p} \phi_j \, E\left[\left(\alpha \circ X_{t-j}\right) \cdot X_{t-k}\right] + \left(1 - \alpha \sum_{j=1}^{p} \phi_j\right) \mu_x^2
\end{aligned}$$

Similarly as defined in Weiβ (2008), we obtain

$$\mu(j,k) = E\left[\left(\alpha \circ_t X_{t-j}\right) \cdot X_{t-k}\right] - \alpha \cdot E\left[X_{t-j} \cdot X_{t-k}\right], \quad k \geq 1$$

Then, the autocovariance function can be determined from the equations

$$\gamma(k) = \alpha \cdot \sum_{j=1}^{p} \phi_j \cdot \gamma(|k-j|) + \sum_{j=k+1}^{p} \phi_j \cdot \mu(j,k)$$

where $\mu(j,k) = 0$ for $j \leq k$,

Remark: Identical and independent thinnings have been considered in Weiβ (2008), who proposed combined INAR(p) models. The autocovariance and autocorrelation functions of MPT(p) process may be given similar consideration as in Weiβ (2008).

Theorem 6.4.3 (Autocorrelation Function of MPT(p) Model): Assuming all the thinning operations are independent, then $cov\left[\alpha \circ_t X_{t-j}, \alpha \circ_{t-k} X_{t-j}\right] = \alpha^2 \sigma_X^2$. The autocorrelation function has the structure

$$\rho(k) = \alpha \cdot \left(\phi_1 \cdot \rho(|k-1|) + \cdots + \phi_p \cdot \rho(|k-p|)\right) \tag{6.24}$$

As mentioned above, similar results have also been provided by Weiβ (2008). Next, we consider the random variable X_t to be a Poisson process, and some important equations will be derived.

Example 6.4.1 (Probability Generating Function Poisson MPT(p) Process): Let X_t be Poisson process with mean λ, which fulfills Definition 6.4.1, the pgf is given by

$$G_{X_t}(z) = \sum_{j=1}^{p} \phi_j \, G_{X_{t-j}}(1 - a + \alpha z) + e^{\lambda(z-1)} - \sum_{j=1}^{p} \phi_j \, e^{\lambda\alpha(z-1)}$$

where the pgf of innovation process $\{\varepsilon_t\}$ is

$$G_{\varepsilon_t}(z) = \frac{1}{1 - \sum_{j=1}^{p} \phi_j} \left(e^{\lambda(z-1)} - \sum_{j=1}^{p} \phi_j \, e^{\lambda\alpha(z-1)} \right)$$

with the corresponding pmf of $\{\varepsilon_t\}$ is

$$P(\varepsilon_t = i) = \frac{1}{1-\sum_{j=1}^{p}\phi_j}\left(\frac{e^{-\lambda}\lambda^i}{i!} - \sum_{j=1}^{p}\phi_j \frac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right)$$

To ensure the pmf of ε_t is strictly positive, the parameters must fulfils $\frac{1}{\sum_{j=1}^{p}\phi_j} > e^{\lambda(1-\alpha)}$ for $i = 0,1,2 \ldots$.

Remark: Refer to Section 7.5 for real illustration of Poisson MPT(p) processes.

6.5 Mixture of Pegram and Thinning of (p,q)th-Order Integer-Valued Autoregressive Moving Average Processes

To end this chapter, we introduce the combination of mixture of integer-valued AR(p) and MA(q), which yields the thinning and Pegram's mixture of integer-valued autoregressive moving average. It is abbreviated by MPT-ARMA(p,q).

Definition 6.5.1 (Mixture of Pegram-INARMA(p,q) Process): Let ε_t be an i.i.d. discrete random variable, the mixing weights $\alpha \in (0,1), i = 1,2,\ldots,p$ and $\beta_j \in (0,1), j = 0,1,2,\ldots,q$

$$\begin{aligned} X_t = {} & (\phi_1, \alpha \circ X_{t-1}) * (\phi_2, \alpha \circ X_{t-2}) * \ldots * (\theta_0, \beta_0 \circ \varepsilon_t) * (\theta_1, \beta_1 \circ \varepsilon_{t-1}) * \ldots \\ & * (\theta_q, \beta_q \circ \varepsilon_{t-q}) \end{aligned} \tag{6.25}$$

with the pgf

$$G_X(z) = \sum_{i=1}^{p}\phi_i G_{X-i}(1-\alpha+\alpha z) + \sum_{j=0}^{q}\theta_j G_\varepsilon(1-\beta_j+\beta_j z) \tag{6.26}$$

where all thinning and mixture operators follow Definition (6.3.1) and (6.4.1). The above definition leads to autocovariance function in the form of

$$\gamma(k) = \alpha\sum_{i=1}^{p}\phi_i\, E(X_{t-i}X_{t-k}) + \sum_{j=0}^{q}\theta_j\beta_j E\left(\varepsilon_{t-j}X_{t-k}\right) \tag{6.27}$$

To reduce the complexity, we consider only when $p = 1$ and $q = 1$.

Example 6.5.1 (MPT-INARMA(1,1)): Consider the random variable X_t as a mixture of Pegram and thinning of first order Integer-valued Autoregressive and Moving Average process which is given by

$$X_t = (\phi_1, \alpha \circ X_{t-1}) * (\theta_0, \beta_0 \circ \varepsilon_t) * (\theta_1, \beta_1 \circ \varepsilon_{t-1}) \tag{6.28}$$

and the autocovariance function is

$$\gamma(k) = \phi_1 \alpha \gamma(k-1) + \theta_0 \beta_0 \gamma_{\varepsilon x}(k) + \theta_0 \beta_0 \gamma_{\varepsilon x}(k-1) \tag{6.29}$$

where $\gamma_{\varepsilon x}(k) = E(\varepsilon_t X_{t-k})$ is the cross covariance function between random variables ε_t and X_{t-k}. Since X_{t-k} occurs only up to $t-k$, it follows that $\gamma_{\varepsilon x}(k) = 0$ for $k > 0$ otherwise $\gamma_{\varepsilon x}(k) \neq 0$ for $k \leq 0$. To obtain the expression for the autocovariance function of the process, some algebra are needed. First, multiply Eq. (6.28) with ε_{t-1} and taking the expectation, we obtain

$$\gamma_{\varepsilon x}(-1) = (\phi_1 \alpha + \theta_1 \beta_1) \sigma_\varepsilon^2$$

Also, multiplying Eq. (6.28) with X_{t-k} and taking expectation before proceeding to set $k = 0$ and $k = 1$, we arrive at the following two equations

$$\gamma(0) = \phi_1 \alpha \gamma(1) + [\theta_0 \beta_0 + \theta_1 \beta_1 (\phi_1 \alpha + \theta_1 \beta_1)] \sigma_\varepsilon^2$$
$$\gamma(1) = \phi_1 \alpha \gamma(1) + \theta_1 \beta_1 \sigma_\varepsilon^2$$

Solving both equations simultaneously, we obtain the autocovariance function for MPT-ARMA (1,1) process.

$$\gamma(0) = \frac{\theta_0\beta_0 + (\theta_1\beta_1)^2 + 2\phi_1\theta_1\beta_1\alpha}{1 - (\phi_1\alpha_1)^2}\sigma_\varepsilon^2 \qquad ; k = 0$$

$$\gamma(1) = \frac{\theta_0\beta_0\phi_1\alpha + (\theta_1\beta_1)^2\phi_1\alpha + (\phi_1\alpha)^2\theta_1\beta_1 + \theta_1\beta_1}{1 - (\phi_1\alpha_1)^2}\sigma_\varepsilon^2 \qquad ; k = 1$$

$$\gamma(k) = \phi_1\alpha\gamma(k-1) \qquad ; k \geq 2$$

The expressions are rather complicated but it may be useful in real application, which can be discussed in the future research.

6.6 Concluding Remarks

Chapter 6 is an extension of Chapter 3. With newly defined discrete-valued operator, it presents other potential candidates for MPT family. In this chapter, we discussed the MPT-MA(q) model and studied the order one process as a special case, which is abbreviated by MPT-MA(1). Discrete marginal distributions such as Poisson and Binomial have been fitted in the models. Also, some important properties like reversibility have been highlighted. One natural extension of the MPT(1) model is a higher order process. This is the thinning and Pegram's mixture of pth-order integer-valued autoregressive (MPT(p)) process. The investigation has been done to MPT(p) model, specifically fitting of Poisson margin for real data and this is illustrated in the next chapter. Last but not least, the combination of AR(p) and MA(q) processes which contains mixing parameters, yields the new MPT-ARMA(p,q) model.

CHAPTER 7: APPLICATION TO REAL DATA SETS

7.0 Introduction

This chapter illustrates the application of the proposed model. One of the aim of this thesis is to show that the proposed model has competitive advantages over existing models. We fitted a variety of real data sets into MPT(1) process and various graphical plots and simple descriptive statistics are given to summarize the analysis. Model parameter estimation is discussed and extensive comparison with existing discrete-valued time series models has been carried out.

7.1 Criminal: Sex Offences

7.1.1 Data Description

An observation of the time series represents a count of sex offences reported in the 21st police car beat in Pittsburgh, during one month. The data consists of 144 observations, starting in January 1990 and ending in December 2001. The data are available on-line at http://www.forecastingprinciples.com under the section of crime data. The partial autocorrelation cuts off after the first lag. This behaviour indicates that MPT(1) is appropriate. See Fig. 7.1.1. A summary of simple descriptive statistics of the data is given in Table 7.1.1. From preliminary observation, the data are overdispersed since the index of dispersion exceeds 1. We further diagnose whether this overdispersion is statistically significant by adopting the dispersion test of Schweer and Weiβ (2014). The null hypothesis is that the data are equi-dispersed from the Poisson INAR(1) process against the alternative of an INAR(1) with an over dispersed marginal distribution. From Schweer and Weiβ (2014, Eq. (9)) the critical value is 1.1977 and the index of dispersion of the data, $\hat{I}_{data} = \frac{sample\ \ variance}{sample\ \ mean}$ is 1.7395. The null hypothesis is rejected as the index of dispersion exceeds the critical value and an overdispersed marginal may be assumed for the data.

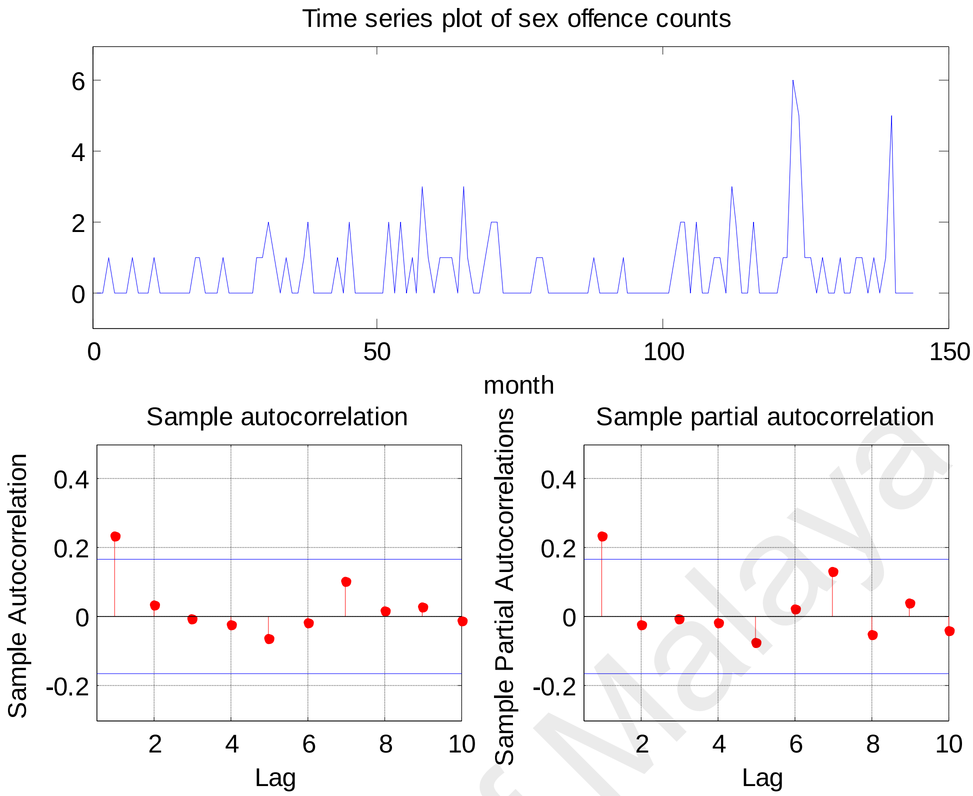


Figure 7.1.1: Time series plot, sample autocorrelation and partial autocorrelation function of sex offence counts

Table 7.1.1: Descriptive statistics of criminal data

Data set	Minimum	Maximum	Median	Mode	Mean	Variance
Sex offence	0	6	0	0	0.5903	1.0268

7.1.2 Comparison with Existing Models

We fitted the proposed MPT(1) model to the criminal offense data and compared with the fit of existing models in the literature. Since the data is overdispersed, marginals such as negative binomial and geometric have been considered. Table 7.1.2 tabulates the models as follow:

(I) MPT(1) with

(a) Poisson, (b) Negative Binomial, (c) Geometric, (d) New Geometric marginals;

(II) Pegram's AR(1) with

(e) Poisson (Biswas and Song 2009), (f) Negative Binomial, (g) Geometric marginals;

(III) INAR(1) with

(h) Geometric (Alzaid and Al-Osh 1988), (i) Poisson (Al-Osh and Alzaid 1987), (j) Negative Binomial (Zhu and Joe 2006), (k) New Geometric (Ristic et al. 2009) marginals;

(IV) Iterated INAR(1) with

(l) Negative Binomial marginal (Al-Osh and Aly 1992);

(V) Random coefficient INAR(1) with

(m) Negative Binomial marginal (Weiβ 2008);

(VI) Quasi-Binomial INAR(1) with

(n) Generalized Poisson marginal (Alzaid and Al-Osh 1993).

Table 7.1.2: Transition probabilities of integer-valued time series models

Model	Conditional probabilities
(I) MPT(1) (a) Poisson	$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + (1-\phi) P(\varepsilon_t = i)$ where $P(\varepsilon_t = i) = \dfrac{1}{1-\phi}\left\{\dfrac{e^{-\lambda}\lambda^i}{i!} - \phi \dfrac{e^{-\lambda\alpha}(\lambda\alpha)^i}{i!}\right\}$ $i = 0,1, \ldots$
(b) Negative Binomial	$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + (1-\phi) P(\varepsilon_t = i)$ where $P(\varepsilon_t = i) = \dfrac{1}{1-\phi}\left[\binom{k+x-1}{k-1} p^k (1-p)^i - \phi \binom{k+x-1}{k-1} \alpha p^k (1-\alpha p)^i\right]$
(c) Geometric	$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + (1-\phi) P(\varepsilon_t = i)$ where $P(\varepsilon_t = i) = \dfrac{1}{1-\phi}\left[p(1-p)^i - \phi\alpha p(1-\alpha p)^i\right]$
(d) New Geometric	$P(X_t = i \mid X_{t-1} = j) = \phi \binom{j}{i} \alpha^i (1-\alpha)^{j-i} + (1-\phi) P(\varepsilon_t = i)$ where $P(\varepsilon_t = 0) = \dfrac{1}{1-\phi}\left\{\dfrac{1}{1+p} - \phi \dfrac{1+\alpha}{1+\alpha(1+p)}\right\}$ $P(\varepsilon_t = i) = \dfrac{1}{1-\phi}\left\{\dfrac{p^i}{(1+p)^{i+1}} - \alpha^i \phi p \left[\dfrac{(1+p)^{i-1}}{\{1+\alpha(1+p)\}^{i+1}}\right]\right\}$ $i = 1,2,3, \ldots$

(II) Pegram's AR(1) (e) Poisson (Biswas and Song, 2009)	$P(X_t = i \mid X_{t-1} = j) = \phi I[X_{t-1} = j] + (1-\phi)P(\varepsilon_t = i)$ where $P(\varepsilon_t = i) = \frac{e^{-\lambda}\lambda^i}{i!}$
(f) Negative Binomial	$P(X_t = i \mid X_{t-1} = j) = \phi I[X_{t-1} = j] + (1-\phi)P(\varepsilon_t = i)$ where $P(\varepsilon_t = i) = \binom{k+x-1}{k-1} p^k (1-p)^i$
(g) Geometric	$P(X_t = i \mid X_{t-1} = j) = \phi I[X_{t-1} = j] + (1-\phi)P(\varepsilon_t = i)$ where $P(\varepsilon_t = i) = p(1-p)^i$
(III) INAR(1) with (h) Poisson (Al-Osh and Alzaid, 1987)	$P(X_t = i \mid X_{t-1} = j) = \sum_{k=0}^{min\,(i,j)} \binom{j}{i} \alpha^i (1-\alpha)^{j-i} \cdot P(\varepsilon_t = i-k)$ where $P(\varepsilon_t = i) = \frac{e^{-\lambda(1-\alpha)}\lambda(1-\alpha)^i}{i!}$
(i) Geometric (Alzaid and Al-Osh, 1988)	$P(X_t = i \mid X_{t-1} = j) = \sum_{k=0}^{min\,(i,j)} \binom{j}{i} \alpha^i (1-\alpha)^{j-i} \cdot P(\varepsilon_t = i-k)$ The innovation distribution of $(\varepsilon_t)_{\mathbb{N}}$ is $P(\varepsilon_t = k) = \begin{cases} \alpha + (1-\alpha)p, & k = 0 \\ (1-\alpha)pq^k, & k > 0 \end{cases}$

(j) Negative Binomial (Zhu and Joe, 2006)	$P(X_t = i \vert X_{t-1} = j) = \sum_{m=0}^{min\,(i,j)} \binom{j}{i} \alpha^i (1-\alpha)^{j-i} \cdot P(\varepsilon_t = i - m)$ The innovation distribution of $(\varepsilon_t)_{\mathbb{N}}$ is $P(\varepsilon(\alpha) = 0) = \dfrac{P(X = 0)}{P(\alpha * X = 0)}$, $P(\varepsilon(\alpha) = j) = \dfrac{P(X = j) - \sum_{l=1}^{j} P(\alpha * X = l) \cdot P(\varepsilon(\alpha) = j - l)}{P(\alpha * X = 0)}, \;\; j = 1,2,\ldots$ where $P(\alpha * X = l) = \sum_{k=l}^{\infty} \binom{k}{l} \alpha^l (1-\alpha)^{k-l} P(X = k), \;\; l = 0,1,2,\ldots$ and $P(X = k) = \binom{r+k-1}{k} p^k (1-p)^r, \;\; k = 0,1,2,\ldots$
(k) New Geometric (Ristic et al., 2009)	$P(X_t = i \vert X_{t-1} = 0) = \left(1 - \dfrac{\alpha\mu}{\mu-\alpha}\right) \dfrac{\mu^j}{(1+\mu)^{j+1}} + \dfrac{\alpha\mu}{\mu-\alpha} \dfrac{\alpha^j}{(1+\alpha)^{j+1}}, \;\; j = 0$ $P(X_t = i \vert X_{t-1} = j) = \dfrac{\mu\alpha^{i+1}}{(\mu-\alpha)(1+\alpha)^{j+i+1}} \binom{i+j}{i}$ $+ \left(1 - \dfrac{\alpha\mu}{\mu-\alpha}\right) \dfrac{\mu^i}{(1+\alpha)^j (1+\mu)^{i+1}} \sum_{k=0}^{i} \binom{j+k-1}{j-1} \left(\dfrac{\alpha(1+\mu)}{\mu(1+\alpha)}\right)^k, \;\; j \geq 1$

(IV) Iterated INAR(1) with Negative Binomial (Al-Osh and Aly, 1992)	$P(X_t = i \mid X_{t-1} = j) = \frac{\alpha^n (1-\alpha\rho)^j}{(1+\alpha)^{n+i}} \cdot \left[\binom{n+i-1}{i} + \sum_{k=1}^{j} \binom{j}{k} \cdot \left(\frac{\alpha\rho}{1-\alpha\rho} \right)^k \cdot \left(\frac{\alpha}{1+\alpha} \right)^k \sum_{m=0}^{i} \binom{n+i-m-1}{i-m} \binom{m+k-1}{m} \right]$
(V) Random coefficient INAR(1) with Negative Binomial marginal (Weiβ, 2008)	$P(X_t = i \mid X_{t-1} = j) = \sum_{k=0}^{min\,(i,j)} \binom{j}{k} \frac{B(k+n\rho, j-k+n(1-\rho))}{B(n\rho, n(1-\rho))} \cdot P(\varepsilon_t = i-k)$ where $(\varepsilon_t)_{\mathbb{N}} \sim NB(n(1-\rho), p)$
(VI) Quasi-Binomial INAR(1) with Generalized Poisson marginal (Alzaid and Al-Osh, 1993)	$P(X_t = i \mid X_{t-1} = j) = \sum_{k=0}^{min\,(i,j)} \binom{j}{k} \frac{\rho(1-\rho)\lambda}{\lambda+j\theta} \left(\frac{\rho\lambda+k\theta}{\lambda+j\theta} \right)^{k-1} \times \left(\frac{(1-\rho)\lambda+(j-k)\theta}{\lambda+j\theta} \right)^{j-k-1} \times P(\varepsilon_t = i-k)$ where $(\varepsilon_t)_{\mathbb{N}} \sim GP(\lambda(1-\rho), \theta)$

For each model from (a)-(g), we obtain the ML estimates and the corresponding AIC values while for models (h)-(n), the values found in Ristic et al. (2009) are quoted. More details about the models can be obtained from the references cited above. Section 3.4 gives the expressions for several discrete-valued marginals, including Negative Binomial, Geometric ($k = 1$), and new Geometric (Ristic et al., 2009), fitted in MPT(1) models.

As seen from Table 7.1.3, MPT(1) model is competitive with the existing models. The AIC value of Poisson MPT(1) provides 315.41, which is lower than Poisson Pegram's AR(1) and Poisson INAR(1). As the data is overdispersed, the Poisson marginal is not appropriate and negative binomial, geometric and new geometric marginal are considered. We applied these marginals in MPT(1) process. Here, the results show that the MPT(1) with new Geometric marginal gives the substantially lowest AIC value of 243.19. It can be concluded that the new geometric MPT(1) performs the best among all discrete time series models.

Table 7.1.3: Estimated parameters of the models and AIC

Model	ML estimates	AIC
(I) MPT(1) with (a) Poisson marginal	$\hat{\lambda} = 0.5749$ $\hat{\alpha} = 0.7936$ $\hat{\phi} = 0.1154$	315.41
(b) Negative Binomial marginal	$\hat{k} = 0.9741$ $\hat{p} = 0.6451$ $\hat{\alpha} = 0.9394$ $\hat{\phi} = 0.0807$	307.05
(c) Geometric marginal	$\hat{p} = 0.6288$ $\hat{\alpha} = 0.9388$ $\hat{\phi} = 0.0848$	304.70
(d) New Geometric marginal	$\hat{p} = 0.6288$ $\hat{\alpha} = 0.9388$ $\hat{\phi} = 0.3304$	243.19
(II) Pegram's AR(1) with (e) Poisson marginal	$\hat{\lambda} = 0.5979$ $\hat{\phi} = 0.0370$	322.98
(f) Negative Binomial marginal	$\hat{p} = 0.6122$	313.29

	$\hat{k} = 0.9371$ $\hat{\phi} = 0.0317$	
(g) Geometric marginal	$\hat{p} = 0.6275$ $\hat{\phi} = 0.0311$	311.78
(III) INAR(1) with (h) Poisson marginal	$\hat{\lambda} = 0.5063$ $\hat{\alpha} = 0.1404$	316.89
(i) Geometric marginal	$\hat{p} = 0.6304$ $\hat{\alpha} = 0.1055$	303.74
(j) Negative Binomial marginal	$\hat{r} = 1.1167$ $\hat{p} = 0.6552$ $\hat{\alpha} = 0.1113$	305.67
(k) New Geometric	$\hat{\mu} = 0.5872$ $\hat{\alpha} = 0.1650$	302.67
(lV) Iterated INAR(1) Negative Binomial marginal	$\hat{\alpha} = 0.9838$ $\hat{n} = 2.0270$ $\hat{\rho} = 0.1753$	303.28
(V) Random coefficient INAR(1) Negative Binomial marginal	$\hat{n} = 1.0858$ $\hat{p} = 0.6529$ $\hat{\rho} = 0.1315$	303.71
(VI) Quasi-Binomial INAR(1) Generalized Poisson marginal	$\hat{\lambda} = 0.4050$ $\hat{\theta} = 0.1940$ $\hat{\rho} = 0.1304$	303.38

7.2 Internet Protocol (IP) Addresses Counts

7.2.1 Data Description

The IP access count data has been analyzed by Weiβ (2007, 2012). The server of the Statistics Department of the University of W$\ddot{u}$rzburg collects log data concerning accesses to pages on the server. Each line of this data corresponds to exactly one such access, containing information like host name of the user accessing a Web site, date and time of the request, the address of the page required, etc. The data set can be arranged in a way that for each minute in the period observed, it is indicated if there was an access to the home directory of one of six particular members of the Department of Statistics. If X_t denotes the sum of these indicator variables from number of different staff members, whose home directory was accessed in minutes t. X_t has a range of $\{0, \ldots, 5\}$. The number of different IP addresses registered within the period of 2-minute length at the server of the Department of Statistics of the University of W$\ddot{u}$rzburg is collected. In particular, the time series data (without the outliers) on November 29th, 2005, between 10 o'clock in the morning and 6 o'clock in the evening, a length of 241 is collected. The lag-1 MPT process is appropriate to analyze the IP count data, as the ACF and PACF both cut off at lag 1. It is also obvious from the graph that the data exhibits serial dependence. The empirical variance is smaller than mean which shows that binomial marginal is suitable.

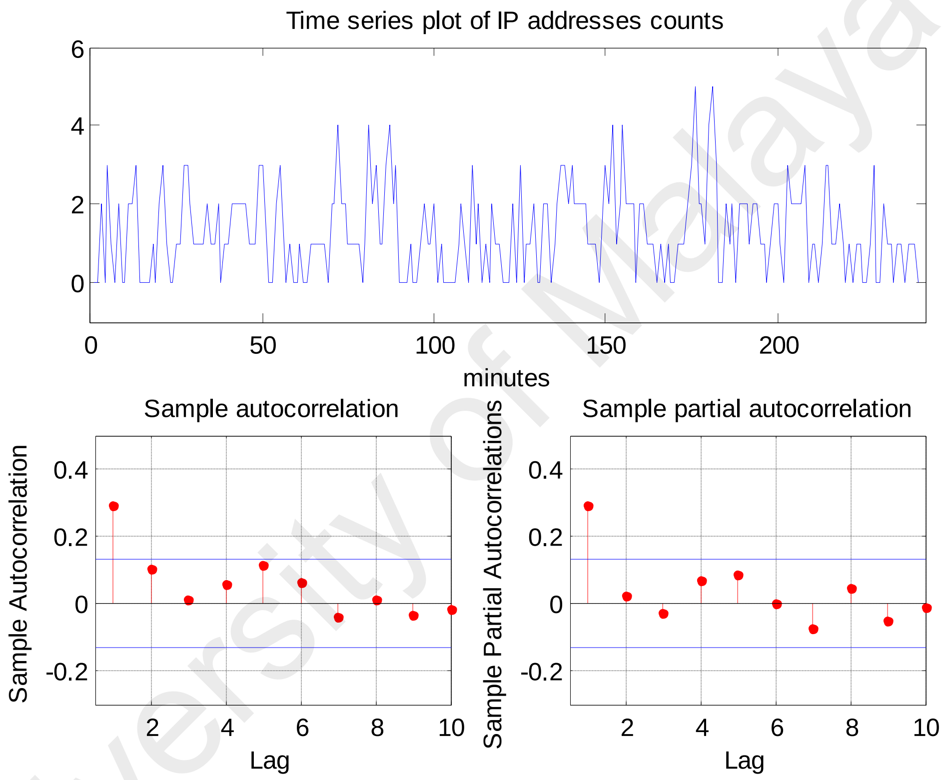


Figure 7.2.1: Time series plot, sample autocorrelation and partial autocorrelation function of IP addresses counts

Table 7.2.2: Descriptive statistics of IP counts

Data set	Minimum	Maximum	Median	Mode	Mean	Variance
Sex offence	0	5	1	1	1.2863	1.2051

7.2.2 Data Implication

We may interpret the binomial MPT(1) in light of conditional expectation. Suppose that we have a server, that can be used to count the number of access to the home directory: Given n members of the Department of Statistics, each of them either access (log-in) or non-access (log-out). The expected number of the member log-in to the home directory at time t now consists of X_{t-1} members with transition probability α who have logged in previously, and the average number μ machines which logged in recently. The parameter ϕ indicates the weight for the number of members logged in recently.

7.2.3 Comparison with Binomial Marginal Models

It is well-known that the limit of a sequence of Binomial distributions with N independent trials tend to infinity and probability of success p tends to zero while Np remains finite and equal to a parameter, say θ, lead to Poisson distribution. In this section, we carried out the marginal comparison; Poisson and Binomial marginals. The result is shown in Table 7.2.3. Then, we compare Binomial MPT(1) with Binomial Pegram's AR(1), and the existing binomial AR(1) model proposed by McKenzie (1985), which was well discussed by Weiβ. See Table 7.2.4 for the result of model comparison.

Table 7.2.3: Comparison of MPT(1) process, estimated parameters and AIC values

MPT(1) model	ML estimates	AIC
Poisson marginal	$\hat{\alpha} = 0.7536$ $\hat{\phi} = 0.1354$ $\hat{\lambda} = 1.2852$	692.12
Binomial marginal	$\hat{\alpha} = 0.6224$ $\hat{\phi} = 0.1524$ $\hat{p} = 0.2593$	666.00

Table 7.2.4: Models comparison, estimated parameters and AIC values

Model	ML estimates	AIC
Binomial MPT(1)	$\hat{\alpha} = 0.6224$ $\hat{\phi} = 0.1524$ $\hat{p} = 0.2593$	666.00
Binomial AR(1) (McKenzie, 1985)	$\hat{\alpha} = 0.4359$ $\hat{\beta} = 0.1965$	681.55
Binomial Pegram's AR(1)	$\hat{\alpha} = 0.2598$ $\hat{\phi} = 0.0845$	698.36

Since Weiβ (2007) suggested applying Poisson marginal to fit the IP counts data, we are keen to compare the fits between Poisson and binomial marginal in MPT(1) model. The result shows that Binomial MPT(1) performs better than the Poisson marginal. Also, the comparison has been made between the existing integer-valued time series models. Here, we provide the transition probability of binomial Pegram's AR(1) model for convenience, which is shown as follows:

$$P(X_t = i | X_{t-1} = j) = \phi I[i = j] + (1 - \phi) P(\varepsilon_t = i)$$

where $\phi \in (0,1), \alpha \in [0,1]$, the innovation distribution is simply

$$P(\varepsilon_t = i) = \binom{n}{i} \alpha^i (1 - \alpha)^{n-i}$$

The transition probability of Binomial AR(1) process can be obtained in McKenzie (1985). Also see Weiβ (2008) for model properties. Binomial MPT(1) process gives the lowest AIC value, 666.00 which outperformed the results computed by Binomial AR(1) and Pegram's AR(1) models. We conclude that the proposed MPT(1) model with a simple interpretation is a good viable alternative to INAR(1) and Pegram's AR(1) for empirical modelling.

7.3 Worker Compensation Burn Claims

7.3.1 Data Description

This section analyses the real time series count data obtained from the Workers Compensation Board (WCB) of British Columbia, Canada. The data consists of monthly counts of claims collecting Short Term Wage Loss Benefit (STWLB) for injuries reported in the workplace. In the selected data set all the claimants are male, between the age of 35 and 54 work in logging industry and reported their claim to the Richmond, BC service delivery location. The considered data consist of 120 observations starting in January 1984 and ending in December 1994. The data are claimants whose injuries are burn related. This data has been analyzed by Freeland and McCabe (2004) with the removing of outliers to be able to fit the first order time series models. Table 7.3.1 provides the basic statistical summary of the data. It is sensible to apply Poisson marginal in the analysis for mean equals to variance. There are two objectives of the analysis. The first is to produce forecasting of the numbers of claimants of 1995 for the Richmond delivery area. We are interested to know the *k*-step convergence of conditional distribution to marginal distribution. The second aim is to compare Poisson MPT(1) model with INAR(1) and Pegram's based AR(1) model, in terms of 95% confidence intervals width and conditional mean.

In this section we show one equi-dispersed claim count data. This set of data has been used in Chapter 5 as an application for coherent forecasting. At first glance the time series plot of burn claims data has significant change in the pattern after the middle of 1993. It is therefore unlikely that an order-one AR process will fit the series well. This is further confirmed by the sample ACF and PACF, whereby the ACF does not decay fast enough for an AR(1) model. Also, the PACF shows that an AR(2) might be a more appropriate suggestion. However, this thesis is concentrated on AR(1) process and may leave this for future research works. See Figure 7.3.1 for the sample plots, ACF and PACF.

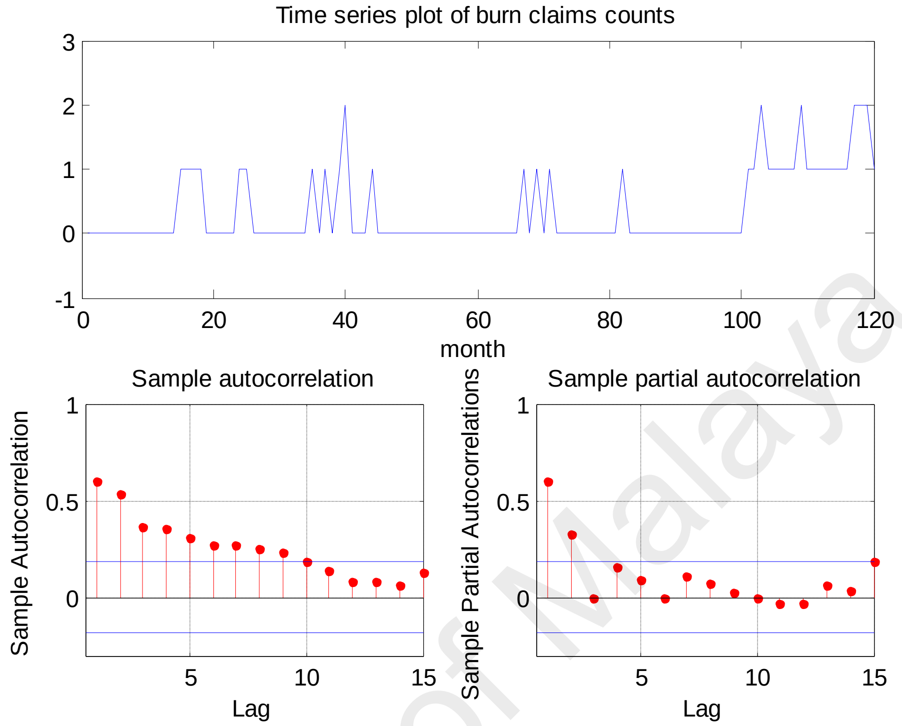


Figure 7.3.1: Time series plots, sample autocorrelation function and sample partial autocorrelation function of burn claim counts

Table 7.3.1: Descriptive statistics of burn claims data

Data set	Minimum	Maximum	Median	Mode	Mean	Variance
Sex offence	0	2	0	0	0.34	0.33

The burn claims data has been handled by Freeland (1998) in his thesis with outliers removed. The new series which is distinguished from the original is referred to as data 1*, in Freeland (1998, Appendix). The new series is shown in Figure 7.3.2. Now, we see that the sample autocorrelation presents a slight seasonal pattern, but it may not be necessarily so, since the correlations at lag-6 and lag-12 are within the 5% confidence intervals. Also, sample partial autocorrelation function indicates that a first-order AR process is appropriate.

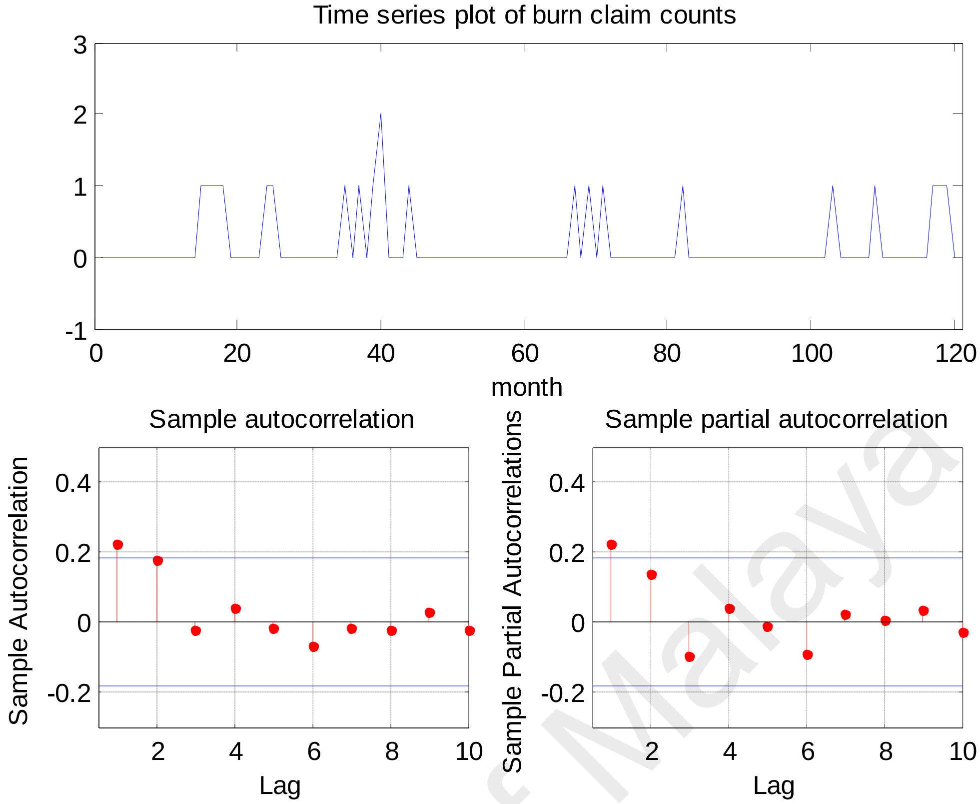


Figure 7.3.2: Time series plot, sample autocorrelation function and sample partial autocorrelation function of burn claim counts (without outlier)

Table 7.3.2: Model comparison, ML estimates and AIC values

Model	ML estimates	AIC
MPT(1)	$\hat{\alpha} = 0.9979$ $\hat{\lambda} = 0.1792$ $\hat{\phi} = 0.1789$	119.30
INAR(1)	$\hat{\alpha} = 0.2397$ $\hat{\lambda} = 0.1342$	115.47
Pegram's AR(1)	$\hat{\phi} = 0.1811$ $\hat{\lambda} = 0.1768$	116.90

Likewise, for each model we obtain the ML estimates and AIC value, which is shown in Table 7.3.2. The Poisson marginal is applied in the respective models. For burn claim counts, the Poisson INAR(1) gives the lowest AIC value. The forecasts from burn claims data have been compared among three models.

7.3.2 Application in Coherent Forecasting Distribution

To compute the 95% confidence intervals, we first estimate the parameters of Poisson MPT(1) process. The MLE elaborated in Section 4.1.1 has been applied. The parameters (α, ϕ, λ) for burn claims were estimated to be $(0.9979, 0.1789, 0.1792)$. Then, the inverse of Fisher information matrix is computed numerically which is given by

$$i^{-1} = \begin{bmatrix} 0.0006 & -0.0042 & -0.0022 \\ -0.0042 & 0.0272 & -0.0006 \\ -0.0022 & -0.0006 & 0.0000 \end{bmatrix}$$

The diagonal of information matrix gives the variance of the parameters and the off-diagonal elements are the covariances between the parameters of the model; i.e. $\left(\sigma_{\alpha}^2, \sigma_{\lambda}^2, \sigma_{\phi}^2\right) = (0.0006, 0.0272, 0.0000)$. Then, we compute the *k*-step-ahead distributions (Theorem 5.1.1) to perform forecast for burn claims data by the MPT(1), and also present the 95% confidence intervals. The results are shown in Table 7.3.3 and 7.3.4 respectively. In this real data analysis, we run $k = 10$ to see the overall performance of the model in burn claims data. However, we only show the analysis results for the first 6 months because we noted that after the 6-step-ahead conditional distribution it is very close to the marginal distribution, thus presenting a consistent pattern beyond the 6th-step predictive distribution. The 10-step-ahead forecast distribution for $x = 0,1,2,3,4$ are 0.835939, 0.149800, 0.013422, 0.00000802, and 0.000036, respectively. It is noted that there is not much difference in the results between the 6-step-ahead and the 10-step-ahead forecasting distributions. Also, the consistent standard errors are observed when *k*-step is increased. Therefore, it can be concluded that if we require forecasting for more than 6 months ahead, it is suggested to simply use the marginal distribution or marginal mean. The table provides more information. The mean (rounded down), median and mode show that there is no injuries to be expected in the following k-step forecast periods. It quantifies that the probability of having zero claim in the first month forecast is 0.8652 (point estimate); that is, the possibility is 87% of having zero claims and 13% of having a single claim. Similar forecast are obtained with INAR(1) and AR(1) processes.

7.3.3 Model Comparison with INAR(1) and Pegram's Based AR(1) Processes

This section discusses the model comparison by using the forecasting distribution of the INAR(1) and Pegram's based AR(1) process. We adopted the formulation of Fisher information matrix by Silva et al. (2005) when $k = 1$ for INAR(1) process. The asymptotic distributions of Pegram's AR(1) model can be obtained from Biswas and Song (2009). For each of the models we develop the numerical computation by using MATLAB R2013a. The 95% confidence interval in this thesis is slightly different from Freeland and McCabe (2004b) due to the difference in numerical software and program implementation. Furthermore, Freeland and McCabe (2004a) considered the Fisher IM in conditional expectations. The predictive results with 95% confidence interval of Pegram's AR(1) are tabulated in Table 7.3.5 and 7.3.6, while Table 7.3.7 and 7.3.8 provide the predictive information and 95% confidence intervals for INAR(1) process. It can be noted that all models are relatively similar in prediction. All conditional means and predictive distributions are approximately equal to the unconditional means and the marginal distributions after the 6th step. Also, the predictions are similar which is to have zero claims for the first 6 months. It should be noted here that the width of the 95% confidence intervals for MPT(1) is narrower compared to the other two models. One particular example for the 6-step-ahead $p_k(0|1)$; the standard deviation (in 3 decimal places) for MPT(1) is 0.013, but for Pegram's AR(1) is 0.075 and INAR(1) is 0.045.

Table 7.3.3: Forecast from burn claims data by Poisson MPT(1) process

k	1	2	3	4	5	6	∞
Mean	0.3257	0.2054	0.1839	0.1800	0.1793	0.1792	0.1792
Median	0	0	0	0	0	0	0
Mode	0	0	0	0	0	0	0
$p_k(0\|1)$	0.865233 (0.01045)	0.841169 (0.01223)	0.836873 (0.01253)	0.836105 (0.01258)	0.835958 (0.01258)	0.835944 (0.01258)	0.835939 (0.01258)
$p_k(1\|1)$	0.123047 (0.00858)	0.145022 (0.01004)	0.148947 (0.01028)	0.149648 (0.01032)	0.149773 (0.01032)	0.149795 (0.01032)	0.149800 (0.01033)
$p_k(2\|1)$	0.011030 (0.00170)	0.012996 (0.00199)	0.013346 (0.00204)	0.013409 (0.00205)	0.013420 (0.00205)	0.013422 (0.00205)	0.013422 (0.00205)
$p_k(3\|1)$	0.000659 (0.00016)	0.000776 (0.00018)	0.000797 (0.00019)	0.000801 (0.00019)	0.000802 (0.00019)	0.000802 (0.00019)	0.000802 (0.00019)
$p_k(4\|1)$	0.000030 (0.00001)	0.000035 (0.00001)	0.000036 (0.00001)	0.000036 (0.00001)	0.000036 (0.00001)	0.000036 (0.00001)	0.000036 (0.00001)
$p_k(5\|1)$	0.000001 (0.00000)	0.000001 (0.00001)	0.000001 (0.00001)	0.000001 (0.00001)	0.000001 (0.00001)	0.000001 (0.00001)	0.000001 (0.00001)

Table 7.3.4: 95% confidence interval for k-step ahead conditional distributions, Poisson MPT(1) process

k	1	2	3	4	5	6	∞
$p_k(0\|1)$	(0.8443, 0.8861)	(0.8167, 0.8656)	(0.8118, 0.8619)	(0.8110, 0.8613)	(0.8108, 0.8611)	(0.8108, 0.8611)	(0.8108, 0.8611)
$p_k(1\|1)$	(0.1059, 0.1402)	(0.1250, 0.1651)	(0.1284, 0.1695)	(0.1290, 0.1703)	(0.1291, 0.1704)	(0.1291, 0.1704)	(0.1291, 0.1704)
$p_k(2\|1)$	(0.0076, 0.0144)	(0.0090, 0.0170)	(0.0093, 0.0174)	(0.0093, 0.0175)	(0.0093, 0.0175)	(0.0093, 0.0175)	(0.0093, 0.0175)
$p_k(3\|1)$	(0.0003, 0.0010)	(0.0004, 0.0010)	(0.0004, 0.0012)	(0.0004, 0.0012)	(0.0004, 0.0012)	(0.0004, 0.0012)	(0.0004, 0.0012)
$p_k(4\|1)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)
$p_k(5\|1)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)

Table 7.3.5: Forecast distribution by Poisson Pegram's AR(1), and SE (in bracket)

k	1	2	3	4	5	6	∞
Mean	0.3254	0.2036	0.1816	0.1776	0.1769	0.1767	0.1767
Median	0	0	0	0	0	0	0
Mode	0	0	0	0	0	0	0
$p_k(0\|1)$	0.86728 (0.01454)	0.84331 (0.04281)	0.83899 (0.06646)	0.83820 (0.07313)	0.83806 (0.07477)	0.83804 (0.07514)	0.83803 (0.07524)
$p_k(1\|1)$	0.12134 (0.01330)	0.14325 (0.01330)	0.14721 (0.01330)	0.14792 (0.01330)	0.14805 (0.01330)	0.14808 (0.01330)	0.14808 (0.01330)
$p_k(2\|1)$	0.01072 (0.00118)	0.01266 (0.00118)	0.01301 (0.00118)	0.01301 (0.00118)	0.01301 (0.00118)	0.01301 (0.00118)	0.01301 (0.00118)
$p_k(3\|1)$	0.00063 (0.00007)	0.00075 (0.00007)	0.00077 (0.00007)	0.00077 (0.00007)	0.00077 (0.00007)	0.00077 (0.00007)	0.00077 (0.00007)
$p_k(4\|1)$	0.00003 (0.00000)	0.00003 (0.00000)	0.00003 (0.00000)	0.00003 (0.00000)	0.00003 (0.00000)	0.00003 (0.00000)	0.00003 (0.00000)

Table 7.3.6: 95% confidence intervals for k-step ahead conditional distributions, Poisson Pegram's AR(1) process

k	1	2	3	4	5	6	∞
$p_k(0\|1)$	(0.8382, 0.8964)	(0.7577, 0.9289)	(0.7061, 0.9719)	(0.6920, 0.9846)	(0.6885, 0.9876)	(0.6878, 0.9883)	(0.6875, 0.9885)
$p_k(1\|1)$	(0.0947, 0.1479)	(0.1167, 0.1698)	(0.1206, 0.1738)	(0.1213, 0.1745)	(0.1215, 0.1746)	(0.1215, 0.1746)	(0.1215, 0.1747)
$p_k(2\|1)$	(0.0084, 0.0131)	(0.0103, 0.0150)	(0.0107, 0.0154)	(0.0107, 0.0154)	(0.0107, 0.0154)	(0.0107, 0.0154)	(0.0107, 0.0154)
$p_k(3\|1)$	(0.0005, 0.0008)	(0.0006, 0.0008)	(0.0006, 0.0009)	(0.0006, 0.0009)	(0.0006, 0.0009)	(0.0006, 0.0009)	(0.0006, 0.0009)
$p_k(4\|1)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)

Table 7.3.7: Forecast distribution by Poisson INAR(1) model, and SE (in bracket)

k	1	2	3	4	5	6	∞
Mean	0.374	0.224	0.188	0.179	0.177	0.176	0.176
Median	0	0	0	0	0	0	0
Mode	0	0	0	0	0	0	0
$p_k(0\|1)$	0.87459 (0.03546)	0.84691 (0.04255)	0.84040 (0.04417)	0.83884 (0.04455)	0.83847 (0.04465)	0.83838 (0.04467)	0.83835 (0.04468)
$p_k(1\|1)$	0.11720 (0.03071)	0.14072 (0.03548)	0.14613 (0.03649)	0.14741 (0.03672)	0.14772 (0.03678)	0.14779 (0.03679)	0.14782 (0.03679)
$p_k(2\|1)$	0.00785 (0.00443)	0.01169 (0.00648)	0.01270 (0.00701)	0.01295 (0.00714)	0.01301 (0.00717)	0.01303 (0.00718)	0.01303 (0.00718)
$p_k(3\|1)$	0.00035 (0.00030)	0.00065 (0.00056)	0.00074 (0.00063)	0.00076 (0.00065)	0.00076 (0.00065)	0.00077 (0.00065)	0.00077 (0.00065)
$p_k(4\|1)$	0.00001 (0.00001)	0.00003 (0.00003)	0.00003 (0.00003)	0.00003 (0.00004)	0.00003 (0.00004)	0.00003 (0.00004)	0.00003 (0.00004)

Table 7.3.8: 95% confidence intervals for k-step ahead conditional distributions, Poisson INAR(1) process

k	1	2	3	4	5	6	∞
$p_k(0\|1)$	(0.8051, 0.9441)	(0.7635, 0.9303)	(0.7538, 0.9269)	(0.7515, 0.9262)	(0.7510, 0.9260)	(0.7508, 0.9259)	(0.7508, 0.9259)
$p_k(1\|1)$	(0.0570, 0.1774)	(0.0712, 0.2103)	(0.0746, 0.2177)	(0.0754, 0.2194)	(0.0756, 0.2198)	(0.0756, 0.2199)	(0.0756, 0.2199)
$p_k(2\|1)$	(0.0000, 0.0165)	(0.0000, 0.0244)	(0.0000, 0.0264)	(0.0000, 0.0269)	(0.0000, 0.0271)	(0.0000, 0.0271)	(0.0000, 0.0271)
$p_k(3\|1)$	(0.0000, 0.0009)	(0.0000, 0.0020)	(0.0000, 0.0020)	(0.0000, 0.0020)	(0.0000, 0.0020)	(0.0000, 0.0020)	(0.0000, 0.0020)
$p_k(4\|1)$	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)	(0.0000, 0.0000)

7.4 Criminal: Drug Offences

7.4.1 Data Description

Finally, we illustrate the proposed Poisson MPT(p) model as defined in Chapter 6 with a real data set. The data was taken from crime data section of the Forecasting Principles; http://forecastingprinciples.com. It is reported in the 33rd police car beat, from January 1990 to December 2001, consisting of 144 observations. The mean and variance of the data are 4.6250 and 12.3059, respectively. The autocorrelation value is 0.379. In Figure 7.4.1 we notice from PACF that the data is suitable to be fitted by the 5th order model. We fitted the data to 6th order Poisson MPT(6) process. However, the lower order models have been considered for comparison.

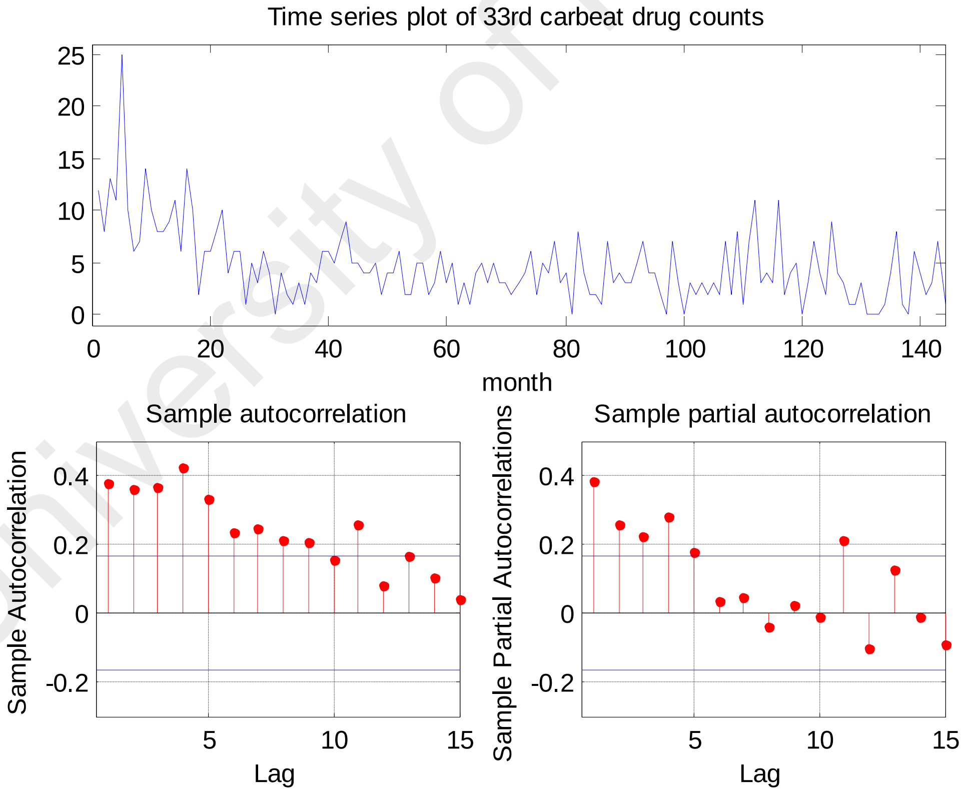


Figure 7.4.1: Time series plot of 33rd carbeat drug counts

7.4.2 Fitting to Poisson MPT(p) Models

All models have been estimated by MLE via EM algorithm. First, the parameter mean $\hat{\alpha} = 0.9977$ and $\hat{\lambda} = 4.6355$ are estimated, then the mixing proportions (as stated above) are estimated recursively until the tolerance level reaches 0.001. Obviously, by the information criteria AIC, it indicates that the model with $p = 5$ has the lowest score. In terms of AIC criterion, it is the most appropriate model to be fitted by the data. Also, the fourth component gives the greatest impact among others.

Table 7.4.1: Parameter estimates and AIC values of the models

Order	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\phi}_4$	$\hat{\phi}_5$	$\hat{\phi}_6$	AIC
$p = 1$	0.03220						799.35
$p = 2$	0.02795	0.01672					795.08
$p = 3$	0.02831	0.01832	0.01659				781.62
$p = 4$	0.03010	0.01858	0.01574	0.04354			767.46
$p = 5$	0.02821	0.02022	0.01619	0.04327	0.03513		752.96
$p = 6$	0.02800	0.01144	0.01231	0.04160	0.01626	0.00485	758.58

7.5 Concluding Remarks

This chapter reveals the practical advantage of MPT(1) model. The proposed MPT(1) process is compared extensively with the current integer-valued time series models (see Chapter 2 for details) for the criminal data set. The new Geometric MPT(1) process is seen to outperform the other models.

For IP counts data set, Binomial MPT(1) shows its flexibility in handling the data set. The transition probability of Binomial MPT(1) has a simple form, is straightforward to apply as well as being easy to construct. This is unlike the Binomial AR(1) (McKenzie, 1985) that requires modification from the conventional INAR(1) and the derivation of the properties are relatively more involved. For underdispersed data set, it is obvious that the binomial marginal model performs better than Poisson marginal model. The Binomial MPT(1) model has the lowest AIC values compared to the Binomial AR(1) and Pegram's AR(1) models.

In the third data set, MPT(1) model fits not better than the other two. However in coherent forecasting the confidence interval is narrower due to smaller standard deviation. From the illustrative fits to real count data sets, it can be concluded that MPT(1) model appear to be a useful alternative model in discrete-valued time series modelling.

CHAPTER 8: CONCLUSION AND FURTHER RECOMMENDATIONS

Discrete-valued time series modelling has been a dynamic research topic since three decades ago. Back in the 80's, the thinning operator which replaces the scalar multiplication in continuous time series models, has been extensively discussed in the literature. However, it is well known that the thinning operator is confined to infinitely divisible distributions. Joe (1996) discussed the thinning operator with convolution-closed infinite divisible univariate margins. Aly and Bouzar (2005) studied the generalized thinning operator with stationarity solution. In the studies, many researchers have showed the significance of thinning operator in describing the real phenomena. However, it is limited to infinitely divisible cases. Weiβ (2008) handled the Binomial case through relatively more complicated expressions. One alternative operator was introduced by Pegram's (1980) to deal with non-infinitely divisible cases. The Pegram's operator was applied by Biswas and Song (2009) for categorical data with discussion on its statistical analysis by Song et al. (2013). In ARMA-type time series model, Pegram's operator emphasizes the mixing concept, which appears to be more flexible than the thinning operator in dealing with Binomial and categorical observations.

The new model proposed in this dissertation inherits the traits from thinning and Pegram's operators, as it was created via the combination of both operators. This dissertation proposed and investigated an entirely new model, which is able to deal with infinitely and non-infinitely divisible distributions in a simpler approach. It has been further revealed to have better performance than the current existing thinning and Pegram's operators. The previous chapters have begun with the construction of the proposed MPT(1) model, followed by a discussion of useful properties; stationarity and regression, as well as simulation studies. Along this line, we consider also the forecasting model and the asymptotical distributions. Some real applications have been studied to illustrate the performance of the MPT(1) model with infinitely and non infinitely divisible distributions. Comparison with existing discrete-valued models has also been carried out and the results show that the proposed model has outstanding performance. The merits of the proposed model have been explored. On top of these, model flexibility and simplicity are distinguished features of the proposed model. This is the main distribution of the study.

It is important to consider the possible extensions from the current research work. One immediate extension from Chapter 3 is to consider the Negative Binomial (NB) and Geometric marginal distributions in MPT(1) model. It is of interest as the NB and Geometric margins are well known for overdispersed observations. Compared to the complication encountered by NB and Geometric in INAR(1) model; see McKenzie (1987), the NB MPT(1) model may be considered in the near future. It is suggested to make comparison with NB and Geometric in MPT(1) model. Real life examples can be presented based upon that to illustrate the model efficiency.

Future development on higher order moments of Poisson MPT(1) model can be considered. Schweer and Weiβ (2014) investigated the statistical analysis in stochastic properties and testing for overdispersion for compound Poisson INAR(1) process. It is recommended to implement a similar analysis for the Poisson, NB and Binomial MPT(1) processes. Weiβ has discussed the Binomial AR(1) model in a series of papers. One particular paper (Weiβ and Kim, 2013) derived explicit expressions for the joint moments and cumulants up to order 4. Thus it is of interest to further explore Binomial MPT(1) process.

Coherent forecasting model of Binomial MPT(1) model is one of the possible piece of future work. As it can be seen, we discussed only for Poisson case in this dissertation. It will be of particular interest to derive the asymptotic distributions for *k*-step-ahead forecasting Binomial MPT(1) process. Besides applying Fisher Information Matrix to estimate the standard errors, one can consider estimation by block-of-block bootstrapping. Both methods can be compared through simulation studies. The coherent forecasting for time series models has been studied by some researchers. Kim and Park (2010a) extended the INAR(2) forecasting results of Jung and Tremayne (2006) to the Binomial AR(p) model. The bootstrap method for prediction intervals is clearly explained in the paper, and empirical application is shown for illustration. Maiti et al. (2014) investigated the coherent forecasting of zero-inflated count time series process.

It is a straightforward to extend the current first order MPT(1) process to MPT(p) process. The MPT(p) model has been discussed in Chapter 6, mainly in model construction and fundamental properties, with other properties left to be investigated. One important aspect is the existence of MPT(p) process. A unique, stationary and ergodic MPT(p) process can be studied in the similar approach as presented by Ristic and Nastic (2012). On the other hand, there are many studies left behind for MPT-MA processes. In Chapter 6, the construction and fundamental properties of MPT-MA processes have been developed. Brännäs and Hall (2001) investigated four different characteristics of INMA processes. Further research works for the MPT-MA processes, can be done in this direction. Parameter estimation and Monte Carlo simulation and potential real life applications can also be considered in further investigation of the MPT-MA processes. The qth-order MPT-MA processes can be examined along with the real examples. This will be a new research area in discrete-valued time series modelling.

Last but not least, a slight modification can be considered over the current proposed model. This dissertation mainly emphasized the mixture of thinning operator ($\alpha \circ X_{t-1}$) and innovation process(ε_t); where we mixed both random variables with parameters ϕ

and $1-\phi$ respectively. In considering the modification, we mix the INAR(1) process $(\alpha \circ X_{t-1}+\varepsilon_t)$ and innovation process (ε_t) with the parameter $1-\phi$ and ϕ respectively. Assuming that the mixing parameter $\phi \in (0,1)$. Bakouch and Ristic (2009) have discussed zero-truncated Poisson mixed INAR(1) model which is the special case of this modified model. Preliminary works are shown to make the modified version possible. The definition is given as such. Let X_t be a NB process with parameter (θ, p), where $\theta > 0$ and $0 < p < 1$, let $\phi \in (0,1)$, then the modified model is defined by

$$X_t = (1-\phi, \alpha \circ X_{t-1}+\varepsilon_t) * (\phi, \varepsilon_t) \tag{8.1}$$

where the INAR(1) process follows Definition 2.1.2 and the marginal distribution of the innovation process follows X_t. A general pgf of Eq. (8.1) is given by

$$G_X(z) = (1-\phi)G_{X_{t-1}}(1-\alpha+\alpha z)G_\varepsilon(z) + \phi G_\varepsilon(z) \tag{8.2}$$

while the pgf of innovation process is given by

$$G_\varepsilon(z) = \frac{G_X(z)}{(1-\phi)G_{X_{t-1}}(1-\alpha+\alpha z)+\phi} \tag{8.3}$$

If we consider $X_t \sim NB(\theta, p)$, then the pgf of ε_t is

$$G_\varepsilon(z) = \frac{(1+p-pz)^{-\theta}}{(1-\phi)(1+\alpha p-\alpha p z)^{-\theta}+\phi} \tag{8.4}$$

First, the pmf of model (8.1) can be obtained via the multiple differentiation at $z=0$. We have

$$G_\varepsilon^{'}(z) = G_\varepsilon(z)\left[\theta p(1+p-pz)^{-1} - (1-\phi)(\theta)(\alpha p)\frac{(1+\alpha p-\alpha p z)^{-\theta-1}}{(1-\phi)(1+\alpha p-\alpha p z)^{-\theta}+\phi}\right]$$

and

$$G_\varepsilon'(z)|_{z=0} = G_\varepsilon(z)|_{z=0}\left[\theta p(1+p)^{-1} - (1-\phi)(\theta)(\alpha p)\frac{(1+\alpha p)^{-\theta-1}}{(1-\phi)(1+\alpha p)^{-\theta}+\phi}\right]$$

Before we move forward to find the second derivative of $G_\varepsilon(z)$, we assume that

$$A = \theta p(1+p-pz)^{-1} - (1-\phi)(\theta)(\alpha p)\frac{(1+\alpha p-\alpha pz)^{-\theta-1}}{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi}$$

$$A' = \theta p^2(1+p-pz)^{-2} - (1-\phi)(\theta)(\alpha p)\frac{(1+\alpha p-\alpha pz)^{-\theta-2}}{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi}$$
$$-(1-\phi)^2(\theta)^2(\alpha p)^2\frac{(1+\alpha p-\alpha pz)^{-2(\theta+1)}}{\{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi\}^2}$$

$$A'|_{z=0} = \theta p^2(1+p)^{-2} - (1-\phi)(\theta)(\alpha p)\frac{(1+\alpha p)^{-\theta-2}}{(1-\phi)(1+\alpha p)^{-\theta}+\phi}$$
$$-(1-\phi)^2(\theta)^2(\alpha p)^2\frac{(1+\alpha p)^{-2(\theta+1)}}{\{(1-\phi)(1+\alpha p)^{-\theta}+\phi\}^2}$$

and so the second derivative of ε_t pgf is

$$G_\varepsilon''(z) = G_\varepsilon(z)\cdot A' + A\cdot G_\varepsilon'(z)$$

We continue to find the third derivative. To avoid confusion, we look at the derivatives of A' one by one

$$\frac{d}{dz}(1+p-pz)^{-2} = 2p(1+p-pz)^{-3}$$

$$\frac{d}{dz}\frac{(1+\alpha p-\alpha pz)^{-\theta-2}}{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi}$$
$$= \frac{(-\theta-2)(-\alpha p)(1+\alpha p-\alpha pz)^{-\theta-3}}{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi}$$
$$-(1-\phi)(\theta)(\alpha p)\frac{(1+\alpha p-\alpha pz)^{-2\theta-3}}{\{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi\}^2}$$

$$\frac{d}{dz}\frac{(1+\alpha p-\alpha pz)^{-2(\theta+1)}}{\{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi\}^2}$$

$$=\frac{2(\theta+1)(\alpha p)(1+\alpha p-\alpha pz)^{-2\theta-3}}{\{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi\}^2}$$

$$-2(1-\phi)(\theta)(\alpha p)\frac{(1+\alpha p-\alpha pz)^{-3(\theta+1)}}{\{(1-\phi)(1+\alpha p-\alpha pz)^{-\theta}+\phi\}^3}$$

the third derivative of ε_t pgf becomes

$$G_\varepsilon^{(3)}(z)=G_\varepsilon(z)\cdot A'' + A'\cdot G_\varepsilon'(z)+A\cdot G_\varepsilon''(z)+G_\varepsilon'(z)\cdot A'$$

$$=A''\cdot G_\varepsilon(z)+2A'\cdot G_\varepsilon'(z)+A\cdot G_\varepsilon''(z)$$

and

$$G_\varepsilon^{(4)}(z)=A^{(3)}\cdot G_\varepsilon(z)+3\left(A''\cdot G_\varepsilon'(z)+A'\cdot G_\varepsilon''(z)\right)+A\cdot G_\varepsilon^{(3)}(z)$$

The multiple derivations can be represented in the form of binomial coefficient and it is generally represented by the formulation

$$G_\varepsilon^{(n)}(z)=\binom{n-1}{i}A^i\, G_\varepsilon^{(n-1)}(z); n=1,2,\dots;\ i=1,2,\dots n-1$$

and so the pmf can be obtained via the expression above.

This modified version can then be extended to consider more details, such as fundamental and regression properties, parameter estimation and consistency problems. Stationarity is immediate. Real life applications should be carried out as well upon the introduction of important properties. The special case of applying Geometric marginal through this approach is possible, by just taking the parameter $\theta=1$. In fact, it is a neat form for Geometric case as the denominator of $G_\varepsilon(0)$ leads to $(1-\phi)(1+\alpha p)^{-1}+\phi$, which can be simplified to 1. Therefore, one can observe that the marginal of X_t is deterministic for Geometric marginal case. Indeed, a simpler formulation is obtained for Geometric marginal. One important result is discovered, that the Poisson marginal may not be applicable or still undetermined for this modified version. The implementation for non-infinitely divisible distributions like Binomial is still under investigation. Comparison with other time series models is of interest, such as the recent works by Nastic and Ristic (2012). The mixture model suggested in the paper seems complicated. An easy interpretation and efficient model is needed for the time series of counts community. More detailed study can be explored in the near future.

This thesis has proposed a useful time series model for count data, which has also been extended to the MA and ARMA processes. It is hoped that this thesis has contributed to discrete-valued time series modelling.

REFERENCES

Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive. *Journal of Time Series Analysis, 8*, 261-275.

Al-Osh, M.A. and Alzaid, A.A. (1988). Integer-valued moving average (INMA) process. *Statistical Papers, 29*, 281-300.

Al-Osh, M.A. and Alzaid, A.A. (1991). Binomial autoregressive moving average models. *Stochastic Models*, *7*(2), 261–282.

Al-Osh, M.A. and Aly, E. E. A. A. (1992). First-order autoregressive time series with negative binomial and geometric marginals. *Communications in Statistics - Theory and Methods, 21*(9), 2483-2492.

Aly, E. E. A. A. and Bouzar, N. (1994). On some integer-valued autoregressive moving average models. *Journal of Multivariate Analysis, 50*, 132-151.

Aly, E. E. A. A. and Bouzar, N. (2005). Stationary solutions for integer-valued autoregressive processes. *International Journal of Mathematics and Mathematical Sciences, 1*, 1-18.

Alzaid, A., Al-Osh, M. (1988). First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties. *Statistics Neerlandica*, *42*(1), 53-61.

Alzaid, A.A. and Al-Osh, M.A. (1993). Some autoregressive moving average processes with generalized poisson marginal distributions. *Annals of the Institute of Statistical Mathematics, 45*(20), 223-232.

Bakouch, H. and Ristic, M. M. (2010). Zero truncated poisson integer-valued AR(1) model. *Metrika*, *72*(2), 265-280.

Barczy, M., Ispany, M., Pap, G., Scotto, M. and Silva, M. E. (2010). Outliers in INAR(1) models. *arXiv:0903.2421 [math.PR]*.

Bhat, B. R. and Adke, S. R. (1981). Maximum likelihood estimation for branching processes with immigration. *Advances in Applied Probability*, *13*(3), 498-509.

Biswas, A., Song, P. X. -K. (2009). Discrete-valued ARMA processes. *Statistics and Probability Letters*, *79*, 1884-1889.

Biswas, A. and Guha, A. (2009). Time series analusis of categorical data using auto-mutual information. *Journal of Statistical Planning and Inference, 139*, 3076-3087.

Brännäs, K. (1994). Estimation and testing in integer-valued AR(1) models. *Umeå Economic Studies. Department of Economics, Umeå University.*

Brännäs, K. (1995). Prediction and control for a time-series count data model. *International Journal of Forecasting, 11*, 263-270.

Brännäs, K. and Hellström, J. (2001). Generalized integer-valued autoregression. *Economic Reviews, 20*(4), 425-443.

Brännäs, K. and Hall, A. (2001). Estimation in integer-valued moving average models. *Applied Stochastic Models in Business and Industry, 17*, 277-291.

Brockwell, P. J., Davis, R. A. (2002). *Introduction to time series and forecasting, second edition*. New York: Springer-Verlag.

Brockwell, P. J and Davis, R. A. (1987). *Time series: theory and methods*. Springer-Verlag, New York.

Box, G. E. P. and Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. San Francisco: Holden-Day.

Bu, R. J, McCabe, B. and Hadri, K. (2008) Maximum likelihood estimation of higher-order integer-valued autoregressive processes. *Journal of Time Series Analysis, 29*(6), 973-994.

Bu, R. J. and McCabe, B. P. M (2008) Model selection, estimation and forecasting in INAR (p) models: a likelihood-based markov chain approach. *International Journal of Forecasting, 24*, 151-162.

Consul, P. C. and Mittal, S. P. (1975). A new urn model with predetermined strategy. *Biometrical Journal, 17*, 67-75.

Cui, Y. and Lund, R. B. (2009). A new look at time series of counts. *Biometrika*, 96, 781-792.

Cui, Y. and Lund, R. B. (2010). Inference for binomial AR(1) models. *Statistics and Probability Letters, 80*: 1985-1990.

Doukhan, P. Fokianos, K. and Li, X. Y. (2012). On weak dependence conditions: The case of discrete valued processes. Statistical and Probability Letters 82, 1941-1948.

Freeland, R. K. and McCabe, B. P. M. (2004a). Analysis of low count time series data by poisson autoregression. *Journal of Time Series Analysis*, *25*(5), 701-722.

Freeland, R. K. and McCabe, B. P. M. (2004b). Forecasting discrete valued low count time series. *International Journal of Forecasting*, *20*, 427-434.

Freeland, R. K. and McCabe, B. P. M. (2005). Asymptotic properties of CLS estimators in poisson AR(1) model. *Statistics and Probability Letters, 73*(2), 147-153.

Freeland, R. K. (1998). Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data. *Ph.D. Thesis, The University of British Columbia, Canada.*

Furman, W. D. and Lindsay, B. G. (1994). Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods. *Computational Statistics & Data Analysis, 17*, 493-507.

Gourieroux, C. Monfort, A. and Trongon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica, 52*(3), 681-700.

Grunwald, G. K., Hyndman, R. J., Tedesco, L., Tweedie, R. L. (2000). Non-gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics, 42*(4), 479-495.

Jacobs, P. A., Lewis, P. A. W. (1977). A mixed autoregressive-moving average exponential sequence and point process (EARMA(1,1)). *Advances in Applied Probability, 9*(1), 87-104.

Jacobs, P. A., Lewis, P. A. W. (1978b). Discrete time series generated by mixture II: asymptotic properties. *Journal of Royal Statistical Society B, 40*, 222-228.

Jacobs, P. A., Lewis, P. A. W. (1978a). Discrete time series generated by mixtures I: correlation and run properties. *Journal of Royal Statistical Society B, 40*, 94-105.

Jazi, M. A., Jones, G. and Lai, C. D. (2012). First-order integer-valued AR processes with zero inflated poisson innovations. *Journal of Time Series Analysis, 33*(6), 954-963.

Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability, 33*(3), 664-677.

Jung, R.C. and Tremayne, A.R. (2006). Coherent forecasting in integer time series models. *International Journal of Forecasting, 22*, 223-238.

Karlis, D. and Xekalaki, E. (1999). Improving the EM algorithm for mixtures. *Statistics and Computing, 9*, 303-307.

Kim, H. Y and Park, Y. S. (2008). A non-stationary integer-valued autoregressive model. *Statistical Papers, 49*(3), 485-502.

Kim, H. Y and Park, Y. S. (2010a). Coherent forecasting in binomial AR(p) model. *Communications of the Korean Statistical Society, 17*(1), 27-37.

Kim, H. Y and Park, Y. S. (2010b). Markov chain approach to forecast in the binomial autoregressive models. *Communications of the Korean Statistical Society, 17*(3), 27-37.

Klimko, L. A and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *The Annals of Statistics, 6*(3), 629-642.

Lawrence, A. J. and Lewis, P. A. (1980). The exponential autoregressive-moving average EARMA(p,q) process. *Journal of the Royal Statistical Society, Series B, 42*(2), 150-161.

Latour, A. (1995). Existence and stochastic structure of a non-negative integer-valued autoregressive process. *Journal of Time Series Analysis, 19*(4), 439-455.

Liboschik, T., Fokianos, K. and Fried, R. (2015). tscount: a R package for analysis of count time series following generalized linear models. *Yet Another Blog in Statistical Computing*.

McCabe, B. P. M. and Martin, G. M. (2005). Bayesian predictions of low count time series. *International Journal of Forecasting, 21*, 315-330.

McLachlan, G. and Peel, D. (2000). *Finite mixture models*. John Wiley & Sons. Inc.

McKenzie, E. (1981). Extending the correlation structure of exponential autoregressive-moving average processes. *Journal of Applied Probability, 18*(1), 181-189.

McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bulletin, 21*(4), 645-650.

McKenzie, E. (1988). Some ARMA models for dependent sequences of poisson counts. *Advances in Applied Probability, 20*(4), 822-835.

McKenzie, E. (1987). Innovation distributions for gamma and negative binomial autoregressions. *Scandivanian Journal of Statistics, 14*(1), 79-85.

Nastic, A. S. and Ristic, M. M. (2012). Some geometric mixed integer-valued autoregressive (INAR) models. *Statistics and Probability Letters, 82*, 805-811.

Neway, W. K. and McFadden, D. (1994). *Large sample estimation and hypothesis testing*. Handbook of Econometrics, volume IV, edited by Engle, R. F and McFadden, D. L., Elsevier Science.

Pegram, G. G. (1980). An autoregressive model for multilag markov chains. *Journal of Applied Probability, 17*, 350-362.

Pedeli, X. and Karlis, D. (2011). A bivariate INAR(1) process with application. *Statistical Modelling, 11*(4), 325-349.

Pedeli, X. and Karlis, D. (2013). Some properties of multivariate INAR(1) processes. *Computational Statistics and Data Analysis, 67*, 213-225.

Ristic, M. M., Bakouch, H. S., Nastic, A. S. (2009). A new geometric first-order integer-valued autoregressive (NGINAR(1)) process. *Journal of Statistical Planning and Inference, 139*, 2218-2226.

Ristic, M. M. and Nastic, A. S. (2012). A mixed INAR(p) model. *Journal of Time Series Analysis, 33*, 903-915.

Ross, S. M. (2000). *Introduction to probability models*. Academic Press.

Schweer, S. and Weiß, C. H. (2014). Compound poisson INAR(1) processes: stochastic properties and testing for overdispersion. *Computational Statistics and Data Analysis*, *77*, 267-284.

Silva, I. M. M. (2005). Contribution to the analysis of discrete-valued time series. *Ph.D Thesis, Universidade do Porto, Portugal*.

Silva, I., Silva, M. E., Pereira, I. and Silva. N. (2005) Replicated INAR(1) process. *Methodology and Computing in Applied Probability, 7*, 517-542.

Silva N., Pereira, I. and Silva, M. E. (2009) Forecasting in INAR (1) model. *Revstat-Statistical Journal, 7*(1), 119-134.

Shenton, L. R. (1986). Quasibinomial distributions. *Encyclopedia of Statistical Sciences (eds. S. Kotz and N. L. Johnson), 7*, 458-460.

Song, P. X. -K., Freeland, R. K., Biswas, A. and Zhang, S. L. (2013). Statistical analysis of discrete-valued time series using categorical ARMA models. *Computational Statistics and Data Analysis, 57*, 112-124.

Sprott, D. A. (1983). Estimating the parameters of a convolution by maximum likelihood. *Journal of the American Statistical Association*, *78*(382), 457-460.

Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *The Annals of Probability, 7*(5), 893-899.

Weiß, C. H. (2007). Controlling correlated processes of poisson counts. *Quality and Reliability Engineering International*, *23,* 741-754.

Weiß, C.H. (2008). The combined INAR(p) models for time series of counts. *Statistics and Probability Letters, 78*,1817-1822.

Weiß, C. H. (2012). Controlling correlated processes of poisson counts. *Quality and Reliability Engineering International*, *23*(6), 741-754.

Weiß, C.H. (2008). Thinning operations for modelling time series of counts - a survey. *AStA Advances in Statistical Analysis, 92*(3), 319-341.

Weiß, C. H. (2009). A new class of autoregressive models for time series of binomial counts. *Communications in Statistics-Theory and Methods, 38*(4), 447-460.

Weiß, C. H. (2012). Fully observed INAR(1) processes. *Journal of Applied Statistics, 39*(3), 581-598.

Weiß, C. H. and Kim, H. Y. (2013). Binomial AR(1) processes: moments, cumulants, and estimation. *Statistic: A Journal of Theoretical and Applied Statistics*, *47*(3), 494-510.

Zhang, H. X., Wang, D. H. and Zhu, F. K. (2012). Generalized RCINAR(1) process with signed thinning operator. *Communications in Statistics - Theory and Methods, 41* (10), 1750-1770.

Zheng, H., Basawa, I.V. and Datta, S. (2007). First-order random coefficient integer-valued autoregressive processes. *Journal of Statistical Planning and Inference*, *173*, 212-229.

Zhu , R., Joe, H. (2006). Modelling count data time series with Markov processes based on binomial thinning. *Journal of Time Series Analysis, 27*(5), 725-738.

LIST OF PUBLICATIONS AND PAPER PRESENTED

International Journals

1. Wooi Chen Khoo, Seng Huat Ong, Atanu Biswas (2015). Modelling Time Series of Counts with a New Class of INAR(1) Models. Statistical Papers. DOI 10.1007/s00362-015-0704-0.

Conference Proceedings

1. Wooi Chen Khoo and Seng Huat Ong (2014). A New Model for Time Series of Counts. AIP Conf. Proc. 1605, 938. http://dx.doi.org/10.1063/1.4887716.
2. Wooi Chen Khoo and Seng Huat Ong (2015). A Mixed Time Series Model of Binomial Counts. AIP Conf. Proc. 1682, 050001. http://dx.doi.org/10.1063/1.4932492.