WADLEY'S PROBLEM WITH OVERDISPERSION

by

KERRY LEIGH LEASK

(B.Sc. Honours)

Submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN SCIENCE

in the

School of Statistics and Actuarial Science
Faculty of Science and Agriculture
University of KwaZulu-Natal
Pietermaritzburg

February 2009

Declaration

The work described in this thesis was carried out in the School of Statistics and Actuarial Science, University of KwaZulu-Natal, Pietermaritzburg, under the supervision of Professor Glenda Matthews and the co-supervision of Professor Linda M. Haines.

The thesis presents original work by the author and has not otherwise been submitted in any form for any degree or diploma to any University. Where use has been made of the work of others it is duly acknowledged in the text.

February, 2009.

Student:

_____

Kerry Leigh Leask

Supervisor:

_____

Professor Glenda Matthews

Co-supervisor:

_____

Professor Linda M. Haines

Acknowledgments

Firstly I would like to thank Professor Linda Haines for her enthusiasm, guidance and patience. She has been an inspiration and a mentor and I am greatly indebted to her. I would also like to thank Professor Glenda Matthews for her encouragement and suggestions.

I am grateful for the seemingly small, yet much appreciated, contributions of others to this work. In particular, I extend my gratitude to Professor Ray Haines for his eloquent suggestions and to Professor Tim Dunne for his assistance with unusual references.

On a personal note, I would like to mention my family. I thank my Dad for his enthusiasm and encouragement along the way. My mum is my heroine and my greatest fan and Pops has always been keenly interested in and supportive of what I have been doing.

Finally, I give honour to God whose grace is sufficient for me.

Abstract

Wadley's problem frequently emerges in dosage-mortality data and is one in which the number of surviving organisms is observed but the number initially treated is unknown. Data in this setting are also often overdispersed, that is the variability within the data exceeds that described by the distribution modelling it. The aim of this thesis is to explore distributions that can accommodate overdispersion in a Wadley's problem setting. Two methods are essentially considered. The first considers adapting the beta-binomial and multiplicative binomial models that are frequently used for overdispersed binomial-type data to a Wadley's problem setting. The second strategy entails modelling Wadley's problem with a distribution that is suitable for modelling overdispersed count data. Some of the distributions introduced can be used for modelling overdispersed count data as well as overdispersed dose-response data from a Wadley context. These models are compared using goodness of fit tests, deviance and Akaike's Information Criterion and their properties are explored.

Contents

1 Introduction **1**

2 General Background **3**

2.1 Introduction 3

2.2 Data Sets 4

2.2.1 Algae Data 4

2.2.2 Malaria Data 5

2.3 The Poisson Model 6

2.3.1 Setup and Notation 6

2.3.2 Poisson Distribution for N 7

2.3.3 Likelihood Function 8

2.3.4 Modelling the Probabilities 9

2.3.5 Score Function and Information Matrix 11

2.3.6 Estimation 15

2.3.7 Examples 19

2.4 Conclusions 29

3 The Beta-Binomial Model **30**

3.1 Introduction 30

3.2 The Model 31

3.2.1 The Binomial Model 31

3.2.2 The Beta-Binomial Model 31

3.2.3 Likelihood Function . . . 34
3.2.4 Modelling the Probability . . . 34
3.3 Score Functions and Observed Fisher Information Matrix . . . 35
3.3.1 Observed Fisher Information . . . 35
3.3.2 Models for Separate Doses . . . 37
3.4 Estimation . . . 37
3.4.1 Inference . . . 37
3.5 Examples . . . 39
3.5.1 Algae Data . . . 39
3.5.2 Malaria Data . . . 41
3.6 Conclusions . . . 44

4 The Beta-Poisson Model 45
4.1 Introduction . . . 45
4.2 Distribution . . . 46
4.2.1 Derivation of the p.m.f. . . . 46
4.2.2 Literature Review . . . 49
4.3 Confluent Hypergeometric Function . . . 50
4.3.1 Modality of the Confluent Hypergeometric Function . . . 51
4.4 Properties of the Distribution . . . 56
4.4.1 Moment Generating Function . . . 56
4.4.2 Modality of the beta-Poisson Distribution . . . 60
4.4.3 Examples . . . 62
4.5 Likelihood, Score Functions and the Information Matrix . . . 66
4.5.1 Likelihood Function . . . 66
4.5.2 Score Functions and Information Matrix . . . 66
4.6 Estimation . . . 67
4.6.1 Inference . . . 68
4.6.2 Goodness of Fit . . . 69
4.7 An Example . . . 70

4.8 Modelling Dose-Response Data . . . 73
4.8.1 Likelihood Function . . . 73
4.8.2 Models for Separate Doses . . . 74
4.8.3 Modelling the Probabilities . . . 75
4.8.4 Estimation . . . 75
4.8.5 Inference . . . 76
4.8.6 Goodness of Fit . . . 77
4.8.7 Examples . . . 77
4.9 Conclusions . . . 81

5 The Multiplicative Binomial Model 83
5.1 Introduction . . . 83
5.2 Distribution . . . 84
5.2.1 Derivation . . . 84
5.2.2 Properties of the Distribution . . . 88
5.2.3 Examples . . . 93
5.3 Likelihood, Score Functions and Information Matrix . . . 95
5.3.1 Likelihood and Score Functions . . . 97
5.3.2 Information Matrix . . . 98
5.4 Estimation . . . 99
5.4.1 Maximum Likelihood Estimation . . . 99
5.4.2 Inference . . . 103
5.4.3 Goodness of Fit . . . 104
5.5 An Example . . . 105
5.6 Modelling Dose-Response Data . . . 108
5.6.1 Model Setting . . . 108
5.6.2 Likelihood, Score Functions and Information Matrix . . . 108
5.6.3 Model Fitting and Checking . . . 111
5.6.4 An Example . . . 112
5.7 Conclusions . . . 114

6 The Altham-Poisson Model **117**

6.1 Introduction . . . 117

6.2 Distribution . . . 118

6.2.1 Derivation of the p.m.f. . . . 118

6.2.2 Infinite Sum . . . 119

6.2.3 Simulation . . . 121

6.3 Moments of the Distribution . . . 122

6.3.1 Selection of a Cutoff . . . 123

6.3.2 Simulation . . . 124

6.3.3 Conditioning . . . 124

6.3.4 Examples . . . 125

6.4 Likelihood, Score Functions and Information Matrix . . . 129

6.4.1 Likelihood Function . . . 129

6.4.2 Score Functions . . . 129

6.4.3 Information Matrix . . . 131

6.5 Estimation . . . 133

6.5.1 EM algorithm . . . 133

6.5.2 Direct Maximization of the Likelihood . . . 138

6.5.3 Inference . . . 139

6.5.4 Goodness of Fit . . . 139

6.6 An Example . . . 140

6.7 Modelling Dose-Response Data . . . 142

6.7.1 Modelling the Probabilities . . . 142

6.7.2 Likelihood, Score Functions and Information Matrix . . . 143

6.7.3 Model Fitting and Checking . . . 145

6.7.4 Results . . . 146

6.8 Conclusions . . . 152

7 The Negative Binomial Model **154**

7.1 Introduction . . . 154

7.2 The Model . . . 155
7.2.1 The Negative Binomial Distribution . . . 155
7.2.2 Negative Binomial distribution for N . . . 156
7.2.3 Gamma Distribution for the Poisson Parameter . . . 157
7.3 Modelling the Probability . . . 158
7.4 Likelihood, Score Functions and Information Matrix . . . 159
7.4.1 Likelihood Function . . . 159
7.4.2 Score Functions . . . 160
7.4.3 Information Matrix . . . 161
7.5 Models for Separate Doses . . . 163
7.6 Estimation . . . 163
7.6.1 Parameter Estimates . . . 163
7.6.2 Standard Errors of Parameter Estimates . . . 165
7.6.3 Parameter Estimates for the Separate Dose Models . . . 166
7.6.4 Goodness of Fit . . . 167
7.7 Examples . . . 167
7.7.1 Algae Data . . . 167
7.7.2 Malaria Data . . . 168
7.8 Conclusions . . . 170

8 The Binomial-EWP Model **171**
8.1 Introduction . . . 171
8.2 The EWP Distribution . . . 172
8.2.1 Distribution . . . 172
8.2.2 Moments of the EWP Distribution . . . 173
8.2.3 Examples . . . 175
8.2.4 Likelihood, Score Functions and Information Matrix . . . 175
8.2.5 Estimation and Inference . . . 175
8.2.6 Goodness of Fit . . . 179
8.2.7 An Example . . . 179

8.3 Derivation of the Binomial-EWP p.m.f. . . . 182
8.4 Moments of the Binomial-EWP Distribution . . . 183
8.5 Examples . . . 186
8.6 Modelling Dose-Response Data . . . 186
8.6.1 Likelihood Function . . . 186
8.6.2 Score Functions and the Information Matrix . . . 189
8.7 Model Fitting and Checking . . . 189
8.7.1 Estimation . . . 189
8.7.2 Inference . . . 190
8.7.3 Goodness of Fit . . . 191
8.8 Examples . . . 191
8.8.1 Algae Data . . . 191
8.8.2 Malaria Data . . . 194
8.9 Conclusions . . . 195

9 Conclusions **197**

Chapter 1

Introduction

Wadley's problem is one that frequently emerges in dosage-mortality studies where the number of organisms that survive treatment with a particular dose of a drug is recorded, but the number initially exposed to the drug is unknown and must be estimated from the data. The counts cannot, therefore, be considered to follow a binomial distribution but are often modelled by a Poisson distribution (Wadley, 1949). Particular data sets in Wadley's problem setting, including the data set for antimalarial drugs which is reported in the thesis by Gouws (1995, p.98) and used extensively in this study, exhibit overdispersion (Gouws, 1995, p.90). That is, the variability in the data set exceeds the variability in the distribution which models it. The aim of this thesis is to explore and derive models that are suitable for modelling overdispersed dose-response data that arise from a Wadley's problem setting. The thesis is comprised of three main themes, each of which is divided into two chapters. The first two of these themes explore models that have been used for overdispersed binomial-type data and which are adapted to a Wadley's problem setting. In the last of these themes, distributions that are suitable for overdispersed count data are considered for modelling the unknown number of organisms initially treated.

Chapter 2 includes a description of the primary data sets considered in this thesis as well as a preliminary investigation of the data. The beta-binomial dis-

tribution is presented in Chapter 3 and in Chapter 4 this model is adapted to a Wadley's problem setting. The multiplicative binomial distribution is investigated in Chapter 5 and its adaptation to a Wadley's problem setting, resulting in the Altham-Poisson distribution, is explored in Chapter 6. The negative binomial model in the Wadley context is presented in Chapter 7 and the two parameter exponentially weighted Poisson distribution is considered in the same context in Chapter 8. Conclusions are drawn and pointers for future research are presented in Chapter 9.

Chapter 2

General Background

2.1 Introduction

Wadley (1949) first considered modelling dose-mortality data in which N, the number of organisms initially exposed to the drug, is unknown and must therefore be estimated from a control sample. This phenomenon frequently emerges and has aptly been termed Wadley's problem. Wadley (1949) showed that if the number of organisms treated follows a Poisson distribution with parameter τ, then the number of survivors will follow a Poisson distribution with parameter $\tau(1-p)$, where p is the probability that an organism dies. He also obtained maximum likelihood estimates of the parameters in this model. Anscombe (1949) introduced the idea of using the negative binomial distribution as a means of accommodating overdispersion in Wadley-type data.

More recently, Baker, Pierce and Pierce (1980) developed a method for analyzing Wadley-type dose-response data when control observations are present. The approach they employed was to introduce separate link functions for control observations and those that arise from the administration of a non-zero dose of the drug. This idea is used by Smith and Morgan (1989), who presented a GLIM macro for modelling data in a Wadley's problem setting. Incorporated into the macro is the use of a Taylor series expansion as a method of overcoming nonlinearity in the link

function.

The primary data sets included in this study are described in Section 2.2. In Section 2.3 the classical Wadley-Poisson model which entails modelling N, the unknown number of organisms initially exposed to the drug, with a Poisson distribution (Wadley, 1949) is examined. A consideration of this model serves to provide a preliminary investigation of the data and to lay a platform for the remainder of this thesis. Concluding remarks are recorded in Section 2.4.

2.2 Data Sets

2.2.1 Algae Data

In order to gauge the effect of chemicals in water on the environment, the growth of unicellular bacteria under exposure to differing concentrations of chemicals was noted. These data were taken from a paper by Baker, Pierce and Pierce (1980). A number of *Selenastrum Capricoruntum* organisms was exposed to a variety of concentrations of a toxicant for seven days, after which the number of survivors was recorded. Each concentration was administered five times. The method of obtaining the counts was based on an optical density technique and the observations are presented in Table 2.1.

Table 2.1: Data for *Selenastrum Capricoruntum* exposed to a toxicant.

Drug conc. (μg/ml)	Number Surviving					Mean	Variance
0	219	228	202	237	228	223	176
1	167	158	158	175	167	165	52
5	105	123	105	105	105	109	65
10	88	88	61	61	88	77	219
50	61	44	35	35	44	44	89

2.2.2 Malaria Data

This data set was extracted from the Masters thesis of Eleanor Gouws (1995, p.98) and was collected by researchers from the Medical Research Council (MRC) in Durban who were involved in its Malaria National Program.

Malaria researchers collected blood samples from suspected malaria sufferers who reported to clinics in the Ubombo district of KwaZulu-Natal during April 1989 and March 1990. The samples were treated with varying concentrations of the antimalarial drug, Halofantrine, and the number of surviving malaria parasites was recorded. Three batches were exposed to each dose of Halofantrine. Due to the technique used to measure the number of parasites (Gouws, 1995, p.51-53), the observations were not integer-valued. They have been rounded off to the nearest whole number in this thesis because they record counts of parasites. The data are presented in Table 2.2.

Table 2.2: Data for malaria parasites exposed to the antimalarial drug Halofantrine.

Drug conc. (μ/l)	Parasitaemia			Mean	Variance
0	4957	5065	5010	5011	2916
1	5193	4897	4816	4969	39384
2	4590	4516	4223	4443	37669
4	3615	3356	3102	3357	65794
8	914	816	657	796	16822
16	49	12	12	24	456
32	23	30	19	24	31
64	33	88	62	61	757

The World Health Organization (WHO) kits used in drug sensitivity include plates that are predosed with a particular drug. The preparation of the plates

is quality controlled and takes place in pharmaceutical laboratories in order to eliminate variation between wells within plates and variation between plates as much as possible (Gouws, 1995, p.48-50). As a consequence of these precautions, a replication effect is not considered for the malaria data of Table 2.2. This decision is in accord with Gouws (1995) who also did not include a replication effect in her analysis of the data.

It is common in the analysis of bioassay data to model the natural mortality of the population, λ. However, in a Wadley's problem setting, the estimate of N incorporates natural mortality since it represents the number of organisms that would have survived treatment with a zero dose of the drug. Thus the estimation of λ is omitted in this study when considering the algae and malaria data.

2.3 The Poisson Model

The aim of this section is to introduce the classical approach to modelling Wadley's problem, which entails modelling N, the unknown number of organisms initially exposed to a drug, with a Poisson distribution (Wadley, 1949).

2.3.1 Setup and Notation

Following Baker, Pierce and Pierce (1980), the control observations and the counts corresponding to non-zero doses of the drug are considered separately. Specifically, let $y_{cj},\ j = 1, \ldots, n_c$, denote an observation from the control group, and let y_{ij} refer to the number of survivors observed for a non-zero concentration d_i of the drug, with $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Associated with a dose d_i of the drug is the probability of death of a parasite p_i and the log-dose $\log d_i = x_i$. The intention is to describe a function of the probability of death as a linear expression in log-dose, i.e. $g(p_i) = \alpha + \beta x_i$, where g is a function of the unknown parameters α and β.

2.3.2 Poisson Distribution for N

Consider a random variable Y which denotes the number of organisms that survive exposure to a drug, where the probability of death for an organism is p and the probability of survival is $1-p$. The survival of one organism is assumed to be independent of the survival of another. As a result, if N, the number of organisms initially treated with the drug, was known, then Y given that $N = n$ would be expected to follow a binomial distribution with parameters n and $(1-p)$, i.e. $Y \mid N = n \sim$Binomial$(n,\ 1-p)$. Wadley (1949) considered modelling the unknown N using a Poisson distribution with parameter τ. Under these circumstances, the marginal probability mass function (p.m.f.) of Y is

$$\begin{aligned}
\Pr(Y=y) &= \sum_{n=y}^{\infty} \Pr(Y=y|N=n)\Pr(N=n) \\
&= \sum_{n=y}^{\infty} \binom{n}{y} (1-p)^y p^{n-y} \frac{e^{-\tau}\tau^n}{n!} \\
&= \frac{[\tau(1-p)]^y e^{-\tau}}{y!} \sum_{n=y}^{\infty} \frac{(\tau p)^{n-y}}{(n-y)!} \\
&= \frac{[\tau(1-p)]^y e^{-\tau}}{y!} \sum_{x=0}^{\infty} \frac{(\tau p)^x}{x!} \\
&= \frac{[\tau(1-p)]^y e^{-\tau(1-p)}}{y!} \ ,
\end{aligned}$$

which indicates that Y follows a Poisson distribution with parameter $\tau(1-p)$, i.e. $Y \sim$Poisson$[\tau(1-p)]$. This result was obtained by Wadley (1949) and is intuitively satisfying.

An alternative derivation of this result can be obtained using probability generating functions (p.g.f.'s). The random variable Y can be thought of as the sum of N Bernoulli trials, $X_1, X_2, \ldots, X_N$ where $X_i \sim$ Binomial$(1, 1-p)$, i.e. $Y = \sum_{i=1}^{N} X_i$. Suppose that N$\sim$ Poisson(τ). Then, since the p.g.f. of X_i is $P_X(s) = p + (1-p)s$ and that of N is $P_N(s) = e^{\tau(1-s)}$, the p.g.f. of Y can be found by invoking a well-known result on compound distributions (Grimmett and Welsh, 1986, p.51)

as

$$\begin{aligned} P_Y(s) &= P_N[P_X(s)] \\ &= e^{\tau[1-p-(1-p)s]} \\ &= e^{\tau(1-p)(1-s)} . \end{aligned}$$

Thus Y is Poisson distributed with parameter $\tau(1-p)$.

2.3.3 Likelihood Function

The likelihood function for the data $y_{cj},\ j = 1, \ldots, n_c$, and $y_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, D$, is algebraically identical to the joint p.m.f. of the random variables Y_{cj} and Y_{ij}. Since the probability of non-natural death of a parasite in the control group is zero, the variable Y_{cj} follows a Poisson distribution with parameter τ and the likelihood and log-likelihood functions are given by

$$L_c = \prod_{j=1}^{n_c} \frac{\tau^{y_{cj}} e^{-\tau}}{y_{cj}!}$$

and

$$\begin{aligned} \ell_c &= \sum_{j=1}^{n_c} y_{cj} \ln \tau - \sum_{j=1}^{n_c} \tau - \sum_{j=1}^{n_c} \ln(y_{cj}!) \\ &= \ln \tau \sum_{j=1}^{n_c} y_{cj} - n_c \tau - \sum_{j=1}^{n_c} \ln(y_{cj}!) \end{aligned}$$

respectively. Observations corresponding to non-zero doses d_i of the drug are a random sample from a Poisson distribution with parameter $\tau(1-p_i),\ i = 1, \ldots, D$. The likelihood and log-likelihood functions for these observations can therefore be written as

$$L_d = \prod_{i=1}^{D} \prod_{j=1}^{n_i} \frac{[\tau(1-p_i)]^{y_{ij}} e^{-\tau(1-p_i)}}{y_{ij}!}$$

and

$$\ell_d = \sum_{i=1}^{D} \sum_{j=1}^{n_i} y_{ij} \ln[\tau(1-p_i)] - \sum_{i=1}^{D} n_i \tau(1-p_i) - \sum_{i=1}^{D} \sum_{j=1}^{n_i} \ln(y_{ij}!)$$

$$= \ln\tau \sum_{i=1}^{D}\sum_{j=1}^{n_i} y_{ij} + \sum_{i=1}^{D}\sum_{j=1}^{n_i} y_{ij}\ln(1-p_i) - \tau\sum_{i=1}^{D} n_i(1-p_i) - \sum_{i=1}^{D}\sum_{j=1}^{n_i}\ln(y_{ij}!)$$

respectively. The log-likelihood for the data, which is denoted by $\ell(\tau, p; \underline{y})$, is then obtained by summing the log-likelihoods for all control observations and all observations arising from treatments with non-zero doses of the drug as

$$\begin{aligned}
\ell(\tau, p; \underline{y}) &= \ell_c + \ell_d \\
&= \ln\tau \sum_{j=1}^{n_c} y_{cj} - n_c\tau - \sum_{j=1}^{n_c}\ln(y_{cj}!) + \ln\tau \sum_{i=1}^{D}\sum_{j=1}^{n_i} y_{ij} + \sum_{i=1}^{D}\sum_{j=1}^{n_i} y_{ij}\ln(1-p_i) \\
&\quad -\tau\sum_{i=1}^{D} n_i(1-p_i) - \sum_{i=1}^{D}\sum_{j=1}^{n_i}\ln(y_{ij}!) \ .
\end{aligned}$$

2.3.4 Modelling the Probabilities

Two drug tolerance distributions were investigated for modelling the binomial probability of death p_i at log-dose x_i, $i = 1, \ldots, D$, namely those based on the logit model and the complementary log-log model. When the logit model is used

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta x_i \ ,$$

which is typically interpreted as the logarithm of the odds ratio. As a result

$$p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}} \ , \text{ for } i = 1, \ldots, D \ .$$

The complementary log-log model can be expressed as

$$\ln[-\ln(1-p_i)] = \alpha + \beta x_i$$

and therefore

$$p_i = 1 - e^{-e^{\alpha+\beta x_i}} \ , \text{ for } i = 1, \ldots, D.$$

The logit function is symmetric in that

$$\ln\left(\frac{p_i}{1-p_i}\right) = -\ln\left(\frac{1-p_i}{p_i}\right)$$

(McCullagh and Nelder, 1989, p.108) and it also gives similar results to the probit function (Dobson, 2002, p.118). The complementary log-log function is an asymmetric function. These two tolerance distributions therefore take various properties of the probability of death into account. The two functions relate the probability of the death of an organism to the linear parameters α and β and are therefore types of link functions. They are not, however, link functions as related to generalized linear models (GLM's) because they do not link p_i linearly with all the parameters in the model, specifically with τ.

It has been shown that the responses $Y_{ij},\ j = 1, \ldots, n_i$ and $i = 1, \ldots, D$, follow a Poisson distribution, which is a member of the exponential family of distributions. To determine whether or not the model Y_{ij} ~Poisson$[\tau(1 - p_i)]$, together with the probabilities, p_i, which are modelled either by a logit or a complementary log-log function is a GLM, consider the expected number of survivors for the ith dose, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. This parameter is given as

$$\begin{aligned} E(Y_{ij}) &= \mu_i \\ &= \frac{\tau}{1 + e^{\alpha+\beta x_i}} \end{aligned}$$

when a logit model is adopted for p_i and as

$$E(Y_{ij}) = \tau e^{-e^{\alpha+\beta x_i}}$$

when p_i is modelled with a complementary log-log function. In order for either model to be a GLM a monotone function of the mean, denoted $g(\mu_i)$, which is linear in all the parameters τ, α and β is required. It would seem that such a function cannot be found. This absence is implied by Smith and Morgan (1989) who linearize the link function by means of a Taylor series expansion in order to overcome the nonlinearity. Therefore there seems to be no way of casting Wadley's problem with the number of parasites surviving modelled by a Poisson$[\tau(1 - p)]$ model and with the probability of death of a parasite modelled by a logit or complementary log-log function, as a GLM. Hence the model is classified as a generalized nonlinear model.

For the remainder of this thesis the Poisson model with parameter $\tau(1-p)$ which models the number of parasites surviving, together with the logit or complementary log-log function adopted for the probability p, will be referred to as the Poisson$[\tau(1-p)]$ model.

2.3.5 Score Function and Information Matrix

Score Function

The score function, denoted by U, is defined as the first derivative of the log-likelihood ℓ with respect to the parameters in the model. Denote the log-likelihood for a single control observation by $\ell_{cj},\ j = 1, \ldots, n_c$, and for a single observation corresponding to a non-zero dose of the drug d_i by ℓ_{ij} where $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Then the scores for a single observation from the control group when the probability of death is modelled with either a logit or complementary log-log function are identical and are given by

$$\begin{aligned} \frac{\partial \ell_{cj}}{\partial \tau} &= \frac{y_{cj}}{\tau} - 1 \ , \\ \frac{\partial \ell_{cj}}{\partial \alpha} &= 0 \ , \\ \frac{\partial \ell_{cj}}{\partial \beta} &= 0 \ . \end{aligned}$$

When the logit function is adopted to model the probability of the death of an exposed organism, the first derivatives of $\ell_{ij},\ j = 1, \ldots, n_i$ and $i = 1, \ldots, D$, with respect to $\tau,\ \alpha$ and β are

$$\begin{aligned} \frac{\partial \ell_{ij}}{\partial \tau} &= \frac{y_{ij}}{\tau} - (1 - p_i) \ , \\ \frac{\partial \ell_{ij}}{\partial \alpha} &= -y_{ij} p_i + \tau p_i (1 - p_i) \ , \\ \frac{\partial \ell_{ij}}{\partial \beta} &= -y_{ij} x_i p_i + \tau x_i p_i (1 - p_i) \ . \end{aligned}$$

The score functions for the data are then found by summing the scores for the individual observations over all observations. Therefore when a logit model is

used to model p_i the score functions for the data are

$$\begin{aligned}
\frac{\partial \ell(\tau, p; \underline{y})}{\partial \tau} &= \sum_{j=1}^{n_c} \frac{y_{cj}}{\tau} + \sum_{i=1}^{D} \sum_{j=1}^{n_i} \frac{y_{ij}}{\tau} - n_c - \sum_{i=1}^{D} n_i(1-p_i) \ , \\
\frac{\partial \ell(\tau, p; \underline{y})}{\partial \alpha} &= -\sum_{i=1}^{D} \sum_{j=1}^{n_i} y_{ij} p_i + \tau \sum_{i=1}^{D} n_i p_i (1-p_i) \ , \\
\frac{\partial \ell(\tau, p; \underline{y})}{\partial \beta} &= -\sum_{i=1}^{D} \sum_{j=1}^{n_i} y_{ij} x_i p_i + \tau \sum_{i=1}^{D} n_i x_i p_i (1-p_i) \ .
\end{aligned}$$

For the complementary log-log distribution with $p_i = 1 - e^{-e^{\alpha+\beta x_i}}$ the first derivatives of $\ell_{ij},\ j = 1, \ldots, n_i$ and $i = 1, \ldots, D$, with respect to $\tau,\ \alpha$ and β are given by

$$\begin{aligned}
\frac{\partial \ell_{ij}}{\partial \tau} &= \frac{y_{ij}}{\tau} - (1-p_i) \ , \\
\frac{\partial \ell_{ij}}{\partial \alpha} &= y_{ij} \ln(1-p_i) - \tau(1-p_i)\ln(1-p_i) \ , \\
\frac{\partial \ell_{ij}}{\partial \beta} &= y_{ij} x_i \ln(1-p_i) + \tau x_i (1-p_i)\ln(1-p_i) \ .
\end{aligned}$$

The scores for the data are therefore

$$\begin{aligned}
\frac{\partial \ell(\tau, p; \underline{y})}{\partial \tau} &= \sum_{j=1}^{n_c} \frac{y_{cj}}{\tau} + \sum_{i=1}^{D} \sum_{j=1}^{n_i} \frac{y_{ij}}{\tau} - n_c - \sum_{i=1}^{D} n_i(1-p_i) \ , \\
\frac{\partial \ell(\tau, p; \underline{y})}{\partial \alpha} &= \sum_{i=1}^{D} \sum_{j=1}^{n_i} y_{ij} \ln(1-p_i) - \tau \sum_{i=1}^{D} n_i (1-p_i)\ln(1-p_i) \ , \\
\frac{\partial \ell(\tau, p; \underline{y})}{\partial \beta} &= \sum_{i=1}^{D} \sum_{j=1}^{n_i} y_{ij} x_i \ln(1-p_i) + \tau \sum_{i=1}^{D} n_i x_i (1-p_i)\ln(1-p_i) \ .
\end{aligned}$$

Note that for both the logit and complementary log-log link functions, the maximum likelihood estimates for the parameters $\tau,\ \alpha$ and β are solutions to the above equations. In order to verify that the solutions found are maxima, the matrix of second derivatives of the log-likelihood function with respect to the parameters can be computed. If the matrix is negative definite then the solutions are indeed the maximum likelihood estimates of the parameters (Dobson, 2008, p.12).

Information Matrix

The information matrix is defined to be minus the matrix of expected values of the second-order derivatives of the log-likelihood function, ℓ, with respect to the parameters in the model. The inverse of the information matrix corresponds to the asymptotic covariance matrix of parameter estimates (Dobson, 2002, p.73). In particular the square roots of the diagonal elements of this inverse are the approximate standard errors of the maximum likelihood estimates.

For the parameters $\tau,\ \alpha$ and β in the present example, the information matrix is found by invoking the expression

$$\mathfrak{I} = -E \begin{bmatrix} \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \tau^2} & \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \tau \partial \alpha} & \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \tau \partial \beta} \\ \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \tau \partial \alpha} & \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \alpha^2} & \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \alpha \partial \beta} \\ \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \tau \partial \beta} & \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \alpha \partial \beta} & \dfrac{\partial^2 \ell(\tau,p;Y)}{\partial \beta^2} \end{bmatrix} .$$

Denote the information for a single observation from the control category by $\mathfrak{I}_{cj},\ j = 1,\ldots,n_c$, and for a single observed count for dose d_i by $\mathfrak{I}_{ij}$, where $j = 1,\ldots,n_i$ and $i = 1,\ldots,D$. Then the information matrix for a single observation from the control group is given by

$$\mathfrak{I}_{cj} = \begin{bmatrix} \dfrac{1}{\tau} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

For an observation $y_{ij},\ j = 1,\ldots,n_i$ and $i = 1,\ldots,D$, arising from the administration of dose d_i of the drug and when a logit function is used to model p_i, the information matrix is

$$\mathfrak{I}_{ij} = \begin{bmatrix} \dfrac{1-p_i}{\tau} & -p_i(1-p_i) & -x_i p_i(1-p_i) \\ -p_i(1-p_i) & \tau p_i^2(1-p_i) & \tau x_i p_i^2(1-p_i) \\ -x_i p_i(1-p_i) & \tau x_i p_i^2(1-p_i) & \tau x_i^2 p_i^2(1-p_i) \end{bmatrix} .$$

Due to the independence of the observations the information matrix for the parameters τ, α and β is found by summing the information matrix for each observation over all of the observations. The information matrix for τ, α and β where p_i is modelled with a logit function is therefore

$$\mathfrak{I} = \begin{bmatrix} \frac{1}{\tau}\left[n_c + \sum_{i=1}^{D} n_i(1-p_i)\right] & -\sum_{i=1}^{D} n_i p_i(1-p_i) & -\sum_{i=1}^{D} n_i x_i p_i(1-p_i) \\ -\sum_{i=1}^{D} n_i p_i(1-p_i) & \tau\sum_{i=1}^{D} n_i p_i^2(1-p_i) & \tau\sum_{i=1}^{D} n_i x_i p_i^2(1-p_i) \\ -\sum_{i=1}^{D} n_i x_i p_i(1-p_i) & \tau\sum_{i=1}^{D} n_i x_i p_i^2(1-p_i) & \tau\sum_{i=1}^{D} n_i x_i^2 p_i^2(1-p_i) \end{bmatrix}, \tag{2.1}$$

which is in accord with that given by Haines, Clarke, Gouws and Rosenberger (2001).

Consider an observation y_{ij}, $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$, resulting from treatment with a dose d_i of the drug on a group of parasites and where p_i is modelled by a complementary log-log function. The information for this observation is

$$\mathfrak{I}_{ij} = \begin{bmatrix} \frac{1-p_i}{\tau} & (1-p_i)\ln(1-p_i) & x_i(1-p_i)\ln(1-p_i) \\ (1-p_i)\ln(1-p_i) & \tau(1-p_i)\ln^2(1-p_i) & \tau x_i(1-p_i)\ln^2(1-p_i) \\ x_i(1-p_i)\ln(1-p_i) & \tau x_i(1-p_i)\ln^2(1-p_i) & \tau x_i^2(1-p_i)\ln^2(1-p_i) \end{bmatrix}.$$

The information matrix for the parameters of the model in this instance is therefore

$$\mathfrak{I} = \begin{bmatrix} \frac{1}{\tau}\left[n_c + \sum_{i=1}^{D} n_i(1-p_i)\right] & \sum_{i=1}^{D} n_i(1-p_i)\ln(1-p_i) & \sum_{i=1}^{D} n_i x_i(1-p_i)\ln(1-p_i) \\ \sum_{i=1}^{D} n_i(1-p_i)\ln(1-p_i) & \tau\sum_{i=1}^{D} n_i(1-p_i)\ln^2(1-p_i) & \tau\sum_{i=1}^{D} n_i x_i(1-p_i)\ln^2(1-p_i) \\ \sum_{i=1}^{D} n_i x_i(1-p_i)\ln(1-p_i) & \tau\sum_{i=1}^{D} n_i x_i(1-p_i)\ln^2(1-p_i) & \tau\sum_{i=1}^{D} n_i x_i^2(1-p_i)\ln^2(1-p_i) \end{bmatrix}$$

.

(2.2)

Models for Separate Doses

A model that fits a separate Poisson distribution to each dose is also considered in the present study and is conveniently referred to as the separate dose model. Specifically the model assumes that observations for the control Y_{cj} are Poisson(μ_c), $j = 1, \ldots, n_c$, and those for dose d_i, namely Y_{ij}, are Poisson(μ_i), $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Note immediately that the maximum likelihood estimates for the means μ_c and μ_i are the sample means $\overline{y}_c$ and $\overline{y}_i$ respectively, where

$$\overline{y}_c = \frac{\sum_{j=1}^{n_c} y_{cj}}{n_c}$$

and

$$\overline{y}_i = \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} .$$

Although this model is not sophisticated, it is a useful base model against which comparisons can be made.

2.3.6 Estimation

Since the likelihood function captures how likely the parameters are for the given data, a meaningful and commonly used approach to estimation is through the maximization of the likelihood function. Note that the logarithm function is monotone increasing and thus maximizing the log-likelihood is equivalent to maximizing the likelihood with respect to the parameters τ, α and β. Three methods of obtaining the maximum likelihood estimates of the parameters are introduced here.

Wadley's Method

Wadley (1949) considered a probit tolerance distribution and estimated τ, α and β for the Poisson dose-response model with parameter $\tau(1-p)$ by equating the score functions to zero and solving them simultaneously, i.e. by finding solutions to

$$\begin{aligned} \frac{\partial \ell(\tau, p; \underline{\underline{y}})}{\partial \tau} &= 0 \ , \\ \frac{\partial \ell(\tau, p; \underline{\underline{y}})}{\partial \alpha} &= 0 \ , \\ \frac{\partial \ell(\tau, p; \underline{\underline{y}})}{\partial \beta} &= 0 \ . \end{aligned}$$

Explicit solutions to these equations cannot be obtained but iterative techniques can be employed to approximate the solutions, as recommended by Finney (1971, p.53). In particular, Wadley (1949) obtained initial estimates from rough graphical or arithmetic estimation and then adopted the Taylor-Maclaurin expansion to improve them until adjustments to estimates on consecutive iterations became appropriately small. This method of estimation was designed for calculation by hand and is old-fashioned. It is therefore not considered in this thesis.

Linearizing the Model and Composite Link Functions

A procedure, WADLEY, which makes use of methods given in a GLIM macro by Smith and Morgan (1989), is available in Genstat. The procedure incorporates composite link functions, one for control observations and the other for observations corresponding to the non-zero doses of the drug, and involves linearizing the link function with respect to the parameters using a Taylor series expansion.

The WADLEY procedure in Genstat was used in the present study but was found to be problematic in a number of ways. For example if Genstat is required to find the log of zero, as is the case for zero doses, it simply excludes the associated observation from the model without warning, thereby decreasing the degrees of freedom for the model. As a result a small positive number needs to be added to

the zero doses in order to retain them in the model. The procedure also requires the user to differentiate between the controls and non-zero doses in order for it to make use of composite link functions, and, as indicated already, the algorithm used involves linearizing the link function with respect to the parameters rather than maximizing the likelihood directly. Further, this procedure acts rather like a "black box" and it is therefore difficult to identify and isolate problems without understanding and modifying the extensive Genstat code.

Constrained Nonlinear Optimization

The log-likelihood function $\ell(\tau, p; \underline{y})$ is a nonlinear function of the parameters τ, α and β. Thus a nonlinear optimization routine can be used to maximize $\ell(\tau, p; \underline{y})$. Such procedures are iterative and require the input of starting values. The maximum observed count provides a reasonable starting value for the parameter τ in the Poisson$[\tau(1-p)]$ model and starting values for α and β can be obtained from fitting an appropriate GLM based on the binomial distribution to the data, with the total number of organisms fixed at the maximum observed count. The constraints $\tau > 0$ and $\beta > 0$ are implicit in the model and must be introduced into the optimization routine. Note that the parameter β is constrained in this manner because the probability of death of a parasite is expected to increase with increasing doses of the drug.

Due to the problems encountered in using the WADLEY procedure in Genstat, the above method of direct optimization was used almost exclusively in this thesis. Specifically the routines were programmed using the language Gauss and incorporate the Gauss procedure OPTMUM.

Inference

The general approach to obtaining the standard errors of the maximum likelihood estimates of the parameters is to substitute these estimates into the inverse of the information matrix and then to take the square roots of the resulting diagonal

elements (Azzalini, 1996, p.91). The approximate standard errors for the parameter estimates $\hat{\tau}$, $\hat{\alpha}$ and $\hat{\beta}$ can be calculated in Gauss, using the inverse of the information matrix specified in (2.1) when the probabilities p_i are modelled with a logit function, and the inverse of the information matrix in expression (2.2) when the complementary log-log function is adopted for p_i, $i = 1, \ldots, D$.

Goodness of Fit

For the maximal model used in the calculation of the deviance, consider $Y_{ij} \sim$Poisson(λ_{ij}), $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Then the maximum likelihood estimate for λ_{ij} is y_{ij}, i.e. $\hat{\lambda}_{ij} = y_{ij}$, and therefore the associated value of the log-likelihood is

$$\begin{aligned} \ell(\underline{\hat{\lambda}}, \underline{y}) &= \sum_{j=1}^{n_c} \{-y_{cj} + y_{cj} \ln y_{cj} - \ln y_{cj}!\} + \\ &\quad + \sum_{i=1}^{D} \sum_{j=1}^{n_i} \{-y_{ij} + y_{ij} \ln y_{ij} - \ln y_{ij}!\} \end{aligned}$$

where $\underline{\hat{\lambda}}$ and $\underline{y}$ are the vectors of parameter estimates and observed responses respectively. Then the deviance for the Poisson model with means $\tau(1 - p_i)$ is

$$\begin{aligned} D &= 2[\ell(\underline{\hat{\lambda}}; \underline{y}) - \ell(\hat{\tau}, \hat{p}; \underline{y})] \\ &= 2 \left\{ \sum_{j=1}^{n_c} (-y_{cj} + y_{cj} \ln y_{cj} - \ln y_{cj}!) + \sum_{i=1}^{D} \sum_{j=1}^{n_i} (-y_{ij} + y_{ij} \ln y_{ij} - \ln y_{ij}!) \right. \\ &\quad \left. - \left[\sum_{j=1}^{n_c} (y_{cj} \ln \tau - \ln y_{cj}! - \tau) + \sum_{i=1}^{D} \sum_{j=1}^{n_i} (y_{ij} \ln \tau(1 - p_i) - \tau(1 - p_i) - \ln y_{ij}!) \right] \right\} . \end{aligned}$$

The aim of introducing the deviance is to determine the adequacy of the Poisson$[\tau(1 - p)]$ model for the data. Since the deviance is approximately chi-squared distributed with degrees of freedom equal to the number of observations minus the number of parameters in the model (Dobson, 2002, p.76), the fit of a model can be assessed by comparing the deviance with its expected value under the associated χ^2 distribution. In general, a large value of the deviance D indicates that a model does not describe the data well. Since the mean of the χ^2 distribution

is its degrees of freedom, it is often convenient to assess the adequacy of a model by comparing the deviance with the degrees of freedom. If the deviance is much greater than the degrees of freedom, the model does not fit the data adequately. Alternatively the mean deviance, which is the ratio of the deviance to its degrees of freedom, can be computed and if a model is satisfactory its mean deviance would be close to one (Collett, 2003, p.70). However, situations arise where the mean deviance for a model is large but plots of the residuals suggest that the model provides a reasonable fit for the data. In these instances the data are said to be overdispersed (Collett, 2003, p.198). Thus the deviance provides a measure of this overdispersion.

The fit of two nested models which have q_1 and q_2 degrees of freedom with $q_1 > q_2$ can be compared by considering the difference in their deviances (Dobson, 2002, p.80-81). If both of the models are suitable for the data the difference in their deviance follows an approximate $\chi^2_{q_1-q_2}$ distribution. A non-significant difference in the deviances is interpreted as indicating that the model with q_2 degrees of freedom should be selected for the data for the sake of parsimony.

Morgan (1992, p.236) gives details regarding a Monte Carlo approach for model checking which was described by Pack (1986) in his unpublished Ph.D. thesis. The method is based on the assumption that the fitted model is the true model for the data. The parameter estimates obtained from this fitted model are used to simulate a number of data sets. The model under consideration is then fitted to each data set and the log-likelihood from the fit is recorded in order to obtain a range of log-likelihoods for the simulated data. A p-value is obtained from this range, as with a randomization test, and a model with a small p-value is considered to provide a poor fit to the data.

2.3.7 Examples

The Poisson$[\tau(1-p)]$ model was fitted to the algae and malaria data sets that were described earlier in Section 2.2. The parameters were estimated using both

the nonlinear optimization routine from Gauss and the Genstat WADLEY procedure, as explained in Section 2.3.6. The results from the Genstat procedure were included for comparison purposes.

Algae Data

The results from fitting the Poisson$[\tau(1-p)]$ models to the algae data in Gauss and Genstat are presented in Table 2.3. The two programs yielded very similar results but Gauss was deemed preferable because it allows for more user control than the Genstat procedure. Details regarding the goodness of fit of the saturated model, denoted $\{\lambda_{ij}\}$, the base model with separate means for each dose, denoted $\{\mu_i\}$, and the Poisson$[\tau(1-p)]$ model with the logit and complementary log-log tolerance distributions used for the probability of death are reported in Tables 2.4 and 2.5.

Under the assumption that the null hypothesis is true, the p-value is the probability of a test statistic being at least as extreme as that calculated for the observed data (Hogg, McKean and Craig, 2005, p.276). The p-values based on the χ^2 distribution for deviances indicate that the Poisson model with a logit link function provides a better fit to the algae data than the model with a complementary log-log link function. Indeed the p-value of 0.006 indicates that the Poisson$[\tau(1-p)]$ model with a complementary log-log tolerance distribution is not suitable for modelling the algae data. These p-values are expected to be similar to the probabilities derived from the Monte Carlo output. For example for the model $\{\mu_i\}$ the Monte Carlo probability for the model for separate doses is $\frac{1000-881}{1000} = \frac{119}{1000} = 0.119$ which can be compared with a p-value of 0.104 associated with the χ^2 value of 28.236, with 20 degrees of freedom.

Plots of observed and expected responses versus log-dose for the Poisson $[\tau(1-p)]$ model with logit and complementary log-log tolerance distributions are shown in Figures 2.1(a) and 2.1(b) respectively. For this data set, the curve of expected responses reaches a maximum asymptotically at -5 and therefore the

Table 2.3: Results from fitting the Poisson$[\tau(1-p)]$ model to the algae data, where Cll refers to the complementary log-log function.

Function for p	Parameter	Gauss		Genstat	
		Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Logit	$\hat{\tau}$	224.852	6.593	224.820	6.450
	$\hat{\alpha}$	-0.902	0.126	-0.907	0.123
	$\hat{\beta}$	1.415	0.089	1.419	0.088
Cll	$\hat{\tau}$	226.519	6.647	226.440	6.530
	$\hat{\alpha}$	-0.932	0.090	-0.936	0.089
	$\hat{\beta}$	0.872	0.058	0.875	0.057

Table 2.4: Details of the deviance of the four models fitted to the algae data, where τ, α, β (logit) and τ, α, β (cll) refer to the Poisson$[\tau(1-p)]$ model with logit and complementary log-log tolerance distributions respectively.

Model	Maximum log-likelihood	Number of parameters	Deviance	p-value
$\{\lambda_{ij}\}$	-81.208	25	0	-
$\{\mu_i\}$	-95.326	5	28.236	0.104
τ, α, β (logit)	-97.776	3	33.136	0.060
τ, α, β (cll)	-102.394	3	42.303	0.006

Table 2.5: Details of the Monte Carlo results of the four models fitted to the algae data, where τ, α, β (logit) and τ, α, β (cll) refer to the Poisson$[\tau(1-p)]$ model with logit and complementary log-log tolerance distributions respectively.

Model	Maximum log-likelihood	Monte Carlo range	Rank of max log-lik. in range	Probability
$\{\lambda_{ij}\}$	-81.208	-	-	-
$\{\mu_i\}$	-95.326	(-84.104; -102.292)	881	0.119
τ, α, β (logit)	-97.776	(-103.201; -85.165)	954	0.046
τ, α, β (cll)	-102.394	(-107.412; -84.638)	985	0.015

x-coordinate at -5 can be taken as the asymptote. The graphs highlight the suitability of the Poisson $[\tau(1-p)]$ model with a logit function for this data set. In particular, this model seems to provide a better fit to the observations corresponding to a log-dose of approximately zero. The two models seem to provide similar fits to the other observed responses. These findings are confirmed by the residual plots which are presented in Figures 2.2(a) and 2.2(b).

As an aside, the LD_{50}, which is the dose required to kill 50% of the organisms, is estimated as $10^{-\hat{\alpha}/\hat{\beta}}$ for the model with a logit tolerance distribution and with log-dose as a predictor variable. A 95% Wald confidence interval for the LD_{50} can be calculated as

$$\hat{LD}_{50} \pm 1.96 \; se(\hat{LD}_{50})$$

where $se(\hat{LD}_{50})$ is the standard error of this estimate (Kutner, Nachtsheim, Neter and Li, 2005, p.579). If dose is used as a predictor variable instead of log-dose, the LD_{50} is estimated as $-\hat{\alpha}/\hat{\beta}$ and its standard error is approximately

$$\sqrt{\frac{\text{Var}(\hat{\alpha}) - 2\frac{\hat{\alpha}}{\hat{\beta}}\text{Cov}(\hat{\alpha},\hat{\beta}) + \left(\frac{\hat{\alpha}}{\hat{\beta}}\right)^2 \text{Var}(\hat{\beta})}{\hat{\beta}^2}}$$

(Collett, 2003, p.108). A 95% Wald interval, denoted $(LD_1; LD_2)$, can be computed for $-\alpha/\beta$. The transformation $(10^{LD_1}; 10^{LD_2})$ then yields a 95% Wald interval for the LD_{50} when a logit tolerance distribution is used with log-dose as a predictor variable. The algae data has an estimated LD_{50} of 4.335 μg/ml when a logit model is used for the probability of death of an organism and the associated 95% Wald confidence interval is (3.327;5.649). When a complementary log-log function models the probability of death of an organism, the LD_{50} is estimated as $10^{\frac{\ln(\ln 2)-\hat{\alpha}}{\hat{\beta}}}$. In this instance, it is estimated as 4.449 μg/ml, which is close to the estimate obtained when a logit tolerance distribution is used. The 95% Wald interval for the LD_{50} with a complementary log-log link function can be obtained in a similar manner as that for the model with a logit link function. For the algae data this interval is (0;9.906).

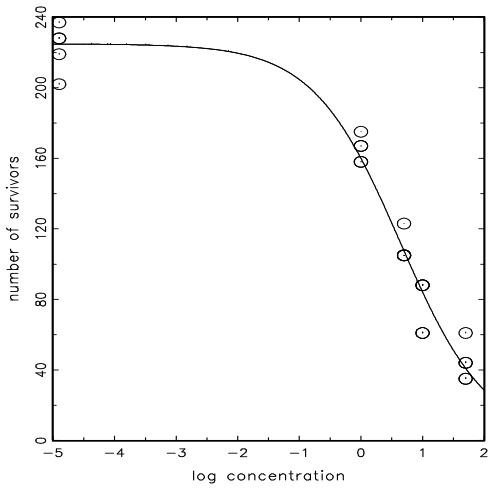


Figure 2.1: (a)

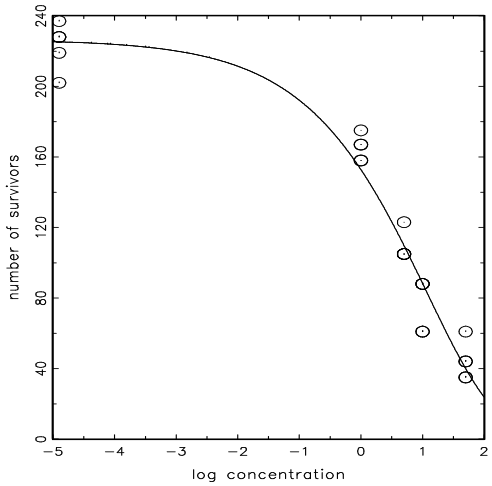


Figure 2.1: (b)

Figure 2.1: Plot of observed counts ($\odot$) and fitted responses (-) against log-dose for (a) the logit model and (b) the complementary log-log model, fitted to the algae data.

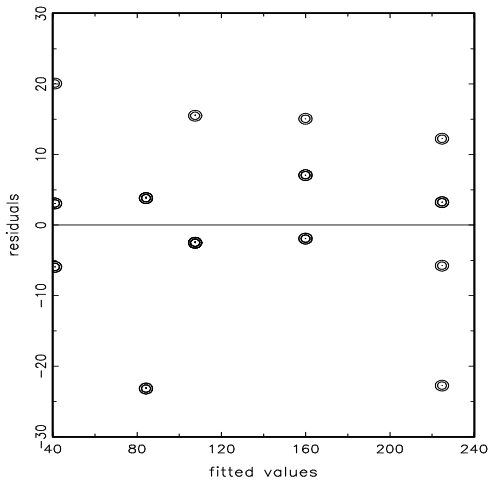


Figure 2.2: (a)

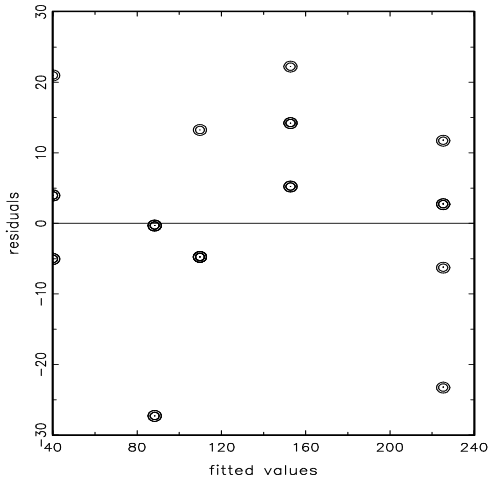


Figure 2.2: (b)

Figure 2.2: Graph of residuals versus fitted values of the Poisson$[\tau(1-p)]$ model for the algae data with (a) the logit function and (b) the complementary log-log function.

Malaria Data

Details regarding the parameter estimates obtained from fitting the Poisson $[\tau(1-p)]$ model to the malaria data are given in Table 2.6. Both of the estimation procedures gave similar parameter estimates and the Genstat output therefore confirms the results obtained from the Gauss program.

Table 2.6: Results from fitting the Poisson$[\tau(1-p)]$ model to the malaria data, where Cll refers to the complementary log-log function.

Function for p	Parameter	Gauss		Genstat	
		Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Logit	$\hat{\tau}$	5011.220	26.919	5012.300	26.700
	$\hat{\alpha}$	-4.523	0.070	-4.535	0.071
	$\hat{\beta}$	6.748	0.079	6.757	0.079
Cll	$\hat{\tau}$	5860.941	39.950	5721.600[1]	41.000
	$\hat{\alpha}$	-1.534	0.026	-1.654	0.024
	$\hat{\beta}$	2.173	0.023	2.280	0.021

[1]The Genstat output contained the message: convergence failure.

Table 2.7: Details of the four models fitted to the malaria data, where τ, α, β (logit) and τ, α, β (cll) refer to the Poisson$[\tau(1-p)]$ model with logit and complementary log-log tolerance distributions respectively.

Model	Maximum log-likelihood	Number of parameters	Monte Carlo range	Deviance from saturated model
$\{\lambda_{ij}\}$	-97.631	24	-	0
$\{\mu_i\}$	-187.338	8	(-115.364; -99.658)	179.414
τ, α, β (logit)	-870.023	3	(-120.860; -98.810)	1544.780
τ, α, β (cll)	-2733.149	3	(-429.792; 299.006)	5721.037

Details of the fits of the saturated model, the separate dose model and the

Poisson$[\tau(1-p)]$ model with both the logit and complementary log-log functions adopted for the probabilities of death are presented in Table 2.7.

The large values of the deviances indicate that the models either provide poor fits compared with the saturated model or that there is significant overdispersion. Specifically, the model for separate doses has a deviance of 179.414, which is highly significant when based on a χ^2 distribution with 16 degrees of freedom. The Poisson models with logit and complementary log-log tolerance distributions have respective deviances of 1544.780 and 5721.037, which are significantly higher than the mean of the χ^2 distribution with 21 degrees of freedom. Furthermore, the deviances indicate that, of the Poisson$[\tau(1-p)]$ models, a better fit is obtained when a logit function is used to model the probability of the death of a parasite. These findings are confirmed by the Monte Carlo results.

Plots of observed and fitted responses against log-dose for the Poisson$[\tau(1-p)]$ model with logit and complementary log-log functions for the probability p are shown in Figures 2.3(a) and 2.3(b) respectively. Since the curve reaches a maximum asymptotically at a log-dose of -2, the y-coordinate at -2 in these diagrams is taken as the asymptote. The graphs indicate that the Poisson$[\tau(1-p)]$ model, with a logit distribution adopted for modelling the probability p, provides a reasonable fit to the data and that the complementary log-log tolerance distribution does not model the probability of death well. The respective residual plots for the logit and the complementary log-log models are shown in Figures 2.4(a) and 2.4(b) and confirm these observations. Thus it can be concluded that the Poisson$[\tau(1-p)]$ model with a logit tolerance distribution fits the data reasonably well, but that the data are significantly overdispersed.

The estimate of the LD_{50} for the malaria data, with a logit model used for the probability of death of a parasite, is 4.681 μmol/l and the 95% Wald interval is (4.594; 4.763). When a complementary log-log tolerance distribution is used, the LD_{50} is estimated as 3.446 μmol/l and the 95% Wald interval is (0;6.233). In this case the estimate of the LD_{50} using a logit tolerance distribution is very different

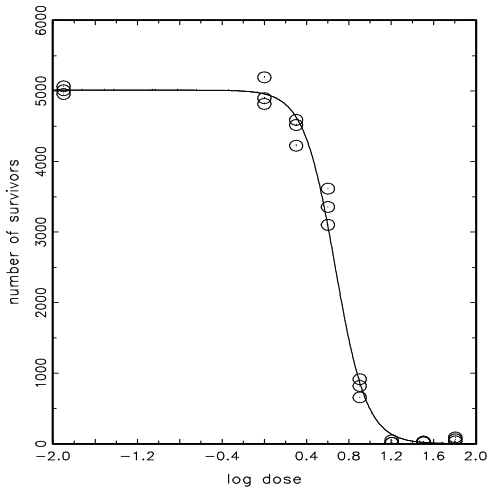


Figure 2.3: (a)

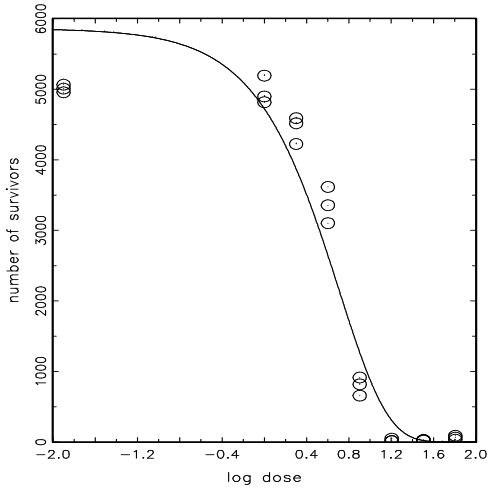


Figure 2.3: (b)

Figure 2.3: Plot of observed counts ($\odot$) and fitted responses (-) against log-dose for (a) the logit model and (b) the complementary log-log model, fitted to the malaria data.

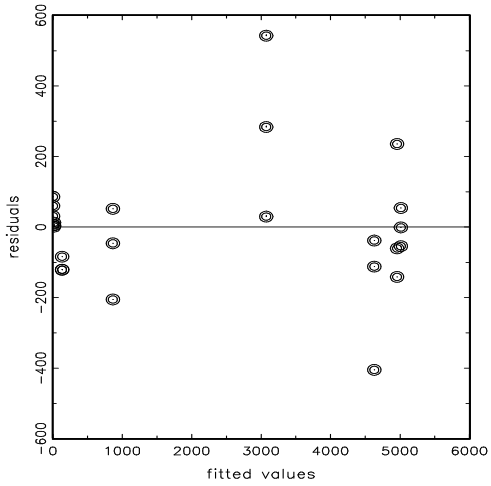


Figure 2.4: (a)

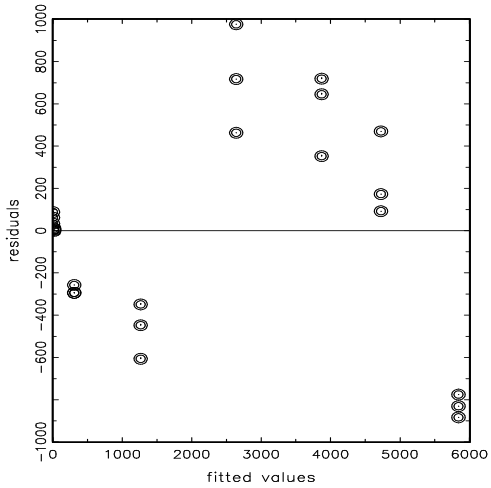


Figure 2.4: (b)

Figure 2.4: Graph of residuals versus fitted values of the Poisson$[\tau(1-p)]$ model for the malaria data with (a) the logit function and (b) the complementary log-log function.

from the estimate obtained when a complementary log-log tolerance distribution is used.

2.4 Conclusions

The Poisson$[\tau(1-p)]$ model is useful for modelling data in a Wadley's problem setting and the residual plots shown in Figures 2.2(a) and 2.4(a) for the algae and malaria data respectively indicate that the distribution provides a reasonable fit to the data when a logit function models the probability of death of an organism. However, values of the deviance of this model compared with the maximal model suggest that it provides a poor fit for the data sets. It can thus be concluded that the apparent lack of fit of this model is due to the presence of overdispersion in the data. As discussed in the examples, the Monte Carlo method of assessing the fit of a model does not give any more information than the other model checking techniques and as a result it is not used in subsequent chapters.

The remainder of this thesis was inspired by the algae and malaria data sets. In particular, the study focuses on developing a theoretical framework that can accommodate overdispersion in a Wadley's problem setting. The subsequent chapters include the logit rather than the complementary log-log tolerance distribution because the logit function is more convenient to work with and because it was found to be more suitable for the data sets included in this study.

Chapter 3

The Beta-Binomial Model

3.1 Introduction

Overdispersion in binomial-type data can be caused by variation in the response probabilities. Skellam (1948) first suggested modelling the variability in the binomial parameter p with a beta distribution, thereby deriving the beta-binomial model.

The beta-binomial distribution has been widely used to model overdispersion. Although Morgan (1992, p.237-254) provides an excellent guide to model fitting and analysis, there do not seem to be many good accounts of the properties of the beta-binomial model. The purpose of this chapter, therefore, is to provide such an account and to adapt the model crudely for a Wadley's problem scenario. This model serves as a prelude to the beta-Poisson model, which is more immediately relevant to Wadley's problem and follows in Chapter 4.

The beta-binomial model is described and its likelihood recorded in Section 3.2. Section 3.3 contains details regarding the score functions and the observed Fisher information matrix, while estimation of the parameters and goodness of fit are discussed in Section 3.4. The results from fitting the model to the algae and malaria data sets are presented in Section 3.5. Section 3.6 contains some concluding remarks.

3.2 The Model

3.2.1 The Binomial Model

In the context of Wadley's problem, the number of organisms initially exposed to a drug, n, is unknown. Therefore the observed number of surviving parasites follows a binomial distribution with unknown n and probability of survival $1-p$, i.e. Y$\sim$Binomial$(n, 1-p)$, where n is unknown. A crude estimate of n can be obtained by using the mean of an untreated (control) sample, $\overline{y}_c$, where

$$\overline{y}_c = \frac{\sum_{j=1}^{n_c} y_{cj}}{n_c},\ j = 1, \ldots, n_c\ .$$

This estimate proves to be problematic if there is at least one observation corresponding to a non-zero dose of the drug that exceeds it. In such cases the maximum observation, denoted by y_{max}, is used as an estimate for n (Olkin, Petkau and Zidek, 1981). Once the number n has been estimated in this ad hoc manner, it can be regarded as known and the binomial model can then be fitted to the data in the usual way. This method of obtaining an estimate for n will lead to further approximations in estimation and particularly in inference.

3.2.2 The Beta-Binomial Model

Consider now a random variable, Y, that corresponds to the number of parasites surviving exposure to a particular dose of a drug and that follows a Binomial $(n, 1-p)$ distribution, where n has been estimated using the methods described in Section 3.2.1. In order to accommodate possible overdispersion in the data, suppose that the probability of death of a parasite, p, follows a beta distribution with parameters a and b. Then the marginal p.m.f. for Y is

$$\Pr(Y = y) \ = \ \int_0^1 \Pr(Y = y|P = p)\Pr(P = p)dp$$

$$
\begin{aligned}
&= \int_0^1 \binom{n}{y} (1-p)^y p^{n-y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} dp \\
&= \int_0^1 \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{n-y+a-1}(1-p)^{b+y-1} dp \\
&= \binom{n}{y} \frac{\Gamma(b+y)\Gamma(n+a-y)}{\Gamma(a+b+n)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \\
&= \binom{n}{y} \frac{B(b+y, n+a-y)}{B(a,b)} \ , \ y = 0, 1, \ldots, n \ ,
\end{aligned}
$$

where $B(a,b)$ refers to the beta function. Thus Y follows a beta-binomial distribution (Williams, 1975) which is unimodal when $a, b > 1$ (Collett, 2003, p.211). Note that a beta distribution which has parameters a and b both equal to 1 corresponds to the uniform distribution on the interval (0,1). Thus when this distribution is used to model the probability of death of a parasite, the resulting distribution for Y is uniform on the set $\{0, 1, \ldots, n\}$, i.e. all values $0, 1, \ldots, n$ are equally likely.

The mean and variance of Y can be obtained using well-known conditioning arguments mentioned in texts such as Ross (2002, p.152) as

$$
\begin{aligned}
\mathrm{E}(Y) &= \mathrm{E}_p[\mathrm{E}(Y|P)] \\
&= \mathrm{E}_p(nP) \\
&= n\frac{a}{a+b}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{Var}(Y) &= \mathrm{E}_p[\mathrm{Var}(Y|P)] + \mathrm{Var}_p[\mathrm{E}(Y|P)] \\
&= \mathrm{E}_p[nP(1-P)] + \mathrm{Var}_p(nP) \\
&= n\mathrm{E}_p(P) - n\mathrm{E}_p(P^2) + n^2\mathrm{Var}_p(P) \\
&= \frac{nab}{(a+b)^2}\left\{1 + \frac{(n-1)}{(a+b+1)}\right\} \ ,
\end{aligned}
$$

where E_p represents the expectation with respect to p and Var_p the variance with respect to p (Griffiths, 1973).

A frequently used reparametrization of this model is in terms of the parameters π and θ where

$$\pi = \frac{a}{a+b}$$

and

$$\theta = \frac{1}{a+b} .$$

The new parametrization is meaningful because π is the expected value of p and the parameter θ determines the shape of the distribution when π is known (Williams, 1975). It is also advantageous in terms of the stability and interpretability of the parameters for dose-response models (Morgan, 1992, p.240). Under the parametrization based on π and θ, the p.m.f. of Y is given by

$$\Pr(Y=y) = \frac{\binom{n}{y}\prod_{r=0}^{y-1}(1-\pi+r\theta)\prod_{r=0}^{n-y-1}(\pi+r\theta)}{\prod_{r=0}^{n-1}(1+r\theta)} .$$

This follows from the fact that for example

$$\begin{aligned}\frac{\Gamma(b+y)}{\Gamma(b)} &= (b+y-1)\ldots(b+1)b \\ &= \prod_{r=0}^{y-1}(b+r) .\end{aligned}$$

The mean and variance of Y are then written respectively as

$$\begin{aligned}\mathrm{E}(Y) &= n\pi \\ \mathrm{Var}(Y) &= n\pi(1-\pi)\left\{1+\frac{\theta}{1+\theta}(n-1)\right\} .\end{aligned}$$

The variance of Y is therefore the usual binomial variance multiplied by the dispersion parameter, $\{1+\frac{\theta(n-1)}{1+\theta}\}$, which accommodates over- or underdispersion in the data. Note that $\mathrm{Var}(Y)$ approaches the binomial variance, $n\pi(1-\pi)$ as θ approaches 0. The beta-binomial distribution with parameters π and θ is referred to as $\mathrm{BetaBin}(n,\pi,\theta)$ for the remainder of this thesis.

3.2.3 Likelihood Function

Since the control observations are used to give a single estimate of the number of parasites, n, initially exposed to the drug and in any case provide no information regarding drug sensitivity, they are excluded from the data set for model fitting. As a result, the likelihood function is only constructed for the observations y_{ij} corresponding to non-zero doses of the drug, $j = 1, \ldots, n_i$ and d_i, $i = 1, \ldots, D$. Associated with each dose of the drug is the log-dose x_i and the probability of death of a parasite, p_i, which follows a beta distribution with parameters defined by π_i and θ_i. Since n has been estimated, it can be taken as a fixed value and the likelihood and log-likelihood functions for the data can thus be written as

$$L(n, \pi, \theta; \underline{y}) = \prod_{i=1}^{D} \prod_{j=1}^{n_i} \left\{ \frac{\binom{n}{y_{ij}} \prod_{r=0}^{y_{ij}-1} (1 - \pi_i + r\theta_i) \prod_{r=0}^{n-y_{ij}-1} (\pi_i + r\theta_i)}{\prod_{r=0}^{n-1} (1 + r\theta_i)} \right\}$$

and

$$\begin{aligned} \ell(n, \pi, \theta; \underline{y}) = \sum_{i=1}^{D} \sum_{j=1}^{n_i} \Bigg\{ & \ln n! - \ln(n - y_i)! - \ln y_i! + \sum_{r=0}^{y_{ij}-1} \ln(1 - \pi_i + r\theta_i) \\ & + \sum_{r=0}^{n-y_{ij}-1} \ln(\pi_i + r\theta_i) - \sum_{r=0}^{n-1} \ln(1 + r\theta_i) \Bigg\} \end{aligned}$$

respectively.

3.2.4 Modelling the Probability

Results presented in Chapter 2 suggest that the logit function is a suitable distribution for modelling the probability of death of an organism for the algae and malaria data sets. As mentioned in Chapter 2, this function is considered in the present study for modelling the expected value of this probability, π_i, and corresponding to a log-dose x_i,

$$\pi_i = \frac{e^{\alpha+\beta x_i}}{1 + e^{\alpha+\beta x_i}}, \; i = 1, \ldots, D \; .$$

Furthermore, a common parameter θ is considered for all the doses of the drug when this model is fitted. Note that the beta-binomial distribution is not a member of the exponential family of distributions (Lindsey and Altham, 1998), and thus the model which incorporates a logit function for π into this distribution is not a GLM.

3.3 Score Functions and Observed Fisher Information Matrix

Denote the log-likelihood function for a single observation y_{ij} by ℓ_{ij}, $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Then the score functions for a single observation, y_{ij}, are simply the first-order derivatives of ℓ_{ij} with respect to the parameters θ, α and β and are given by

$$\begin{aligned}
\frac{\partial \ell_{ij}}{\partial \theta} &= \sum_{r=0}^{y_{ij}-1} \frac{r}{1-\pi_i+r\theta} + \sum_{r=0}^{n-y_{ij}-1} \frac{r}{\pi_i+r\theta} - \sum_{r=0}^{n-1} \frac{r}{1+r\theta} \\
\frac{\partial \ell_{ij}}{\partial \alpha} &= -\sum_{r=0}^{y_{ij}-1} \frac{\pi_i(1-\pi_i)}{1-\pi_i+r\theta} + \sum_{r=0}^{n-y_{ij}-1} \frac{\pi_i(1-\pi_i)}{\pi_i+r\theta} \\
\frac{\partial \ell_{ij}}{\partial \beta} &= \sum_{r=0}^{y_{ij}-1} \frac{x_i\pi_i(1-\pi_i)}{1-\pi_i+r\theta} + \sum_{r=0}^{n-y_{ij}-1} \frac{x_i\pi_i(1-\pi_i)}{\pi_i+r\theta} \, .
\end{aligned}$$

The score functions for the data are then found by summing the score functions for each observation over all observations.

3.3.1 Observed Fisher Information

In order to derive the information matrix of the beta-binomial distribution, the second-order derivatives of the log-likelihood function with respect to the parameters θ, α and β are required. The nature of these second-order derivatives for a beta-binomial model with a logistic tolerance distribution leads to difficulties

in computing the information matrix. Consider, for example, the expected value with respect to Y_{ij} of the second-order derivative of ℓ_{ij} with respect to θ for the observation y_{ij}. This is written

$$\begin{aligned} -\mathrm{E}\left\{\frac{\partial^2 \ell_{ij}}{\partial \theta^2}\right\} &= E\left\{\sum_{r=0}^{Y_{ij}-1} \frac{r^2}{(1-\mu_i+r\theta)^2} + \sum_{r=0}^{n_i-Y_{ij}-1} \frac{r^2}{(\mu_i+r\theta)^2} + \sum_{r=0}^{n_i-1} \frac{r^2}{(1+r\theta)^2}\right\} \\ &= \sum_{y=0}^{\infty} \Pr(Y=y)\left\{\sum_{r=0}^{y_{ij}-1} \frac{r^2}{(1-\mu_i+r\theta)^2}\right. \\ &\quad \left. + \sum_{r=0}^{n_i-y_{ij}-1} \frac{r^2}{(\mu_i+r\theta)^2} + \sum_{r=0}^{n_i-1} \frac{r^2}{(1+r\theta)^2}\right\}. \end{aligned} \tag{3.1}$$

The computation of this expectation involves summing the appropriate function of y_{ij} over all values of y_{ij} from 0 to ∞. The information $-\mathrm{E}\left(\dfrac{\partial^2 \ell(n,\pi,\theta;Y)}{\partial\theta^2}\right)$ can then be obtained by summing the terms for each observation over all observations. This is clearly a laborious task and can be problematic.

In cases where the information matrix is difficult to calculate, the standard errors of the maximum likelihood estimates of the parameters can be estimated using the inverse of the observed Fisher information matrix. This matrix is simply minus the matrix of the second-order derivatives of the log-likelihood with respect to the parameters in the model. For the beta-binomial model with a logit function modelling π, the observed Fisher information matrix, which is denoted by $\hat{\mathfrak{I}}$ is

$$\hat{\mathfrak{I}} = -\begin{bmatrix} \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\theta^2} & \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\theta\partial\alpha} & \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\theta\partial\beta} \\ \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\theta\partial\alpha} & \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\alpha^2} & \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\alpha\partial\beta} \\ \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\theta\partial\beta} & \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\alpha\partial\beta} & \dfrac{\partial^2 \ell(n,\pi,\theta;\underline{y})}{\partial\beta^2} \end{bmatrix}.$$

The diagonal elements of $\hat{\mathfrak{I}}^{-1}$ are the approximate variances of the parameter estimates. The formulae for these elements were obtained using Mathematica but they are very tedious and are therefore not presented explicitly here.

3.3.2 Models for Separate Doses

Following Morgan (1992, p.254), two base models with large numbers of parameters are considered for comparative purposes. The first model fits a separate beta-binomial distribution to each dose of the drug, i.e. for observations corresponding to dose d_i of the drug, $i = 1, \ldots, D$, it assumes that $Y_{ij} \sim \text{BetaBin}(n, \pi_i, \theta_i)$ for $j = 1, \ldots, n_i$ and is referred to as the $\{\pi_i, \theta_i\}$ model. The second base model is denoted by $\{\pi_i, \theta\}$ and is a nested version of the $\{\pi_i, \theta_i\}$ model in that it adopts a common value of θ for all of the doses of the drug.

3.4 Estimation

The log-likelihood of the data for the beta-binomial model, with a logit function modelling π, is a nonlinear function in the parameters θ, α and β and can be maximized using a nonlinear optimization routine such as the iterative OPTMUM routine in Gauss. This procedure requires the input of starting values for the parameters that are to be estimated. The estimates of the parameters obtained from fitting the binomial model to the data provide starting values for α and β, while small starting values for θ seem to be most suitable. The constraints $\theta > 0$ and $\beta > 0$ need to be incorporated into the routine.

The square roots of the diagonal elements of the inverse of the observed Fisher information matrix, evaluated at the parameter estimates, provide the asymptotic standard errors of these estimates (Azzalini, 1996, p.91) and for the present example they were obtained from a Gauss program.

3.4.1 Inference

The fit of the beta-binomial model with a logit tolerance distribution can be compared with those of the $\{\pi_i, \theta_i\}$ and $\{\pi_i, \theta\}$ models which were introduced in Section 3.3.2 using the deviance. Morgan (1992, p.254) used the deviance approach to good effect for the beta-binomial model with natural mortality and it is therefore

adopted here as a means of model checking. Akaike's Information Criterion (AIC) which is calculated as -2ℓ+2q, where ℓ is the log-likelihood and q is the number of parameters in the model, is also included here as a method of comparing the fits of different models. The AIC provides a balance between the number of parameters in a model of interest and the value of its maximum log-likelihood. The model with the smallest AIC is thought to provide the best fit for the data. The AIC for the beta-binomial model is calculated as $AIC = -2\ell(\hat{n}, \hat{\pi}, \hat{\theta}; \underline{y}) + 6$, where $.\ell(\hat{n}, \hat{\pi}, \hat{\theta}; \underline{y})$ is the maximized log-likelihood and $q = 3$.

95% Wald confidence intervals can be obtained for the parameters. For example, the 95% Wald interval for the parameter θ is given by $\hat{\theta} \pm 1.96\ se(\hat{\theta})$, where $se(\hat{\theta})$ is the standard error of the estimate of θ (Kutner, Nachtsheim, Neter and Li, 2005, p.579). Profile likelihood plots can also be produced for each of the parameters θ, α and β. A range of values of the parameter under consideration is specified and for each value within this range the likelihood function is maximized with respect to the remaining parameters in the model. The profile likelihood plots can be used to construct approximate 95% confidence intervals for the parameters. In particular, the 95% confidence interval for the parameter θ is given by $[\theta_1, \theta_2]$, where θ_1 and θ_2 satisfy

$$\ell_P(\theta_1) = \ell_P(\theta_2) = \ell(\hat{n}, \hat{\pi}, \hat{\theta}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05}$$

and where $\ell_P(\theta)$ is the profile log-likelihood for θ evaluated at θ (Azzalini, 1996, p.146). It is possible, however, to obtain one-sided confidence intervals when using profile likelihoods. In these instances the Wald intervals are used to estimate the 95% confidence intervals of the parameters.

3.5 Examples

3.5.1 Algae Data

The mean of the control observations for the algae data was 223 and was used as an estimate of the number of organisms initially treated with the drug. A binomial model was fitted to this data and the deviance obtained was 47.97 with 18 degrees of freedom, indicating that the data may well be overdispersed. The beta-binomial model with π specified by a logistic function was thus fitted to the algae data by maximizing the likelihood function directly. 95% Wald confidence intervals, which assume that the distributions of the parameters are approximately normal, were computed for θ, α and β. Profile likelihood plots were also produced for θ, α and β and were used to obtain 95% confidence intervals for these parameters. These plots are presented in Figures 3.1 (a),(b) and (c) respectively. The estimates of the parameters, along with their standard errors, 95% Wald intervals and 95% profile likelihood intervals are recorded in Table 3.1. Although the 95% Wald

Table 3.1: Results from fitting the beta-binomial model to the algae data, where π is modelled with a logit function.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
θ	0.006	0.002	(0.002;0.010)	(0.002;0.017)
α	-0.977	0.094	(-1.161;-0.793)	(-1.160;-0.800)
β	1.477	0.107	(1.268;1.687)	(1.290;1.670)

intervals are very similar to the 95% profile likelihood intervals, it is clear that the distribution of θ is skew.

The fit of the beta-binomial model with a logit tolerance distribution was compared with that of beta-binomial models which considered each dose separately and the results are presented in Table 3.2. It is usual to perform a residual analysis to verify the goodness of fit of a model and in the examples considered in this

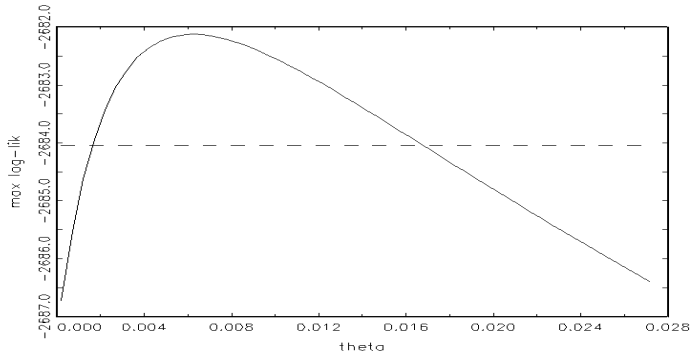


Figure 3.1: (a)

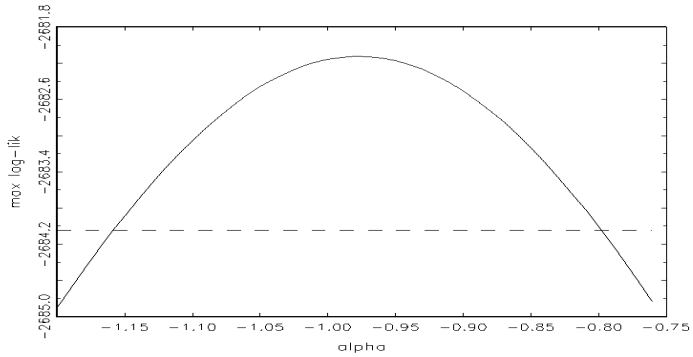


Figure 3.1: (b)

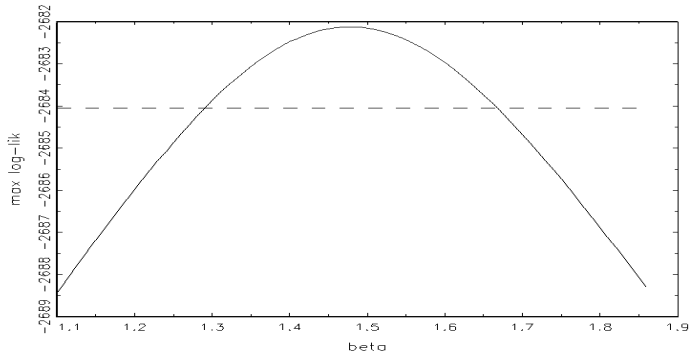


Figure 3.1: (c)

Figure 3.1: Profile likelihood plots for the parameters (a) θ, (b) α and (c) β of the beta-binomial model with a logit tolerance distribution fitted to the algae data.

thesis, the overdispersion has been identified in a residual analysis. A compari-

Table 3.2: A comparison of the beta-binomial models fitted to the algae data.

Model	Maximum log-likelihood	Number of parameters	AIC	Deviance
$\{\pi_i, \theta_i\}$	-2678.183	8	5372.366	-
$\{\pi_i, \theta\}$	-2679.871	5	5369.742	3.376
θ, α, β	-2682.123	3	5370.246	7.880

son of the models indicates that a common value for θ can be adopted, since the deviance of the $\{\pi_i, \theta\}$ model from the $\{\pi_i, \theta_i\}$ base model is 3.376 which is close to the mean 3 of a χ^2_3 distribution. In fact, the AIC for the $\{\pi_i, \theta\}$ model was lower than that of the $\{\pi_i, \theta_i\}$ model. The beta-binomial distribution, which has a logit function modelling the parameter π, also seems to be a reasonably adequate model for the algae data since it has a deviance of 7.880 from the $\{\pi_i, \theta_i\}$ model. This can be compared with a χ^2 distribution with 5 degrees of freedom. The AIC value of 5370.246 for this model was only slightly higher than 5369.742, the AIC of the $\{\pi_i, \theta\}$ model. This indicates that the $\{\pi_i, \theta\}$ model provides the best fit but that the beta-binomial model with a logistic tolerance distribution, which is more parsimonious and captures dose dependence, is adequate for the algae data.

3.5.2 Malaria Data

The maximum observed count of 5193 parasites surviving for this data set corresponded to the dose of 1 *nmol/l* of the drug and exceeded the control observations and thus their mean of 5011. This estimate of n was increased slightly to allow for the death of some of the parasites due to exposure to a non-zero dose of the drug and also to assess the sensitivity of this change. It was found that incrementing the estimate of n in this manner did not have a large impact on the estimation of the remaining parameters. Therefore the highest observation of 5193 was used as the estimate for n.

A binomial distribution, using the estimate of 5193 for the number of parasites treated with the drug, was fitted to the malaria data. The deviance obtained from fitting this model was 3022 with 19 degrees of freedom, indicating that the data could be severely overdispersed. The beta-binomial model with parameters θ, α and β was therefore fitted to the data by directly optimizing the log-likelihood function. 95% Wald intervals were obtained for each of the parameters in this model. Profile likelihood plots were also produced for the parameters and are presented in Figures 3.2 (a), (b) and (c). These plots were used to obtain 95% profile likelihood intervals for the parameters. The estimates of the parameters, together with their standard errors, 95% Wald confidence intervals and 95% profile likelihood intervals are presented in Table 3.3.

Table 3.3: Results from fitting the beta-binomial model to the malaria data, where the expected probability, π, is modelled with a logit function.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
θ	0.037	0.014	(0.008;0.064)	(0.018;0.092)
α	-3.539	0.356	(-4.237;-2.841)	(-4.200;-2.750)
β	5.479	0.486	(4.527;6.431)	(4.350;6.400)

Observe that the distribution of θ is very skew and that the distribution of α is symmetric, while the distribution of β is slightly skew. As a result of the shapes of these distributions, the 95% Wald interval for θ is very different from the 95% profile likelihood interval. The 95% Wald and 95% profile likelihood intervals for the remaining parameters are reasonably similar.

The fit of this model was compared with those of the base models which are described in Section 3.3.2 and which consider each dose of the drug separately. Table 3.4 contains a summary of the comparisons of the different models that were considered. The change in deviance of the $\{\pi_i, \theta\}$ model, with a common θ parameter for all doses, from the $\{\pi_i, \theta_i\}$ model was 33.942 which is very high

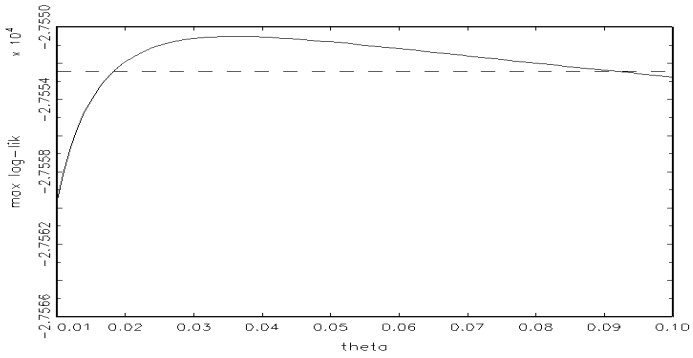


Figure 3.2: (a)

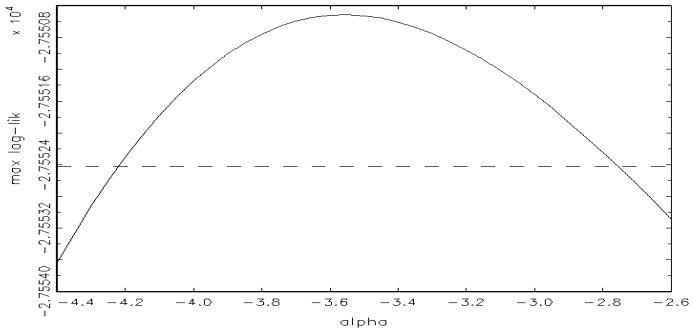


Figure 3.2: (b)

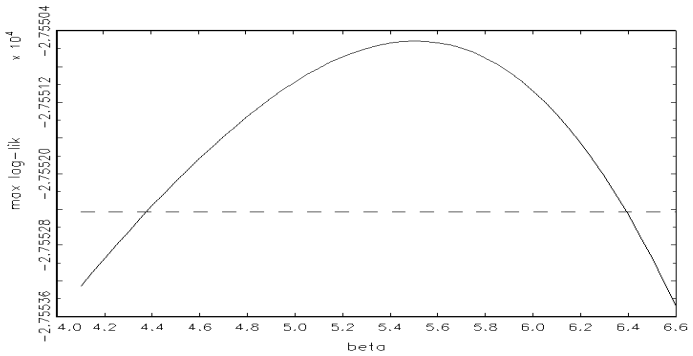


Figure 3.2: (c)

Figure 3.2: Profile likelihood plots for the parameters (a) θ, (b) α and (c) β of the beta-binomial model with a logit tolerance distribution fitted to the malaria data.

compared with that expected for a χ^2_6 distribution. This suggests that a single θ for all doses of the drug is not suitable. The beta-binomial model with a logit tolerance distribution for π had a deviance of 111.144 from the $\{\pi_i, \theta_i\}$ model. A comparison of this deviance with the expected value of the χ^2_{11} distribution indicated that the model did not provide a satisfactory fit for the malaria data. These findings are confirmed by the AIC values. Therefore the beta-binomial model, with a fixed value selected for n, provides a disappointing fit to the malaria data.

Table 3.4: A comparison of the beta-binomial models fitted to the malaria data.

Model	Maximum log-likelihood	Number of parameters	AIC	Deviance
$\{\pi_i, \theta_i\}$	-27494.940	14	55017.880	-
$\{\pi_i, \theta\}$	-27511.911	8	55039.822	33.942
θ, α, β	-27550.512	3	55107.024	111.144

3.6 Conclusions

The beta-binomial distribution was considered in this chapter as a crude model for overdispersed data that arise from a Wadley's problem setting. The overdispersion in the algae data seems to have been well modelled by this distribution with a logistic function used for the parameter π. In contrast, however, the beta-binomial distribution did not adequately account for the overdispersion in the malaria data. These findings support the notion that the beta-binomial model can provide a valuable, but somewhat limited, tool for modelling overdispersion.

Overall it is clear that the estimates for the number of organisms initially exposed to the treatments used in this chapter were crude but nevertheless provided a simple solution to Wadley's problem. These results lead to an investigation of more holistic models for overdispersed data and in particular, more elegant distributions that incorporate the estimation of n into the model.

Chapter 4

The Beta-Poisson Model

4.1 Introduction

The beta-binomial distribution described in Chapter 3 is now extended to a Wadley's problem setting by modelling the unknown number of trials as a random variable, N, with a Poisson distribution. The resulting distribution is termed the beta-Poisson distribution. Properties of this model are explored and the distribution is examined as a means of accommodating overdispersion in count data. The model is then extended to a dose-response setting and its suitability for modelling overdispersed dose-response data that arise from a Wadley problem setting is investigated.

Section 4.2 comprises an introduction to the beta-Poisson distribution and a literature review. The confluent hypergeometric function and methods of approximating the p.m.f. of the beta-Poisson distribution are discussed in Section 4.3. Properties of the distribution, together with some examples, are presented in Section 4.4. The likelihood, score function and information matrix are discussed in Section 4.5 and estimation in Section 4.6. Section 4.7 contains an example of overdispersed count data and the adaptation of the beta-Poisson distribution for dose-response data, together with examples, is discussed in Section 4.8. Section 4.9 reports some concluding remarks.

4.2 Distribution

4.2.1 Derivation of the p.m.f.

Consider a random variable Y which follows a Poisson$[\tau(1-p)]$ distribution as discussed in Chapter 2. Suppose that the parameter p varies according to a beta distribution with parameters a and b where $a > 0$ and $b > 0$, i.e. $p \sim$Beta(a, b). Then the p.m.f. of Y can be derived from first principles as follows:

$$\begin{aligned}
\Pr(Y=y) &= \int_0^1 \Pr(Y=y|P=p)\Pr(P=p)\ dp \\
&= \int_0^1 \frac{[\tau(1-p)]^y e^{-\tau(1-p)}}{y!}\ \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ p^{a-1}(1-p)^{b-1}\ dp \\
&= \frac{\tau^y e^{-\tau}\Gamma(a+b)}{y!\ \Gamma(a)\Gamma(b)} \int_0^1 e^{\tau p} p^{a-1}(1-p)^{b+y-1}\ dp\ . \qquad (4.1)
\end{aligned}$$

Consider expanding $e^{\tau p}$ using a Taylor series expansion as $\sum_{s=0}^{\infty} \frac{(\tau p)^s}{s!}$ and interchanging the summation and the integral. The infinite sum and the integral can be interchanged because this sum converges (Grattan-Guinness, 1970, p.65). Then expression (4.1) can be written

$$\begin{aligned}
\Pr(Y=y) &= \frac{\tau^y e^{-\tau}}{y!}\ \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \sum_{s=0}^{\infty} \frac{\tau^s}{s!} \int_0^1 p^{a+s-1}(1-p)^{b+y-1} dp \\
&= \frac{\tau^y e^{-\tau}}{y!}\ \frac{\Gamma(a+b)\Gamma(b+y)}{\Gamma(a)\Gamma(b)} \sum_{s=0}^{\infty} \frac{\Gamma(a+s)}{\Gamma(a+s+b+y)} \frac{\tau^s}{s!}\ . \qquad (4.2)
\end{aligned}$$

The confluent hypergeometric function, which is often referred to as the Kummer function and is denoted by ${}_1F_1$, is defined in the present context as

$$ {}_1F_1(a, a+b+y; \tau) = \sum_{s=0}^{\infty} \frac{(a)_s}{(a+b+y)_s} \frac{\tau^s}{s!}\ , \qquad (4.3)$$

where $(a)_n$ is the Pochhammer symbol and is given by $(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}$ (Arfken and Weber, 1995, p.797). The p.m.f. of Y can therefore be rewritten as

$$Pr(Y=y) = \frac{\tau^y e^{-\tau}}{y!} \frac{\Gamma(a+b)\Gamma(b+y)}{\Gamma(b)\Gamma(a+b+y)} \sum_{s=0}^{\infty} \frac{\Gamma(a+s)/\Gamma(a)}{\Gamma(a+b+y+s)/\Gamma(a+b+y)} \frac{\tau^s}{s!}$$

$$= \frac{\tau^y e^{-\tau}}{y!} \frac{\Gamma(a+b)\Gamma(b+y)}{\Gamma(a+b+y)\Gamma(b)} \, {}_1F_1(a, a+b+y; \tau) \tag{4.4}$$

which is the p.m.f. of a weighted Poisson distribution with weight

$$w_y = \frac{\Gamma(a+b)\Gamma(b+y)}{\Gamma(a+b+y)\Gamma(b)} \, {}_1F_1(a, a+b+y; \tau) \; .$$

An integral form of the Kummer function exists whereby

$${}_1F_1(a, a+b+y; \tau) = \frac{\Gamma(a+b+y)}{\Gamma(a)\Gamma(b+y)} \int_0^1 e^{\tau x} x^{a-1} (1-x)^{b+y-1} \, dx$$

(Arfken and Weber, 1995, p.801). This integral could have been introduced when $\int_0^1 e^{\tau p} p^{a-1}(1-p)^{b+y-1} dp$ was written down in equation (4.1) but it was instructive to develop the derivation of the p.m.f. from first principles. The distribution defined by (4.4) for the random variable Y is referred to as the beta-Poisson distribution and is denoted by $Y \sim$BetaPoisson(τ, a, b) throughout this thesis.

In terms of the (π, θ)-parametrization which was introduced for the beta-binomial model with $\pi = \dfrac{a}{a+b}$ and $\theta = \dfrac{1}{a+b}$, the p.m.f. of Y can be expressed as

$$\Pr(Y = y) = \frac{\tau^y e^{-\tau}}{y!} \prod_{r=0}^{y-1} \frac{(1 - \pi + r\theta)}{(1 + r\theta)} \sum_{s=0}^{\infty} \left\{ \frac{\prod_{r=0}^{s-1} (\pi + r\theta)}{\prod_{r=y}^{y+s-1} (1 + r\theta)} \frac{\tau^s}{s!} \right\} \; .$$

The calculation of the p.m.f. for a single observation written in this form is computationally awkward since it involves working with a product within an infinite sum. As a result, preference is given to the (a, b)-parametrization invoked in expressions (4.2) and (4.4) when fitting the beta-Poisson distribution in the present study.

The beta-Poisson p.m.f. can also be obtained by considering a random variable Y which follows a beta-binomial distribution as described in Chapter 3 and modelling the parameter N with a Poisson(τ) distribution. In this instance the p.m.f.

of Y can be derived as

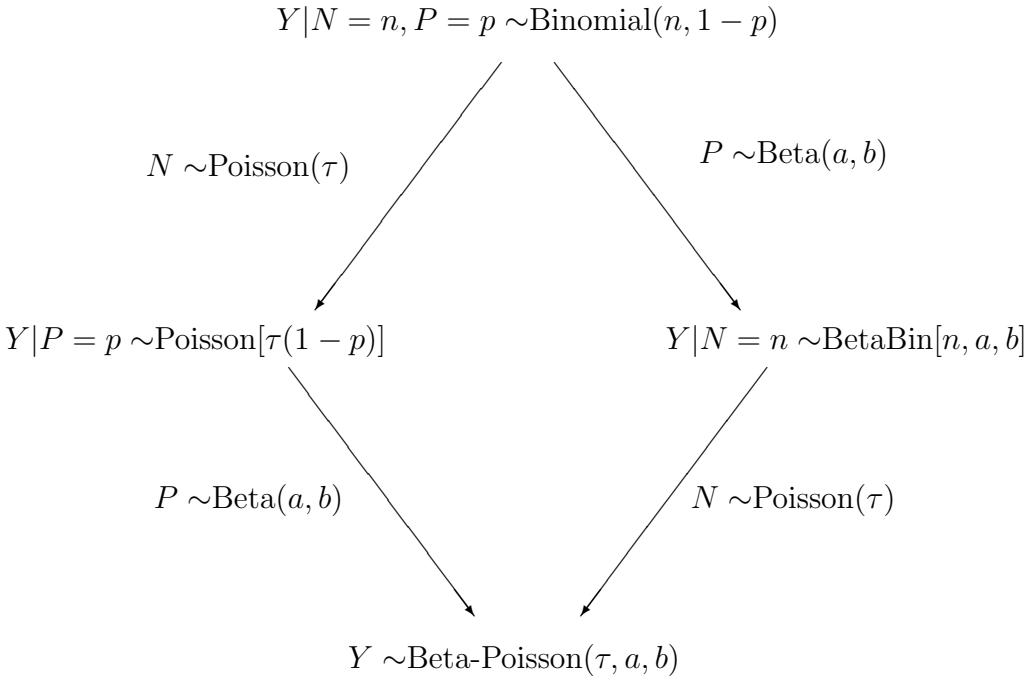
$$\begin{aligned}
\Pr(Y=y) &= \sum_{n=y}^{\infty} \Pr(Y=y|N=n)\Pr(N=n) \\
&= \sum_{n=y}^{\infty} \binom{n}{y} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(b+y)\Gamma(n-y+a)}{\Gamma(a+b+n)} \frac{\tau^n e^{-\tau}}{n!} \\
&= \frac{e^{-\tau}}{y!} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \Gamma(b+y) \sum_{n=y}^{\infty} \frac{\tau^n}{(n-y)!} \frac{\Gamma(n-y+a)}{\Gamma(a+b+n)} \\
&= \frac{e^{-\tau}}{y!} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \Gamma(b+y) \sum_{s=0}^{\infty} \frac{\tau^{y+s}}{s!} \frac{\Gamma(a+s)}{\Gamma(a+b+y+s)}
\end{aligned}$$

where $s = n - y$, and therefore

$$\Pr(Y=y) = \frac{e^{-\tau}\tau^y \Gamma(a+b)\Gamma(b+y)}{y!\ \Gamma(b)\Gamma(a+b+y)} {}_1F_1(a, a+b+y; \tau)\ .$$

Thus Y follows a beta-Poisson distribution with parameters $\tau,\ a$ and b.

The diagram below provides a summary of how the BetaPoisson(τ, a, b) distribution for a random variable Y can be derived.



The special case where p follows a beta distribution with parameters a and

b both equal to 1 is of interest. This case is equivalent to p following a uniform distribution on the interval (0,1). In this instance the p.m.f. of Y is

$$\begin{aligned}
\Pr(Y=y) &= \int_0^1 \Pr(Y=y|P=p)\Pr(P=p)\ dp \\
&= \int_0^1 \frac{e^{-\tau(1-p)}[\tau(1-p)]^y}{y!}\ dp \\
&= \frac{\Gamma(y+1)-\Gamma(1+y,\tau)}{\tau y!} \\
&= \frac{1}{\tau}\left\{1-\frac{\Gamma(y+1,\tau)}{y!}\right\}\ ,
\end{aligned}$$

where $\Gamma(1+y,\tau)$ is the incomplete gamma function and satisfies the equation

$$\begin{aligned}
\frac{\Gamma(1+y,\tau)}{y!} &= \frac{1}{y!}\int_\tau^\infty x^y e^{-x}dx \\
&= e^{-\tau}\sum_{k=0}^{y}\frac{\tau^k}{k!}
\end{aligned}$$

(Hogg, McKean and Craig, 2005, p.150). The p.m.f. of Y can therefore be expressed in terms of a finite sum as

$$\Pr(Y=y) = \frac{1}{\tau}\left\{1-e^{-\tau}\sum_{k=0}^{y}\frac{\tau^k}{k!}\right\}\ .$$

4.2.2 Literature Review

The beta-Poisson distribution introduced in Section 4.2.1 is a variant of the Poisson-beta distribution which was formulated by Holla and Bhattacharya (1965) in the context of examples such as multiple accidents or repeated incidents of a disease. In their study, the number of accidents experienced by an individual in time period t is modelled by Y$\sim$Poisson(λt), where λ measures the susceptibility of an individual to accidents per unit time and $0 \leq \lambda \leq 1$. Holla and Bhattacharya (1965) place particular emphasis on the distribution of accidents in a second time period given the number sustained in a first time period, i.e. the primary focus is accident proneness. They argue that accident proneness varies from individual to individual and consider modelling the variability in the Poisson parameter

λ using a beta distribution. As a result, the p.m.f. of the Poisson-beta model, which is the same as the p.m.f. described in Section 4.2.1, is obtained. Holla and Bhattacharya (1965) obtained the moment generating function (m.g.f.) of this distribution, as well as its mean and variance, and their results are drawn upon in Section 4.4.1. Bhattacharya and Holla (1965) also considered modelling the parameter λ using a uniform distribution and derived the p.m.f. of what they termed the Poisson-rectangular distribution. They do not, however, give any details regarding parameter estimation for the Poisson-beta distribution. In fact the Poisson-beta distribution does not seem to have arisen in any other practical applications and as a result little work has been reported on it. The model appears in the Encyclopedia of Statistical Sciences (Kotz and Johnson, 1986, p.19) and in the text by Johnson, Kotz and Kemp (1992) but no other references to the papers by Holla and Bhattacharya (1965) or indeed by Bhattacharya and Holla (1965) have been located.

It is clear that the context of the beta-Poisson distribution introduced in the present study is very different from the problem considered by Holla and Bhattacharya (1965). The distribution is introduced here as a means of accommodating overdispersion in the Poisson$[\tau(1-p)]$ model from Chapter 2, which is frequently used for data in a Wadley's problem setting. This distribution will be referred to as the beta-Poisson distribution for the present study, in line with the term beta-binomial distribution but deviating from the terminology Poisson-beta introduced by Holla and Bhattacharya (1965).

4.3 Confluent Hypergeometric Function

The p.m.f. of the beta-Poisson distribution from expression (4.4) includes the confluent hypergeometric function except when p is modelled by a uniform distribution. This confluent hypergeometric function is, by definition, an infinite sum. As a result this p.m.f. cannot be calculated explicitly. For a sufficiently large cutoff

value for s, namely s_{cut}, the infinite sum in the confluent hypergeometric function can be approximated by

$$\sum_{s=0}^{s_{\text{cut}}} \frac{(a)_s}{(a+b+y)_s} \frac{\tau^s}{s!}$$

and the probabilities associated with the p.m.f. of the beta-Poisson distribution can then be estimated accordingly. It is necessary to consider the choice of s_{cut} carefully, in order to ensure that the error resulting from the approximation is negligible. In this section the modality of the confluent hypergeometric function is considered because this will give an indication of an appropriate choice for s_{cut}.

4.3.1 Modality of the Confluent Hypergeometric Function

Consider the mode of the generic confluent hypergeometric function, expressed as

$$ {}_1F_1(a,b;z) \;=\; \sum_{s=0}^{\infty} \frac{\Gamma(a+s)/\Gamma(a)}{\Gamma(b+s)/\Gamma(b)} \frac{z^s}{s!} \;, \tag{4.5}$$

where $a,\ b$ and $z > 0$ as in the p.m.f. of the beta-Poisson distribution.

Let t_s denote the sth term in the summation in expression (4.5), which can be written as

$$t_s \;=\; \frac{\Gamma(a+s)/\Gamma(a)}{\Gamma(b+s)/\Gamma(b)} \frac{z^s}{s!} \;.$$

The confluent hypergeometric function is decreasing if and only if $t_s \geq t_{s+1}$ and thus if and only if

$$\frac{\Gamma(a+s)}{\Gamma(b+s)} \frac{z^s}{s!} \geq \frac{\Gamma(a+s+1)}{\Gamma(b+s+1)} \frac{z^{s+1}}{(s+1)!}$$

which is true if and only if

$$s^2 + (1+b-z)s + (b-az) \geq 0 \;. \tag{4.6}$$

An investigation of the roots of the quadratic equation in expression (4.6) will indicate the modal form of the confluent hypergeometric function and hence an appropriate choice of s_{cut}. Although s is a non-negative integer, it is treated

as a non-negative real number for simplicity. As a result, if the mode of the function under consideration occurs at a non-integer value of s, the confluent hypergeometric function will have at most two modes, which will occur at $\lfloor s \rfloor$ and $\lfloor s+1 \rfloor$, where $\lfloor s \rfloor$ denotes the integer part of the value of s. Let the two roots of the quadratic function in (4.6) be denoted by s_- and s_+ where $s_- < s_+$. Then

$$s_- = \frac{(z-1-b) - \sqrt{(z-1-b)^2 - 4(b-az)}}{2}$$

and

$$s_+ = \frac{(z-1-b) + \sqrt{(z-1-b)^2 - 4(b-az)}}{2} \; . \tag{4.7}$$

and the nature of the change in t_s as s changes can be inferred from these roots. Specifically, suppose that z is fixed and consider the relationship between a and b and the nature of the roots of the quadratic function in expression (4.6). Figures 4.1 (a), (b), (c) and (d) include an illustration of the relationship between the form of the terms of the confluent hypergeometric function and the nature of the roots s_- and s_+.

Case I: Complex Roots:

The confluent hypergeometric function is decreasing if the roots s_- and s_+ are complex. This is depicted in Figure 4.1 (a) and occurs if and only if

$$(z-1-b)^2 - 4(b-az) < 0$$

and thus if and only if

$$b^2 - 2b(1+z) + (z^2 + 2z(2a-1) + 1) < 0 \; . \tag{4.8}$$

The quadratic function in expression (4.6) therefore has complex roots if and only if b lies between the roots b_- and b_+ of expression (4.8), where $b_- < b_+$,

$$b_- = (1+z) - 2\sqrt{z(1-a)}$$

and

$$b_+ = (1+z) + 2\sqrt{z(1-a)} \; .$$

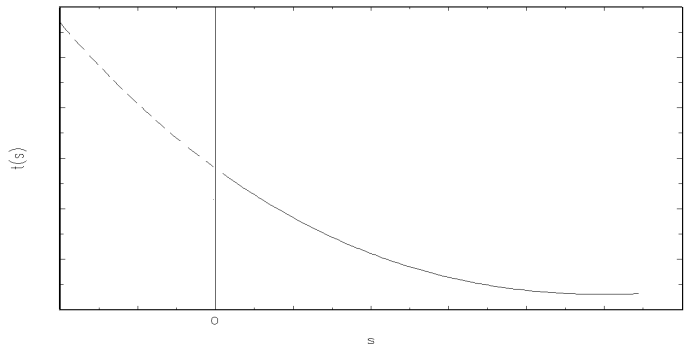


Figure 4.1: (a)

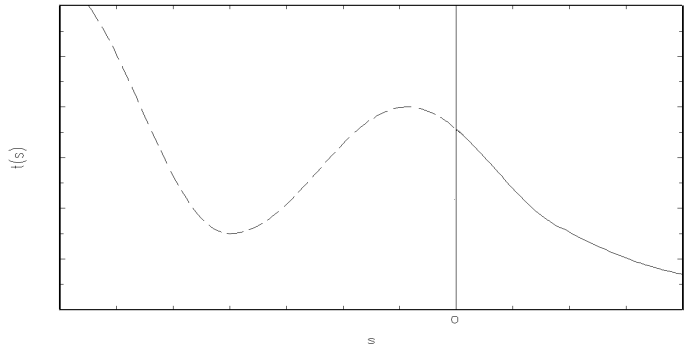


Figure 4.1: (b)

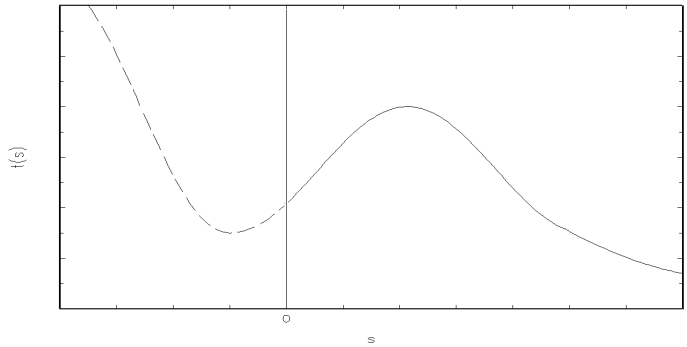


Figure 4.1: (c)

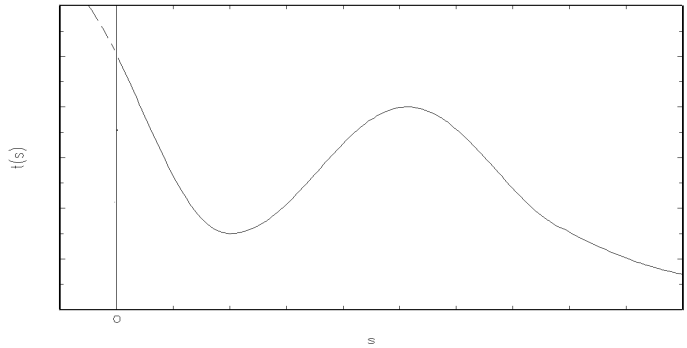


Figure 4.1: (d)

Figure 4.1: Form of the terms of the confluent hypergeometric function for (a) complex roots, (b) two negative roots, (c) one positive and one negative root and (d) two positive roots of expression (4.6), taking $t(s)$ and s as continuous.

Observe that a must be less than or equal to one in order for b_- and b_+ to exist.

Consider the (a, b)-space represented in Figure 4.2. Since the parameters a and b of the beta-Poisson distribution are both positive, only the first quadrant of the (a, b)-space is considered. The area labelled as I represents the range of values of a and b for which the quadratic function in expression (4.6) has complex roots. Observe that the curve defining the roots b_- and b_+ intersects with the b-axis, i.e. $a = 0$, at the points $z + 1 + 2\sqrt{z}$ and $z + 1 - 2\sqrt{z}$. Note that these points of intersection are both non-negative when $z + 1 - 2\sqrt{z} \geq 0$ and thus when $z \geq 1$. Suppose that $b = az$ and consider the intersection between the line az and the

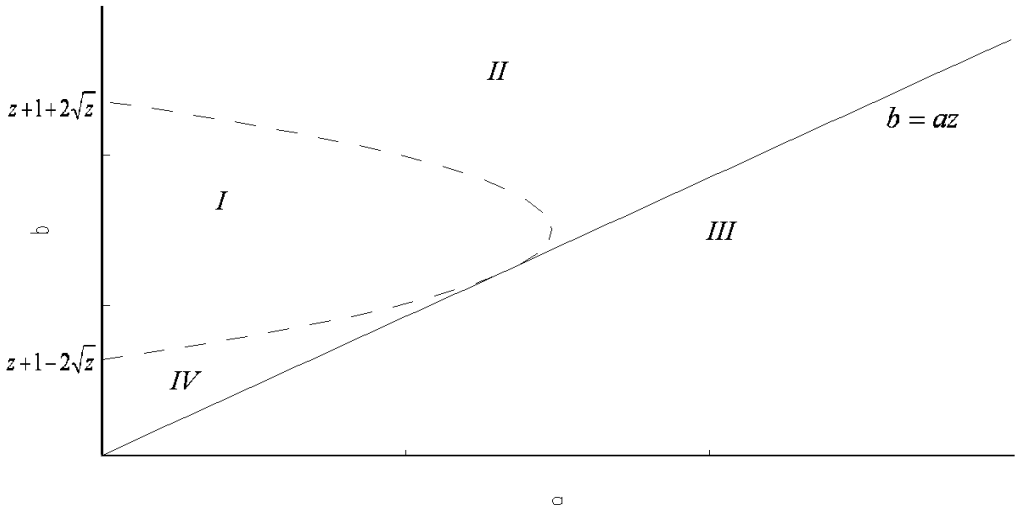


Figure 4.2: Diagram showing the (a, b) parameter space.

curve describing b_- and b_+:

$$1 + z - 2\sqrt{z(1-a)} \;=\; az \;,$$

which implies that

$$z(1-a) - 2\sqrt{z(1-a)} + 1 \;=\; 0$$

and therefore that $a = 1 - \dfrac{1}{z}$. Now consider the partial derivative of b_- with respect to a

$$\frac{\partial b_-}{\partial a} \;=\; \frac{z}{\sqrt{z(1-a)}} \;,$$

which is equal to z when $a = 1 - \frac{1}{z}$. It can therefore be deduced that $b = az$ is a tangent to the function defining b_- and b_+ which meets this curve at the point $(a = 1 - \frac{1}{z}, b = z - 1)$.

Case II: Two Negative Roots:

When the roots of the quadratic function in expression (4.6) are negative, the confluent hypergeometric function is decreasing for all positive values of s. This is illustrated in Figure 4.1 (b). The roots of the quadratic equation (4.6) are negative when $s_+ < 0$. This occurs if and only if $z - 1 - b < 0$ and thus if and only if $(z-1-b)^2 > (z-1-b)^2 - 4(b-az)$ which implies that $4(b-az) > 0$.

Therefore the quadratic function in expression (4.6) has two negative roots if and only if $b > z - 1$ and $b > az$, i.e. $b >$max$\{z-1, az\}$. The range of values of a and b for which the quadratic function in expression (4.6) has two negative roots is represented by area II in Figure 4.2.

Case III: One Positive and One Negative Root:

Figure 4.1 (c) illustrates that the confluent hypergeometric function is unimodal when the quadratic function in expression (4.6) has one positive and one negative root. Observe that the root $s_+ > 0$ when $z-b-1 > 0$ and therefore when $b < z-1$. Recall that

$$s_- \quad = \quad \frac{(z-1-b) - \sqrt{(z-1-b)^2 - 4(b-az)}}{2} \ .$$

Thus s_- is negative when $z-b-1 < 0$ or when $4(b-az) < 0$, which can be simplified as $b > z-1$ or $b < az$. The region labelled III in Figure 4.2 represents the portion of the (a, b)-parameter space for which the quadratic function in expression (4.6) has one positive and one negative root.

Case IV: Two Positive Roots:

When s_- and s_+ are both positive the confluent hypergeometric function is bimodal. This is presented graphically in Figure 4.1 (d). The quadratic function in equation (4.6) has two positive roots when s_- is positive. This occurs when $b < z - 1$ and $b > az$, i.e. when $az < b < z - 1$. Recall that the root s_- is only

non-negative when z is greater than or equal to one. Thus the area labelled IV in Figure 4.2 represents the values of a and b for which the quadratic in expression (4.6) has two positive roots.

From the investigation of the modality of the confluent hypergeometric function a method for approximating this infinite sum can be devised. The first step entails evaluating the root s_+ from expression (4.7). If s_+ is negative or complex, the confluent hypergeometric function is decreasing for all positive values of s. Thus, beginning with $s = 0$ the terms t_s of the confluent hypergeometric function can be evaluated and every 50th term can be considered to determine whether that term is less than a suitably small value, say 10^{-6}. If $t_k < 10^{-6}$ then k is considered a suitable choice for s_{cut}.

If, however, s_+ is positive and $\lfloor s_+ \rfloor$ denotes the integer part of s_+, then $\sum_{s=0}^{\lfloor s_+ \rfloor} t_s$ can be evaluated and incremented by adding terms in the sum for $s > \lfloor s_+ \rfloor$. Every 50th term for $s > \lfloor s_+ \rfloor$ can be considered individually and when $t_k < 10^{-6}$ for $k > \lfloor s_+ \rfloor$ an appropriate choice for s_{cut} is k.

4.4 Properties of the Distribution

4.4.1 Moment Generating Function

In order to derive the m.g.f. of the beta-Poisson distribution, two properties of the confluent hypergeometric function, namely

$$ {}_1F_1(a, c; z) \;=\; e^z \; {}_1F_1(c - a, c; -z) \tag{4.9} $$

and

$$ {}_1F_1(a, b; \lambda z) \;=\; \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \cdot \frac{1}{n!} (\lambda - 1)^n z^n \; {}_1F_1(a + n, b + n; z) \; . \tag{4.10} $$

are required. Expression (4.9) is known as the Kummer transformation (Erdélyi, Magnus, Oberhettinger and Tricomi, 1953, p.253) and the relationship introduced

in equation (4.10) is a multiplication formula of the confluent hypergeometric function (Erdélyi, Magnus, Oberhettinger and Tricomi, 1953, p.283).

The m.g.f. of the beta-Poisson distribution can be obtained using conditioning arguments. Since the m.g.f. of a Beta(a, b) distributed random variable is ${}_1F_1(a, a+b; t)$ (Johnson, Kotz and Balakrishnan, 1995, p.218), the m.g.f. of Y is

$$\begin{aligned} M(t) = \mathrm{E}(e^{tY}) &= \mathrm{E}_p[\mathrm{E}(e^{tY}|P=p)] \\ &= \mathrm{E}_p\{\exp[\tau(1-p)(e^t-1)]\} \\ &= \exp[\tau(e^t-1)]\mathrm{E}_p\{\exp[-\tau p(e^t-1)]\} \\ &= \exp[\tau(e^t-1)]\mathrm{E}_p\{\exp[p\ (-\tau(e^t-1))]\} \\ &= \exp[\tau(e^t-1)]\ {}_1F_1[a, a+b; -\tau(e^t-1)]. \end{aligned}$$

By invoking the Kummer transformation in expression (4.9), it then follows that

$$M(t) = {}_1F_1[b, a+b; \tau(e^t-1)]\ .$$

Using the method outlined by Holla and Bhattacharya (1965), the m.g.f. of the beta-Poisson distribution can also be calculated as the expected value of e^{tY} directly. Specifically

$$M(t) = \mathrm{E}(e^{tY}) = \sum_{y=0}^{\infty} \frac{e^{ty}e^{-\tau}\tau^y}{y!} \frac{\Gamma(a+b)\Gamma(b+y)}{\Gamma(a+b+y)\Gamma(b)}\ {}_1F_1(a, a+b+y; \tau)\ .$$

Using the Kummer transformation from (4.9), this becomes

$$\begin{aligned} M(t) = \mathrm{E}(e^{tY}) &= \sum_{y=0}^{\infty} \frac{(b)_y}{(a+b)_y\ y!}(e^t)^y(-\tau)^y\ {}_1F_1(b+y, a+b+y; -\tau) \\ &= \sum_{y=0}^{\infty} \frac{(b)_y}{(a+b)_y}\frac{1}{y!}[1-(e^t-1)]^y(-\tau)^y\ {}_1F_1(b+y, a+b+y; -\tau) \end{aligned}$$

and by the multiplication formula (4.10) it then follows that

$$M(t) = {}_1F_1(b, a+b; \tau(e^t-1))\ .$$

Moments of the Distribution

Mean and Variance

The mean and variance of the beta-Poisson distributed random variable Y can be obtained by using well-known conditioning arguments presented for example by Ross (2002, p.152). These expectations are written as

$$\begin{aligned} \mathrm{E}(Y) &= \mathrm{E}_p[\mathrm{E}(Y|P)] \\ &= \mathrm{E}_p[\tau(1-p)] \\ &= \frac{\tau b}{a+b} \end{aligned}$$

and

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{E}_p[\mathrm{Var}(Y|P)] + \mathrm{Var}_p[\mathrm{E}(Y|P)] \\ &= \mathrm{E}_p[\tau(1-p)] + \mathrm{Var}_p[\tau(1-p)] \\ &= \frac{\tau b}{a+b} + \frac{\tau^2 ab}{(a+b)^2(a+b+1)} \\ &= \frac{\tau b}{a+b}\left\{1 + \frac{\tau a}{(a+b)(a+b+1)}\right\} \end{aligned}$$

respectively. The mean and variance of Y can also be obtained from the m.g.f. by noting that $M^{(n)}(0) = E(Y^n)$, where $M^{(n)}(0)$ is the nth derivative of the m.g.f. evaluated at $t = 0$. In terms of the (π, θ)-parametrization of the beta distribution (Morgan, 1992, p.240), the mean of Y is given by

$$\mathrm{E}(Y) = \tau(1-\pi)$$

and the variance by

$$\mathrm{Var}(Y) = \tau(1-\pi)\left\{1 + \frac{\tau\pi\theta}{1+\theta}\right\} ,$$

which is comprised of the usual Poisson variance of $\tau(1-\pi)$ and a dispersion parameter $\{1 + \frac{\tau\pi\theta}{1+\theta}\}$. Since τ, π and θ are non-negative, the variance of the beta-Poisson distribution exceeds the mean when $\theta > 0$ and reduces to the Poisson variance when $\theta = 0$. This model is not suitable for underdispersed data

because the variance can never be less than the mean.

Skewness and Kurtosis

The measure of the skewness of a distribution, denoted by γ, is calculated as $\mathrm{E}[(Y-\mu)^3]/\sigma^3$, where μ and σ refer to the mean and standard deviation of the distribution respectively. A zero skewness indicates that the distribution is symmetric. The skewness for the beta-Poisson distribution can be obtained by evaluating $\mathrm{E}(Y^2) = M^{(2)}(0)$ and $\mathrm{E}(Y^3) = M^{(3)}(0)$ and is calculated as

$$\begin{aligned}
\gamma &= \frac{\mathrm{E}(Y^3) - 3\mu\mathrm{E}(Y^2) + 2\mu^3}{\sigma^3} \\
&= \{(a+b)^3(1+a+b)^2[a^4 + b^2(2+3b+b^2) + a^3(3+4b+3\tau) + \\
&\quad a^2(2+9b+6b^2+6\tau+6b\tau+2\tau^2) + ab(4+4b^2+6\tau-2\tau^2+3b(3+\tau))]\}/ \\
&\quad \{2b^2\tau^2(2+a+b)(a+a^2+b+2ab+b^2+a\tau)^3\} \ .
\end{aligned}$$

Kurtosis, which is denoted κ, means peakedness and is measured as $\mathrm{E}[(Y-\mu)^4]/\sigma^4$. A large value of this measure is associated with a distribution with a high central peak. For the beta-Poisson distribution, the measure of kurtosis is given by

$$\begin{aligned}
\kappa &= \frac{\mathrm{E}(Y^4) + 6\mu^2\mathrm{E}(Y^2) - 4\mu\mathrm{E}(Y^3) - 3\mu^4}{\sigma^4} \\
&= \left\{ b\left[(a+b)^3 - 3b^3\tau^3 + \frac{(7\tau+7\tau b)(a+b)^3}{1+a+b} \right.\right. \\
&\quad + \frac{6b^2\tau^2(a+b)(1+a+b+\tau+b\tau)}{1+a+b} + \frac{6\tau^2(1+b)(2+b)(a+b)^3}{(1+a+b)(2+a+b)} \\
&\quad - \frac{4b\tau(a+b)^2[a^2+a(3+2b+3\tau+3b\tau)+(2+3b+b^2)(1+3\tau+\tau^2)]}{(1+a+b)(2+a+b)} \\
&\quad \left.\left. + \frac{\tau^3(1+b)(2+b)(3+b)(a+b)^3}{(1+a+b)(2+a+b)(3+a+b)} \right] \right\} \Big/ \\
&\quad \left\{ (a+b)^4\tau^2 \left[1 - \frac{a}{a+b} + \frac{ab\tau}{(a+b)^2(1+a+b)} \right] \right\} \ .
\end{aligned}$$

The expressions for the skewness and kurtosis of the beta-Poisson distribution are somewhat intractable and not very informative but can be readily used to calculate values of γ and κ for given values of $a,\ b$ and τ.

4.4.2 Modality of the beta-Poisson Distribution

Holgate (1970) investigated the conditions under which a compound Poisson distribution is unimodal. Specifically, he proved a theorem which states:

> "Let $f(\lambda)$ be the probability density function of a positive, unimodal absolutely continuous random variable. Then the non-negative integer-valued random variable with probability function
>
> $$p_n \ = \ (n!)^{-1} \int_0^\infty e^{-\lambda} \lambda^n f(\lambda) d\lambda \ , \quad (n \geq 0)$$
>
> is a unimodal lattice variable".

Since the Beta(a, b) distribution is unimodal for $a > 1$ and $b > 1$ (Johnson, Kotz and Balakrishnan, 1995, p.217), the proof of Holgate's theorem can be used to provide a proof for the beta-Poisson distribution.

Let $f(p)$ denote the p.d.f. of the beta distribution which was adopted for the parameter p in deriving the Poisson(τp) distribution. The Poisson(τp) distribution is considered here in place of the Poisson$[\tau(1-p)]$ distribution which was derived in Chapter 2. This is for simplicity and is acceptable because $0 < p < 1$ implies that $0 < 1 - p < 1$ and if p follows a Beta(a, b) distribution then $(1 - p)$ follows a Beta(b, a) distribution. Since $f(p)$ is differentiable on the interval (0,1) and by integration by parts for $n \geq 1$ it is seen that

$$\begin{aligned} p_n &= \int_0^1 \frac{e^{-\tau p}(\tau p)^n}{n!} f(p) dp \\ &= \left[\frac{-e^{-\tau p}}{\tau} \frac{(\tau p)^n}{n!} f(p) \right]_0^1 + \int_0^1 \frac{e^{-\tau p}}{\tau} \left\{ \frac{n\tau(\tau p)^{n-1} f(p)}{n!} + \frac{(\tau p)^n f'(p)}{n!} \right\} dp \\ &= 0 + (n-1)!^{-1} \int_0^1 e^{-\tau p}(\tau p)^{n-1} f(p) dp + \int_0^1 \frac{e^{-\tau p}(\tau p)^n}{\tau n!} f'(p) dp \end{aligned} \tag{4.11}$$

$$= \quad p_{n-1} \; + \; \int_0^1 \frac{e^{-\tau p}(\tau p)^n}{\tau n!} f'(p)dp \; . \tag{4.12}$$

Let p_0 denote the mode of the beta distribution and let $\Delta p_n \; = \; p_{n+1} - p_n$. Then expression (4.12) can be written as

$$p_n \quad = \quad p_{n-1} + \int_0^{p_0} \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^n}{n!} f'(p)dp \; + \; \int_{p_0}^1 \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^n}{n!} f'(p)dp$$

and therefore

$$\Delta p_{n-1} \quad = \quad \int_0^{p_0} \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^n}{n!} f'(p)dp \; + \; \int_{p_0}^1 \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^n}{n!} f'(p)dp \; . \tag{4.13}$$

Since $f'(p) > 0$ when $p < p_0$ and $f'(p) < 0$ when $p > p_0$, the first term on the right hand side of expression (4.13) is non-negative and the second term is non-positive.

Observe that

$$\Delta p_n \quad = \quad \int_0^{p_0} \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^{n+1}}{(n+1)!} f'(p)dp \; + \; \int_{p_0}^1 \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^{n+1}}{(n+1)!} f'(p)dp \; .$$

and hence

$$\begin{aligned}
\frac{(n+1)\Delta p_n}{(\tau p_0)^{n+1}} \quad &= \quad \frac{1}{n!} \int_0^{p_0} \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^{n+1}}{(\tau p_0)^{n+1}} f'(p)dp \; + \; \frac{1}{n!} \int_{p_0}^1 \frac{e^{-\tau p}}{\tau} \frac{(\tau p)^{n+1}}{(\tau p_0)^{n+1}} f'(p)dp \\
&\leq \quad \frac{1}{n!} \int_0^{p_0} \frac{e^{-\tau p}}{\tau} \left(\frac{\tau p}{\tau p_0} \right)^n f'(p)dp \; + \; \frac{1}{n!} \int_{p_0}^1 \frac{e^{-\tau p}}{\tau} \left(\frac{\tau p}{\tau p_0} \right)^n f'(p)dp \qquad (4.14) \\
&= \quad \frac{1}{n!} \int_0^1 \frac{e^{-\tau p}}{\tau} \left(\frac{\tau p}{\tau p_0} \right)^n f'(p)dp \\
&= \quad \frac{\Delta p_{n-1}}{(\tau p_0)^n} \qquad (4.15)
\end{aligned}$$

This implies that

$$\frac{(n+1)}{(\tau p_0)} \Delta p_n \quad \leq \quad \Delta p_{n-1} \; . \tag{4.16}$$

The inequality in expression (4.14) follows because when $p < p_0$, $\dfrac{p}{p_0} < 1$ and as a result $\left(\dfrac{\tau p}{\tau p_0} \right)^{n+1} \leq \left(\dfrac{\tau p}{\tau p_0} \right)^n$. Similarly, when $p > p_0$ the magnitude of the negative component of the integral is not increased by dividing by $\dfrac{\tau p}{\tau p_0}$.

From the inequality in expression (4.16), $\Delta p_{n-1} \leq 0$ implies that $\Delta p_n \leq 0$ i.e. $p_n \leq p_{n-1}$ suggests that $p_{n+1} \leq p_n$ and $\Delta p_n \geq 0$ implies that $\Delta p_{n-1} \geq 0$ i.e. $p_{n+1} \geq p_n$ indicates that $p_n \geq p_{n-1}$. Since the distribution defined by p_n is a probability function, p cannot increase indefinitely when n approaches infinity and therefore $\{p_n\}$ is the probability function of a discrete, unimodal random variable.

Recall that a and b are the parameters of the beta distribution and that $p \sim$Beta(a, b). The theorem cannot be used when $a = 1$ and $b \neq 1$ or when $a \neq 1$ and $b = 1$. This is because in these instances the first term in (4.11) does not equate to zero and the arguments for positive and negative gradients cannot be used when the distribution is always increasing ($a = 1,\ b \neq 1$) or always decreasing ($a \neq 1,\ b = 1$). Since p denotes the probability of death of an organism, it can be represented by an S-shaped curve and thus not by a function that is always increasing or always decreasing.

Consider the scenario $a < 1$ and $b < 1$. When these conditions are satisfied, the beta distribution is U-shaped (Johnson, Kotz and Balakrishnan, 1995, p.217) and when $p = 0$ or 1, $f(p)$ approaches infinity. Holgate's (1970) theorem therefore cannot be applied to this instance of the beta-Poisson distribution because the first term in expression (4.11) does not disappear. Since p in the present context refers to the probability of death of an organism, it is unlikely that this probability will be high for both low and high doses of the drug. Therefore in this context the distribution of p will not be U-shaped and thus both a and b will not be less than one.

4.4.3 Examples

Figure 4.3 presents graphic representations, together with the mean, variance, skewness (γ) and kurtosis (κ), of the p.m.f. of Y when the parameters a and b are equal, that is when p was taken from a symmetric beta distribution. Examples of the p.m.f. when p is from skew beta distributions are presented in Figures 4.4 and 4.5. The infinite sum included in the p.m.f. of the beta-Poisson distribution was

estimated by a finite sum with a cutoff value selected using the methods outlined in Section 4.3. It was observed that when the terms in the sum for the confluent hypergeometric function were less than 10^{-6}, and provided they were beyond the mode, a suitable cutoff for the infinite sum was obtained. The results for a cutoff value of 10^{-6} were compared with smaller values and the estimate of the confluent hypergeometric function remained unchanged.

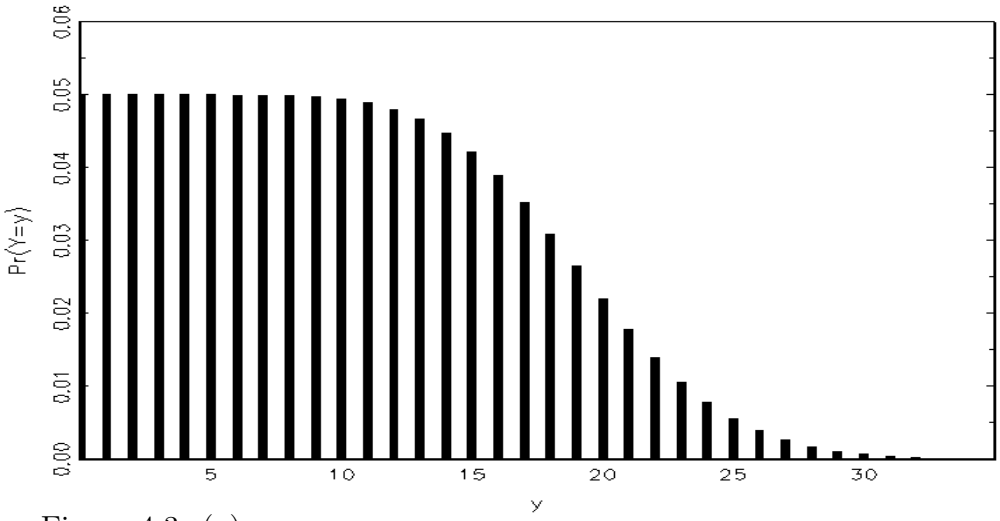


Figure 4.3: (a)$\tau = 20,\ a = 1,\ b = 1$, E($Y$)=10, Var($Y$)=43.333, γ=0.001, κ=-3.384 .

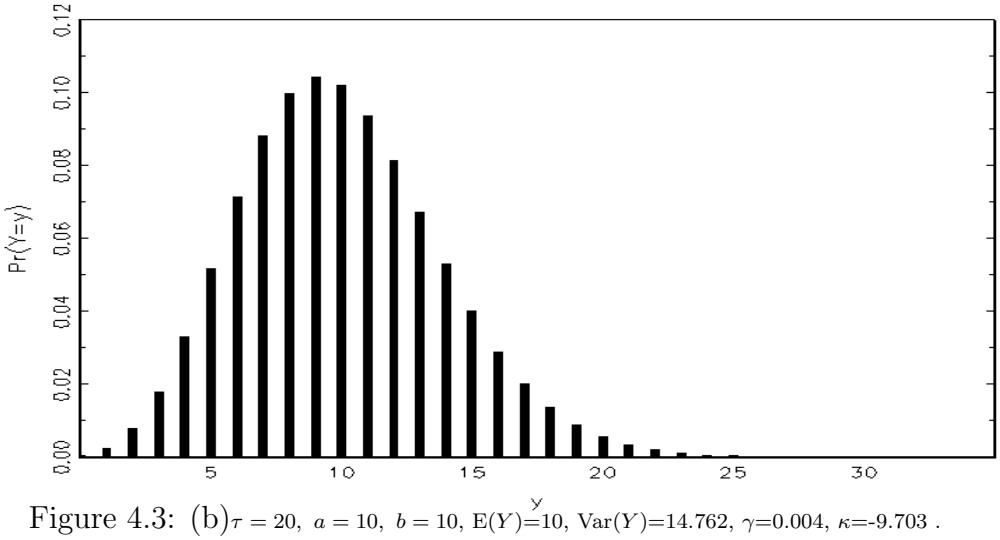


Figure 4.3: (b)$\tau = 20,\ a = 10,\ b = 10$, E($Y$)=10, Var($Y$)=14.762, γ=0.004, κ=-9.703 .

Figure 4.3: Plots showing the shapes of various beta-Poisson distributions, with $\tau = 20$ and $a = b$.

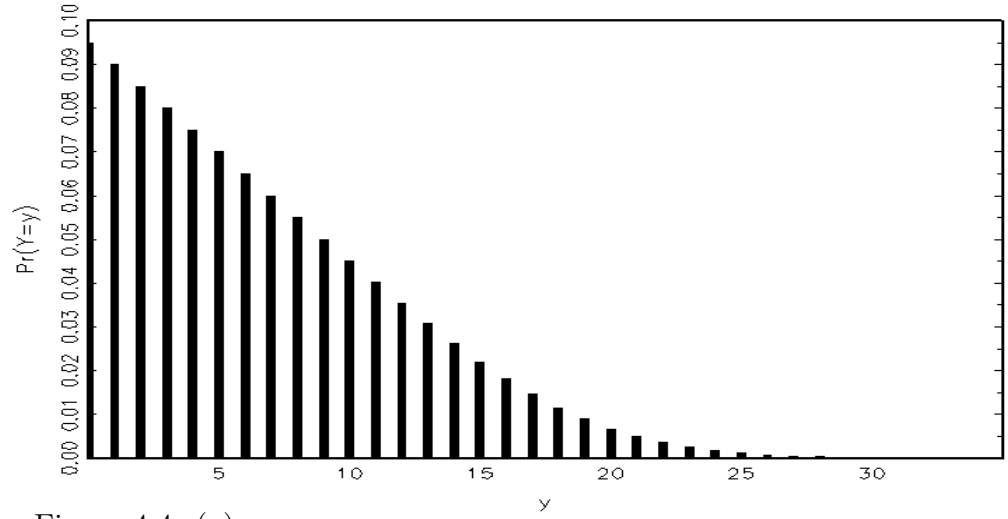


Figure 4.4: (a)$\tau = 20,\ a = 2,\ b = 1$, E($Y$)=6.667, Var($Y$)=28.889, γ=0.003, κ=-1.318 .

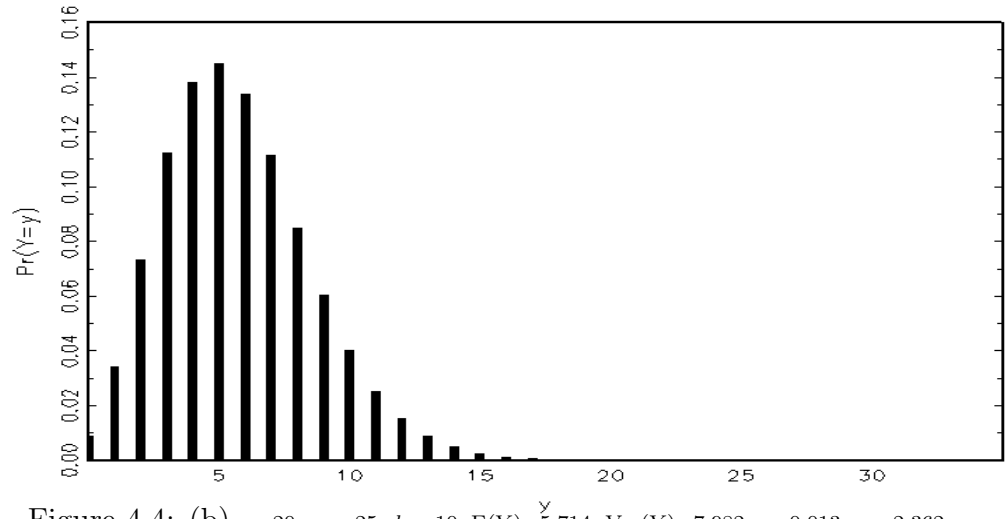


Figure 4.4: (b)$\tau = 20,\ a = 25,\ b = 10$, E($Y$)=5.714, Var($Y$)=7.982, γ=0.013, κ=-2.362 .

Figure 4.4: Plots showing the shapes of various beta-Poisson distributions, with $\tau = 20$ and $a > b$.

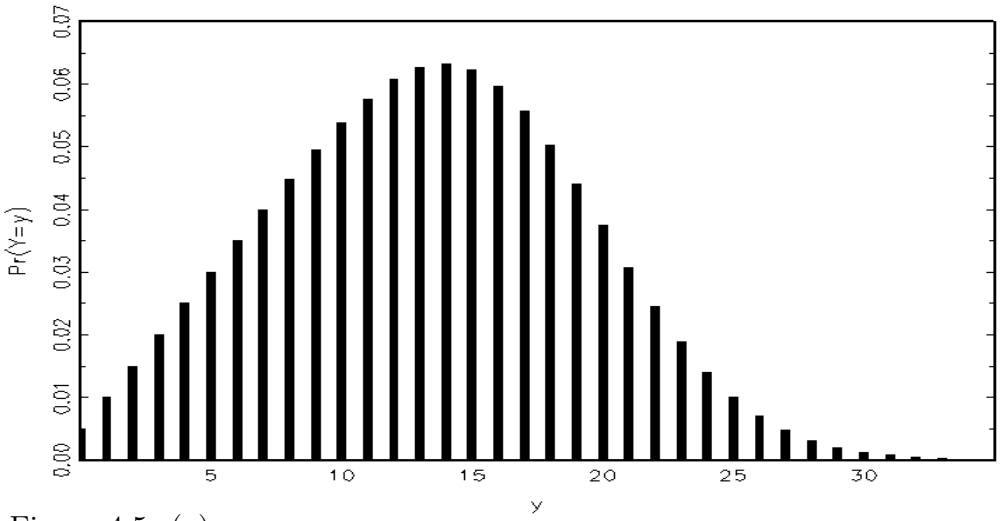


Figure 4.5: (a)$\tau = 20,\ a = 1,\ b = 2$, E($Y$)=13.333, Var($Y$)=35.556, γ=0.0002, κ=-11.886 .

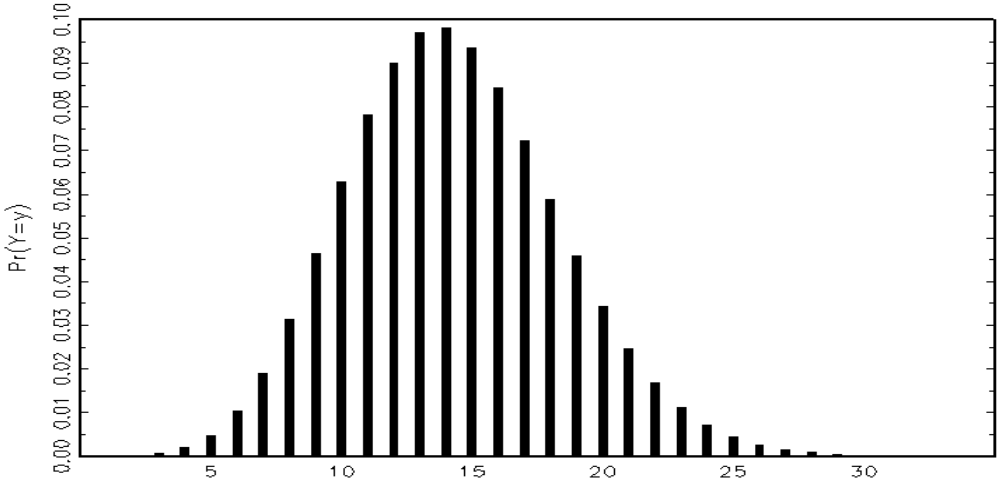


Figure 4.5: (b)$\tau = 20,\ a = 10,\ b = 25$, E($Y$)=14.286, Var($Y$)=16.553, γ=0.002, κ=-31.169 .

Figure 4.5: Plots showing the shapes of various beta-Poisson distributions, with $\tau = 20$ and $a < b$.

4.5 Likelihood, Score Functions and the Information Matrix

4.5.1 Likelihood Function

The likelihood function of the beta-Poisson distribution is given by

$$L(\tau, a, b; y) = \frac{\tau^y e^{-\tau}\Gamma(a+b)\Gamma(b+y)}{y!\Gamma(a+b+y)\Gamma(b)} {}_1F_1(a, a+b+y; \tau)$$

and the log-likelihood function can therefore be written as

$$\begin{aligned}\ell(\tau, a, b; y) &= y\ln\tau - \tau + \ln[\Gamma(a+b)] + \ln[\Gamma(b+y)] + \ln[{}_1F_1(a, a+b+y; \tau)] \\ &\quad - \ln(y!) - \ln[\Gamma(a+b+y)] - \ln[\Gamma(b)] \ .\end{aligned}$$

4.5.2 Score Functions and Information Matrix

The score functions for τ, a and b are obtained by differentiating the log-likelihood function with respect to these parameters and the information matrix is the expectation of the matrix of second derivatives of the log-likelihood function with respect to the parameters. However, the confluent hypergeometric function ${}_1F_1(a, a+b+y; \tau)$ is only differentiable explicitly with respect to τ (Erdélyi, Magnus, Oberhettinger and Tricomi, 1953, p.254). Therefore, since the confluent hypergeometric function is included in the log-likelihood, only the score function for the parameter τ can be derived explicitly and is given by

$$\frac{\partial\ell(\tau, a, b; y)}{\partial\tau} = \frac{a\,{}_1F_1(a+1, a+b+y+1, \tau)}{(a+b+y)\,{}_1F_1(a, a+b+y, \tau)} + \frac{y}{\tau} - 1 \ .$$

The derivatives of the confluent hypergeometric function with respect to the parameters a and b cannot be written down explicitly. As a result, the score functions for these parameters as well as the information matrix of the beta-Poisson distribution cannot be obtained explicitly. However, the information matrix can be estimated as minus the Hessian matrix which can be obtained numerically using commands such as `hessp` in the programming language Gauss.

4.6 Estimation

Consider a random sample of observations from the beta-Poisson distribution where the count y_i is observed k_i times for $i = 1, \ldots, d$. The log-likelihood function is then written as

$$\begin{aligned}\ell(\tau, a, b; \underline{y}) &= \sum_{i=1}^{d} k_i\{y_i \ln \tau - \tau + \ln[\Gamma(a+b)] + \ln[\Gamma(b+y_i)] + \ln[{}_1F_1(a, a+b+y_i; \tau)] \\ &\quad - \ln(y_i!) - \ln[\Gamma(a+b+y_i)] - \ln[\Gamma(b)]\} \ ,\end{aligned}$$

where $\underline{y}$ denotes the vector of observed counts. Since the score functions for the parameters a and b cannot be derived explicitly, the maximum likelihood estimates of the parameters cannot be obtained directly. The constraints $\tau,\ a$ and $b > 0$ need to be imposed on the parameter space and thus a constrained nonlinear optimization routine, such as the OPTMUM routine in Gauss, can be used to maximize the log-likelihood function numerically. The infinite sum included in the likelihood function can be approximated by a finite sum with a cutoff value selected using the techniques described in Section 4.3. The cutoff value s_{cut} is selected as the value of k for which the term t_k in the confluent hypergeometric function is less than 10^{-6} and where t_k has surpassed the mode of the confluent hypergeometric function.

It was observed, during preliminary investigations of the beta-Poisson distribution, that there were instances in which the optimization routine used to maximize the log-likelihood function had difficulty converging and where the estimates of the parameters were unstable. In order to explore this observation, data were simulated from the beta-Poisson distribution. Specifically, if the random variable Y follows a beta-Poisson distribution then Y given $P = p$ is Poisson distributed. Thus observations from the beta-Poisson distribution were simulated by first simulating a value for p from the beta distribution with parameters a and b and then simulating an observation, given this value of p, from a Poisson distribution with mean $\tau(1-p)$. The beta-Poisson distribution was then fitted to the simulated data and the estimates of the parameters were considered.

In order to investigate the cause of the convergence problems and apparent instabilities in the parameter estimates the situation was simplified by setting a equal to b. Furthermore, since the method of moments estimates for τ and a can be obtained explicitly when $a = b$, these estimates from the simulated data sets were explored. The method of moments estimates of τ and a are $\tilde{\tau} = 2\bar{y}$ and $\tilde{a} = \dfrac{\bar{y}^2 + \bar{y} - s^2}{2(s^2 - \bar{y})}$ respectively, so that the respective sample mean and variance in terms of $\tilde{\tau}$ and $\tilde{a}$ are $\bar{y} = \dfrac{\tilde{\tau}}{2}$ and $s^2 = \dfrac{\tilde{\tau}(\tilde{\tau} - 2 + 4\tilde{a})}{4(2\tilde{a} + 1)}$.

In some of the simulated data sets the method of moments estimate of the parameter a was more than thirteen times the original value from which the data were simulated. The data sets that yielded unusual estimates of a were considered individually and it was observed that in each case the sample variance was very close to the sample mean. These data are therefore Poisson-like and thus it would seem that fitting the beta-Poisson distribution to such data is equivalent to attempting to model non-existent overdispersion. As a consequence of this numerical problems seem to arise.

4.6.1 Inference

The information matrix of the beta-Poisson distribution cannot be derived explicitly but, as noted earlier, it can be estimated by minus the Hessian matrix which can be obtained numerically using the Gauss command, `hessp`. The inverse of this matrix is the approximate covariance matrix of the parameter estimates. The square roots of the diagonal elements of the covariance matrix provide the approximate marginal standard errors of the maximum likelihood estimates, $\hat{\tau}$, $\hat{a}$ and $\hat{b}$. 95% Wald intervals can be computed for the parameters where, for example, the 95% Wald interval for the parameter τ is $\hat{\tau} \pm 1.96\ se(\hat{\tau})$ and where $se(\hat{\tau})$ is the standard error of the estimate of τ (Kutner, Nachtsheim, Neter and Li, 2005, p.579).

Profile likelihood plots can also be constructed for each of the parameters

τ, a and b of the beta-Poisson distribution and 95% confidence intervals for the parameters can be obtained from these plots. Specifically, the 95% confidence interval for the parameter τ is given by $[\tau_1; \tau_2]$, where the limits τ_1 and τ_2 satisfy the condition

$$\ell_P(\tau_1) = \ell_P(\tau_2) = \ell(\hat{\tau}, \hat{a}, \hat{b}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05} \ ,$$

with $\ell_P(\tau)$ the profile likelihood for the parameter τ, $\chi^2_{1,0.05}$ the critical χ^2 value at a 5% level of significance with 1 degree of freedom, where $\chi^2_{1,0.05}$ is such that $\Pr(\chi^2_1 > \chi^2_{1,0.05})$=0.05, and $\ell(\hat{\tau}, \hat{a}, \hat{b}; \underline{y})$ is the maximum log-likelihood of the beta-Poisson distribution (Azzalini, 1996, p.146). Confidence intervals for a and b can be obtained from the profile likelihood plots in a similar manner.

4.6.2 Goodness of Fit

The adequacy of the fit of the beta-Poisson distribution to overdispersed count data can be assessed by a chi-squared goodness of fit test. This test compares the observed frequencies of counts with those expected if the data constitute a random sample of observations from the model of interest. The test statistic is calculated as

$$\chi^2 \ = \ \sum_{i=1}^{d} \frac{(k_i - e_i)^2}{e_i} \ ,$$

where k_i is the ith observed frequency and e_i the ith expected frequency of counts. A significant test statistic leads to a rejection of the hypothesis that the data arise from the beta-Poisson distribution.

The goodness of fit of the beta-Poisson distribution can also be compared with that of the Poisson model and other distributions suitable for overdispersed count data using AIC, with the AIC of the beta-Poisson distribution calculated as AIC $= -2\ell(\hat{\tau}, \hat{a}, \hat{b}; \underline{y}) + 6$. Note that the best fitting model is the one with the lowest AIC value.

4.7 An Example

Shmueli, Minka, Kadane, Borle and Boatwright (2005) introduced an overdispersed data set based on the quarterly sales of a particular item of clothing. The data consist of the number of articles sold per quarter and are summarized in a frequency distribution in Table 4.1. Shmueli, Minka, Kadane, Borle and Boatwright (2005) concluded that the sales data are overdispersed and fitted the Conway-Maxwell Poisson (CMP) distribution to the data as a means of accommodating the overdispersion.

Table 4.1: Frequency distribution of the sales data.

No. of Sales	Frequency	No. of Sales	Frequency	No. of Sales	Frequency
0	514	11	40	22	0
1	503	12	37	23	0
2	457	13	22	24	0
3	423	14	9	25	0
4	326	15	7	26	0
5	233	16	10	27	0
6	195	17	9	28	0
7	139	18	3	29	0
8	101	19	2	30	1
9	77	20	2		
10	56	21	2		

The CMP distribution was introduced by Conway and Maxwell in 1962 in the context of queueing systems and later revisited by Shmueli, Minka, Kadane, Borle and Boatwright in 2005 who showed that the distribution is suitable for modelling over- and underdispersed count data. The p.m.f. of the CMP model is

$$\Pr(Y = y) \;=\; \frac{\lambda^y}{(y!)^\nu \; Z(\lambda, \nu)} \;,$$

where $Z(\lambda, \nu)$ is the normalizing constant written as

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu} .$$

The beta-Poisson distribution was fitted to the sales data by directly optimizing the appropriate log-likelihood function. The information matrix was estimated by minus the Hessian matrix and the resulting matrix was used to obtain the standard errors of the parameter estimates. Profile likelihood plots were produced for each of the parameters τ, a and b and 95% confidence intervals of the parameters were obtained from these plots. The respective profile likelihood plots for the parameters τ, a and b are presented in Figures 4.6 (a), (b) and (c), where the points of intersection between the curve and the dashed line yield the 95% confidence intervals for the parameters.

The parameter estimates, together with standard errors, 95% Wald intervals and 95% profile likelihood intervals are presented in Table 4.2. Observe that only

Table 4.2: Results from fitting the beta-Poisson model to the sales data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	55.230	24.005	(8.180;102.280)	(25.500;*)
a	21.225	10.719	(0.216;42.234)	(8.250;*)
b	1.462	0.080	(1.305;1.619)	(1.291;1.668)

*These limits of the confidence intervals cannot be computed

one-sided 95% profile likelihood intervals could be computed for τ and a and that the standard errors of these estimates are very large, resulting in wide 95% Wald intervals for these parameters.

The adequacy of the fit of the beta-Poisson distribution to this data was assessed and compared with that of the Poisson and CMP models using chi-squared goodness of fit tests as well as the AIC values. The p-values from the goodness of fit tests for the Poisson, CMP and beta-Poisson distributions, together with the AIC

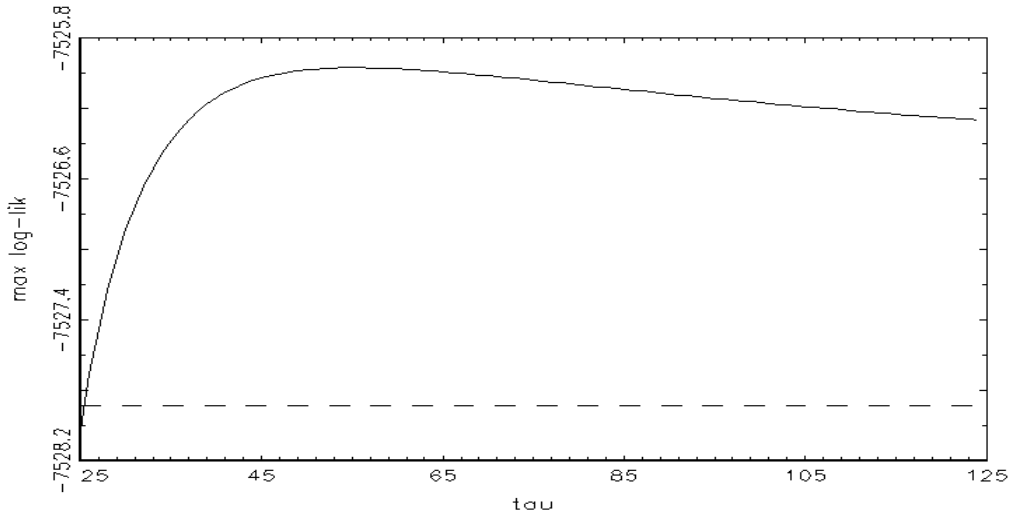


Figure 4.6: (a)

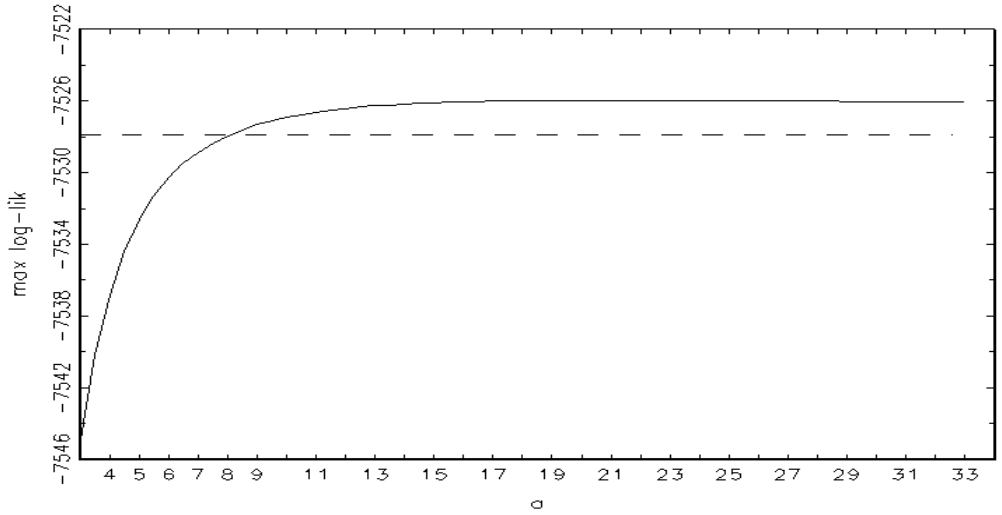


Figure 4.6: (b)

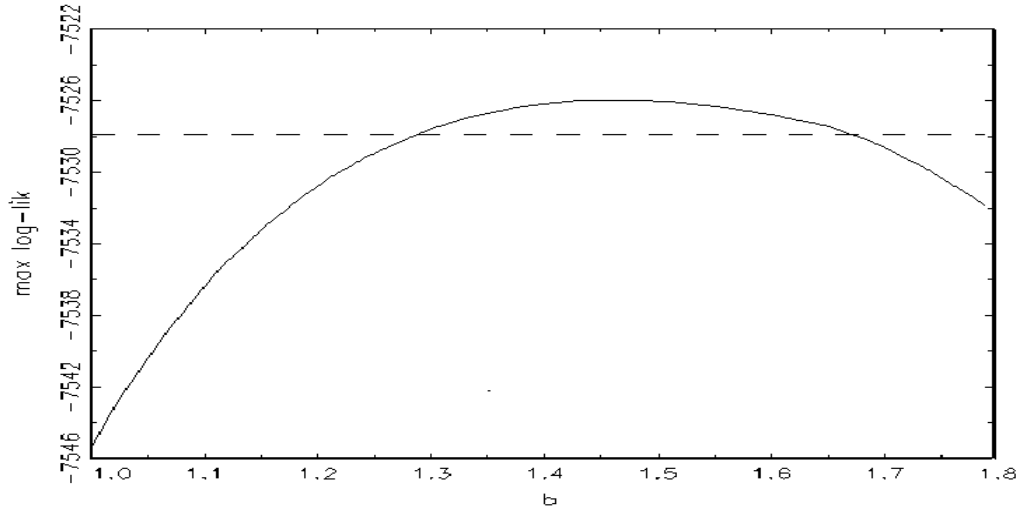


Figure 4.6: (c)

Figure 4.6: Profile likelihood plots for the parameters (a) τ, (b) a and (c) b of the beta-Poisson model fitted to the sales data.

Table 4.3: A comparison of the various models fitted to the sales data.

Model	χ^2 p-value	AIC
Poisson	0.000	17 921.622
CMP	0.242	15 056.712
Beta-Poisson	0.002	15 057.934

values are presented in Table 4.3. The p-values for the Poisson and beta-Poisson distributions suggest that these models are not suitable for the data. However, the beta-Poisson distribution does provide a better fit than the Poisson distribution. The AIC of the beta-Poisson distribution is extremely close to that of the CMP distribution and is much lower than that of the Poisson model, indicating that the beta-Poisson distribution seems to accommodate some of the overdispersion in the sales data.

4.8 Modelling Dose-Response Data

4.8.1 Likelihood Function

In the context of dose-response studies, consider the random variable Y which measures the number of parasites that survive exposure to particular doses of a drug. As introduced earlier, denote an observation from the control group by $y_{cj},\ j = 1, \ldots, n_c$, and let y_{ij} refer to the observed number of parasites that survive exposure to a dose d_i of the drug where $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Associated with the ith dose of the drug is the log-dose x_i and the probability of the death of a parasite, which is denoted by p_i and which is considered to follow a Beta(a_i, b_i) distribution for $i = 1, \ldots, D$.

Since the probability of non-natural death for untreated parasites is assumed to be zero, the number of survivors in the control group follows a Poisson(τ)

distribution with likelihood and log-likelihood functions given respectively by

$$L_c = \prod_{j=1}^{n_c} \frac{e^{-\tau}\tau^{y_{cj}}}{y_{cj}!}$$

and

$$\ell_c = \sum_{j=1}^{n_c} y_{cj} \ln \tau - n_c\tau - \sum_{j=1}^{n_c} \ln y_{cj}! \ .$$

The likelihood and log-likelihood functions for observations corresponding to non-zero doses of the drug based on the appropriate beta-Poisson distribution are written as

$$L_d = \prod_{i=1}^{D}\prod_{j=1}^{n_i} \left\{ \frac{e^{-\tau}\tau^{y_{ij}}\Gamma(a_i+b_i)\Gamma(b_i+y_{ij})}{y_{ij}!\ \Gamma(b_i)\Gamma(a_i+b_i+y_{ij})}\ {}_1F_1(a_i, a_i+b_i+y_{ij};\tau) \right\}$$

and

$$\ell_d = \sum_{i=1}^{D}\sum_{j=1}^{n_i} \ln \left\{ \sum_{s=0}^{\infty} \frac{\tau^{y_{ij}}e^{-\tau}}{y_{ij}!} \frac{\Gamma(a_i+b_i)\Gamma(b_i+y_{ij})\Gamma(a_i+s)}{\Gamma(a_i+s+b_i+y_{ij})\Gamma(a_i)\Gamma(b_i)}\ \frac{\tau^s}{s!} \right\} \tag{4.17}$$

$$\begin{aligned} = \sum_{i=1}^{D}\sum_{j=1}^{n_i} \{ & y_{ij}\ln\tau - \tau + \ln\Gamma(a_i+b_i) + \ln\Gamma(b_i+y_{ij}) + \ln[{}_1F_1(a_i, a_i+b_i+y_{ij};\tau)] \\ & - \ln y_{ij}! - \ln\Gamma(b_i) - \ln\Gamma(a_i+b_i+y_{ij})\} \ , \end{aligned}$$

respectively. The log-likelihood for the data is then found by summing the log-likelihood functions for the individual observations over all of the observations and is therefore given by $\ell_c + \ell_d$.

4.8.2 Models for Separate Doses

Two models that consider each dose $d_i,\ i = 1, \ldots, D$, of the drug separately are introduced in order to facilitate the assessment of the adequacy of the beta-Poisson model. The first model is referred to as the $\{\tau, \pi_i, \theta_i\}$ model and fits a separate beta-Poisson distribution to each non-zero dose of the drug, where $\pi_i = \dfrac{a_i}{a_i+b_i}$

and $\theta_i = \dfrac{1}{a_i + b_i}$ but where a common value of τ is used for all doses. The second model, a nested version of the first, adopts common parameters τ and θ and is denoted as the $\{\tau, \pi_i, \theta\}$ model.

4.8.3 Modelling the Probabilities

Following the beta-binomial distribution as applied to dose-response data in Chapter 3, a logit function is used to model the expected probability of death of a parasite π_i for dose d_i, $i = 1, \ldots, D$, and a common parameter θ is adopted for all doses of the drug. When the logistic tolerance distribution is used,

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) \;=\; \alpha + \beta x_i$$

and therefore

$$\pi_i \;=\; \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\ , i = 1, \ldots, D\ .$$

Note that although the (π, θ)-parametrization is more meaningful, it is convenient to work with the p.m.f. in terms of a and b when doing computations by using the simple conversions $a_i = \dfrac{\pi_i}{\theta}$ and $b_i = \dfrac{1-\pi_i}{\theta}$.

4.8.4 Estimation

Since the log-likelihood function is nonlinear in the parameters a nonlinear optimization routine must be used to maximize it. Consider the component

$$\sum_{s=0}^{\infty}\left\{\frac{\tau^{y_{ij}}e^{-\tau}}{y_{ij}!}\,\frac{\Gamma(a_i+b_i)\Gamma(b_i+y_{ij})\Gamma(a_i+s)}{\Gamma(a_i+s+b_i+y_{ij})\Gamma(a_i)\Gamma(b_i)}\,\frac{\tau^s}{s!}\right\} \tag{4.18}$$

within the summation (4.17) corresponding to a single observation treated with a non-zero dose of the drug. The calculation of this expression can be problematic because it entails computing a generic gamma function $\Gamma(r)$, which can be awkward for large values of r. A method of overcoming this difficulty is to first calculate the logarithm of the summand in expression (4.18) and then to exponentiate it,

thereby making use of the `lnfact` command in Gauss. The command `lnfact(r)` uses Stirling's formula to approximate $\ln \Gamma(r)$ and avoids the overflow that can arise from computing the gamma function of large numbers. Specifically expression (4.18) can be approximated by a finite sum

$$\sum_{s=0}^{s_{\text{cut}}} \exp\{y_{ij} \ln \tau - \tau + \ln \Gamma(a_i + b_i) + \ln \Gamma(b_i + y_{ij}) + \ln \Gamma(a_i + s) + s \ln \tau \\ - \ln y_{ij}! - \ln \Gamma(a_i + s + b_i + y_{ij}) - \ln \Gamma(a_i) - \ln \Gamma(b_i) - \ln s!\}$$

where s_{cut} is chosen empirically for each example using the methods outlined in Section 4.3. The parameter estimates obtained from fitting the beta-binomial model in Chapter 3 can be used as starting values for the optimization routine.

This method for maximizing the log-likelihood function can be adapted for the $\{\tau, \pi_i, \theta\}$, $\{\tau, \pi_i, \theta_i\}$ and $\{\tau, \alpha, \beta, \theta\}$ models using the following equations where necessary:

$$\pi_i = \frac{a_i}{a_i + b_i} = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}}$$

and

$$\theta_i \quad = \quad \frac{1}{a_i + b_i} \; .$$

4.8.5 Inference

Since the information matrix of the beta-Poisson distribution cannot be derived explicitly it can be estimated by minus the Hessian matrix which can be obtained numerically using the `hessp` Gauss command, as noted in Section 4.6.1. The inverse of this matrix can then be used to approximate the standard errors of the parameter estimates. 95% Wald confidence intervals can be computed for the parameters. Wald intervals, however, have limited usefulness because they make a symmetric approximation to the sampling distributions of the estimators and they do not respect the boundaries of the parameter space. For the beta-Poisson distribution the restrictions on the parameter space are $\theta > 0$ and $\beta > 0$.

Furthermore, profile likelihood plots can be obtained for each of the parameters τ, α, β and θ and these plots can be used to construct 95% confidence intervals for the parameters using the method described in Section 4.6.1.

4.8.6 Goodness of Fit

The methods of model checking described for the beta-binomial setting were used to assess the fit of the beta-Poisson model. In particular, the fit of the $\{\tau, \alpha, \beta, \theta\}$ model can be compared with that of the separate dose models $\{\tau, \pi_i, \theta_i\}$ and $\{\tau, \pi_i, \theta\}$ by considering the appropriate deviance. The AIC can also be used as a means of goodness of fit assessment.

4.8.7 Examples

Algae Data

The beta-Poisson model with a logistic tolerance distribution was fitted to the algae data described in Section 2.2.1. The profile likelihood plots for the parameters of the beta-Poisson distribution fitted to the algae data are presented in Figures 4.7 (a), (b), (c) and (d). Table 4.4 contains the parameter estimates, together with their standard errors and 95% Wald and profile likelihood intervals.

Table 4.4: Results from fitting the beta-Poisson model to the algae data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	223.966	6.560	(211.110;236.823)	(233.050;237.100)
θ	0.005	0.004	(-0.003;0.0119)	(0.000;0.017)
α	-0.936	0.141	(-1.213;-0.660)	(-1.240;-0.670)
β	1.443	0.110	(1.227;1.659)	(1.230;1.680)

The adequacy of the beta-Poisson model for the algae data was assessed by comparing its fit with the fits of the separate dose models described in Section 4.8.2.

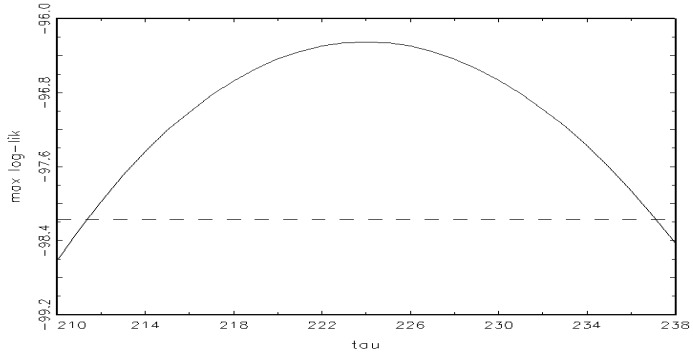


Figure 4.7: (a)

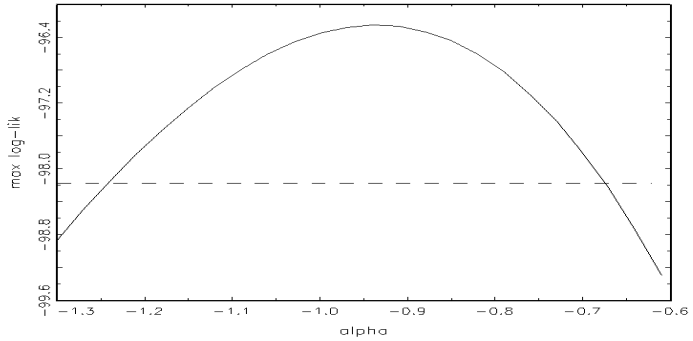


Figure 4.7: (b)

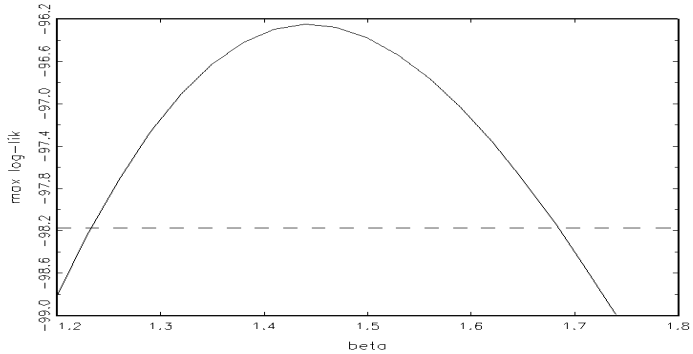


Figure 4.7: (c)

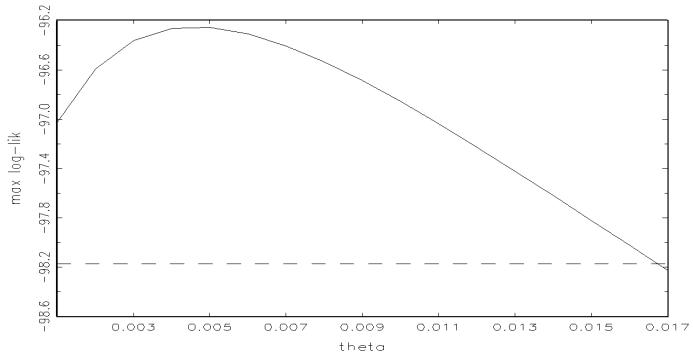


Figure 4.7: (d)

Figure 4.7: Profile likelihood plots for the parameters (a) τ, (b) α, (c) β and (d) θ of the beta-Poisson model with a logit tolerance distribution fitted to the algae data.

Details of these fits are presented in Table 4.5. The deviance of the $\{\tau, \pi_i, \theta\}$ model

Table 4.5: A comparison of the beta-Poisson models fitted to the algae data.

Model	Maximum log-likelihood	Number of parameters	AIC	Deviance
$\{\tau, \pi_i, \theta_i\}$	-93.348	9	204.696	0
$\{\tau, \pi_i, \theta\}$	-94.529	6	201.058	2.362
$\tau, \theta, \alpha, \beta$	-96.251	4	200.501	5.806

from the $\{\tau, \pi_i, \theta_i\}$ model is 2.362, which can be compared with a χ^2_3 distribution to indicate that a common value of θ can be used for the different doses of the drug. Further for this data, the beta-Poisson model with a logistic tolerance distribution yields a deviance of 5.806 from the $\{\tau, \pi_i, \theta_i\}$ model, which is close to the mean of the χ^2_5 distribution. This comparison indicates that the beta-Poisson model, with a logit function modelling π, provides a good fit for the algae data. These findings are reinforced by the AIC values for the different models. The beta-Poisson distribution, with a logit function modelling the probability of death of a parasite, has the smallest AIC, indicating that this parsimonious model provided the best fit for the algae data.

Malaria Data

The $\{\tau, \alpha, \beta, \theta\}$ beta-Poisson model, with a logit function modelling the probability of death of a parasite, was fitted to the malaria data summarized in Section 2.2.2. 95% Wald intervals were computed for each of the parameters τ, α, β and θ. Profile likelihood plots were also obtained for the parameters and are presented in Figures 4.8 (a), (b), (c) and (d). These plots were used to obtain 95% confidence intervals for the parameters. The estimates of the parameters, along with the standard errors, 95% Wald confidence intervals and 95% profile likelihood intervals are presented in Table 4.6.

The beta-Poisson model fitted to the malaria data was compared with the

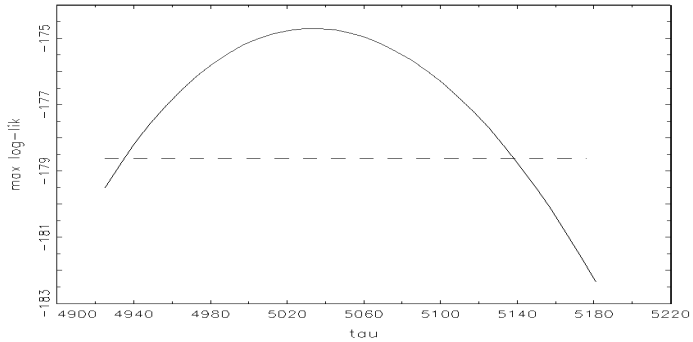


Figure 4.8: (a)

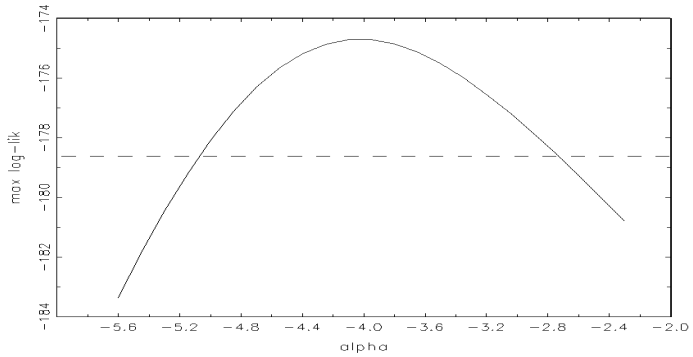


Figure 4.8: (b)

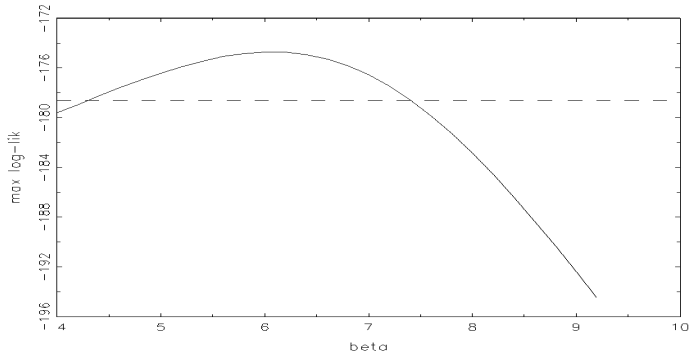


Figure 4.8: (c)

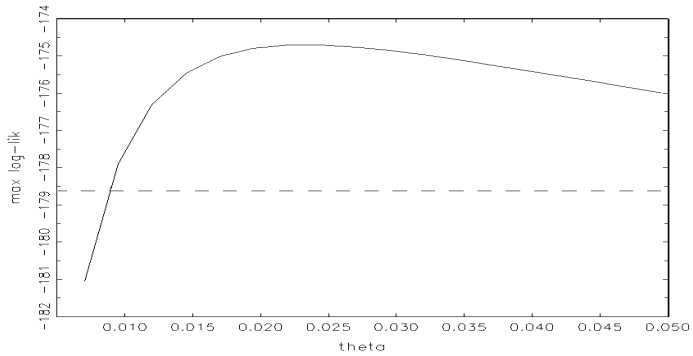


Figure 4.8: (d)

Figure 4.8: Profile likelihood plots for the parameters (a) τ, (b) α, (c) β and (d) θ of the beta-Poisson model with a logit tolerance distribution fitted to the malaria data.

Table 4.6: Results from fitting the beta-Poisson model to the malaria data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	5033.448	35.375	(4966.391;5105.061)	(4933.000;5141.000)
θ	0.023	0.003	(0.006;0.018)	(0.009;*)
α	-4.027	0.290	(-4.615;-3.479)	(-5.152;-2.750)
β	6.072	0.367	(6.060;7.550)	(4.205;7.412)

*This limit of the confidence interval cannot be computed

$\{\tau, \pi_i, \theta_i\}$ and $\{\tau, \pi_i, \theta\}$ models which consider each dose of the drug separately and details regarding the fits are presented in Table 4.7. The deviance of the

Table 4.7: A comparison of the beta-Poisson models fitted to the malaria data.

Model	Maximum log-likelihood	Number of parameters	AIC	Deviance
$\{\tau, \pi_i, \theta_i\}$	-130.276	15	290.551	0
$\{\tau, \pi_i, \theta\}$	-143.820	9	299.640	27.088
$\tau, \theta, \alpha, \beta$	-174.670	4	357.340	88.842

$\{\tau, \pi_i, \theta\}$ model from the $\{\tau, \pi_i, \theta_i\}$ model was 27.088 which, when compared with the mean of a χ^2_6 distribution, suggests that a common parameter θ is not suitable for the data. The beta-Poisson model with a logit tolerance function has a deviance of 49.089 from the $\{\tau, \pi_i, \theta_i\}$ model, which can be compared with the mean of the χ^2_{11} distribution to indicate that it also provides a disappointingly poor fit for the malaria data. This outcome is confirmed by the AIC values.

4.9 Conclusions

The beta-Poisson distribution was introduced and the p.m.f. of the model was investigated by considering approximating the confluent hypergeometric function.

Some properties of the distribution were explored such as the moments and modality. The suitability of the beta-Poisson distribution for modelling overdispersed count data was investigated and the model seemed to accommodate some of the overdispersion in this type of data but for the example chosen it seemed quite poor. The model was then adapted to a dose-response setting and was considered for modelling overdispersed dose-response data from a Wadley's problem setting. The results presented in Section 4.8.7 indicate that the beta-Poisson model adequately accounted for the overdispersion in the algae data. When fitted to the malaria data set, however, it failed to provide a good fit.

Overall, the beta-Poisson distribution was algebraically tractable. In addition to modelling overdispersed count data, the model can be considered a useful tool for accommodating overdispersion in dose-response data arising from a Wadley's problem setting.

Chapter 5

The Multiplicative Binomial Model

5.1 Introduction

A common cause of over- and underdispersion in a binomial experiment is correlation among the responses (Collett, 2003, p.196). Altham (1978) introduced a generalization of the binomial distribution in the context of toxicology studies on litters of rabbits for which the binary outcomes for rabbits in the same litter are correlated. The resulting distribution is the multiplicative binomial model, also termed the Altham distribution by authors such as Zelterman (2004, p.212).

This chapter provides a critique of the multiplicative binomial distribution and builds on the knowledge base of the model as it has not been explored in great detail in the literature. The model has been cited by authors such as Engel and te Brake (1993) and Madden and Hughes (1995), who merely stated that the distribution can be used for modelling overdispersed binary data. Lovison (1997) and Zelterman (2004, Chapter 8) investigated some of the properties of the distribution but generally the model has not been used very much practically. This chapter also provides a preamble for the Altham-Poisson distribution which follows in Chapter 6 and which is an adaptation of the multiplicative binomial

distribution to a Wadley's problem scenario.

A derivation of the p.m.f. and some properties and examples of the multiplicative binomial distribution are presented in Section 5.2. The likelihood function, score functions and information matrix are included in Section 5.3, while methods of estimation are outlined in Section 5.4. Results from fitting the multiplicative binomial distribution to overdispersed binomial-type data are presented in Section 5.5. In Section 5.6 the model is adapted for a dose-response setting and fitted to dose-mortality data. Concluding remarks are contained in Section 5.7.

5.2 Distribution

5.2.1 Derivation

The derivation presented here is that given by Altham (1978) but it has been expanded and clarified to provide insight. Consider a toxicology experiment where the outcomes recorded are the numbers of rabbits in litters of various sizes that are alive at the end of the experiment. For a litter of size n let the binary outcome for the ith rabbit in the litter, $i = 1, \ldots, n$, be

$$X_i = \begin{cases} 0 & \text{if the rabbit lives} \\ 1 & \text{if the rabbit dies.} \end{cases}$$

Then the random variable $Y = \sum_{i=1}^{n} X_i$ represents the number of survivors in that litter and would typically be modelled by a binomial distribution. However, the situation frequently emerges in which the variation in the responses exceeds that of the binomial distribution. Altham (1978) derived the multiplicative binomial distribution for modelling over- and underdispersion by considering a binomial-type experiment in which the binary outcomes $X_1, \ldots, X_n$ for rabbits in a litter of size n have a symmetric joint distribution and are correlated. Altham (1978) regarded the binary responses as generating a 2^n contingency table where the ith margin label is 0 or 1 corresponding to the outcome X_i, $i = 1, \ldots, n$. The

probabilities associated with $X_i,\ i = 1, \ldots, n$, are therefore the probabilities of membership of the 2^n cells of the contingency table and can be modelled using a log-linear model with all third and higher order interactions ignored.

For example, consider a litter of size 3. The 2^3 contingency table can be represented as

Level 0 of X_3

	X_2	
X_1	000	010
	100	110

Level 1 of X_3

	X_2	
X_1	001	011
	101	111

and the following probability statement can be written down:

$$\begin{aligned} \Pr(X_1 = a,\ X_2 = b,\ X_3 = c) &= p_{abc} \text{ for } a, b, c = 0, 1 \\ &\neq \Pr(X_1 = a)\Pr(X_2 = b)\Pr(X_3 = c) \end{aligned}$$

where $\sum_{a=0}^{1}\sum_{b=0}^{1}\sum_{c=0}^{1} p_{abc} = 1$.

Altham (1978) followed Darroch (1974) and used the log-linear multiplicative definition of no three-variable interactions to obtain

$$\ln p_{abc} = \psi_{bc} + \psi_{ac} + \psi_{ab}$$

where for example ψ_{bc} represents the first-order interaction of X_2 and X_3. In generalizing this statement to a litter of size n Altham (1978) obtained the following expression

$$\Pr(X_1 = i_1, \ldots, X_n = i_n) = K \prod_{1 \le r < s \le n} \phi_{i_r i_s}\ , \tag{5.1}$$

where K is a normalizing constant ensuring that all of the probabilities sum to 1, $\phi_{i_r i_s}$ are multiplicative interaction terms and $i_1, \ldots, i_n = 0$ or 1.

In order to derive the p.m.f. of the multiplicative binomial distribution consider the random variable $Y = X_1 + X_2 + \ldots + X_n$ and suppose that a realization of Y yields y ones and hence $n - y$ zeros. Since the 1-1 interactions must be selected

from the y ones and the 0-0 interactions from the $n-y$ zeros and it follows from (5.1) that

$$\begin{aligned}\Pr(Y=y) &= K\begin{pmatrix} n \\ y \end{pmatrix}\phi_{00}^{\begin{pmatrix} n-y \\ 2 \end{pmatrix}}\phi_{01}^{y(n-y)}\phi_{11}^{\begin{pmatrix} y \\ 2 \end{pmatrix}} \\ &= K\begin{pmatrix} n \\ y \end{pmatrix}\phi_{00}^{\frac{(n-y)(n-y-1)}{2}}\phi_{01}^{y(n-y)}\phi_{11}^{\frac{y(y-1)}{2}} \ .\end{aligned}$$

Altham (1978) introduced a reparametrization of this distribution in terms of θ and p where

$$\theta = \frac{\phi_{01}}{\sqrt{\phi_{00}\phi_{11}}} \tag{5.2}$$

and

$$p = \frac{\phi_{11}^{(n-1)/2}}{\phi_{00}^{(n-1)/2}+\phi_{11}^{(n-1)/2}} \ ,$$

which implies that

$$\begin{aligned} q &= 1-p \\ &= \frac{\phi_{00}^{(n-1)/2}}{\phi_{00}^{(n-1)/2}+\phi_{11}^{(n-1)/2}} \ .\end{aligned}$$

This parametrization is useful because p measures the prevalence of a characteristic and θ the strength of the correlation among the responses (Zelterman, 2004, p.219). From the definition of θ in (5.2),

$$\phi_{01}^{y(n-y)} = \theta^{y(n-y)}\phi_{00}^{\frac{y(n-y)}{2}}\phi_{11}^{\frac{y(n-y)}{2}} \ .$$

Hence

$$\begin{aligned}\Pr(Y=y) &= K\begin{pmatrix} n \\ y \end{pmatrix}\phi_{00}^{\frac{(n-y)(n-y-1)}{2}}\theta^{y(n-y)}\phi_{00}^{\frac{y(n-y)}{2}}\phi_{11}^{\frac{y(n-y)}{2}}\phi_{11}^{\frac{y(y-1)}{2}} \\ &= K\begin{pmatrix} n \\ y \end{pmatrix}\phi_{00}^{\frac{(n-1)(n-y)}{2}}\theta^{y(n-y)}\phi_{11}^{\frac{y(n-1)}{2}} \ . \end{aligned} \tag{5.3}$$

Substituting for p and $q = 1 - p$ in (5.3) gives

$$\Pr(Y = y) \;=\; \frac{\left(\begin{array}{c} n \\ y \end{array} \right) p^y q^{n-y} \theta^{y(n-y)}}{f(p, \theta, n)} \;, \tag{5.4}$$

where $f(p, \theta, n)$ is a normalizing constant given by

$$f(p, \theta, n) \;=\; \sum_{y=0}^{n} \left(\begin{array}{c} n \\ y \end{array} \right) p^y q^{n-y} \theta^{y(n-y)} \tag{5.5}$$

with $0 \leq p \leq 1$ and $\theta > 0$. The p.m.f. of Y is of a similar form to that of a binomial random variable but with an attenuating parameter θ which captures the correlation among the responses and therefore allows for under- and overdispersion. Note that the distribution reduces to the binomial distribution when $\theta = 1$. Since the p.m.f. of the multiplicative binomial distribution can be written in the form

$$\Pr(Y = y) \;=\; \exp\left\{ y \ln\left(\frac{p}{1-p} \right) + y(n-y) \ln\theta - A(p, \theta, n) \right\} \tag{5.6}$$

where $A(p, \theta, n) = \ln f(p, \theta, n) - n \ln(1-p) - \ln \left(\begin{array}{c} n \\ y \end{array} \right)$, it is a member of the multivariate exponential family of distributions (McCullagh and Nelder, 1989, p.28).

Molenberghs and Ryan (1999) derived a distribution to model clustered binary data in developmental toxicity studies by conditioning on the value of n_i, $i = 1, \ldots, N$, where n_i represents the number of individuals in the ith cluster and can vary from cluster to cluster. The model they obtained was written in the form (5.6) and is identical to the multiplicative binomial distribution but they did not comment on this fact. Lovison (1997) described an alternative derivation of the multiplicative binomial distribution which is similar to the technique used by Molenberghs and Ryan (1999) and which was based on methods introduced by Cox (1972).

5.2.2 Properties of the Distribution

Factorial Moments

Altham (1978) derived the following expressions for the first and second factorial moments of the multiplicative binomial distribution:

$$\begin{aligned} \mathrm{E}(Y) &= \frac{np(p+q\theta)^{n-1}}{f(p,\theta,n)} f\left(\frac{p}{p+q\theta}, \theta, n-1\right) \\ &= np\tau_1 \ , \end{aligned} \tag{5.7}$$

where $\tau_1 = \dfrac{(p+q\theta)^{n-1}}{f(p,\theta,n)} f\left(\dfrac{p}{p+q\theta}, \theta, n-1\right)$ and

$$\mathrm{E}[Y(Y-1)] = \frac{n(n-1)p^2(p+q\theta^2)^{n-2}}{f(p,\theta,n)} f\left(\frac{p}{p+q\theta^2}, \theta, n-2\right) \ . \tag{5.8}$$

A simple manipulation of these moments yields the variance of the distribution as

$$\begin{aligned} \mathrm{Var}(Y) &= \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2 \\ &= \mathrm{E}[Y(Y-1)] + \mathrm{E}(Y) - [\mathrm{E}(Y)]^2 \\ &= np\tau_1 \left[1 + p\left(\frac{-\tau_2}{\tau_1}\right)\right] \ , \end{aligned} \tag{5.9}$$

where

$$\tau_2 = \frac{n(p+q\theta)^{2n-2}}{f^2(p,\theta,n)} f^2\left(\frac{p}{p+q\theta}, \theta, n-1\right) - \frac{(n-1)(p+q\theta^2)^{n-2}}{f(p,\theta,n)} f\left(\frac{p}{p+q\theta^2}, \theta, n-2\right) \ .$$

Similar expressions were also obtained by Lovison (1997) and are not very tractable since they are expressed in terms of the normalizing constant.

More general formulae for the factorial moments of the multiplicative binomial distribution are now considered here for completeness. Concise expressions for the skewness and kurtosis of the distribution are also sought in order to determine how these statistics depend on the parameters. A general expression for the factorial moments of the multiplicative binomial distribution has not been derived in previous studies.

Consider the expectation

$$E[Y^{(r)}(n-Y)^{(s)}] = E[Y(Y-1)\dots(Y-r+1)(n-Y)(n-Y-1)\dots(n-Y-s+1)]$$

for the three scenarios $r = s$, $s > r$ and $r > s$.

When $r = s$

$$
\begin{aligned}
& E[Y^{(r)}(n-Y)^{(r)}] = E[Y(Y-1)\dots(Y-r+1)(n-Y)(n-Y-1)\dots(n-Y-r+1)] \\
= & \sum_{y=0}^{n} \frac{y(y-1)\dots(y-r+1)(n-y)(n-y-1)\dots(n-y-r+1)n!}{y!(n-y)!} \frac{p^y q^{(n-y)}\theta^{y(n-y)}}{f(p,\theta,n)} \\
= & \frac{1}{f(p,\theta,n)} \sum_{y=r}^{n-r} \frac{n!}{(y-r)!(n-r-y)!} p^y q^{n-y}\theta^{y(n-y)} \\
= & \frac{1}{f(p,\theta,n)} \sum_{y^*=0}^{n-2r} \frac{n!}{y^*!(n-2r-y^*)!} p^{r+y^*} q^{n-y^*-r}\theta^{(y^*+r)(n-y^*-r)} \text{ , where } y^* = y-r \\
= & \frac{n(n-1)\dots(n-2r+1)p^r q^r \theta^{r(n-r)}}{f(p,\theta,n)} \sum_{y^*=0}^{n-2r} \frac{(n-2r)!}{y^*!(n-2r-y^*)!} p^{y^*} q^{n-2r-y^*}\theta^{y^*(n-2r-y^*)} \\
= & n(n-1)\dots(n-2r+1)p^r q^r \theta^{r(n-r)} \frac{f(p,\theta,n-2r)}{f(p,\theta,n)} \ .
\end{aligned}
$$

If $s > r$ the expectation becomes

$$
\begin{aligned}
& E[Y^{(r)}(n-Y)^{(s)}] = E[Y(Y-1)\dots(Y-r+1)(n-Y)(n-Y-1)\dots(n-Y-s+1)] \\
= & \sum_{y=0}^{n} \frac{y(y-1)\dots(n-r+1)(n-y)(n-y-1)\dots(n-y-s+1)n!}{y!(n-y)!} \frac{p^y q^{n-y}\theta^{y(n-y)}}{f(p,\theta,n)} \\
= & \frac{1}{f(p,\theta,n)} \sum_{y=r}^{n-s} \frac{n!}{(y-r)!(n-y-s)!} p^y q^{n-y}\theta^{y(n-y)} \\
= & \frac{1}{f(p,\theta,n)} \sum_{y^*=0}^{n-r-s} \frac{n!}{y^*!(n-r-s-y^*)!} p^{r+y^*} q^{n-y^*-r}\theta^{(y^*+r)(n-y^*-r)} \ , \text{ where } y^* = y-r \\
= & \frac{n(n-1)\dots(n-s-r+1)p^r q^s \theta^{r(n-r)}}{f(p,\theta,n)} \sum_{y^*=0}^{n-r-s} \frac{(n-r-s)!(p\theta^{(s-r)})^{y^*} q^{n-r-s-y^*}\theta^{y^*(n-r-s-y^*)}}{y^*!(n-r-s-y^*)!} \\
= & \frac{n(n-1)\dots(n-r-s+1)p^r q^s \theta^{r(n-r)}(p\theta^{s-r}+q)^{n-s-r}}{f(p,\theta,n)} \\
& \sum_{y^*=0}^{n-s-r} \binom{n-s-r}{y^*} \left(\frac{p\theta^{s-r}}{p\theta^{s-r}+q}\right)^{y^*} \left(\frac{q}{p\theta^{s-r}+q}\right)^{n-s-r-y^*} \theta^{y^*(n-s-r-y^*)} \\
= & \frac{n(n-1)\dots(n-s-r+1)p^r q^s \theta^{r(n-r)}(p\theta^{s-r}+q)^{n-s-r}}{f(p,\theta,n)} f\left(\frac{p\theta^{s-r}}{p\theta^{s-r}+q}, \theta, n-s-r\right) \ .
\end{aligned}
$$

Finally, when $r > s$

$$
E[Y^{(r)}(n-Y)^{(s)}] = E[Y(Y-1)\dots(Y-r+1)(n-Y)(n-Y-1)\dots(n-Y-s+1)]
$$

$$
\begin{aligned}
&= \sum_{y=0}^{n} \frac{y(y-1)\dots(n-r+1)(n-y)(n-y-1)\dots(n-y-s+1)n!}{y!(n-y)!} \frac{p^y q^{n-y}\theta^{y(n-y)}}{f(p,\theta,n)} \\
&= \frac{1}{f(p,\theta,n)} \sum_{y=r}^{n-s} \frac{n!}{(y-r)!(n-y-s)!} p^y q^{n-y}\theta^{y(n-y)} \\
&= \frac{1}{f(p,\theta,n)} \sum_{y^*=0}^{n-r-s} \frac{n!}{y^*!(n-r-s-y^*)!} p^{r+y^*} q^{n-y^*-r}\theta^{(y^*+r)(n-y^*-r)} \text{ , where } y^* = y-r \\
&= \frac{n(n-1)\dots(n-r-s+1)p^r q^s \theta^{s(n-s)}(p+q\theta^{r-s})^{n-s-r}}{f(p,\theta,n)} \\
&\quad \sum_{y^*=0}^{n-r-s} \frac{(n-r-s)!}{y^*!(n-r-s-y^*)!} \left(\frac{p}{p+q\theta^{r-s}}\right)^{y^*} \left(\frac{q\theta^{r-s}}{p+q\theta^{r-s}}\right)^{n-r-s-y^*} \theta^{y^*(n-r-s-y^*)} \\
&= \frac{n(n-1)\dots(n-r-s+1)p^r q^s \theta^{s(n-y)}(p+q\theta^{r-s})^{n-s-r}}{f(p,\theta,n)} f\left(\frac{p}{p+q\theta^{r-s}}, \theta, n-r-s\right)
\end{aligned}
$$

The factorial moments can be used to calculate the skewness and kurtosis of the multiplicative binomial distribution. The skewness of the multiplicative binomial distribution is calculated as

$$
\begin{aligned}
\frac{\mathrm{E}[(Y-\mu)]^3}{\sigma^3} &= \frac{\mathrm{E}(Y^3) - 3\mu\mathrm{E}(Y^2) + 2\mu^3}{\sigma^3} \\
&= \left\{ \frac{2n^3p^2(p+q\theta)^{3(n-1)} f^3\left(\frac{p}{p+q\theta}, \theta, n-1\right)}{f^3(p,\theta,n)} - \frac{2n^2p^2(p+q\theta)^{2(n-1)} f^2\left(\frac{p}{p+q\theta}, \theta, n-1\right)}{f^2(p,\theta,n)} \right. \\
&\quad - \frac{np^2(n-1)(n-2)q\theta^{n-1}(p+q\theta)^{n-3} f\left(\frac{p}{p+q\theta}, \theta, n-3\right)}{f(p,\theta,n)} \\
&\quad \left. \frac{n^2p^2(n-1)(p+q\theta^2)^{n-2} f\left(\frac{p}{p+q\theta^2}, \theta, n-2\right)}{f(p,\theta,n)} \right\} / \\
&\quad \left\{ f^2(p,\theta,n) \left[\frac{np(p+q\theta)^{n-1} f\left(\frac{p}{p+q\theta}, \theta, n-1\right)}{f(p,\theta,n)} + \frac{n^2p^2(p+q\theta)^{2(n-1)} f^2\left(\frac{p}{p+q\theta}, \theta, n-1\right)}{f^2(p,\theta,n)} \right. \right. \\
&\quad \left. \left. + \frac{n(n-1)p^2(p+q\theta^2)^{n-2} f\left(\frac{p}{p+q\theta^2}, \theta, n-2\right)}{f(p,\theta,n)} \right]^{3/2} \right\} .
\end{aligned}
$$

Similarly the kurtosis of the distribution is

$$
\frac{\mathrm{E}[(Y-\mu)]^4}{\sigma^4} = \frac{\mathrm{E}(Y^4) + 6\mu\mathrm{E}(Y^2) - 4\mu\mathrm{E}(Y^3) - 3\mu^4}{\sigma^4}
$$

$$
\begin{aligned}
= \; & -f(p,\theta,n)\left\{(n^3-6n^2+11n-6)p^2q\theta^n(p+q\theta)^4(p+q\theta^2)^n f\left(\frac{p}{p+q\theta^2},\theta,n-4\right)\right. \\
& f^4(p,\theta,n)+(n-1)p\theta(p+q\theta)^3(p+q\theta^2)^{n+2}f\left(\frac{p}{p+q\theta^2},\theta,n-2\right) \\
& f^3(p,\theta,n)\left[2n(2n-3)p(p+q\theta)^n f\left(\frac{p}{p+q\theta},\theta,n-1\right)-(n^2-2)(p+q\theta)f(p,\theta,n)\right] \\
& +(p+q\theta)^n(p+q\theta^2)^4\left[3n^3p^3\theta(p+q\theta)^{3n}f^4\left(\frac{p}{p+q\theta},\theta,n-1\right)+4n^2p^2\theta(p+q\theta)^{2n+1}\right. \\
& f^3\left(\frac{p}{p+q\theta},\theta,n-1\right)f(p,\theta,n)-n(n+9)p\theta(p+q\theta)^{n+2}f^2\left(\frac{p}{p+q\theta},\theta,n-1\right) \\
& f^2(p,\theta,y)+(n^3-7n+6)pq\theta^n(p+q\theta)f\left(\frac{p}{p+q\theta},\theta,n-3\right)f^3(p,\theta,n) \\
& f\left(\frac{p}{p+q\theta},\theta,n-1\right)f^2(p,\theta,n)\left[-4n(n^2-3n+2)p^2q\theta^n(p+q\theta)^n f\left(\frac{p}{p+q\theta},\theta,n-3\right)\right. \\
& \left.\left.+2(n+2)\theta(p+q\theta)^3 f(p,\theta,n)\right]\right]f(p,\theta,n)\Big\}\Big/ \\
& \left\{np\theta\left[(n-1)p(p+q\theta)^2(p+q\theta^2)^n f\left(\frac{p}{p+q\theta^2},\theta,n-2\right)f^2(p,\theta,n)\right.\right. \\
& (p+q\theta)^n(p+q\theta^2)^2 f\left(\frac{p}{p+q\theta},\theta,n-1\right)\left[np(p+q\theta)^n f\left(\frac{p}{p+q\theta},\theta,n-1\right)\right. \\
& \left.\left.\left.-(p+q\theta)f(p,\theta,n)\right]f(p,\theta,n)\right]^2\right\} \; .
\end{aligned}
$$

It is disappointing that the expressions for the skewness and kurtosis of the multiplicative binomial distribution are not algebraically tractable. They can, however, be easily programmed and computed via the factorial moments.

Generating Functions

The probability generating function of a distribution is defined as the expectation $\mathrm{E}(s^Y)$ where s is an indeterminant and for the multiplicative binomial distribution it is derived as

$$
\begin{aligned}
E(s^Y) &= \sum_{y=0}^{n}\left(\begin{array}{c} n \\ y \end{array}\right)s^y p^y q^{n-y}\theta^{y(n-y)}/f(p,\theta,n) \\
&= (ps+q)^{-n}\sum_{y=0}^{n}\left(\begin{array}{c} n \\ y \end{array}\right)\left(\frac{ps}{ps+q}\right)^y\left(\frac{q}{ps+q}\right)^{n-y}\theta^{y(n-y)}/f(p,\theta,n) \\
&= (ps+q)^{-n}f\left(\frac{ps}{ps+q},\theta,n\right)/f(p,\theta,n) \; .
\end{aligned}
$$

The moment generating function $\mathrm{E}(e^{tY})$, where t is an indeterminant, follows by setting $s = e^t$. These generating functions can be used to obtain the moments of the multiplicative binomial distribution but, since they contain the normalizing constant, it is easier to obtain the moments directly from the expressions for the factorial moments.

Modality

The modality of the multiplicative binomial distribution is investigated in this study for the purpose of completeness. Although this modality was observed by Zelterman (2004, p.213), it has not been explored in detail in any of the previous studies. Consider a random variable Y which follows a multiplicative binomial distribution with parameters p and θ and let y and $y+1$ denote two consecutive realizations of Y. Then the ratio of the probabilities of y and $y+1$ is written as

$$\frac{\Pr(Y=y)}{\Pr(Y=y+1)} \;=\; \frac{(y+1)(1-p)}{(n-y)p\theta^{n-2y-1}} \;,$$

which exceeds unity if and only if

$$\frac{p(n-y)}{(1-p)(y+1)}\theta^{n-2y-1} \;<\; 1 \;.$$

Taking the logarithm of the above expression results in the following inequality:

$$(2y+1-n)\ln\theta \;>\; \ln\left(\frac{n-y}{y+1}\right) + \ln\left(\frac{p}{1-p}\right) \;. \tag{5.10}$$

For the purpose of establishing the modality of the multiplicative binomial distribution the functions on each side of the inequality in expression (5.10) are treated as continuous functions of y. Observe that if a continuous distribution is unimodal its discrete version will have at most two consecutive modes, which effectively equates to unimodality. Consider the inequality in equation (5.10) for values of θ greater than one and values of θ less than one. When θ exceeds unity $\ln\theta$ is positive, indicating that $(2y+1-n)\ln\theta$ is a linearly increasing function of y. Since $(n-y)$ decreases and $(y+1)$ increases as y increases, the function on the

right hand side of the inequality in expression (5.10) is a decreasing function of y. As a result the functions on the left and right hand sides of equation (5.10) have at most one point of intersection. Figure 5.1(a) includes a graphical representation of these two functions for $n = 20,\ p = 0.48$ and $\theta = 1.1$. The solid line depicts $(2y + 1 - n)\ln\theta$ while the dashed line illustrates the function on the right hand side of the inequality in (5.10). Since the two functions intersect at most once it can be deduced that the multiplicative binomial distribution is unimodal when θ is greater than one.

When θ is less than one, $(2y+1-n)\ln\theta$ is a decreasing function of y. The two functions on either side of the inequality in expression (5.10) are therefore both decreasing and as a result they can intersect more than once. This is illustrated in Figure 5.1(b) for $n = 20,\ p = 0.48$ and $\theta = 0.85$. Since it is possible for the two functions on either side of the inequality to have more than one point of intersection, the multiplicative binomial distribution can have more than one mode when θ is less than one.

5.2.3 Examples

Graphical representations of the p.m.f., together with the mean, variance, skewness (γ) and kurtosis (κ) of Y are considered for various values of p and θ. Figure 5.2 includes an illustrative example of the unimodal p.m.f. of Y when $\theta > 1$. Various forms of the p.m.f. of Y when $\theta < 1$ are presented in Figure 5.3.

Preliminary investigations of the p.m.f. of the multiplicative binomial distribution were carried out and the examples presented in Figures 5.2 and 5.3 are typical representations of the p.m.f. for various values of p and θ. It was observed that for large values of n the p.m.f. of the multiplicative binomial distribution seems to be distinctly U-shaped when θ is much less than one. This is illustrated in Figures 5.3 (a) and 5.3 (b). Zelterman (2004, p.213) plotted the p.m.f. of the multiplicative binomial distribution on a log scale for various values of θ when $n = 10$ and

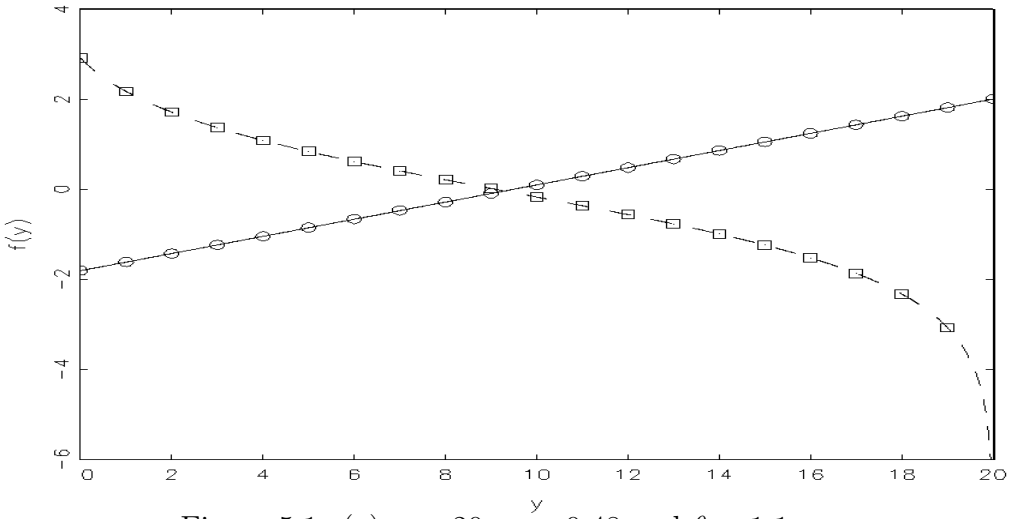


Figure 5.1: (a) $n = 20,\ p = 0.48$ and $\theta = 1.1$

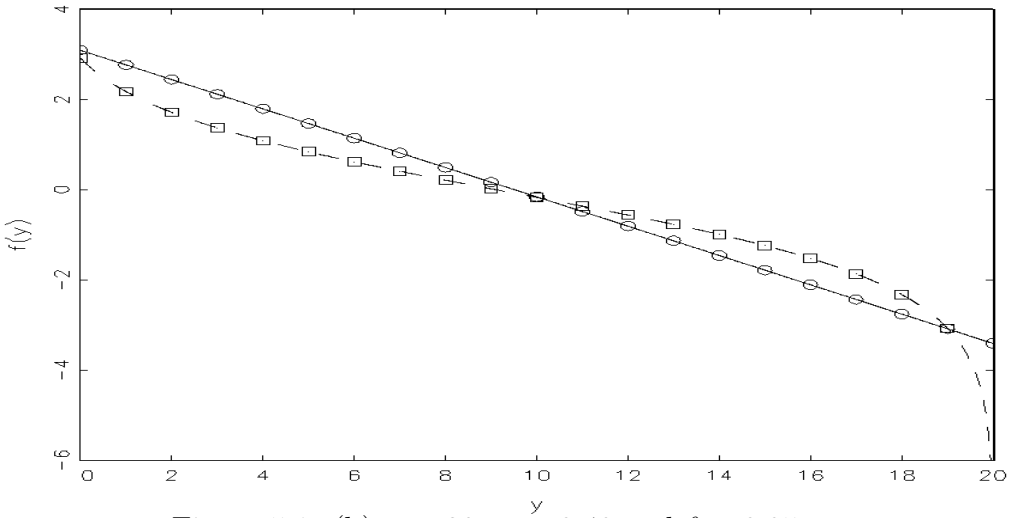


Figure 5.1: (b) $n = 20,\ p = 0.48$ and $\theta = 0.85$

Figure 5.1: Illustration of the modality of the multiplicative binomial distribution (a) when θ is greater than 1 and (b) when θ is less than 1.

$p = 0.3$. The examples considered by Zelterman (2004, p.213) seem to support this tendency of the p.m.f. to be U-shaped when θ is much less than one. It was noted that the probability of observing values between 0 and n seemed to increase when θ was closer to one.

Altham (1978) stated that the multiplicative binomial distribution exhibits

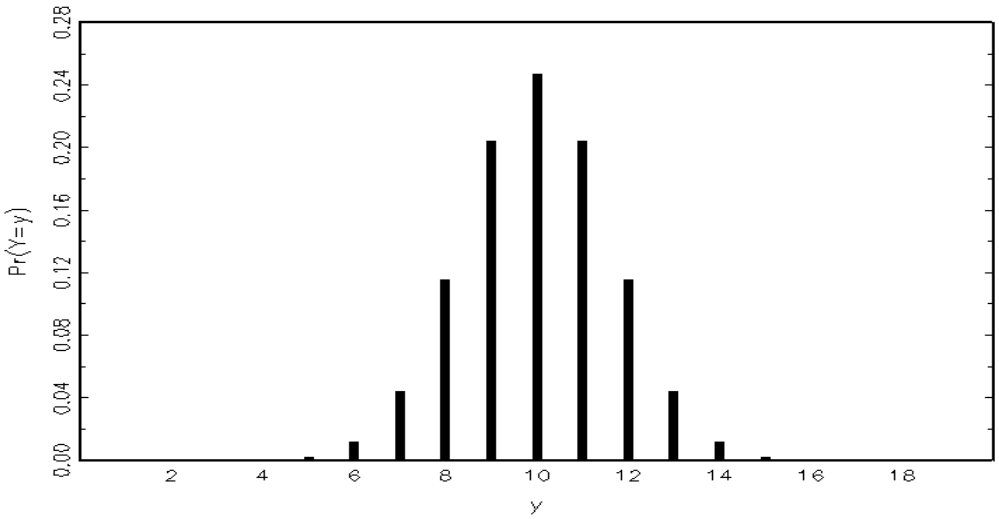


Figure 5.2: $n = 20,\ p = 0.48,\ \theta = 1.1$, E($Y$)=9.792, Var($Y$)=2.592, γ=-0.618, κ=-7563.59 .

underdispersion when θ is greater than one and overdispersion when θ is less than one. This relationship between the value of θ and the dispersion of the data seems intuitively attractive since the multiplicative binomial distribution is reduced to the binomial distribution when $\theta = 1$. Lovison (1997) investigated this claim and remarked that it is unsubstantiated. Figure 5.4 contains a plot of the variance of Y over a range of values of θ for $n = 5$ and $p = 0.01$. The dashed line represents the binomial variance for these values of n and p. The graph demonstrates that for values of θ exceeding 1 the variance of the multiplicative binomial distribution can exceed that of the binomial distribution, thereby supporting the claim made by Lovison (1997).

5.3 Likelihood, Score Functions and Information Matrix

The likelihood and score functions are presented since they are required for estimating the parameters of the multiplicative binomial distribution and also for completeness. The information matrix of the multiplicative binomial distribu-

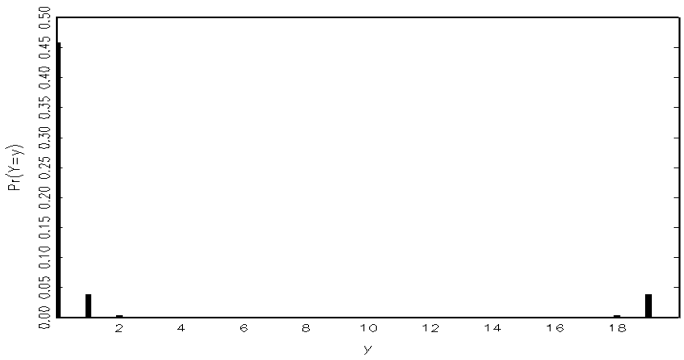


Figure 5.3: (a) $n = 20,\ p = 0.48,\ \theta = 0.75$, E($Y$)=3.450, Var($Y$)=55.486, γ=1.620, κ=2.812.

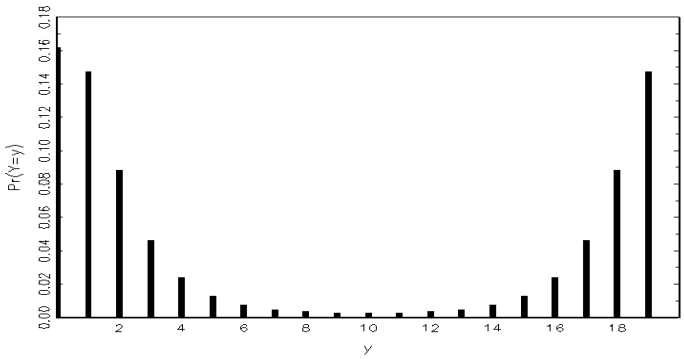


Figure 5.3: (b) $n = 20,\ p = 0.48,\ \theta = 0.85$, E($Y$)=4.777, Var($Y$)=49.159, γ=1.244, κ=-0.158.

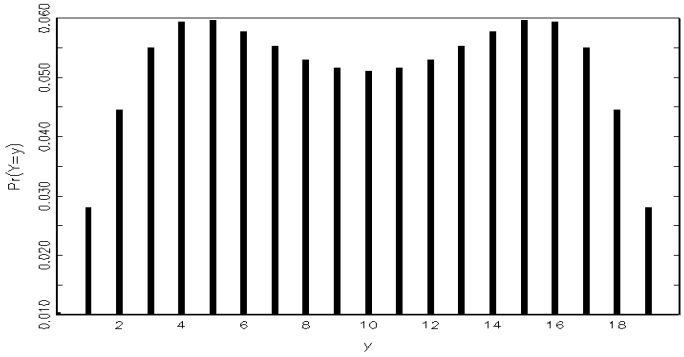


Figure 5.3: (c) $n = 20,\ p = 0.48,\ \theta = 0.9$, E($Y$)=7.808, Var($Y$)=25.581, γ=0.292, *kappa*=-39.669.

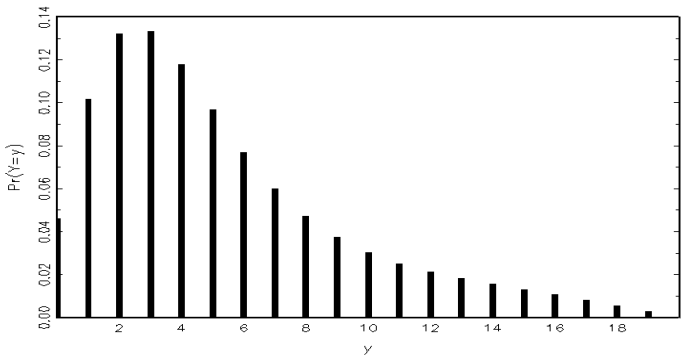


Figure 5.3: (d) $n = 20,\ p = 0.55,\ \theta = 0.9$, E($Y$)=14.733, Var($Y$)=16.357, γ=-1.367, κ=-1052.96.

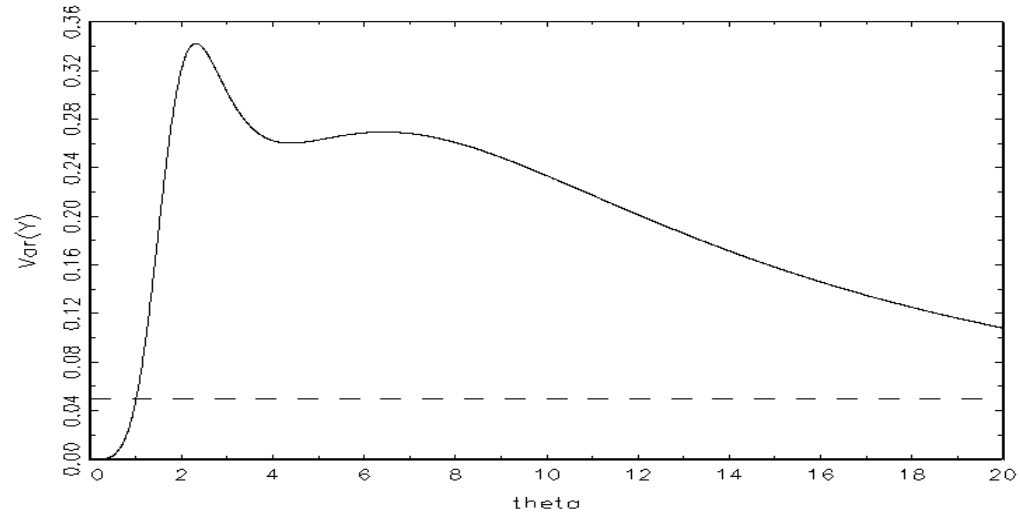


Figure 5.4: Plot indicating the changes of the variance of the multiplicative binomial random variable, Y, over different values of θ, with $n = 5$ and $p = 0.01$.

tion has not been derived in previous studies and is considered here to determine whether it has an attractive form and because it can be used to obtain the standard errors of the parameter estimates.

5.3.1 Likelihood and Score Functions

The likelihood function of the multiplicative binomial distribution is

$$L(p, \theta; y) = \frac{\binom{n}{y} p^y (1-p)^{n-y} \theta^{y(n-y)}}{f(p, \theta, n)} ,$$

where $f(p, \theta, n)$ is the normalizing constant previously defined in expression (5.5). The log-likelihood function is therefore written as

$$\ell(p, \theta; y) = \ln \binom{n}{y} + y \ln p + (n-y) \ln(1-p) + y(n-y) \ln \theta - \ln f(p, \theta, n) . \tag{5.11}$$

The first derivatives of the log-likelihood function with respect to the parameters p and θ give the score functions for the multiplicative binomial distribution.

For a single observation from this distribution the score functions for p and θ are derived below. Thus

$$\begin{aligned}\frac{\partial \ell(p,\theta;y)}{\partial p} &= \frac{\partial}{\partial p}\left\{\ln\binom{n}{y} + y\ln p + (n-y)\ln(1-p) + y(n-y)\ln\theta - \ln f(p,\theta,n)\right\} \\ &= \frac{y}{p} - \frac{n-y}{1-p} - \frac{\partial f(p,\theta,n)}{\partial p}\ .\end{aligned}$$

However, since E(U)=0,

$$\mathrm{E}\left(\frac{\partial \ell(p,\theta;Y)}{\partial p}\right) = \frac{\mathrm{E}(Y)-np}{p(1-p)} - \frac{\partial \ln f(p,\theta,n)}{\partial p} = 0$$

and therefore

$$\frac{\partial \ell(p,\theta;y)}{\partial p} = \frac{y-\mathrm{E}(Y)}{p(1-p)}\ . \tag{5.12}$$

Similarly

$$\frac{\partial \ell(p,\theta;y)}{\partial \theta} = \frac{y(n-y)}{\theta} - \frac{\partial f(p,\theta,n)}{\partial \theta}$$

and since E(U) = 0 it follows that

$$\frac{\partial \ell(p,\theta;y)}{\partial \theta} = \frac{y(n-y)}{\theta} - \frac{\mathrm{E}[Y(n-Y)]}{\theta}\ . \tag{5.13}$$

These score functions were derived by Altham (1978) for the purpose of estimating the parameters of the multiplicative binomial distribution.

5.3.2 Information Matrix

The information matrix can be written as the expectation of the pairwise products of the score functions. For the parameters p and θ in the present model the information matrix can be obtained by using

$$\mathfrak{I} = E\begin{bmatrix} \left(\frac{\partial \ell(p,\theta;Y)}{\partial p}\right)^2 & \frac{\partial \ell(p,\theta;Y)}{\partial p}\frac{\partial \ell(p,\theta;Y)}{\partial \theta} \\ \frac{\partial \ell(p,\theta;Y)}{\partial p}\frac{\partial \ell(p,\theta;Y)}{\partial \theta} & \left(\frac{\partial \ell(p,\theta;Y)}{\partial \theta}\right)^2 \end{bmatrix}\ .$$

Consider the individual entries of the above matrix for a single observation. Then

$$\begin{aligned} \mathrm{E}\left[\left(\frac{\partial \ell(p,\theta;Y)}{\partial p}\right)^2\right] &= \mathrm{E}\left[\frac{[Y-\mathrm{E}(Y)]^2}{p^2(1-p)^2}\right] \\ &= \frac{\mathrm{Var}(Y)}{p^2(1-p)^2} \end{aligned}$$

$$\begin{aligned} \mathrm{E}\left(\frac{\partial \ell(p,\theta;Y)}{\partial p}\frac{\partial \ell(p,\theta;Y)}{\partial \theta}\right) &= \frac{-n[\mathrm{E}(Y)]^2+\mathrm{E}(Y^2)\mathrm{E}(Y)+n\mathrm{E}(Y^2)-\mathrm{E}(Y^3)}{\theta p(1-p)} \\ &= \frac{\mathrm{E}[Y^2(n-Y)]-\mathrm{E}[Y(n-Y)]\mathrm{E}(Y)}{\theta p(1-p)} \\ &= \frac{\mathrm{Cov}[Y(n-Y),Y]}{\theta p(1-p)} \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}\left[\left(\frac{\partial \ell(p,\theta;Y)}{\partial \theta}\right)^2\right] &= \mathrm{E}\left\{\frac{\{[Y(n-Y)]-\mathrm{E}[Y(n-Y)]\}^2}{\theta^2}\right\} \\ &= \frac{\mathrm{Var}[Y(n-Y)]}{\theta^2}\,. \end{aligned}$$

Thus the information matrix for the multiplicative binomial distribution can be written as

$$\mathfrak{I} = \begin{bmatrix} \dfrac{\mathrm{Var}(Y)}{p^2(1-p)^2} & \dfrac{\mathrm{Cov}[Y(n-Y),Y]}{\theta p(1-p)} \\ \dfrac{\mathrm{Cov}[Y(n-Y),Y]}{\theta p(1-p)} & \dfrac{\mathrm{Var}[Y(n-Y)]}{\theta^2} \end{bmatrix}. \tag{5.14}$$

This matrix has a simple and elegant form and it is preferable to use the information matrix rather than the observed Fisher information matrix for computing the standard errors of the parameter estimates.

5.4 Estimation

5.4.1 Maximum Likelihood Estimation

Methods for estimating the parameters of the multiplicative binomial distribution are explored for completeness and in order to advise practitioners on which method

is most suitable. Consider a random sample of observations from the multiplicative binomial distribution and assume that the response y_i is observed k_i times, where $\sum_{i=1}^{d} k_i = k$ and $i = 1, \ldots, d$. Then the log-likelihood function for this random sample is written as

$$\ell(p, \theta; \underline{y}) = \sum_{i=1}^{d} k_i \left[\ln \begin{pmatrix} n \\ y_i \end{pmatrix} + y_i \ln p + (n - y_i) \ln(1 - p) + y_i(n - y_i) \ln \theta - \ln f(p, \theta, n) \right] .$$

The maximum likelihood estimates of the parameters p and θ cannot be written down explicitly due to the intractability of the normalizing constant in the log-likelihood function. As a result three methods for estimating the parameters of the distribution are introduced and investigated. These methods include Altham's (1978) method, a method proposed by Lindsey and Mersch (1992) for estimating the normalizing constant and direct optimization of the likelihood function.

Altham's Method

Consider the score functions from expressions (5.12) and (5.13) for a random sample of observations and let ℓ_i denote the log-likelihood for response y_i, $i = 1, \ldots, d$. Altham (1978) derived the maximum likelihood equations for the parameters p and θ by equating the score functions to zero. Consider the score function for parameter p. Then

$$\sum_{i=1}^{d} k_i \frac{\partial \ell_i}{\partial p} = 0 ,$$

which implies that

$$\sum_{i=1}^{d} k_i [y_i - \mathrm{E}(Y)] = 0$$

and thus

$$\begin{aligned} \mathrm{E}(Y) &= \sum_{i=1}^{d} \frac{y_i k_i}{k} \\ &= \frac{np(p + q\theta)^{n-1}}{f(p, \theta, n)} f\left(\frac{p}{p + q\theta}, \theta, n - 1 \right) . \end{aligned} \tag{5.15}$$

Equating the score function for θ to zero gives

$$\sum_{i=1}^{d} k_i \frac{\partial \ell_i}{\partial \theta} \quad = \quad 0 \ ,$$

which yields

$$\begin{aligned} & \mathrm{E}[Y(n-Y)] = \sum_{i=1}^{d} \frac{y_i(n-y_i)k_i}{k} \\ = \quad & \frac{n^2 p(p+q\theta)^{n-1}}{f(p,\theta,n)} f\left(\frac{p}{p+q\theta}, \theta, n-1\right) - \frac{np(p+q\theta)^{n-1}}{f(p,\theta,n)} f\left(\frac{p}{p+q\theta}, \theta, n-1\right) \\ & - \frac{np(n-1)(p+q\theta)^{n-2}}{f(p,\theta,n)} f\left(\frac{p}{p+q\theta^2}, \theta, n-2\right) \ . \end{aligned} \tag{5.16}$$

Altham (1978) thus demonstrated that maximum likelihood estimation of p and θ for the multiplicative binomial distribution is equivalent to the method of moments. The parameters p and θ are estimated by equating the moments $\mathrm{E}(Y)$ and $\mathrm{E}[Y(n-Y)]$ of the multiplicative binomial distribution to the equivalent observed moments. She suggested an iterative procedure for estimating the parameters with a starting value of 1 for θ and the estimated binomial probability of success as a starting value for p. The initial value for θ is substituted into (5.15) to yield an updated estimate of p, which is in turn substituted into (5.16) to find an updated estimate of θ. The algorithm oscillates between estimates of p and θ until it converges.

Method of Lindsey and Mersch

Lindsey and Mersch (1992) discussed a method for fitting discrete probability distributions which are members of the exponential family of distributions. Their discrete distribution version of estimation is drawn upon in the present study.

The general method proceeds as follows: Suppose that y_i is a realization of the discrete variable Y and that it is observed k_i times, where $\sum_{i=1}^{d} k_i = k$ and $i = 1, \ldots, d$. Then according to Lindsey and Mersch (1992) the likelihood function

is proportional to

$$L(\underline{\psi}; y) \quad \propto \quad \prod p_i^{y_i} \ , \tag{5.17}$$

where $p_i = \Pr(Y = y_i)$ and $\underline{\psi}$ is the m-dimensional vector of parameters in the distribution of Y. The likelihood function in expression (5.17) is simply the p.m.f. of a general multinomial distribution. For a member of the exponential family the probability p_i can be written as

$$p_i \quad = \quad \exp\{\sum_{j=1}^{p} t_j(y_i)\psi_j + c(\underline{\psi}) + d(y_i)\} \ ,$$

where $c(\underline{\psi})$ is the normalizing constant, $t_j(.)$ are the sufficient statistics for the parameters and $d(y_i)$ is an arbitrary function of the observations. Since the multinomial distribution can be maximized as a log-linear model provided that an intercept term is included in the model (Dobson, 2002, p.136), Lindsey and Mersch (1992) propose maximizing the likelihood in expression (5.17) as a Poisson likelihood with parameter

$$\lambda_i \quad = \quad \exp\{\sum_{j=1}^{p} t_j(y_i)\psi_j + \psi_0 + d(y_i)\} \ ,$$

where ψ_0 replaces $c(\underline{\psi})$. They then draw on the fact that Y is a member of the exponential family of distributions to fit the sufficient statistics as explanatory variables in the Poisson regression.

The method proposed by Lindsey and Mersch (1992) can be used to estimate the parameters of the multiplicative binomial distribution by writing the parameter λ_i in the Poisson likelihood in terms of p and θ as

$$\ln \lambda_i \quad = \quad \ln \begin{pmatrix} n \\ y_i \end{pmatrix} + y_i \ \text{logit}(p) + y_i(n - y_i) \ln \theta + \theta_0 \ ,$$

where θ_0 is the intercept term which corresponds to $n \ln(1 - p) - f(p, \theta, n)$. The multiplicative binomial distribution can therefore be fitted using a Poisson regression with explanatory variables y_i and $y_i(n - y_i)$, $i = 1, \ldots, d$ and parameters logit p, $\ln \theta$ and θ_0.

Lindsey and Altham (1998) used the method for estimating the normalizing constant presented by Lindsey and Mersch (1992) to fit the multiplicative binomial distribution as a log-linear model, with the parameters p and θ both modelled as functions of n.

Direct Maximization of the Likelihood Function

The log of the likelihood function can be maximized directly using numerical methods such as those employed by the optimization routine OPTMUM in the programming language Gauss. Since this routine requires starting values for the parameter estimates the starting values proposed by Altham (1978) can be utilized.

The three methods introduced in this study will be compared and evaluated using an example in Section 5.5.

5.4.2 Inference

The standard errors of the estimates of p and θ can be found by substituting these estimates into the inverse of the information matrix and then taking the square roots of the resulting diagonal elements.

Confidence intervals for the parameters can be computed using Wald intervals. The 95% Wald interval for the parameter θ is $\hat{\theta} \pm 1.96\ se(\hat{\theta})$, where $se(\hat{\theta})$ is the standard error of the estimate of θ (Kutner, Nachtsheim, Neter and Li, 2005, p.579). Similarly the 95% Wald interval can be constructed for the parameter p. Note that for the multiplicative binomial distribution the restrictions on the parameter space are $0 \leq p \leq 1$ and $\theta > 0$.

An alternative method for obtaining 95% confidence intervals for the parameter estimates is to use the profile likelihood functions. Profile likelihoods for the parameters of the multiplicative binomial distribution have not been considered in any of the earlier literature. Suppose that $\ell(\hat{p}, \hat{\theta}; \underline{y})$ is the maximum log-likelihood of the multiplicative binomial distribution and let $\ell_P(\theta)$ denote the profile log-likelihood for the parameter θ. The 95% confidence interval for the parameter θ is

then given by $[\theta_1, \theta_2]$ where θ_1 and θ_2 satisfy the condition

$$\ell_P(\theta_1) = \ell_P(\theta_2) = \ell(\hat{p}, \hat{\theta}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05}$$

(Azzalini, 1996, p.146). Similarly the 95% confidence interval for p can be obtained by plotting the profile likelihood $\ell_P(p)$ for p.

5.4.3 Goodness of Fit

The adequacy of the fit of the multiplicative binomial distribution to count data can be assessed in a number of ways. Firstly, a chi-squared goodness of fit statistic can be used to determine whether the data constitute a random sample from the multiplicative binomial distribution. A significant statistic indicates that the multiplicative binomial distribution is not appropriate for the data.

In order to determine if the binomial would be a more suitable model for a given data set, the following hypothesis test can be performed

$$\begin{aligned} \mathrm{H}_0 &: \theta = 1 \\ \mathrm{H}_1 &: \theta \neq 1 \ . \end{aligned}$$

The appropriate test statistic is

$$W = -2[\ell(\tilde{p}, \tilde{\theta}; \underline{y}) - \ell(\hat{p}, \hat{\theta}; \underline{y})] \ ,$$

where $\ell(\tilde{p}, \tilde{\theta}; \underline{y})$ is the maximum log-likelihood under H_0 and $\ell(\hat{p}, \hat{\theta}; \underline{y})$ is the maximum log-likelihood for the model of interest. This test statistic can then be compared with the χ^2_1 distribution (Azzalini, 1996, p.116). Tarone (1979) introduced an alternative test for determining whether θ is one, which is somewhat involved and thus the test statistic, W, described here is preferred because of its simple form.

Finally, to compare the fit of the multiplicative binomial distribution with that of the binomial distribution, AIC can also be used. The AIC value for the multiplicative binomial distribution is calculated as AIC $= -2\ell(\hat{p}, \hat{\theta}; \underline{y}) + 4$, with

$q = 2$, and the model with the smallest AIC value is deemed most suitable for the data.

5.5 An Example

An example is considered to illustrate the above ideas. In 1889 Geissler published data on the distribution of the sexes of children in families up to size 13. Parents of children born in Saxony from 1876 to 1885 recorded the size of their family as well as the numbers of male and female children upon registering the birth of their "last child".

It is thought that the ratio of boys to girls in a family, as well as the size of the family, would impact on a couple's decision to have more children. As a result of this it has been argued that the last recorded birth is unreliable and should not be used to estimate the human sex ratio. Lindsey and Altham (1998) therefore extracted the frequency distribution of male children in families with the most recently recorded birth omitted. A portion of this data corresponding to families of size 13 was investigated in the present study and is given in Table 5.1. Note that the maximum number of boys in a family is twelve because the last recorded birth was omitted.

Table 5.1: Human sex ratio data for families of size 13.

Number of boys	0	1	2	3	4	5	6	7	8	9	10	11	12
Frequency	3	24	104	286	670	1033	1343	1112	829	478	181	45	7

Lindsey (1995, p.132) considered modelling the data recorded in Table 5.1 with Efron's (1986) double binomial model and concluded that the data are overdispersed. The multiplicative binomial model was therefore fitted to this data in the present study in an attempt to account for the overdispersion. The parameters were estimated using Altham's (1978) method, the method proposed by Lindsey and Mersch (1992) and direct optimization of the log-likelihood function. The

estimates obtained from the three methods were identical and therefore direct optimization was selected for the examples in the remainder of this chapter because of its ease of implementation.

Profile likelihood plots of the parameters are presented in Figures 5.5 (a) and (b). The estimates of p and θ are presented with their standard errors, 95% Wald

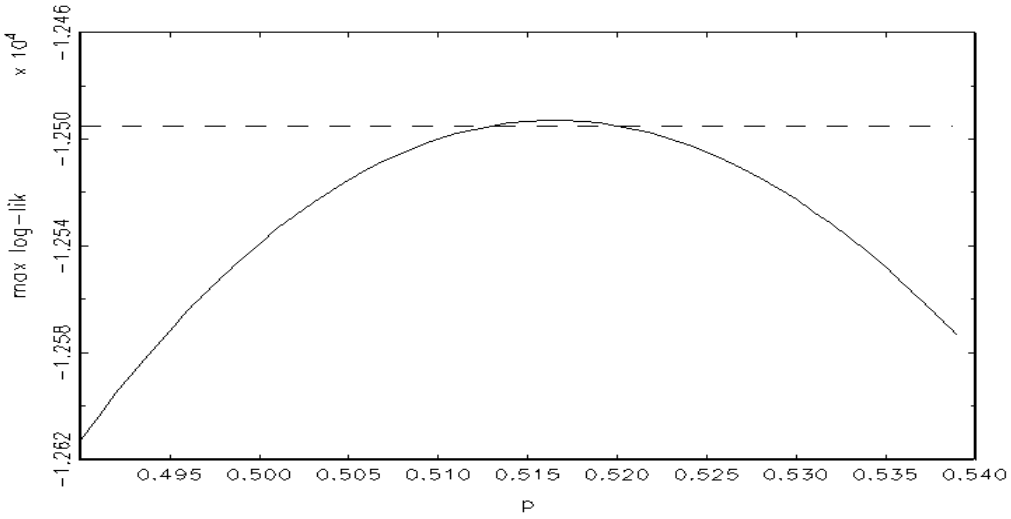


Figure 5.5: (a)

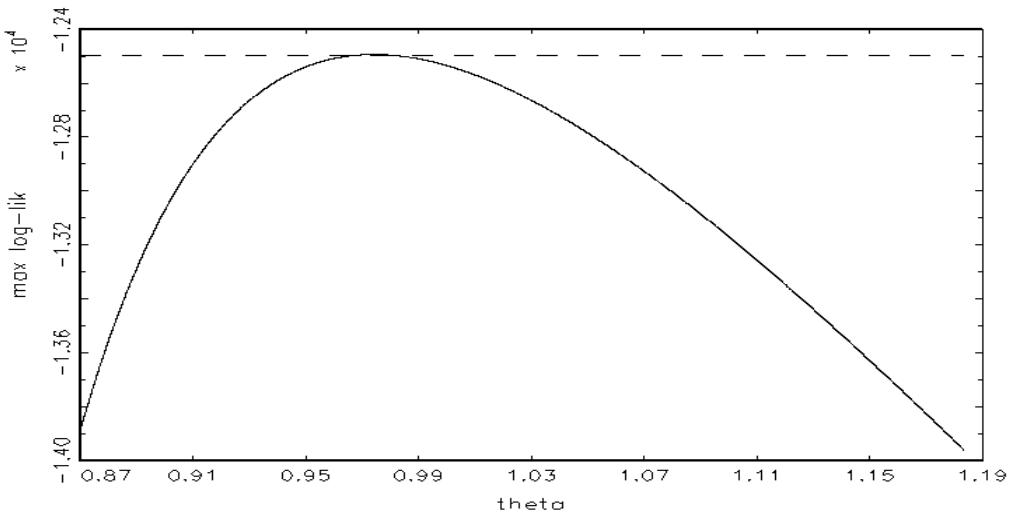


Figure 5.5: (b)

Figure 5.5: Profile likelihood plots for the parameters (a) p and (b) θ of the multiplicative binomial model fitted to the human sex ratio data.

and 95% profile likelihood confidence intervals in Table 5.2.

Table 5.2: Results from fitting the multiplicative binomial model to the human sex ratio data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
p	0.517	0.002	(0.513;0.520)	(0.513;0.520)
θ	0.974	0.003	(0.969;0.980)	(0.969;0.980)

Observe that the 95% Wald intervals for the parameters p and θ were identical to those obtained from the profile likelihoods in this instance, indicating the symmetry of the distributions of these parameters.

A chi-squared goodness of fit statistic was used to assess the fits of the binomial and multiplicative binomial models to the human sex ratio data. The p-values from these goodness of fit tests, as well as the AIC statistics, are presented in Table 5.3. The p-values from the χ^2 tests indicate that the binomial model is inadequate

Table 5.3: A comparison of the various models fitted to the human sex ratio data.

Model	χ^2 p-value	AIC
Binomial	0.000	25070.340
Multiplicative Binomial	0.150	24989.900

for the human sex ratio data and that the multiplicative binomial distribution provides a suitable fit to this data. The superiority of the fit of the multiplicative binomial to the human sex ratio data is reinforced by the AIC values.

Finally, in order to test whether the fit of the multiplicative binomial distribution was superior to that of the binomial distribution the test to determine whether θ is one described in Section 5.4.3 was performed. The value of the test statistic for this hypothesis was 82.538 which is highly significant when compared with a χ^2 value with 1 degree of freedom. Therefore the null hypothesis is rejected and

it is concluded that $\theta \neq 1$, indicating that the binomial model is not suitable for these data.

5.6 Modelling Dose-Response Data

5.6.1 Model Setting

In this section the results for the basic multiplicative binomial distribution are extended to a dose-response setting. In accordance with the previous chapters in this thesis, the logit model is considered as a drug tolerance distribution and, following Morgan (1992, p.253), the parameter λ is introduced as the proportion of natural mortality, where $0 \leq \lambda \leq 1$. This parameter is introduced here because of the example that is considered later in this chapter. Thus for a log-dose x the parameter p can be written in terms of $\alpha,\ \beta$ and λ as

$$p \;=\; \lambda + \frac{1-\lambda}{1+e^{-(\alpha+\beta x)}}\ , \tag{5.18}$$

where $\alpha,\ \beta$ and λ are unknown parameters. Lindsey and Altham (1998) used a logit link function to model the parameter p as a function of n and modelled θ as a function of n using a log link function. Molenberghs and Ryan (1999) considered modelling data arising from developmental toxicity studies and in effect modelled the parameter p of the multiplicative binomial distribution as a function of the dose administered using a linear link function. Natural mortality has not been incorporated into the tolerance distribution of the multiplicative binomial model in previous studies.

5.6.2 Likelihood, Score Functions and Information Matrix

The score functions for the multiplicative binomial model with the link function specified by (5.18) are the first derivatives of the log-likelihood function with re-

spect to the parameters $\lambda,\ \alpha,\ \beta$ and θ and are written as

$$\begin{aligned}
\frac{\partial \ell(\lambda,\alpha,\beta,\theta;y)}{\partial \lambda} &= \frac{\partial \ell(p,\theta;y)}{\partial p}\frac{\partial p}{\partial \lambda} \\
&= \frac{y-\mathrm{E}(Y)}{p(1-p)}\ \frac{1-p}{1-\lambda} \\
&= \frac{y-\mathrm{E}(Y)}{p(1-\lambda)}\ , \\
\frac{\partial \ell(\lambda,\alpha,\beta,\theta;y)}{\partial \alpha} &= \frac{\partial \ell(\alpha,\beta,\theta;y)}{\partial p}\frac{\partial p}{\partial \alpha} \\
&= \left\{\frac{(1-p)(p-\lambda)}{1-\lambda}\right\}\left\{\frac{y-\mathrm{E}(Y)}{p(1-p)}\right\} \\
&= \frac{(p-\lambda)[y-\mathrm{E}(Y)]}{p(1-\lambda)}\ , \\
\frac{\partial \ell(\lambda,\alpha,\beta,\theta;y)}{\partial \beta} &= \frac{\partial \ell(\alpha,\beta,\theta;y)}{\partial p}\frac{\partial p}{\partial \beta} \\
&= \left\{\frac{x(1-p)(p-\lambda)}{1-\lambda}\right\}\left\{\frac{y-\mathrm{E}(Y)}{p(1-p)}\right\} \\
&= \frac{x(p-\lambda)[y-\mathrm{E}(Y)]}{p(1-\lambda)}
\end{aligned}$$

and

$$\frac{\partial \ell(\lambda,\alpha,\beta,\theta;y)}{\partial \theta} = \frac{y(n-y)-\mathrm{E}[Y(n-Y)]}{\theta}\ .$$

The information matrix derived in Section 5.3.2 can be adapted for a dose-response setting by considering the relationship between p and the parameters $\alpha,\ \beta$ and λ from equation (5.18) where x is the log of the dose of the drug administered to a subject. To avoid computing second-order partial derivatives, the information matrix can be calculated from the expected values of the pairwise products of the score functions. Thus the information matrix for the parameters $\lambda,\ \alpha,\ \beta$ and θ of the multiplicative binomial distribution with a logit link function modelling p and incorporating natural mortality is written as

$$\mathcal{J} = E\begin{bmatrix}
\left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \lambda}\right)^2 & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \alpha} & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \beta} & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \theta} \\
\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \alpha} & \left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \alpha}\right)^2 & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \alpha}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \beta} & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \alpha}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \theta} \\
\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \beta} & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \alpha}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \beta} & \left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \beta}\right)^2 & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \beta}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \theta} \\
\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \theta} & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \alpha}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \theta} & \frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \beta}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \theta} & \left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial \theta}\right)^2
\end{bmatrix}.$$

The entries of this information matrix are calculated as

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\lambda}\right)^2\right] &= \mathrm{E}\left[\frac{Y-\mathrm{E}(Y)}{p(1-\lambda)}\right]^2 \\
&= \frac{\mathrm{Var}(Y)}{p^2(1-\lambda)^2}\ , \\
\mathrm{E}\left[\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\alpha}\right] &= \mathrm{E}\left[\frac{Y-\mathrm{E}(Y)}{p(1-\lambda)}\frac{(p-\lambda)[Y-\mathrm{E}(Y)]}{p(1-\lambda)}\right] \\
&= \frac{(p-\lambda)\mathrm{Var}(Y)}{p^2(1-\lambda)^2}\ , \\
\mathrm{E}\left[\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\beta}\right] &= \mathrm{E}\left[\frac{Y-\mathrm{E}(Y)}{p(1-\lambda)}\ \frac{x(p-\lambda)(Y-\mathrm{E}(Y))}{p(1-\lambda)}\right] \\
&= \frac{x(p-\lambda)\mathrm{Var}(Y)}{p^2(1-\lambda)^2}\ , \\
\mathrm{E}\left[\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\lambda}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\theta}\right] &= \mathrm{E}\left\{\frac{Y-E(Y)}{p(1-\lambda)}\ \frac{Y(n-Y)-\mathrm{E}[Y(n-Y)]}{\theta}\right\} \\
&= \frac{\mathrm{E}[Y^2(n-Y)]-\mathrm{E}(Y)\mathrm{E}[Y(n-Y)]}{\theta p(1-\lambda)} \\
&= \frac{\mathrm{Cov}[Y(n-Y),Y]}{\theta p(1-\lambda)}\ , \\
\mathrm{E}\left[\left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\alpha}\right)^2\right] &= \mathrm{E}\left[\frac{(p-\lambda)(Y-\mathrm{E}(Y))}{p(1-\lambda)}\right]^2 \\
&= \frac{(p-\lambda)^2\mathrm{Var}(Y)}{p^2(1-\lambda)^2}\ , \\
\mathrm{E}\left[\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\alpha}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\beta}\right] &= \mathrm{E}\left[\frac{(p-\lambda)(Y-\mathrm{E}(Y))}{p(1-\lambda)}\frac{x(p-\lambda)(Y-\mathrm{E}(Y))}{p(1-\lambda)}\right] \\
&= \frac{x(p-\lambda)^2\mathrm{Var}(Y)}{p^2(1-\lambda)^2}\ , \\
\mathrm{E}\left[\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\alpha}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\theta}\right] &= \mathrm{E}\left[\frac{(p-\lambda)(Y-\mathrm{E}(Y))}{p(1-\lambda)}\ \frac{Y(n-Y)-\mathrm{E}[Y(n-Y)]}{\theta}\right] \\
&= \frac{(p-\lambda)\mathrm{E}\{[\mathrm{E}[Y(Y+n)]+Y(Y+n)][\mathrm{E}(Y)-Y]\}}{\theta p(1-\lambda)}\ , \\
\mathrm{E}\left[\left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\beta}\right)^2\right] &= \mathrm{E}\left[\frac{x(p-\lambda)(Y-\mathrm{E}(Y))}{p(1-\lambda)}\right]^2 \\
&= \frac{x^2(p-\lambda)^2\mathrm{Var}(Y)}{p^2(1-\lambda)^2}\ ,
\end{aligned}$$

$$\mathrm{E}\left[\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\beta}\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\theta}\right] = \mathrm{E}\left[\frac{x(p-\lambda)(Y-\mathrm{E}(Y))}{p(1-\lambda)}\ \frac{Y(n-Y)-\mathrm{E}[Y(n-Y)]}{\theta}\right]$$
$$= \frac{x(p-\lambda)\mathrm{E}\{[\mathrm{E}[Y(Y+n)]+Y(Y+n)][\mathrm{E}(Y)-Y]\}}{\theta p(1-\lambda)}$$

and

$$\mathrm{E}\left[\left(\frac{\partial \ell(\lambda,\alpha,\beta,\theta;Y)}{\partial\theta}\right)^2\right] = \frac{\mathrm{Var}[Y(n-Y)]}{\theta^2}\ .$$

5.6.3 Model Fitting and Checking

Consider a random variable Y that refers to the number of organisms that survive treatment with various doses of a particular drug. Suppose that y_{ij} refers to the number of individuals surviving a dose d_i and log-dose x_i of the drug, $j = 1,\ldots,n_i$ and $i = 1,\ldots,D$. Furthermore, let p_i denote the probability of an organism surviving exposure to the ith dose of the drug and consider writing this probability in terms of the log-dose using expression (5.18). Then the likelihood and log-likelihood functions for the multiplicative binomial model with p modelled with a logit link function can be written as

$$L(\lambda,\alpha,\beta,\theta;\underline{y}) = \prod_{i=1}^{D}\prod_{j=1}^{n_i}\left\{\frac{\binom{n}{y_{ij}} p_i^{y_{ij}}(1-p_i)^{n-y_{ij}}\theta^{y_{ij}(n-y_{ij})}}{f(p_i,\theta,n)}\right\}$$

and

$$\ell(\lambda,\alpha,\beta,\theta;\underline{y}) = \sum_{i=1}^{D}\sum_{j=1}^{n_i}\left\{\ln\binom{n}{y_{ij}} + y_{ij}\ln(p_i) + (n-y_{ij})\ln(1-p_i)\right.$$
$$\left. + y_{ij}(n-y_{ij})\ln(\theta) - lnf(p_i,\theta,n)\right\}$$

respectively. The parameters $\lambda,\ \alpha,\ \beta$ and θ can then be estimated by optimizing the log of the likelihood function directly using an optimization routine such as OPTMUM in Gauss. A starting value of 1 can be used for θ and starting values

for the parameters λ, α and β can be obtained from fitting a binomial distribution with p modelled with a logit link function that includes natural mortality as in expression (5.18).

The information matrix for the data can be obtained by summing the information for a single observation over all of the observed responses. The standard errors of the parameter estimates are then the square roots of the diagonal entries of the inverted information matrix. 95% Wald intervals can be obtained for each parameter using the standard errors obtained from the information matrix. Profile likelihood plots can also be obtained for each parameter and these plots can be used to compute 95% confidence interval estimates for the parameters using the method outlined in Section 5.4.2.

The superiority of the fit of the multiplicative binomial distribution to that of the binomial model can be determined by testing $\mathrm{H}_0 : \theta = 1$ using the χ^2 test previously described in Section 5.4.3. A comparison can also be made between the multiplicative binomial distribution, the binomial distribution and the beta-binomial distribution using AIC statistics. The AIC for the multiplicative binomial distribution with the parameter p modelled with a logit tolerance distribution and incorporating natural mortality is AIC $= -2\ell(\hat{\lambda}, \hat{\alpha}, \hat{\beta}, \hat{\theta}; \underline{y}) + 8$, where $\ell(\hat{\lambda}, \hat{\alpha}, \hat{\beta}, \hat{\theta}; \underline{y})$ is the maximum log-likelihood and $q = 4$. The model with the smallest AIC value provides the best fit.

5.6.4 An Example

Morgan (1992, p.252) presented a data set in which varying doses of trichloromethane were administered to litters of mice seven days after birth and the number that died within 14 days of exposure to the drug was recorded. The data are presented in Table 5.4.

There was a common litter size of 8 for which Morgan (1992, p252) commented that there was no explanation. Morgan (1992, p.253) observed that the variation in the responses exceeds the variation that would be explained by a binomial

Table 5.4: Data for mice exposed to trichloromethane.

Dosage (mg/kg)	Number dead per litter of 8					Total dead
Control	0	0	0	2	2	4
250	0	0	1	3	6	10
300	0	0	0	1	8	9
350	0	2	2	5	8	17
400	1	2	4	6	7	20
450	1	4	5	6	8	24
500	1	7	8	8	8	32

model and concluded that the data are overdispersed. Therefore the multiplicative binomial distribution with the probability p modelled by a logit link function including natural mortality was fitted to Morgan's (1992) mice data using a direct and straightforward optimization routine. The maximum likelihood estimates of the parameters, together with their standard errors and 95% confidence intervals are presented in Table 5.5.

Table 5.5: Results from fitting the multiplicative binomial model to the mice data, where p is modelled with a logit function and λ is the rate of natural mortality.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
$\hat{\lambda}$	0.3572	0.0800	(0.200;0.514)	(0;0.410)
$\hat{\alpha}$	-14.8232	9.7414	(-33.915;4.269)	(*;-4.251)
$\hat{\beta}$	5.2287	3.5624	(-1.753;12.211)	(1.301;*)
$\hat{\theta}$	0.7231	0.0234	(0.678;0.768)	(0.679;0.771)

*These limits of the confidence intervals cannot be computed

The profile likelihood plots for each of the parameters λ, α, β and θ are presented in Figures 5.6 (a), (b), (c) and (d) respectively. The profile likelihood plots

for the parameters α and β exhibited erratic behaviour at some of the extremes and as a result profile likelihood intervals could not be obtained for these parameters. Since confidence intervals for the parameters α and β could not be obtained from the profile likelihoods, 95% Wald intervals for each of the parameters have also been included.

The fit of the multiplicative binomial model was compared with that of the binomial model by testing the two-sided hypothesis $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$. The value of the test statistic W for this hypothesis was 349.954 which, when compared with a χ^2_1 distribution, yielded a p-value less than 0.0001. This suggests that the multiplicative binomial model provides a better fit to this data set than the binomial distribution.

Finally the fit of the multiplicative binomial distribution was compared with that of the beta-binomial and binomial models using the AIC statistic. The AIC values for these models are presented in Table 5.6. The AIC values indicate that the multiplicative binomial distribution provides the best fit to the data and hence accounts for the overdispersion more adequately than the beta-binomial model.

Table 5.6: A comparison of the models fitted to the mice data.

Model	Maximum log-likelihood	Number of parameters	AIC
Multiplicative Binomial	-64.90	4	137.80
Beta-Binomial	-126.43	4	260.86
Binomial	-157.24	3	320.48

5.7 Conclusions

In this chapter some new properties of the multiplicative binomial distribution have been introduced. General expressions for the factorial moments of the distribution were derived and are useful for obtaining statistics such as the skewness

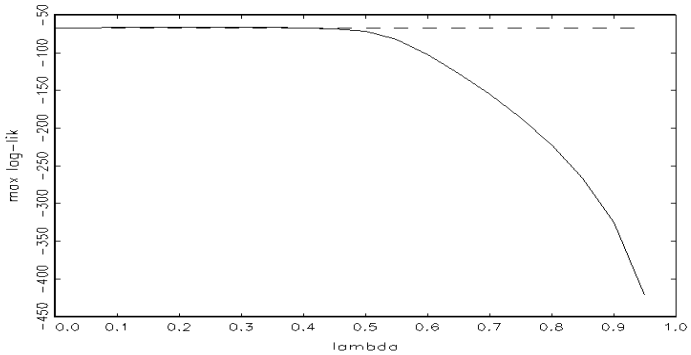


Figure 5.6: (a)

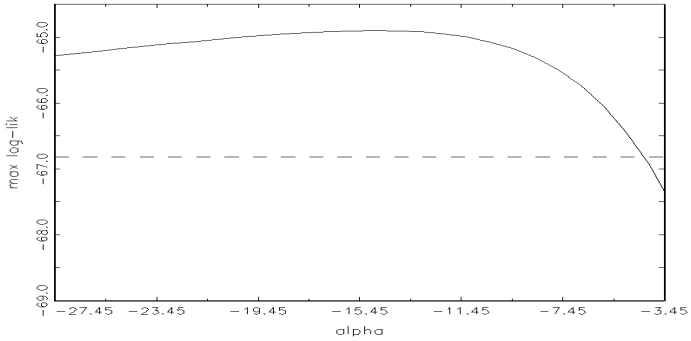


Figure 5.6: (b)

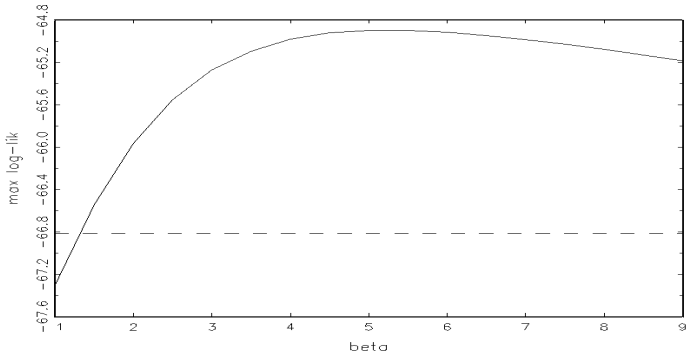


Figure 5.6: (c)

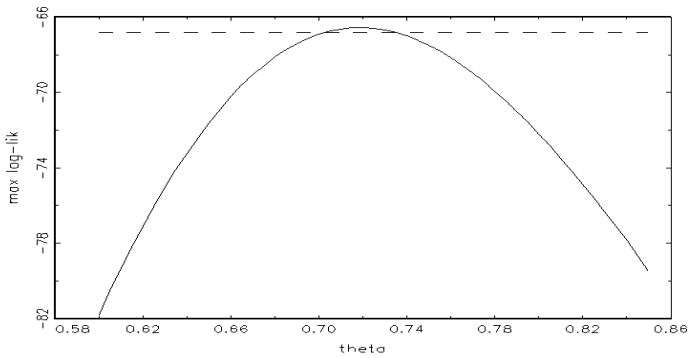


Figure 5.6: (d)

Figure 5.6: Profile likelihood plots for the parameters (a) λ, (b) α, (c) β and (d) θ of the multiplicative binomial model with a logit tolerance distribution fitted to the mice data.

and kurtosis. The information matrix of the multiplicative binomial distribution proved to have an elegant form and was adapted for a dose-response setting incorporating natural mortality. The matrix was also used to obtain the standard errors of the parameter estimates. Various methods of estimation were discussed and direct optimization of the likelihood function was considered preferable because it is simple and easy to implement. Profile likelihoods for the parameters of the multiplicative binomial model were plotted and used to obtain interval estimates associated with the parameters of the distribution.

The model initially introduced by Altham (1978) has been shown to be useful for modelling overdispersed binomial-type data as well as overdispersed binary data arising from dose-mortality studies. The multiplicative binomial model provided a better fit to the mice data than the beta-binomial model. An adaptation of this distribution to a Wadley's problem scenario is investigated in the next chapter.

Chapter 6

The Altham-Poisson Model

6.1 Introduction

The investigation of the multiplicative binomial distribution in Chapter 5 indicated that the distribution effectively models overdispersion in binomial-type data. An extension of the multiplicative binomial distribution to a Wadley's problem setting is explored in the present chapter by modelling the number of trials from the multiplicative binomial distribution with a Poisson distribution. The resulting distribution is termed the Altham-Poisson distribution. It is examined as a means of modelling overdispersed Poisson data in its own right and its suitability for overdispersed dose-response data from a Wadley's problem setting is also investigated.

The p.m.f. of the Altham-Poisson distribution is derived and an approximation suitable for calculating the resultant probabilities is explored and presented in Section 6.2. Section 6.3 includes an investigation of the moments of the distribution together with some examples. The likelihood and log-likelihood functions as well as the score functions and the information matrix are discussed in Section 6.4. Methods for estimating the parameters of the Altham-Poisson distribution are outlined in Section 6.5 and some results from fitting the model to overdispersed Poisson data are presented in Section 6.6. In Section 6.7 the distribution is adapted

to a dose-response setting and concluding remarks are recorded in Section 6.8.

6.2 Distribution

6.2.1 Derivation of the p.m.f.

Consider a random variable Y which follows a multiplicative binomial distribution with number of trials n and parameters p and θ, where p is the probability of a success. Suppose that the number of trials is unobserved and consider modelling this unknown number as a random variable N with a Poisson(τ) distribution. Using the notation introduced in Chapter 5, let $f(p, \theta, n)$ denote the normalizing constant from the multiplicative binomial distribution where

$$f(p,\theta,n) \;=\; \sum_{y=0}^{n} \left(\begin{array}{c} n \\ y \end{array} \right) p^y (1-p)^{n-y} \theta^{y(n-y)} \; .$$

Then the marginal p.m.f. of Y can be obtained from first principles as follows:

$$\begin{aligned}
\Pr(Y=y) &= \sum_{n=y}^{\infty} \Pr(Y=y|N=n)\Pr(N=n) \\
&= \sum_{n=y}^{\infty} \frac{\left(\begin{array}{c} n \\ y \end{array} \right) p^y (1-p)^{n-y} \theta^{y(n-y)}}{f(p,\theta,n)} \frac{\tau^n e^{-\tau}}{n!} \\
&= \frac{e^{-\tau}\tau^y p^y}{y!} \sum_{n=y}^{\infty} \frac{\tau^{n-y}(1-p)^{n-y}\theta^{y(n-y)}}{f(p,\theta,n)(n-y!)} \\
&= \frac{\tau^y e^{-\tau} p^y}{y!} \sum_{s=0}^{\infty} \frac{\tau^s (1-p)^s \theta^{ys}}{s! f(p,\theta,s+y)} \\
&= \frac{(\tau p)^y e^{-\tau p}}{y!} \times e^{-\tau(1-p)} \sum_{s=0}^{\infty} \frac{[\tau\theta^y(1-p)]^s}{s! f(p,\theta,s+y)} \; . \qquad (6.1)
\end{aligned}$$

Thus Y follows a weighted Poisson distribution with weight

$$w_y = e^{-\tau(1-p)} \sum_{s=0}^{\infty} \frac{[\tau\theta^y(1-p)]^s}{s! f(p,\theta,s+y)} \; ,$$

which includes an infinite sum that cannot be expressed explicitly. The p.m.f. in expression (6.1) is a probability and since $\sum_{y=0}^{\infty} \Pr(Y = y) = 1$, the infinite sum included in the expression must converge. When $\theta = 1$ the distribution of Y reduces to a Poisson distribution with mean τp, which is the classic distribution for modelling Wadley's problem (Wadley, 1949). The distribution of Y will be referred to as the Altham-Poisson distribution with parameters $\tau,\ p$ and θ and will be denoted $Y \sim \text{AlthamPoisson}(\tau,\ p,\ \theta)$ for the remainder of this thesis.

6.2.2 Infinite Sum

The p.m.f. in expression (6.1) indicates that evaluating a probability for the Altham-Poisson distribution entails the approximation of an infinite sum. It is desirable to approximate the infinite sum to a specified degree of accuracy. In particular, a cutoff value for the index s, denoted s_{cut}, is sought so that the infinite sum in expression (6.1) can be approximated by the finite series

$$\sum_{s=0}^{s_{\text{cut}}} \frac{[\tau\theta^y(1-p)]^s}{s!f(p,\theta,s+y)} .$$

to a desired degree of accuracy.

Shmueli, Minka, Kadane, Borle and Boatwright (2005) considered the Conway-Maxwell Poisson distribution which is a weighted Poisson distribution and includes an infinite sum. In accordance with their work, an upper bound is sought on the error that results from approximating the infinite sum with a finite sum. If an upper bound on the error term can be determined, the accuracy of the approximation can be controlled. Since the model reduces to a Poisson distribution when $\theta = 1$, consider the component $f(p,\theta,s+y)$ in expression (6.1) for the two cases $\theta > 1$ and $\theta < 1$.

When $\theta > 1$ it follows that

$$\theta^{k(s+y-k)} \geq 1,$$

where k is a positive integer and as a result

$$\begin{aligned} f(p,\theta,s+y) &= \sum_{k=0}^{s+y} \left(\begin{array}{c} s+y \\ k \end{array} \right) p^k (1-p)^{s+y-k} \theta^{k(s+y-k)} \\ &\geq \sum_{k=0}^{s+y} \left(\begin{array}{c} s+y \\ k \end{array} \right) p^k (1-p)^{s+y-k} \\ &= 1 \ . \end{aligned}$$

Thus

$$\frac{[\tau\theta^y(1-p)]^s}{s!f(p,\theta,s+y)} \ \leq \ \frac{[\tau\theta^y(1-p)]^s}{s!} \ . \tag{6.2}$$

Now consider again writing the infinite sum from expression (6.1) as

$$\sum_{s=0}^{\infty} \frac{[\tau\theta^y(1-p)]^s}{s!f(p,\theta,s+y)} \ = \ \sum_{s=0}^{k} \frac{[\tau\theta^y(1-p)]^s}{s!f(p,\theta,s+y)} + R_k \ ,$$

where R_k is the error or remainder term given by

$$R_k \ = \ \sum_{s=k+1}^{\infty} \frac{[\tau\theta^y(1-p)]^s}{s!f(p,\theta,s+y)} \ .$$

From (6.2) it follows that

$$\begin{aligned} R_k &\leq \sum_{s=k+1}^{\infty} \frac{[\tau\theta^y(1-p)]^s}{s!} \\ &= e^{[\tau\theta^y(1-p)]} - \sum_{s=0}^{k} \frac{[\tau\theta^y(1-p)]^s}{s!} \ . \end{aligned}$$

The remainder term R_k therefore has an upper bound that is computable when $\theta > 1$.

The inequality $\theta^{k(s+y-k)} \geq \theta^{k(s+y)}$ holds for values of θ that are less than one. Therefore

$$\begin{aligned} f(p,\theta,s+y) &\geq \sum_{k=0}^{s+y} \left(\begin{array}{c} s+y \\ k \end{array} \right) p^k (1-p)^{s+y-k} \theta^{k(s+y)} \\ &= (1-p+p\theta^{s+y})^{s+y} \ , \end{aligned}$$

and hence

$$\begin{aligned}
\frac{[\tau(1-p)\theta^y]^s}{s!f(p,\theta,s+y)} &\leq \frac{[\tau(1-p)\theta^y]^s}{s!(1-p+p\theta^{s+y})^{s+y}} \\
&= \frac{[\tau\theta^y]^s}{s!(1+\frac{p}{1-p}\theta^{s+y})^{s+y}(1-p)^y} \\
&\leq \frac{[\tau\theta^y]^s}{s!(1-p)^y} \qquad \text{since } \frac{p}{1-p}\theta^{s+y} > 0 \\
&\leq \frac{[\tau\left(\frac{\theta}{1-p}\right)^y]^s}{s!} \ .
\end{aligned}$$

Consider partitioning the infinite sum in the following way

$$\sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s!f(p,\theta,s+y)} = \sum_{s=0}^{k} \frac{[\tau(1-p)\theta^y]^s}{s!f(p,\theta,s+y)} + R_k \ ,$$

where the remainder, R_k, is written as

$$\begin{aligned}
R_k &= \sum_{s=k+1}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s!f(p,\theta,s+y)} \\
&\leq \sum_{s=k+1}^{\infty} \frac{[\tau\left(\frac{\theta}{1-p}\right)^y]^s}{s!} \\
&= e^{\tau\left(\frac{\theta}{1-p}\right)^y} - \sum_{s=0}^{k} \frac{[\tau\left(\frac{\theta}{1-p}\right)^y]^s}{s!} \ .
\end{aligned}$$

Thus when $\theta < 1$, R_k is bounded above and the bound is computable.

Therefore the remainder term R_k has an upper bound for all values of θ and as a result the infinite sum in (6.1) can be approximated by a finite sum with cutoff s_{cut}. For a predetermined upper bound on the remainder term an associated value for s_{cut} can be obtained thereby ensuring that the finite sum approximates the infinite sum with a desired degree of accuracy.

6.2.3 Simulation

An alternative method for estimating the probability in expression (6.1) is by simulation. A 'very large' number of observations can be simulated from the

Altham-Poisson distribution by first simulating a value of N from the Poisson(τ) distribution and then a value of Y given $N = n$ from the multiplicative binomial model with parameters p and θ. The probability that the Altham-Poisson distributed random variable Y assumes a particular value y is then estimated by the proportion of simulated observations in the random sample that are equal to y.

A disadvantage of this method is that its accuracy cannot be readily controlled or readily assessed. In other words the number of simulated observations required to estimate the probabilities adequately is unknown. However, an indication of a suitable number of simulated observations can be obtained by computing and comparing the probabilities described above for a range of sample sizes. When the number of simulated observations is sufficiently large, the estimates of the probabilities will begin to stabilize.

The preferred method for approximating the probabilities comprising the p.m.f. of the Altham-Poisson distribution is thus to select a suitable cutoff value according to a desired degree of accuracy and to then estimate the infinite sum with a finite sum, as described in Section 6.2.2.

6.3 Moments of the Distribution

Since the p.m.f. of the random variable Y includes an infinite sum, the moments of the Altham-Poisson distribution cannot be written down explicitly. The first two moments are considered in this thesis.

If Y follows an Altham-Poisson distribution with parameters τ, p and θ the expected values of Y and Y^2 are written respectively as

$$\begin{aligned} \mathrm{E}(Y) &= \sum_{y=0}^{\infty} y \Pr(Y = y) \\ &= \sum_{y=0}^{\infty} y \sum_{s=0}^{\infty} \frac{(\tau p)^y e^{-\tau}}{y!} \frac{[\tau \theta^y (1-p)]^s}{s! f(p, \theta, s+y)} \\ &= \sum_{y=0}^{\infty} y \frac{(\tau p)^y e^{-\tau}}{y!} \sum_{s=0}^{\infty} \frac{[\tau \theta^y (1-p)]^s}{s! f(p, \theta, s+y)} \end{aligned} \tag{6.3}$$

and

$$\begin{aligned} \mathrm{E}(Y^2) &= \sum_{y=0}^{\infty} y^2 \mathrm{Pr}(Y=y) \\ &= \sum_{y=0}^{\infty} y^2 \sum_{s=0}^{\infty} \frac{(\tau p)^y e^{-\tau}}{y!} \frac{[\tau(1-p)\theta^y]^s}{s! f(p,\theta,s+y)} \\ &= \sum_{y=0}^{\infty} y^2 \frac{(\tau p)^y e^{-\tau}}{y!} \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s! f(p,\theta,s+y)} \ . \end{aligned} \tag{6.4}$$

Expressions (6.3) and (6.4) demonstrate the emergence of an outer infinite sum. As y increases $\mathrm{Pr}(Y=y)$ decreases since $\sum_{y=0}^{\infty} \mathrm{Pr}(Y=y) = 1$. However, since y may increase faster than $\mathrm{Pr}(Y=y)$ decreases, the infinite sums in expressions (6.3) and (6.4) will not converge and as a result the moments of the Altham-Poisson distribution may not exist.

Three techniques for estimating the moments of the Altham-Poisson distribution are assessed. The methods include the selection of a cutoff for the infinite sum, simulation and conditioning.

6.3.1 Selection of a Cutoff

If the moments exist their approximation by the finite expressions

$$\mathrm{E}(Y) = \sum_{y=0}^{y_{\mathrm{cut}}} \frac{y(\tau p)^y e^{-\tau}}{y!} \sum_{s=0}^{s_{\mathrm{cut}}} \frac{[\tau(1-p)\theta^y]^s}{s! f(p,\theta,s+y)}$$

and

$$\mathrm{E}(Y^2) = \sum_{y=0}^{y_{\mathrm{cut}}} \frac{y^2(\tau p)^y e^{-\tau}}{y!} \sum_{s=0}^{s_{\mathrm{cut}}} \frac{[\tau(1-p)\theta^y]^s}{s! f(p,\theta,s+y)}$$

entails the selection of a cutoff value, y_{cut}, for y and nested within that a cutoff value for s, namely s_{cut}. Determining these cutoff values for y and s is not trivial and as a result other methods of estimating the moments of the Altham-Poisson distribution are considered.

6.3.2 Simulation

A large number of observations can be simulated from the Altham-Poisson distribution using the simulation technique described in Section 6.2.3. The mean and variance of the simulated observations provide estimates of the mean and variance of the Altham-Poisson distribution with parameters τ, p and θ.

If the moments exist and the number of simulated observations is sufficiently large the estimates of the moments will be stable. The existence of the moments can therefore be assessed by observing whether their estimates stabilize and an appropriate number of simulated observations can be determined by considering the number of observations required for these estimates to be stable.

6.3.3 Conditioning

The first two factorial moments, and hence the mean and variance, of the Altham-Poisson distribution can be approximated by conditioning on the distribution of N in the following way:

$$\begin{aligned} \mathrm{E}(Y) &= \mathrm{E}_N[\mathrm{E}(Y|N=n)] \\ &= \mathrm{E}_N\left[np(p+q\theta)^{n-1} f\left(\frac{p}{p+q\theta}, \theta, n-1\right)/f(p,\theta,n)\right] \end{aligned} \tag{6.5}$$

and

$$\begin{aligned} \mathrm{E}[Y(Y-1)] &= \mathrm{E}_N\{\mathrm{E}[Y(Y-1)|N=n]\} \\ &= \mathrm{E}_N\left[n(n-1)p^2(p+q\theta^2)^{n-2} f\left(\frac{p}{p+q\theta^2}, \theta, n-2\right)/f(p,\theta,n)\right], \end{aligned} \tag{6.6}$$

(Ross, 2002, p.152) where $\mathrm{E}[(Y|N=n)]$ and $\mathrm{E}[(Y(Y-1)|N=n)]$ are the expectations associated with a random variable following a multiplicative binomial distribution as described in Chapter 5. The moments can then be estimated by simulating a large number of values of N from a Poisson distribution with parameter τ and computing the expected values of Y and $Y(Y-1)$ for each value of N

using the relevant moments from the multiplicative binomial distribution. The averages of these expectations then provide estimates of the required moments. The variance of the Altham-Poisson distribution can be calculated from the moments in equations (6.5) and (6.6) in the usual way from

$$\text{Var}(Y) \;=\; \text{E}[Y(Y-1)] + \text{E}(Y) - [\text{E}(Y)]^2 \; .$$

The number of simulated values of N can once again be selected by determining the number of observations that are required to stabilize the estimates of the moments. This method of conditioning is a more elegant way of simulating the moments of the Altham-Poisson distribution.

The least preferred method of estimating the moments of the Altham-Poisson distribution is the method that involves selecting a cutoff for the infinite sums in expressions (6.3) and (6.4). This is because there is no indication of appropriate values for y_{cut} and s_{cut}. The method of conditioning is better than the method of selecting a cutoff because a suitable number of simulated observations can easily be determined. A limitation of this method, however, is that evaluating the expectations $\text{E}(Y|N = n)$ and $\text{E}[(Y(Y-1)|N = n)]$ can be time consuming. Direct simulation was therefore selected as the most preferred method of estimating the moments of the Altham-Poisson distribution because it is straightforward and quick to implement. The moments of the examples of the distribution in the present study were estimated by simulating observations from the Altham-Poisson distribution and computing the moments of the sample data. Each case was considered separately and the moments seemed to exist. It was further observed that when one million observations were simulated, the estimates of the moments were stable.

6.3.4 Examples

Illustrations of the p.m.f. of the random variable $Y \sim$Altham-Poisson$(\tau,\ p,\ \theta)$ are presented in Figures 6.1 and 6.2. Figure 6.1 includes examples in which the

parameter θ is allowed to vary and the examples in Figure 6.2 have different values of p. The values used for τ, p and θ correspond to the values of n, p and θ for the examples in Section 2 of Chapter 5. The probabilities related to the p.m.f. were estimated by selecting a cutoff for the infinite sum in the p.m.f. according to a specified degree of accuracy, as described in Section 6.2.2.

The mean and variance of these examples were estimated by simulating observations according to the procedure outlined in Section 6.3.2. Since $\mathrm{E}(N) = \tau$, the moments of the Altham-Poisson distribution were compared with those of the multiplicative binomial distribution with the same values of p and θ and with the value of the parameter τ of the Altham-Poisson distribution corresponding to the value of n of the multiplicative binomial distribution. The moments were compared to determine whether the mean-variance relationship of the multiplicative binomial model is carried through to the Altham-Poisson distribution. Table 6.1 contains a comparison of the moments of the multiplicative binomial and Altham-Poisson distributions.

Table 6.1: Moments of the multiplicative binomial and Altham-Poisson distributions.

			Multiplicative Binomial		Altham-Poisson	
$\mathrm{E}(N) = \tau$	p	θ	Mean	Variance	Mean	Variance
20	0.48	1.1	9.792	2.592	9.795	7.458
20	0.48	0.85	4.777	49.159	4.640	45.632
20	0.48	0.75	3.450	55.486	3.399	52.631
20	0.48	0.9	7.808	25.581	7.099	30.063
20	0.55	0.9	14.733	16.357	15.321	41.847

Although it is reasonable to expect the variances of the examples of the Altham-Poisson model to be greater than those of the multiplicative binomial examples because extra variability has been introduced, this does not seem to always be the case. There are two instances in Table 6.1 in which the variance of the multiplica-

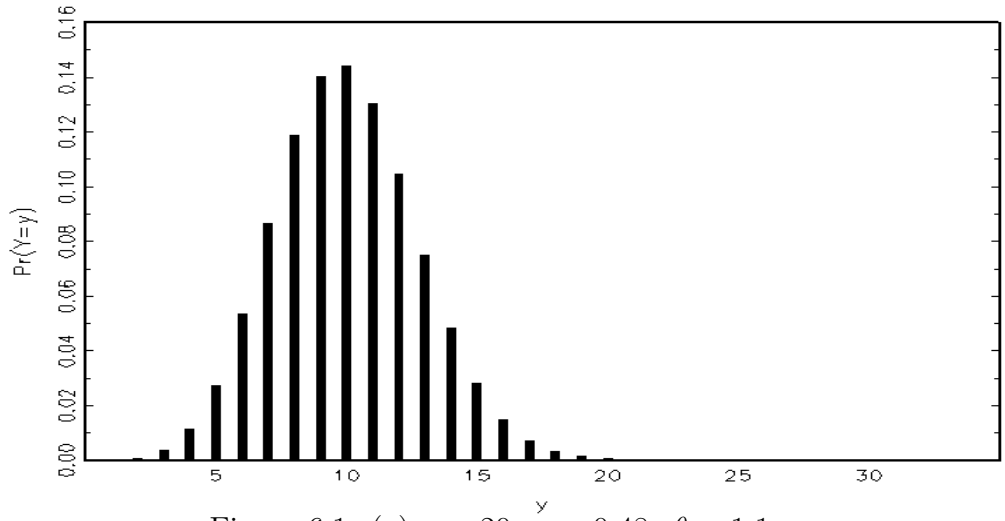


Figure 6.1: (a) $\tau = 20,\ p = 0.48,\ \theta = 1.1$

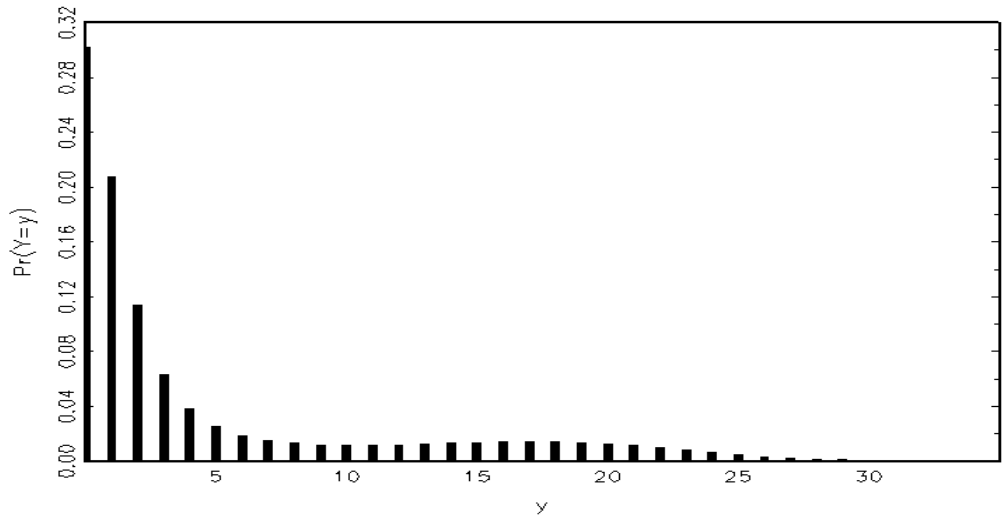


Figure 6.1: (b) $\tau = 20,\ p = 0.48,\ \theta = 0.85$

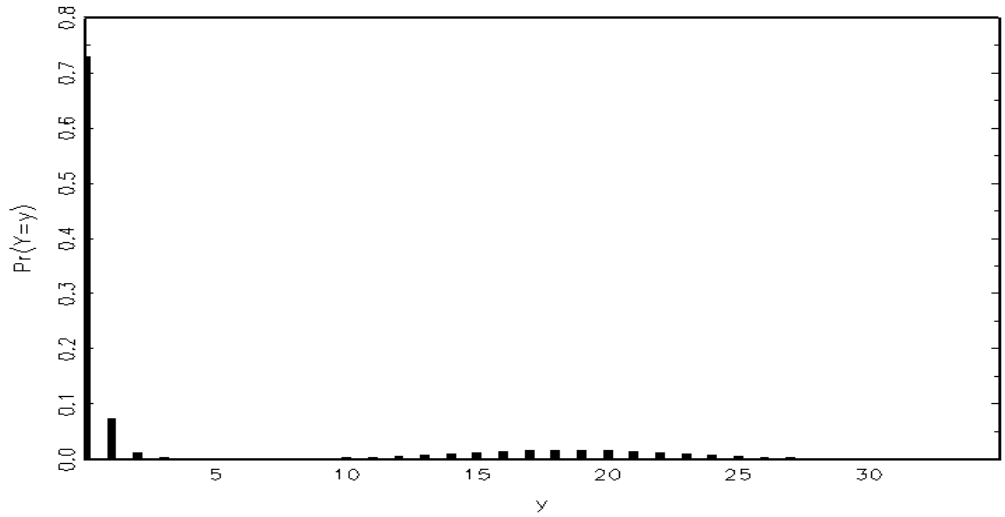


Figure 6.1: (c) $\tau = 20,\ p = 0.48,\ \theta = 0.75$

Figure 6.1: Plots showing the shapes of various Altham-Poisson distributions, with changing values of θ.

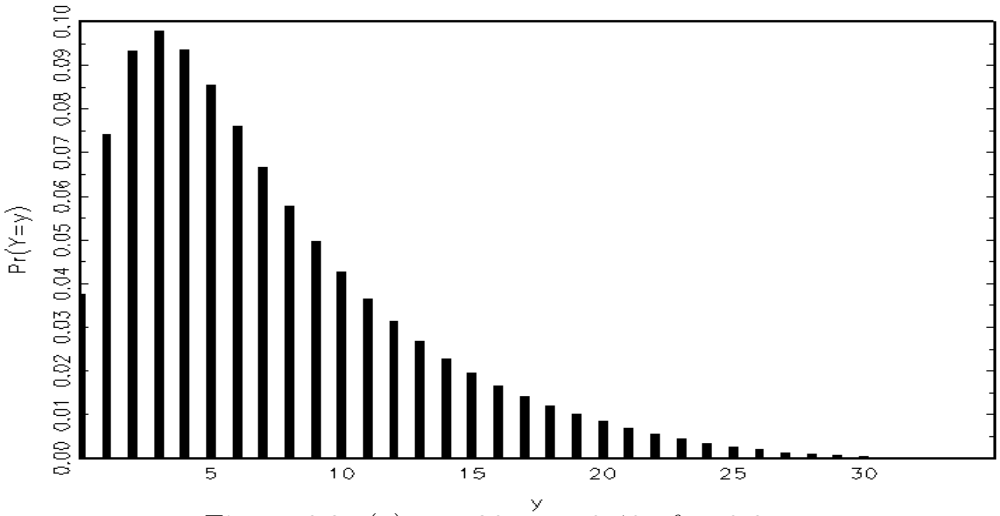


Figure 6.2: (a) $\tau = 20,\ p = 0.48,\ \theta = 0.9$

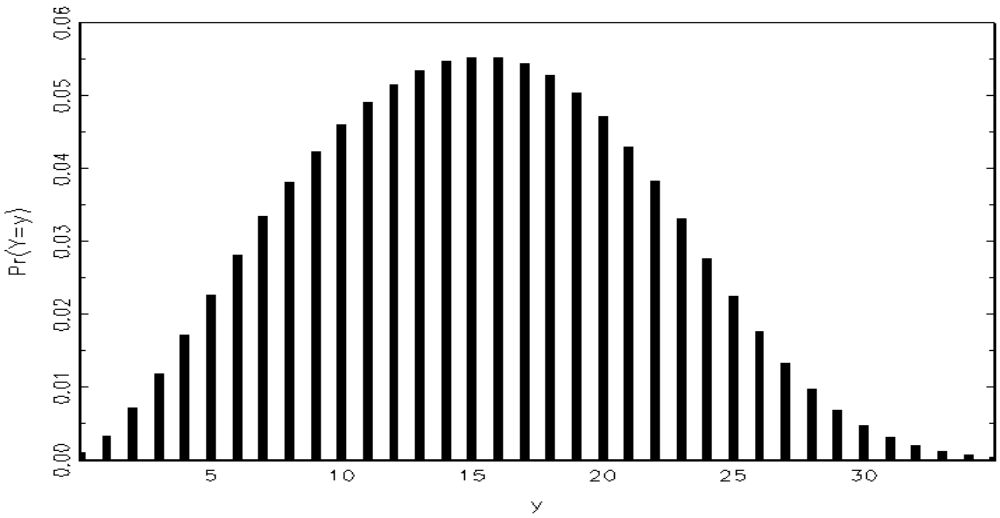


Figure 6.2: (b) $\tau = 20,\ p = 0.55,\ \theta = 0.9$

Figure 6.2: Plots showing the shapes of various Altham-Poisson distributions, with changing values of p.

tive binomial distribution exceeds that of the Altham-Poisson distribution. Recall from Figure 5.3 of Chapter 5 that these two multiplicative binomial models were bimodal. The two Altham-Poisson models, however, are unimodal as depicted in Figure 6.1. Therefore it can be deduced that modelling the parameter n with a Poisson distribution reduced the variability in the model and hence the variance of the Altham-Poisson distribution is indeed less than that of the multiplicative binomial distribution in these instances. However, it was observed that when the variance of the examples of the multiplicative binomial distribution exceeds the mean, the same pattern exists for the examples of the Altham-Poisson distribution.

6.4 Likelihood, Score Functions and Information Matrix

6.4.1 Likelihood Function

The likelihood function for the Altham-Poisson distribution is written as

$$L(\tau, p, \theta; y) = \frac{(\tau p)^y e^{-\tau}}{y!} \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s! f(p, \theta, s+y)}$$

where $f(p, \theta, s+y)$ is the normalizing constant of the multiplicative binomial distribution given by

$$f(p, \theta, s+y) = \sum_{k=0}^{s+y} \binom{s+y}{k} p^k (1-p)^{s+y-k} \theta^{k(s+y-k)} .$$

The log-likelihood function is therefore

$$\ell(\tau, p, \theta; y) = y \ln(\tau p) - \tau - \ln(y!) + \ln \left\{ \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s! f(p, \theta, s+y)} \right\} .$$

6.4.2 Score Functions

The score functions for the Altham-Poisson distribution are obtained by differentiating the log-likelihood function with respect to the parameters τ, p and θ. For

convenience let the function $h(p, \theta, \tau, y)$ denote the infinite sum in the log-likelihood function, i.e.

$$h(p, \theta, \tau, y) \;=\; \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s! f(p, \theta, s+y)} \; .$$

Then the score functions for $\tau,\ p$ and θ can be expressed as follows:

$$\begin{aligned} \frac{\partial \ell(\tau, p, \theta; y)}{\partial \tau} &= \frac{y}{\tau} - 1 + \frac{1}{h(p, \theta, \tau, y)} \frac{\partial h(p, \theta, \tau, y)}{\partial \tau} \\ &= \frac{y}{\tau} - 1 + \frac{1}{h(p, \theta, \tau, y)} \sum_{s=0}^{\infty} \frac{s[\tau(1-p)\theta^y]^s}{\tau s! f(p, \theta, s+y)} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial \ell(\tau, p, \theta; y)}{\partial p} &= \frac{y}{p} + \frac{1}{h(p, \theta, \tau, y)} \frac{\partial h(p, \theta, \tau, y)}{\partial p} \\ &= \frac{y}{p} + \frac{1}{h(p, \theta, \tau, y)} \left\{ - \sum_{s=0}^{\infty} \frac{s[\tau(1-p)\theta^y]^s}{(1-p) s! f(p, \theta, s+y)} \right. \\ &\quad \left. - \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s!} \frac{1}{[f(p, \theta, s+y)]^2} \frac{\partial f(p, \theta, s+y)}{\partial p} \right\} \\ &= \frac{y}{p} - \frac{1}{h(p, \theta, \tau, y)} \left\{ \sum_{s=0}^{\infty} \frac{s[\tau(1-p)\theta^y]^s}{(1-p) s! f(p, \theta, s+y)} \right. \\ &\quad \left. + \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s! f(p, \theta, s+y)} \frac{\partial \ln f(p, \theta, s+y)}{\partial p} \right\} \; . \end{aligned} \tag{6.7}$$

Suppose that K is a random variable that follows a multiplicative binomial distribution with parameters p and θ and number of trials $s+y$. Then expression (6.7) can be written more succinctly as

$$\begin{aligned} \frac{\partial \ell(\tau, p, \theta; y)}{\partial p} &= \frac{y}{p} - \frac{1}{h(p, \theta, \tau, y)} \left\{ \sum_{s=0}^{\infty} \frac{s[\tau(1-p)\theta^y]^s}{(1-p) s! f(p, \theta, s+y)} \right. \\ &\quad \left. + \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s! f(p, \theta, s+y)} \left[\frac{\mathrm{E}(K)}{p} - \frac{s+y-\mathrm{E}(K)}{1-p} \right] \right\} \; . \end{aligned} \tag{6.8}$$

Finally,

$$\begin{aligned}
\frac{\partial \ell(\tau,p,\theta;y)}{\partial \theta} &= \frac{1}{h(p,\theta,\tau,y)}\frac{\partial h(p,\theta,\tau,y)}{\partial \theta} \\
&= \frac{1}{h(p,\theta,\tau,y)}\left\{\sum_{s=0}^{\infty}\frac{ys[\tau(1-p)\theta^y]^s}{\theta s! f(p,\theta,s+y)} - \sum_{s=0}^{\infty}\frac{[\tau(1-p)\theta^y]^s}{s![f(p,\theta,s+y)]^2}\frac{\partial f(p,\theta,s+y)}{\partial \theta}\right\} \\
&= \frac{1}{h(p,\theta,\tau,y)}\left\{\sum_{s=0}^{\infty}\frac{ys[\tau(1-p)\theta^y]^s}{\theta s! f(p,\theta,s+y)} - \sum_{s=0}^{\infty}\frac{[\tau(1-p)\theta^y]^s}{s! f(p,\theta,s+y)}\frac{\partial \ln f(p,\theta,s+y)}{\partial \theta}\right\} \\
&= \frac{1}{h(p,\theta,\tau,y)}\left\{\sum_{s=0}^{\infty}\frac{ys[\tau(1-p)\theta^y]^s}{\theta s! f(p,\theta,s+y)} - \sum_{s=0}^{\infty}\frac{[\tau(1-p)\theta^y]^s}{s! f(p,\theta,s+y)}\frac{\mathrm{E}[K(s+y-K)]}{\theta}\right\}, \qquad (6.9)
\end{aligned}$$

where K again follows a multiplicative binomial distribution with parameters p and θ and number of trials given by $s+y$.

6.4.3 Information Matrix

The information matrix for the Altham-Poisson distribution is minus the expected value of the second derivative of the log-likelihood function with respect to the parameters $\tau,\ p$ and θ and is written as

$$\mathfrak{I} = -E\begin{bmatrix} \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial \tau^2} & \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial \tau \partial p} & \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial \tau \partial \theta} \\ \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial \tau \partial p} & \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial p^2} & \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial p \partial \theta} \\ \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial \tau \partial \theta} & \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial p \partial \theta} & \frac{\partial^2 \ell(\tau,p,\theta;Y)}{\partial \theta^2} \end{bmatrix}.$$

Consider the second derivative of the log-likelihood function with respect to τ given by

$$\begin{aligned}
&\frac{\partial^2 \ell(\tau,p,\theta;y)}{\partial \tau^2} \\
&= -\frac{y}{\tau^2} - \frac{1}{[h(p,\theta,\tau,y)]^2}\left(\frac{\partial h(p,\theta,\tau,y)}{\partial \tau}\right)^2 + \frac{1}{h(p,\theta,\tau,y)}\frac{\partial^2 h(p,\theta,\tau,y)}{\partial \tau^2} \\
&= -\frac{y}{\tau^2} - \frac{1}{\tau^2[h(p,\theta,\tau)]^2}\left\{\sum_{s=0}^{\infty}\frac{s[\tau(1-p)\theta^y]^s}{s! f(p,\theta,s+y)}\right\}^2 + \frac{1}{h(p,\theta,\tau)}\sum_{s=0}^{\infty}\frac{s(s-1)[\tau(1-p)\theta^y]^s}{\tau^2 s! f(p,\theta,s+y)}
\end{aligned}$$

The expression for this second derivative includes a number of infinite sums, and indeed the square of an infinite sum. These sums would need to be approximated by finite sums. The expected value of this second derivative is then required and cannot be obtained explicitly. The second derivatives of the log-likelihood function with respect to the parameters p and θ are more awkward computationally than that for τ and, as with the second derivative of $\ell(\tau, p, \theta; y)$ with respect to τ, can only be estimated numerically.

An alternative method for obtaining the information matrix is to compute the expected value of the matrix of the pairwise products of the score functions. Consider, for example, the score functions of the parameters τ and p from Section 6.4.2:

$$\frac{\partial \ell(\tau, p, \theta; y)}{\partial \tau} = \frac{y}{\tau} - 1 + \frac{1}{h(p, \theta, \tau, y)} \sum_{s=0}^{\infty} \frac{s[\tau(1-p)\theta^y]^s}{\tau s! f(p, \theta, s+y)}$$

and

$$\begin{aligned} \frac{\partial \ell(\tau, p, \theta; y)}{\partial p} = \frac{y}{p} - \frac{1}{h(p, \theta, \tau, y)} \Bigg\{ & \sum_{s=0}^{\infty} \frac{s[\tau(1-p)\theta^y]^s}{(1-p) s! f(p, \theta, s+y)} \\ & + \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^y]^s}{s! f(p, \theta, s+y)} \left[\frac{\mathrm{E}(K)}{p} - \frac{s + y - \mathrm{E}(K)}{1-p} \right] \Bigg\} . \end{aligned}$$

The product of these score functions introduces a product of infinite sums which cannot be readily approximated. The expectation of this product is then required and cannot be computed. Thus the information matrix of the Altham-Poisson distribution cannot be derived explicitly.

The observed Fisher information matrix can be used to estimate the information matrix but that too is difficult to compute and presents the problem of convergence of the infinite sums. Since numerical procedures such as `hessp` in the Gauss programming language are available for estimating the observed information matrix, this seems to be the most suitable way of approximating the Hessian matrix of the Altham-Poisson distribution. The information matrix is then estimated as minus the Hessian matrix.

6.5 Estimation

6.5.1 EM algorithm

The expectation-maximization (EM) algorithm is frequently utilized when estimates of parameters are sought for models that depend on incomplete data. Under ideal circumstances the complete likelihood function which depends on complete data would be maximized but since some data is unavailable the expectation of the log-likelihood function, $\mathrm{E}[\ell(\tau, p, \theta; y)]$ is considered instead of the complete log-likelihood function. The EM algorithm is iterative and each iteration is comprised of an expectation (E) step and a maximization (M) step. In iteration $(m+1)$ of the algorithm the E step entails computing the expected value of the log-likelihood function for the complete data set, conditional on the observed data and the parameter estimates obtained in iteration (m). The initial E step of the algorithm can be implemented by using the largest observation as a starting value for τ and fitting the multiplicative binomial distribution to the data using this value of τ in place of n to obtain starting values for p and θ. The E step is then followed by the M step in which the expected log-likelihood from the E step is maximized with respect to the parameters to yield updated parameter estimates. A new expected log-likelihood function is then computed based on the parameters from iteration $(m+1)$ and the observed data. The algorithm continues to alternate between these two steps until the parameter estimates obtained from successive iterations are satisfactorily close or the change in the log-likelihood in successive iterations is negligible. A detailed discussion of this algorithm is presented by Dempster, Laird and Rubin (1977).

Consider utilizing the EM algorithm to estimate the parameters of the Altham-Poisson distribution. In the present study the random variable Y is observed and the number of trials initially in the system, N, constitutes the missing component of the data. The joint p.m.f. of Y and N is written as

$$\begin{aligned}\Pr(Y=y,N=n) &= \Pr(Y=y|N=n)\Pr(N=n)\\ &= \frac{\left(\begin{array}{c} n \\ y \end{array}\right) p^y(1-p)^{n-y}\theta^{y(n-y)}}{f(p,\theta,n)}\frac{\tau^n \mathrm{e}^{-\tau}}{n!}\ ,\end{aligned}$$

so that the complete log-likelihood, which is denoted ℓ_{cp}, is

$$\begin{aligned}\ell_{cp}(\tau,p,\theta;y,n) &= \sum_{i=1}^{k}\left\{\ln\left(\begin{array}{c} n_i \\ y_i \end{array}\right) + y_i\ln p + (n_i-y_i)\ln(1-p) + y_i(n_i-y_i)\ln\theta\right.\\ &\quad \left. +n_i\ln\tau - \tau - \ln(n_i)! - \ln f(p,\theta,n_i)\right\}\ ,\end{aligned}$$

where k is the number of observed responses. Since the terms that exclude the parameters $p,\ \theta$ and τ will not affect their estimates they can be excluded from the log-likelihood to give

$$\ell^*_{cp}(\tau,p,\theta;y,n) = \sum_{i=1}^{k}\left\{y_i\ln p + (n_i-y_i)\ln(1-p) + y_i(n_i-y_i)\ln\theta + n_i\ln\tau - \tau - \ln f(p,\theta,n_i)\right\}\ . \tag{6.10}$$

E Step

Let $\tau^{(m)},\ p^{(m)}$ and $\theta^{(m)}$ denote the estimates of the parameters obtained from iteration (m) of the EM algorithm. In order to compute the expected log-likelihood function with respect to N for iteration $(m+1)$ the conditional log-likelihood which is written as

$$\begin{aligned}\ell^*_{cp(Y|N)}(\tau,p,\theta;y|N) &= \sum_{i=1}^{k}\{y_i\ln p + [\mathrm{E}(N_i)-y_i]\ln(1-p) + y_i[\mathrm{E}(N_i)-y_i]\ln\theta\\ &\quad + \mathrm{E}(N_i)\ln\tau - \tau - \mathrm{E}[\ln f(p,\theta,N_i)]\}\ .\end{aligned}$$

is sought. For convenience N and Y are considered corresponding to a single dose and the subscript $i,\ i=1,\ldots,k$ is removed. The conditional distribution of N given the observed data, Y, and the parameter estimates $\tau^{(m)},\ p^{(m)}$ and $\theta^{(m)}$ is

proportional to the following expression:

$$\begin{aligned}\Pr(N|Y) &\propto \frac{(1-p^{(m)})^{n-y}\ {\tau^{(m)}}^{n-y}\ {\theta^{(m)}}^{y(n-y)}}{(n-y)!}\frac{1}{f(p^{(m)},\theta^{(m)},n)} \\ &= \frac{[(1-p^{(m)})\tau^{(m)}{\theta^{(m)}}^{y}]^{n-y}}{(n-y)!}\frac{1}{f(p^{(m)},\theta^{(m)},n)}\end{aligned}$$

and therefore

$$\Pr(N-y=n^*) \propto \frac{[(1-p^{(m)})\tau^{(m)}{\theta^{(m)}}^{y}]^{n^*}}{n^*!}\frac{1}{f(p^{(m)},\theta^{(m)},n^*+y)}\ . \quad (6.11)$$

Thus the conditional distribution of N given Y, $\tau^{(m)}$, $p^{(m)}$ and $\theta^{(m)}$ is a weighted Poisson distribution with weight $e^{-[(1-p^{(m)})\tau^{(m)}{\theta^{(m)}}^{y}]^n}\dfrac{1}{f(p^{(m)},\theta^{(m)},n+y)}$. In the E step of the EM algorithm for the Altham-Poisson distribution, expressions for $\mathrm{E}[N|y,\tau^{(m)},p^{(m)},\theta^{(m)}]$ and $\mathrm{E}[\ln f(p,\theta,N)|y,\tau^{(m)},p^{(m)},\theta^{(m)}]$ are sought. The expected value of N conditional on the data and the current parameter estimates can be obtained from expression (6.11) as

$$E[N|y,\tau^{(m)},p^{(m)},\theta^{(m)}] = \frac{\displaystyle\sum_{n=y}^{\infty}\frac{n[\tau^{(m)}(1-p^{(m)})\theta^{(m)y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}}{\displaystyle\sum_{n=y}^{\infty}\frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}}\ , \quad (6.12)$$

where the denominator in (6.12) is a normalizing constant for the conditional distribution of N given $Y=y$ ensuring that the probabilities sum to 1. The conditional expectation in (6.12) is therefore a ratio of infinite sums which can be approximated by selecting a large value n_{cut} and evaluating a ratio of finite sums. Suppose that $\mathrm{E}[N|y,\tau^{(m)},p^{(m)},\theta^{(m)}]$ is denoted by $\hat{N}^{(m+1)}$, then

$$\hat{N}^{(m+1)} \approx \frac{\displaystyle\sum_{n=y}^{n_{\mathrm{cut}}}\frac{n[\tau^{(m)}(1-p^{(m)})\theta^{(m)y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}}{\displaystyle\sum_{n=y}^{n_{\mathrm{cut}}}\frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}}\ .$$

Note that for the examples considered in this study, an exceedingly large value of n_{cut} that was sufficient for both expressions in the numerator and denominator was

selected. The conditional expectation of $\ln[f(p,\theta,N)]$, however, which is written as

$$E[\ln f(p,\theta,N)|y,\tau^{(m)},p^{(m)},\theta^{(m)}] = \frac{\sum_{n=y}^{\infty}\frac{\ln f(p,\theta,n)[\tau^{(m)}(1-p^{(m)})\theta^{(m)^y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}}{\sum_{n=y}^{\infty}\frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)^y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}} \tag{6.13}$$

is a function of the parameters that require estimating and cannot be evaluated. As a result the E step cannot be completed and the conditional expectation of $\ln[f(p,\theta,N)]$ is addressed in the M step of the algorithm. A similar problem was encountered by Zhu, Eickhoff and Kaiser (2003) who used the EM algorithm in a similar context but for a continuous distribution. In personal correspondence with Eickhoff it was determined that implementing the EM algorithm entailed differentiating within an integral when using the Newton-Raphson routine in the M-step.

M Step

Consider maximizing the expected log-likelihood

$$\begin{aligned} E^*_{cp} &= \sum_{i=1}^{k}\{y_i\ln p+(\hat{N}_i^{(m+1)}-y_i)\ln(1-p)+y_i(\hat{N}_i^{(m+1)}-y_i)\ln\theta-E(\ln f(p,\theta,N_i)) \\ &\quad +\hat{N}_i^{(m+1)}\ln\tau-\tau\} \end{aligned} \tag{6.14}$$

with respect to τ, where $\mathrm{E}(N_i)$ is replaced by $\hat{N}_i^{(m+1)}$ from the E step. Then

$$\frac{\partial E^*_{cp}}{\partial\tau} = \sum_{i=1}^{k}\frac{\hat{N}_i^{(m+1)}}{\tau}-k=0$$

which yields

$$\hat{\tau}^{(m+1)}=\frac{\sum_{i=1}^{k}\hat{N}_i^{(m+1)}}{k}\ ,$$

where $\hat{N}_i^{(m+1)}$ is the most recent expected value of N_i, $i=1,\ldots,k$ and depends on the parameters estimated in iteration (m) of the algorithm.

Since the expected value of $\ln[f(p,\theta,N)]$ cannot be evaluated in the E step it is incorporated now into the M step and two methods for estimating p and θ are considered.

Method 1

An optimization routine can be used to maximize the expected log-likelihood function E_c^* with respect to p and θ. If this method is adopted the expected value of $\ln[f(p,\theta,N)]$ in expression (6.13) must itself be maximized with respect to p and θ for that particular iteration of the algorithm and using the parameter estimates $\tau^{(m)}$, $p^{(m)}$ and $\theta^{(m)}$ from the previous iteration to yield updated parameter estimates. A particularly large cut-off value n_{cut} is once again required so that the ratio of infinite sums in expression (6.13) can be approximated by a ratio of finite sums

$$E[\ln f(p,\theta,N)|y,\tau^{(m)},p^{(m)},\theta^{(m)}] = \frac{\displaystyle\sum_{n=y}^{n_{\text{cut}}} \frac{\ln f(p,\theta,n)[\tau^{(m)}(1-p^{(m)})\theta^{(m)y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}}{\displaystyle\sum_{n=y}^{n_{\text{cut}}} \frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)y}]^n}{n!f(p^{(m)},\theta^{(m)},n+y)}}$$

and then maximized accordingly.

Method 2

Alternatively consider differentiating the expected log-likelihood with respect to p and θ and solving the two maximum likelihood equations simultaneously. The maximum likelihood equations from (6.14) for p and θ are

$$\begin{aligned}\frac{\partial E_{cp}^*}{\partial p} &= \sum_{i=1}^{k}\left\{\frac{y_i}{p}-\frac{E(N_i-y_i)}{1-p}-\frac{\displaystyle\sum_{n=y}^{\infty}\frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)y_i}]^{n_i}}{n_i!f(p^{(m)},\theta^{(m)},n_i+y_i)}\frac{\partial \ln f(p,\theta,n_i)}{\partial p}}{\displaystyle\sum_{n=y}^{\infty}\frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)y_i}]^{n_i}}{n_i!f(p^{(m)},\theta^{(m)},n_i+y_i)}}\right\}\\ &= 0\end{aligned}$$

and

$$\frac{\partial E^*_{cp}}{\partial \theta} = \sum_{i=1}^{k} \left\{ \frac{y_i(E(N_i) - y_i)}{\theta} - \frac{\sum_{n=y}^{\infty} \frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)y_i}]^{n_i}}{n_i! f(p^{(m)}, \theta^{(m)}, n_i + y_i)} \frac{\partial \ln f(p, \theta, n_i)}{\partial \theta}}{\sum_{n=y}^{\infty} \frac{[\tau^{(m)}(1-p^{(m)})\theta^{(m)y_i}]^{n_i}}{n_i! f(p^{(m)}, \theta^{(m)}, n_i + y_i)}} \right\}$$
$$= 0$$

respectively. Solutions to these simultaneous equations cannot be written down explicitly. Therefore, since numerical methods are required to estimate p and θ, it is sensible to utilize the optimization procedure described as Method 1.

The algorithm can be initialized using the starting values of the parameters described at the beginning of this section. Once updated parameter estimates have been obtained in the E step, these estimates can be used to approximate the conditional expectation of N which will then in turn be used to update the estimates of the parameters. The algorithm oscillates between the E and M steps until the selected stopping criterion is satisfied.

6.5.2 Direct Maximization of the Likelihood

The parameters $\tau,\ p$ and θ of the Altham-Poisson distribution can be estimated by maximizing the log-likelihood function directly. Consider a random sample from the Altham-Poisson distribution with parameters $\tau,\ p$ and θ and suppose that the response y_i is observed k_i times, $i = 1, \ldots, d$. Then the log-likelihood function, which was previously derived in Section 6.4.1, can be written as

$$\ell(\tau, p, \theta; \underline{y}) = \sum_{i=1}^{d} k_i \left\{ y_i \ln(\tau p) - \tau - \ln(y_i!) + \ln \left\{ \sum_{s=0}^{\infty} \frac{[\tau(1-p)\theta^{y_i}]^s}{s! f(p, \theta, s + y_i)} \right\} \right\} .$$

Due to the infinite sum in the above expression maximum likelihood estimates for $\tau,\ p$ and θ cannot be obtained explicitly. Furthermore the log-likelihood function cannot be computed because it includes an infinite sum. The infinite sum can, however, be approximated for a given set of parameter values by a finite sum,

where a cutoff value is selected to meet a specified degree of accuracy using the method outlined in Section 6.2.2.

Once the log-likelihood function has been approximated it can be optimized numerically with respect to $\tau,\ p$ and θ using procedures such as OPTMUM which is available in the programming language Gauss. The optimization routine requires the constraints $0 \leq p \leq 1$ and $\tau,\ \theta > 0$ to be imposed on the parameters.

6.5.3 Inference

The information matrix of the Altham-Poisson distribution cannot be determined explicitly as noted in Section 6.4.3. However, the Hessian matrix can be approximated numerically and can be used to estimate the standard errors of the parameter estimates. 95% Wald intervals can then be obtained for the parameters. For example, the 95% Wald interval for τ is given by $\hat{\tau} \pm 1.96se(\hat{\tau})$, where $se(\hat{\tau})$ is the standard error of the estimate of τ (Kutner, Nachtsheim, Neter and Li, 2005, p.579).

For each of the parameters $\tau,\ p$ and θ of the Altham-Poisson distribution, the profile likelihood function can be plotted and used to construct approximate confidence intervals. The 95% confidence interval for the parameter τ is given by $[\tau_1,\tau_2]$ which satisfy

$$\ell_P(\tau_1) = \ell_P(\tau_2) = \ell(\hat{\tau}, \hat{p}, \hat{\theta}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05}$$

where $\ell_P(\tau_1)$ is the profile likelihood for τ evaluated at τ_1 and $\ell(\hat{\tau}, \hat{p}, \hat{\theta}; \underline{y})$ is the maximum log-likelihood of the Altham-Poisson distribution (Azzalini, 1996, p.146). Similarly, 95% confidence intervals can be obtained for p and θ.

6.5.4 Goodness of Fit

Two methods can be used to assess the adequacy of the fit of the Altham-Poisson distribution to overdispersed count data. A chi-squared goodness of fit test can be

performed, where a significantly large χ^2 value indicates that the Altham-Poisson distribution provides a poor fit to the data.

The AIC statistic is also available for comparing the fit of the Altham-Poisson distribution with those of the Poisson and Poisson-like distributions. For the Altham-Poisson distribution it is calculated as $AIC = -2\ell(\hat{\tau}, \hat{p}, \hat{\theta}; \underline{y}) + 6$, where $q = 3$. The model with the lowest AIC value is the one that provides the best fit to the data under consideration.

6.6 An Example

The Altham-Poisson distribution was fitted to the sales data described in Section 4.7 using both the EM algorithm and direct optimization of the log-likelihood function. The EM algorithm was found to be disappointing for a number of reasons. Firstly, the algorithm is slow to converge (Nelder, 1977) and in this context it is particularly slow to implement because it requires repeated optimization of a complicated log-likelihood function. Furthermore, the EM algorithm does not immediately give the Hessian matrix (Little, 1977). Therefore, since the parameter estimates obtained by directly optimizing the likelihood function were similar to those obtained by the EM algorithm, direct optimization was selected as the desired method for estimating the parameters of the Altham-Poisson distribution for the remainder of this thesis.

The Hessian matrix was obtained numerically using the `hessp` command in Gauss and minus the Hessian was used to estimate the information matrix. The estimated information matrix was inverted and the square roots of the diagonal elements of the resulting matrix were used to estimate the standard errors of the parameter estimates. 95% Wald intervals were constructed for the parameters using these standard errors. Profile likelihood plots were obtained for each of the parameters τ, p and θ and these plots were used to construct 95% profile likelihood intervals for the parameters. Figures 6.2 (a), (b) and (c) contain the

respective profile likelihood plots for τ, p and θ. The dashed line in the plots represents the value of $\ell(\hat{\tau}, \hat{p}, \hat{\theta}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05}$ which was used to obtain the 95% profile likelihood intervals for the parameters. The estimates of the parameters, as well as their standard errors, 95% Wald confidence intervals and 95% profile likelihood intervals are presented in Table 6.2.

Table 6.2: Results from fitting the Altham-Poisson model to the sales data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	12.618	1.181	(10.303;14.933)	(10.815;14.250)
p	0.458	0.005	(0.448;0.468)	(0.450;0.467)
θ	0.843	0.012	(0.819;0.867)	(0.823;0.862)

The fit of the Altham-Poisson distribution to the sales data was compared with those of the ordinary Poisson distribution, the CMP and the beta-Poisson distributions using the chi-squared goodness of fit test as well as AIC values. The p-values resulting from the χ^2 test and the AIC statistics are presented in Table 6.3.

Table 6.3: A comparison of the various models fitted to the sales data.

Model	χ^2 p-value	AIC
Poisson	0.000	17 921.622
CMP	0.242	15 056.712
beta-Poisson	0.002	15 057.934
Altham-Poisson	0.005	15 083.308

The p-values from the chi-squared goodness of fit tests indicate that the CMP distribution provides a good fit to the sales data and that none of the other models seem suitable for these data. The AIC values suggest that the CMP, beta-Poisson and Altham-Poisson distributions provide superior fits to the sales data than the Poisson distribution. The CMP provides the best fit to the data but its AIC

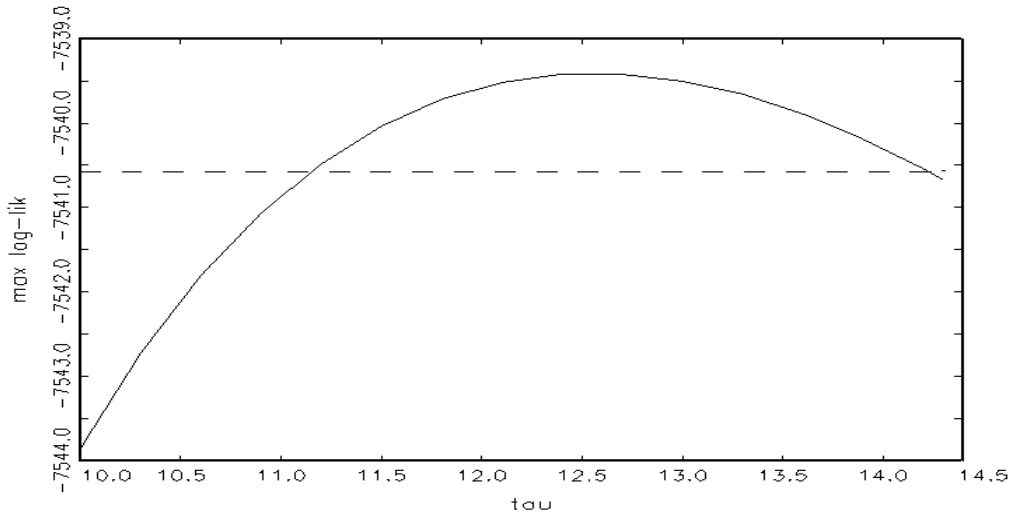


Figure 6.2: (a)

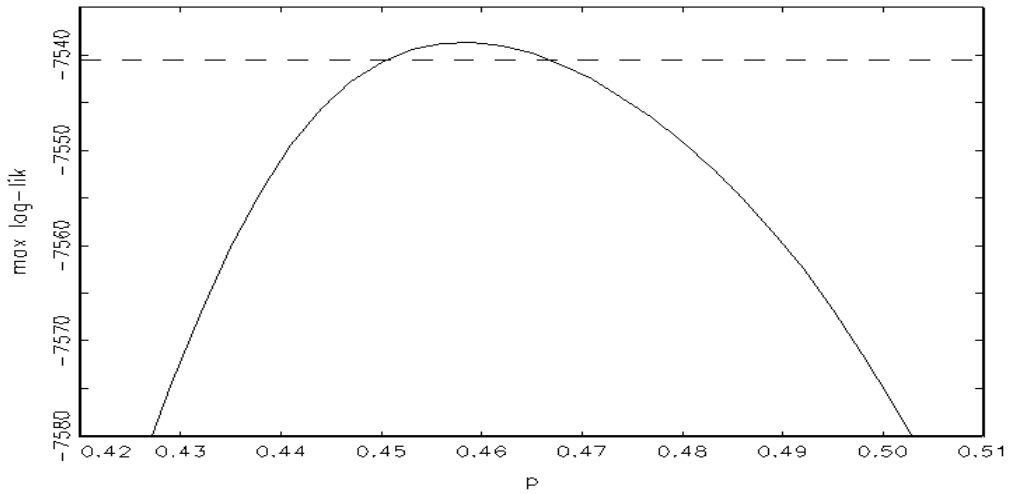


Figure 6.2: (b)

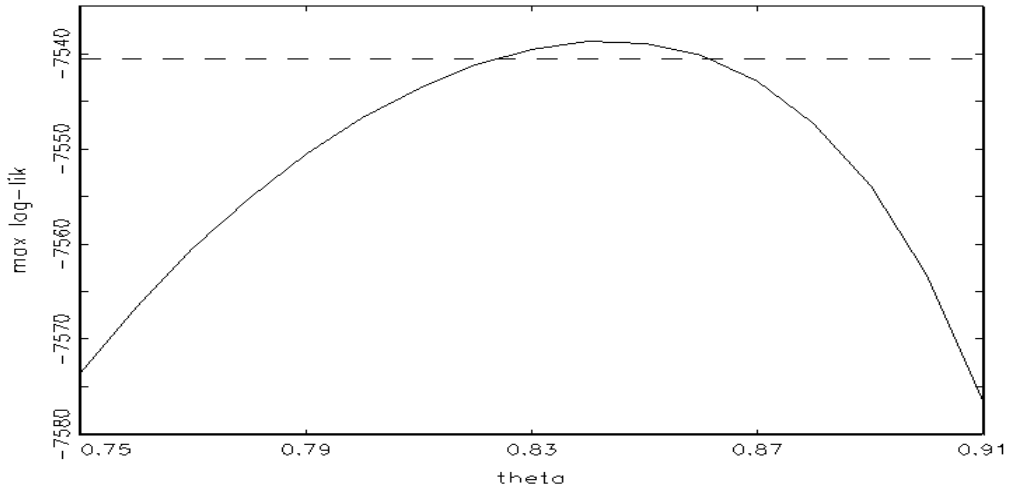


Figure 6.2: (c)

Figure 6.2: Profile likelihood plots for the parameters (a) τ, (b) p and (c) θ of the Altham-Poisson model fitted to the sales data.

value is only marginally lower than those of the beta-Poisson and Altham-Poisson distributions.

6.7 Modelling Dose-Response Data

6.7.1 Modelling the Probabilities

In the context of a dose-response setting, suppose that the parameter p refers to the probability of the death of an organism. The definition of p in this setting is therefore equivalent to the probability of a failure, $1 - p$, of the multiplicative binomial distribution. Consider modelling the probability of death as a function of the log of the dose administered to a subject using a logit link function. The parameter p can be written in terms of log-dose x and the parameters α and β as

$$p = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} . \tag{6.15}$$

6.7.2 Likelihood, Score Functions and Information Matrix

Likelihood

Consider a dose-response study in which the random variable Y denotes the number of organisms that survive treatment with a particular dose of a drug. Following the notation introduced in Chapter 2, let $y_{cj},\ j = 1, \ldots, n_c$ denote an observation from the control group and let y_{ij} refer to the number of organisms that survive exposure to a non-zero dose d_i of the drug, with log-dose x_i and where $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Furthermore suppose that p_i represents the probability that an organism does not survive exposure to the ith dose of the drug. Since the increase in probability of death of an organism exposed to a zero dose of the drug is zero, the number of survivors in the control group will simply follow a Poisson distribution with parameter τ. Thus the likelihood and log-likelihood functions for the

control group are given by

$$L_c = \prod_{j=1}^{n_c} \frac{e^{-\tau}\tau^{y_{cj}}}{y_{cj}!}$$

and

$$\ell_c = \sum_{j=1}^{n_c} y_{cj} \ln \tau - n_c\tau - \sum_{j=1}^{n_c} \ln y_{cj}!$$

respectively.

Consider the number of organisms that survive treatment with non-zero doses of the drug and suppose that these responses follow an Altham-Poisson distribution. The likelihood and log-likelihood functions are written respectively as

$$L_d = \prod_{i=1}^{D}\prod_{j=1}^{n_i} \frac{[\tau(1-p)]^{y_{ij}}e^{-\tau}}{y_{ij}!} \sum_{s=0}^{\infty} \frac{(\tau p\theta^{y_{ij}})^s}{s!f(1-p,\theta,s+y_{ij})}$$

and

$$\ell_d = \sum_{i=1}^{D}\sum_{j=1}^{n_i} y_{ij} \ln[\tau(1-p)] - \tau - \ln(y_{ij}!) + \ln\left\{\sum_{s=0}^{\infty} \frac{(\tau p\theta^{y_{ij}})^s}{s!f(1-p,\theta,s+y_{ij})}\right\},$$

where p is expressed in terms of α and β using a logit link function from equation (6.15). The log-likelihood for the data is then found by summing the log-likelihood function for each observation over all of the observed responses and is given by

$$\begin{aligned}\ell(\tau,\alpha,\beta,\theta;\underline{y}) &= \ell_c(\tau,\underline{y}) + \ell_d(\tau,\alpha,\beta,\theta;\underline{y}) \\ &= \sum_{j=1}^{n_c} y_{cj} \ln \tau - n_c\tau - \sum_{j=1}^{n_c} \ln y_{cj}! \\ &\quad + \sum_{i=1}^{D}\sum_{j=1}^{n_i} y_{ij} \ln[\tau(1-p)] - \tau - \ln(y_{ij}!) + \ln\left\{\sum_{s=0}^{\infty} \frac{(\tau p\theta^{y_{ij}})^s}{s!f(1-p,\theta,s+y_{ij})}\right\}.\end{aligned} \tag{6.16}$$

Score Functions and Information Matrix

Since the score functions and information matrix for the observations from the control group have been derived in Chapter 2, consider the score functions for the

observed number of organisms surviving exposure to a non-zero dose of the drug. The score functions for τ and θ of the Altham-Poisson distribution are not affected by the adaptation of the model to a dose-response setting and are therefore given by

$$\frac{\partial \ell(\tau,\alpha,\beta,\theta;y)}{\partial \tau} = \frac{y}{\tau} - 1 + \frac{1}{h(1-p,\theta,\tau,y)} \sum_{s=0}^{\infty} \frac{s(\tau p \theta^y)^s}{\tau s! f(1-p,\theta,s+y)}$$

and

$$\begin{aligned}\frac{\partial \ell(\tau,\alpha,\beta,\theta;y)}{\partial \theta} &= \frac{1}{h(1-p,\theta,\tau,y)} \left\{ \sum_{s=0}^{\infty} \frac{ys(\tau p \theta^y)^s}{\theta s! f(1-p,\theta,s+y)} \right. \\ &\quad \left. - \sum_{s=0}^{\infty} \frac{(\tau p \theta^y)^s}{s! f(1-p,\theta,s+y)} \frac{\mathrm{E}[K(s+y-K)]}{\theta} \right\}\end{aligned}$$

respectively, where K is a random variable following a multiplicative binomial distribution with parameters p and θ and number of trials $s+y$. These functions were derived in Section 6.4.2.

Consider, therefore, the score functions for the parameters α and β which are written respectively as

$$\begin{aligned}\frac{\partial \ell(\tau,\alpha,\beta,\theta;y)}{\partial \alpha} &= \frac{\partial \ell(\tau,\alpha,\beta,\theta;y)}{\partial p} \frac{\partial p}{\partial \alpha} \\ &= -py + \frac{p(1-p)}{h(1-p,\theta,\tau,y)} \left\{ \sum_{s=0}^{\infty} \frac{s(\tau p \theta^y)^s}{ps! f(1-p,\theta,s+y)} \right. \\ &\quad \left. + \sum_{s=0}^{\infty} \frac{(\tau p \theta^y)^s}{s! f(1-p,\theta,s+y)} \left[\frac{s+y-\mathrm{E}(K)}{p} - \frac{\mathrm{E}(K)}{1-p} \right] \right\}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial \ell(\tau,\alpha,\beta,\theta;y)}{\partial \beta} &= \frac{\partial \ell(\tau,\alpha,\beta,\theta;y)}{\partial p} \frac{\partial p}{\partial \beta} \\ &= -xpy + \frac{xp(1-p)}{h(1-p,\theta,\tau,y)} \left\{ \sum_{s=0}^{\infty} \frac{s(\tau p \theta^y)^s}{ps! f(1-p,\theta,s+y)} \right. \\ &\quad \left. + \sum_{s=0}^{\infty} \frac{(\tau p \theta^y)^s}{s! f(1-p,\theta,s+y)} \left[\frac{s+y-\mathrm{E}(K)}{p} - \frac{\mathrm{E}(K)}{1-p} \right] \right\} .\end{aligned}$$

Following Section 6.4.3 the information matrix of the Altham-Poisson distribution cannot be derived explicitly. The Gauss command `hessp`, however, yields

a numerical approximation of the Hessian matrix and minus this matrix can be used to estimate the information matrix of the Altham-Poisson distribution.

6.7.3 Model Fitting and Checking

The parameters of the Altham-Poisson distribution used to model dose-response data can be estimated by maximizing the log of the likelihood function numerically using the OPTMUM routine in Gauss. Suppose that y_{cj} denotes the number of survivors in the control group, $j = 1, \ldots, n_c$ and that y_{ij} refers to the number of individuals surviving exposure to a non-zero dose d_i of the drug, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. Then the log-likelihood function from expression (6.16) is written as

$$\begin{aligned}
\ell(\tau, \alpha, \beta, \theta; \underline{y}) &= \ell_c + \ell_d \\
&= \sum_{j=1}^{n_c} y_{cj} \ln \tau - n_c \tau - \sum_{j=1}^{n_c} \ln y_{cj}! \\
&\quad + \sum_{i=1}^{D} \sum_{j=1}^{n_i} y_{ij} \ln[\tau(1-p)] - \tau - \ln(y_{ij}!) + \ln \left\{ \sum_{s=0}^{\infty} \frac{(\tau p \theta^{y_{ij}})^s}{s! f(1-p, \theta, s + y_{ij})} \right\},
\end{aligned}$$

where the probability of survival p_i is written in terms of log-dose x_i using the logit link function from expression (6.15). The infinite sum included in the log-likelihood function can be approximated by a finite sum where the cutoff is chosen to achieve a desired level of accuracy.

The information matrix can be estimated by minus the Hessian matrix which can be obtained numerically using methods described in Section 6.7.2. This matrix can be used to approximate the standard errors of the parameter estimates. 95% Wald intervals can be obtained for the parameters τ, α, β and θ of the Altham-Poisson distribution. Profile likelihoods can also be plotted for each of the parameters. The profile likelihood plots can be used to obtain 95% confidence intervals for the parameters using the method outlined in Section 6.5.3.

The fit of the Altham-Poisson model can compared with that of the Poisson and the beta-Poisson distributions using AIC. The AIC of the Altham-Poisson

model, where p is modelled with a logit tolerance distribution, is calculated as $AIC = -2\ell(\hat{\tau}, \hat{\alpha}, \hat{\beta}, \hat{\theta}; \underline{y}) + 8$, where the number of parameters in the model is 4. The best fitting model has the smallest AIC value.

6.7.4 Results

Algae Data

The Altham-Poisson distribution with a logit function modelling the probability of death was fitted to the algae data introduced in Section 2.2.1. The approximation of the likelihood function was complicated by the large values of y and hence of τ. This is due to the fact that the values of τ^n as well as the terms in $f(p, \theta, n + y)$ contained in the infinite sum become too large to be evaluated by programs such as Gauss. A method of overcoming this computational difficulty is to take the logarithm of each term in the sum and in the function $f(p, \theta, n + y)$ and then exponentiate that log-term before adding it to the next term.

The standard errors of the parameter estimates were used to compute 95% Wald intervals for each of the parameters τ, α, β and θ. The profile likelihood plots for the parameters are presented in Figures 6.3 (a), (b), (c) and (d). The parameter estimates, together with their standard errors, 95% Wald confidence intervals and 95% profile likelihood intervals are recorded in Table 6.4.

Table 6.4: Results from fitting the Altham-Poisson model to the algae data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	224.272	6.504	(211.524;237.020)	(211.600;237.512)
α	-0.428	0.159	(-0.740;-0.116)	(*;-0.200)
β	0.660	0.241	(0.188;1.132)	(0.347;1.502)
θ	0.995	0.002	(0.991;0.999)	(0.992;*)

*These limits of the confidence intervals cannot be computed

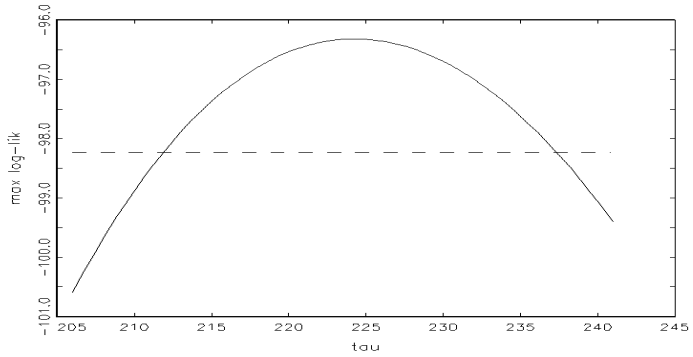


Figure 6.3: (a)

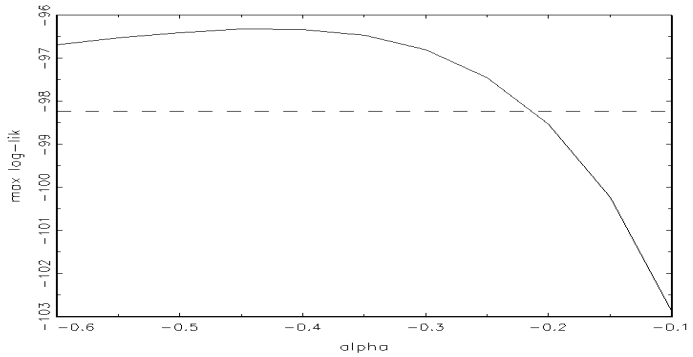


Figure 6.3: (b)

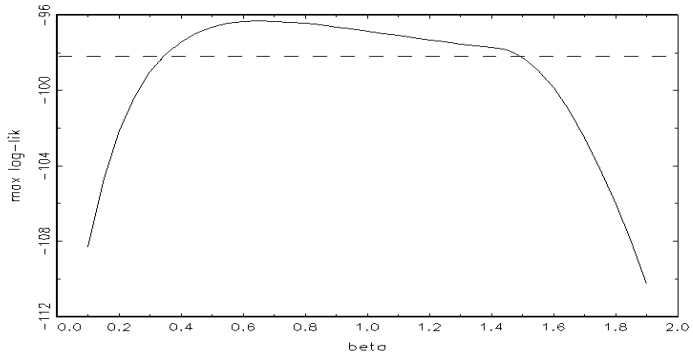


Figure 6.3: (c)

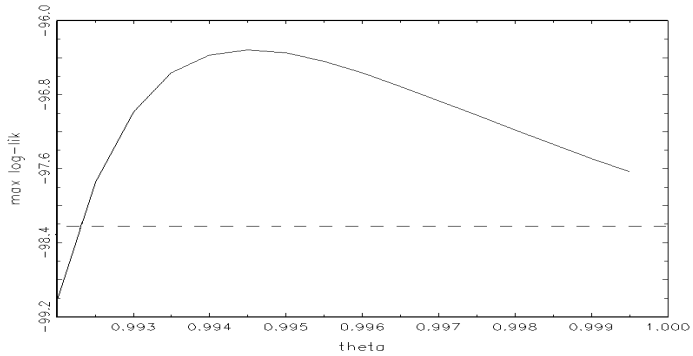


Figure 6.3: (d)

Figure 6.3: Profile likelihood plots for the parameters (a) τ, (b) α, (c) β and (d) θ of the Altham-Poisson model with a logit tolerance distribution fitted to the algae data.

The fit of the Altham-Poisson distribution to the algae data was compared with those of the Poisson and the beta-Poisson distributions. The comparison was made using the AIC values for each of the models and these values are presented in Table 6.5.

Table 6.5: A comparison of the various models fitted to the algae data.

Model	Maximum log-likelihood	Number of parameters	AIC
Poisson	-97.776	3	201.552
Beta-Poisson	-96.2507	4	200.501
Altham-Poisson	-96.314	4	200.628

A comparison of the AIC values shows that the worst fit is provided by the Poisson distribution. The AIC values for the beta-Poisson and Altham-Poisson distributions are very similar and are smaller than that of the Poisson distribution. Therefore the beta-Poisson and Altham-Poisson distributions seem to accommodate the overdispersion in the algae data more effectively than the Poisson distribution, with the beta-Poisson distribution providing arguably the best fit to the algae data.

Malaria Data

The discussion of the examples of the multiplicative binomial distribution seems to indicate that for large values of n the parameter θ must be close to one in order to increase the probability of observing responses between 0 and n, thereby making the distribution less U-shaped. A similar property has been observed for large values of the parameter τ of the Altham-Poisson distribution, as is indicated by the estimate of θ for the algae data. Thus an obvious limitation of the Altham-Poisson distribution is that it cannot easily model large values of y. Attempts to fit the Altham-Poisson distribution to the malaria data were not successful since the values of y, and hence of τ, were extremely large. An alternative example

of fitting the Altham-Poisson distribution to overdispersed dose-response data is therefore considered.

Bovine Data

Trajstman (1989) presented data obtained from a study on bovine tuberculosis in Australia. Samples of bovine tissue were placed on culture plates and the growth of *Mycobacterium bovis* observed. *Mycobacterium bovis* grows slowly and is often overtaken by contaminants. Thus culture plates need to be decontaminated prior to a study and a suitable dose of a decontaminant that kills as few *Mycobacterium bovis* organisms as possible is sought. Table 6.6 contains the colony counts of *Mycobacterium bovis* exposed to varying doses of the decontaminant oxalic acid for 12 weeks.

Table 6.6: Data for *Mycobacterium bovis* exposed to the decontaminant oxalic acid.

% weight/ volume of oxalic acid	Colony Count									
0	52	80	55	50	58	50	43	50	53	54
0	44	51	34	37	46	56	64	51	67	40
5	14	15	6	13	4	1	9	6	12	13
0.5	27	33	31	30	26	41	33	40	31	20
0.05	33	26	32	24	30	52	28	28	26	22
0.005	36	54	31	37	50	73	44	50	37	

Morgan (1992, p.265) considered the bovine data and mentioned that it was overdispersed. As a result the Altham-Poisson distribution was fitted to this data in an attempt to accommodate the overdispersion. A logit tolerance distribution was used to model the probability of death as a function of the log of the dose of the decontaminant. Profile likelihoods were plotted for each of the parameters τ, α, β and θ and the respective plots are presented Figures 6.4 (a) and (b). The

estimates of the parameters, their approximate standard errors and 95% Wald and profile likelihood intervals are presented in Table 6.7.

Table 6.7: Results from fitting the Altham-Poisson model to the bovine data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	53.928	2.822	(48.397;59.459)	(49.400;60.805)
α	0.090	0.042	(0.008;0.172)	(0.015;0.190)
β	0.211	0.088	(0.039;0.383)	(0.090;0.440)
θ	0.967	0.003	(0.961;0.973)	(0.961;0.971)

For purposes of comparison the beta-Poisson distribution was fitted to the bovine data and the fit of the Altham-Poisson distribution was compared with that of the Poisson and beta-Poisson distributions. Table 6.8 contains a comparison of the AIC values for the three models.

Table 6.8: A comparison of the various models fitted to the bovine data.

Model	Maximum log-likelihood	Number of parameters	AIC
Poisson	-236.415	3	478.830
Beta-Poisson	-219.447	4	446.894
Altham-Poisson	-221.427	4	450.854

The AIC values indicate that the beta-Poisson distribution provided the best fit to the bovine data and that the Altham-Poisson distribution provided a reasonable fit. The AIC value for the Altham-Poisson distribution was only slightly higher than that for the beta-Poisson distribution. Thus the Altham-Poisson distribution seemed to account reasonably well for some of the overdispersion in the bovine data.

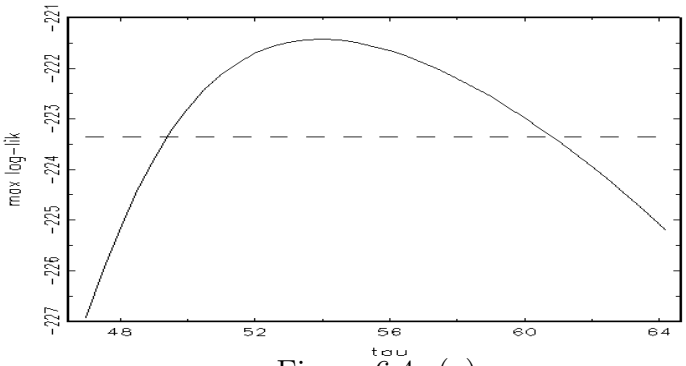


Figure 6.4: (a)

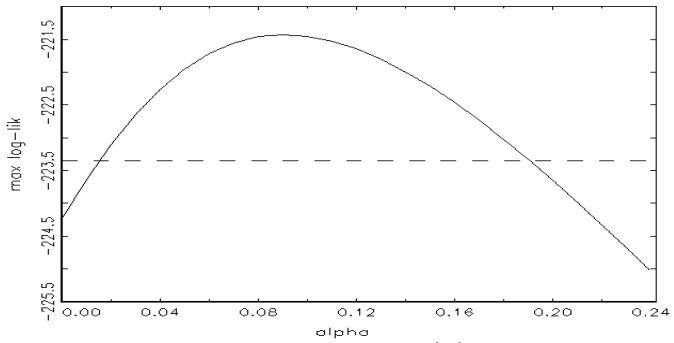


Figure 6.4: (b)

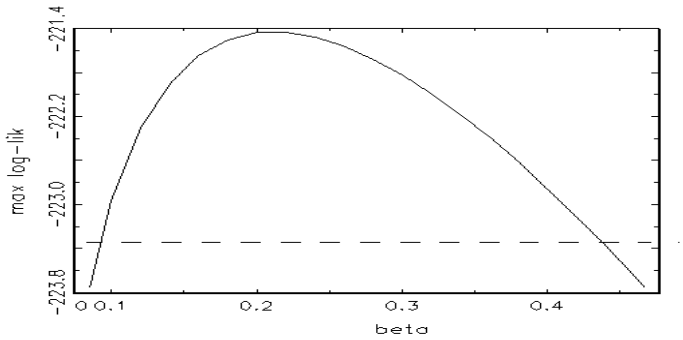


Figure 6.4: (c)

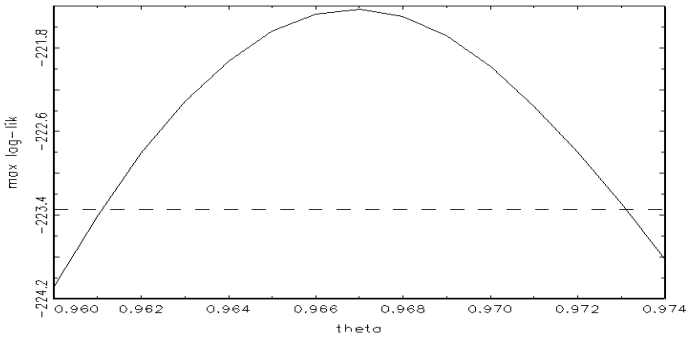


Figure 6.4: (d)

Figure 6.4: Profile likelihood plots for the parameters (a) τ, (b) α, (c) β and (d) θ of the Altham-Poisson model with a logit tolerance distribution to the bovine data.

6.8 Conclusions

The Altham-Poisson distribution was introduced in this chapter and it was observed that the p.m.f. of this distribution is that of a weighted Poisson model. Properties of the Altham-Poisson distribution were investigated. In particular, the p.m.f. includes an intractable infinite sum and methods for approximating the probabilities associated with the distribution were explored.

An EM algorithm was developed for estimating the parameters of the Altham-Poisson distribution but it was disappointingly slow in comparison to direct optimization of the likelihood function. The information matrix of the distribution cannot be readily calculated since expressions for the entries of this matrix include infinite sums. Thus minus the Hessian matrix, which can be obtained numerically, was used to estimate the information matrix and the standard errors of the parameter estimates were derived from its inverse.

The suitability of the Altham-Poisson distribution for modelling overdispersed count data was investigated by considering the sales data. Although the CMP model provided a better fit to the data, the Altham-Poisson distribution seemed to account reasonably well for the overdispersion.

The Altham-Poisson model was then extended in a straightforward way to a dose-response setting. It was observed that when the observations, and hence the value of τ, are very large the estimate of θ is very close to one and the distribution cannot easily model large observations. However, when the observations are not exceptionally large the Altham-Poisson distribution adequately accounts for overdispersion in a Wadley's problem setting.

Chapter 7

The Negative Binomial Model

7.1 Introduction

In the present chapter, the negative binomial distribution, which is suitable for modelling overdispersed count data (Young, Campbell and Capuano, 1999), is used to model dose-response data in a Wadley's problem setting. Anscombe (1949) first considered using this distribution to model the unknown number of organisms initially exposed to a drug. This chapter serves as a prelude to the binomial-EWP distribution which follows in Chapter 8 and which focuses on using an alternative distribution for overdispersed count data to model Wadley's problem.

Section 7.2 contains a description of the model and an adaptation of the model to a dose-response setting is considered in Section 7.3. The likelihood, score functions and information matrix are discussed in Section 7.4 and models for separate doses are described in Section 7.5. Details regarding parameter estimation are included in Section 7.6. A discussion of the fit of the model to the algae and the malaria data sets is presented in Section 7.7 and conclusions are drawn in Section 7.8.

7.2 The Model

7.2.1 The Negative Binomial Distribution

Consider a sequence of independent Bernoulli trials with a constant probability of success denoted by p^* and let Y denote the number of failures until the occurrence of the rth success. Then Y is said to follow a negative binomial distribution (Hogg, McKean and Craig, 2005, p.136) and its p.m.f. is given by

$$\Pr(Y = y) = \begin{pmatrix} y + r - 1 \\ r - 1 \end{pmatrix} (1 - p^*)^y (p^*)^r, \; y = 0, 1, 2, \ldots \tag{7.1}$$

The mean and variance of Y are

$$\mathrm{E}(Y) = \mu = \frac{r(1 - p^*)}{p^*}$$

and

$$\mathrm{Var}(Y) = \frac{r(1 - p^*)}{(p^*)^2}$$

respectively. In the present context, it is convenient to consider a reparametrization of the negative binomial distribution in terms of the mean μ and the parameter r. Substituting $p^* = \dfrac{r}{r + \mu}$ into (7.1) gives

$$\Pr(Y = y) = \begin{pmatrix} y + r - 1 \\ r - 1 \end{pmatrix} \left(\frac{\mu}{r + \mu}\right)^y \left(\frac{r}{r + \mu}\right)^r, \; y = 0, 1, 2, \ldots$$

and the distribution is written as $Y \sim \mathrm{NegBin}(r, \mu)$, with $\mathrm{E}(Y) = \mu$ and $\mathrm{Var}(Y) = \mu + \dfrac{\mu^2}{r}$ (McCullagh and Nelder, 1989, p.373). In the limit as r approaches infinity, the Poisson model for Y is recovered (Anscombe, 1949).

Due to the origin of the model, the parameter r is usually restricted to be a positive integer. However, a general form of the negative binomial distribution exists in which this restriction is relaxed and r is simply taken to be a positive real number. This is the distribution of interest in the present study and under these

conditions, the p.m.f. of Y can be rewritten

$$\Pr(Y=y) \;=\; \frac{\Gamma(y+r)}{y!\,\Gamma(r)}\left(\frac{\mu}{r+\mu}\right)^{y}\left(\frac{r}{r+\mu}\right)^{r},\; y=0,1,2,\ldots$$

with $\mu>0$ and $r>0$.

7.2.2 Negative Binomial distribution for N

Consider a dosage-mortality study carried out in a Wadley's problem setting and suppose that a negative binomial distribution with parameters r and μ is used to model the unknown number of organisms initially exposed to a drug, N. Recall that if N were known, $Y|N=n$ would be expected to follow a Binomial$(n,1-p)$ distribution. The marginal p.m.f. of Y when $N\sim$NegBin(r,μ) is then given by

$$\begin{aligned}
\Pr(Y=y) &= \sum_{n=y}^{\infty}\Pr(Y=y|N=n)\Pr(N=n)\\
&= \sum_{n=y}^{\infty}\binom{n}{y}(1-p)^{y}p^{n-y}\frac{\Gamma(n+r)}{n!\,\Gamma(r)}\left(\frac{r}{r+\mu}\right)^{r}\left(\frac{\mu}{r+\mu}\right)^{n}\\
&= \frac{\Gamma(r+y)}{y!\,\Gamma(r)}\left[\frac{r}{r+\mu(1-p)}\right]^{r}\left[\frac{\mu(1-p)}{r+\mu(1-p)}\right]^{y}.
\end{aligned}$$

Thus $Y\sim$NegBin $[r,\mu(1-p)]$, where the mean of Y is $\mu(1-p)$ and the variance of Y can be written as

$$\mathrm{Var}(Y) \;=\; \mu(1-p)\left\{1+\frac{\mu(1-p)}{r}\right\},$$

(Anscombe, 1949). This variance function is comprised of the usual Poisson variance component, $\mu(1-p)$, and the multiplying factor $\left\{1+\frac{\mu(1-p)}{r}\right\}$, which accommodates overdispersion in the model.

The above result for the distribution of Y can also be obtained using the p.g.f. technique. Consider the p.g.f. of the random variable Y when Y is considered as a sum of N independent Bernoulli random variables, $X_1,X_2,\ldots,X_N$. Then $Y=\sum_{i=1}^{N}X_i$, where $X_i\sim$Binomial$(1,p)$ with p.g.f. $P_X(s)=p+(1-p)s$, $i=1,\ldots,N$. If the total number of organisms follows a NegBin(r,μ) distribution, then

$P_N(s) = \left(\dfrac{r}{\mu + r - \mu s}\right)^r$ and from results pertaining to the p.g.f.'s of compound distributions (Grimmett and Welsh, 1986, p.51) it follows that

$$\begin{aligned} P_Y(s) &= P_N(P_X(s)) \\ &= \left[\frac{r}{\mu + r - \mu(p + qs)}\right]^r \\ &= \left[\frac{r}{\mu(1-p) + r - \mu(1-p)s}\right]^r . \end{aligned}$$

Thus $Y \sim$NegBin$[r, \mu(1-p)]$ in accordance with the earlier result.

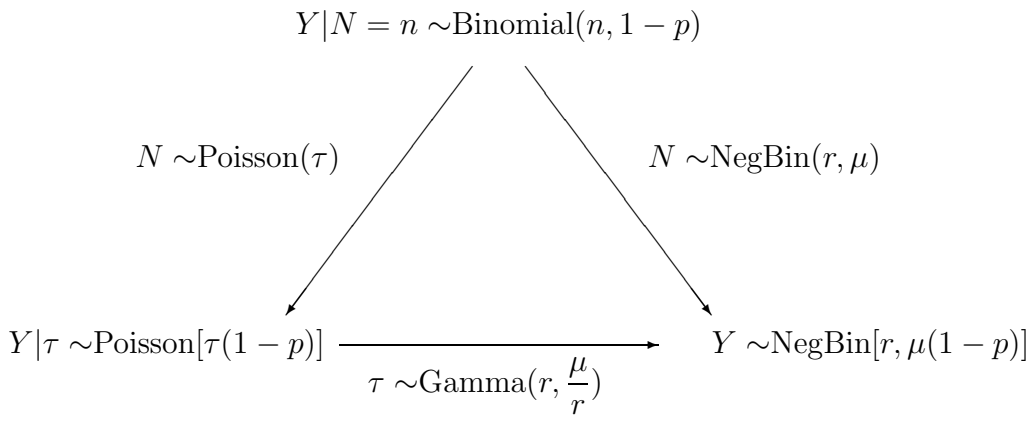
7.2.3 Gamma Distribution for the Poisson Parameter

McCullagh and Nelder (1989, p.374) note that the negative binomial distribution can be obtained by modelling the mean of a Poisson distribution as a gamma variable. In the present context, suppose that the parameter τ in the Poisson $[\tau(1-p)]$ model of Chapter 2 is Gamma(a, b). Then the marginal p.m.f. of Y is given by

$$\begin{aligned} \Pr(Y = y) &= \int_0^\infty \Pr(Y = y|\tau) f(\tau)\, d\tau \\ &= \int_0^\infty \frac{[\tau(1-p)]^y e^{-\tau(1-p)}}{y!} \frac{\tau^{a-1} e^{-\tau/b}}{\Gamma(a) b^a}\, d\tau \\ &= \frac{\Gamma(y+a)}{y!\,\Gamma(a)} b^{-a} (1-p)^y \left(\frac{b+1-pb}{b}\right)^{-a-y} \\ &= \frac{\Gamma(y+a)}{y!\,\Gamma(a)} \left[\frac{1}{b+1-pb}\right]^a \left[\frac{b(1-p)}{b+1-pb}\right]^y . \end{aligned}$$

Thus Y follows a negative binomial distribution with mean $ab(1-p)$ and exponent parameter a, i.e. $Y \sim$NegBin$[a, ab(1-p)]$. Relating this notation to the previous section, it follows that $a = r$ and $b = \dfrac{\mu}{r}$ and thus that τ is taken to be Gamma$\left(r, \dfrac{\mu}{r}\right)$.

The following diagram illustrates ways in which a negative binomial distribution for the random variable describing the number of surviving parasites can be derived.



7.3 Modelling the Probability

Recall that the probability of a parasite dying, p_i, is associated with dose d_i (and log dose x_i) of the drug and thus with observations $y_{ij},\ j = 1, \ldots, n_i$ and $i = 1, \ldots, D$. When a logit tolerance distribution is used to model this probability,

$$\ln\left(\frac{p_i}{1-p_i}\right) \quad = \quad \alpha + \beta x_i$$

and hence

$$p_i \quad = \quad \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}\ ,\ i = 1, \ldots, D.$$

The negative binomial model is a member of the exponential family of distributions when the exponent parameter r is known (Dobson, 2002, p.53). For the negative binomial distribution considered here, with the probability of the death of a parasite modelled by a logit function, consider the expected value of Y_{ij},

$$\begin{aligned} E(Y_{ij}) &= \mu(1-p_i) \\ &= \frac{\mu}{1+e^{\alpha+\beta x_i}}\ . \end{aligned}$$

In order for this model to be a GLM, r must be known and a monotone function of the mean that is linear in the parameters $r,\ \mu,\ \alpha$ and β is required. Such a function

does not appear to exist and since r is not fixed, the model under consideration is treated as a generalized nonlinear model.

7.4 Likelihood, Score Functions and Information Matrix

7.4.1 Likelihood Function

Suppose again that the random variable, $Y_{cj},\ j = 1, \ldots, n_c$, denotes the responses in the control group and that $Y_{ij},\ j = 1, \ldots, n_i$ and $i = 1, \ldots, D$, refers to the number of surviving organisms that are treated with a non-zero dose d_i and log-dose x_i of the drug. Suppose further that these variables follow a negative binomial distribution of the form NegBin$[r, \mu(1-p)]$.

Since the probability of death for subjects in the control group is zero, the likelihood and log-likelihood functions for the observed counts can be written as

$$L_c = \prod_{j=1}^{n_c} \frac{\Gamma(y_{cj}+r)}{y_{cj}!\,\Gamma(r)} \left(\frac{r}{r+\mu}\right)^r \left(\frac{\mu}{r+\mu}\right)^{y_{cj}}$$

and

$$\ell_c = \sum_{j=1}^{n_c} \left\{ \ln\Gamma(y_{cj}+r) - \ln y_{cj}! - \ln\Gamma(r) + r\ln\left(\frac{r}{r+\mu}\right) + y_{cj}\ln\left(\frac{\mu}{r+\mu}\right) \right\}$$

respectively. For observations corresponding to the administration of non-zero doses of the drug, the likelihood and log-likelihood functions are given respectively as

$$L_d = \prod_{i=1}^{D}\prod_{j=1}^{n_i} \frac{\Gamma(y_{ij}+r)}{y_{ij}!\,\Gamma(r)} \left[\frac{r}{r+\mu(1-p_i)}\right]^r \left[\frac{\mu(1-p_i)}{r+\mu(1-p_i)}\right]^{y_{ij}}$$

and

$$\ell_d = \sum_{i=1}^{D}\sum_{j=1}^{n_i} \left\{ \ln\Gamma(y_{ij}+r) - \ln y_{ij}! - \ln\Gamma(r) + r\ln\left[\frac{r}{r+\mu(1-p_i)}\right] + y_{ij}\ln\left[\frac{\mu(1-p_i)}{r+\mu(1-p_i)}\right] \right\}.$$

The log-likelihood for the data is then found by summing the log-likelihoods for single observations over all the observations and can be written as

$$\begin{aligned}
\ell(r,\mu,p;\underline{y}) &= \ell_c + \ell_d \\
&= \sum_{j=1}^{n_c}\left\{\ln\Gamma(y_{cj}+r) - \ln y_{cj}! - \ln\Gamma(r) + r\ln\left(\frac{r}{r+\mu}\right) + y_{cj}\ln\left(\frac{\mu}{r+\mu}\right)\right\} \\
&\quad + \sum_{i=1}^{D}\sum_{j=1}^{n_i}\left\{\ln\Gamma(y_{ij}+r) - \ln y_{ij}! - \ln\Gamma(r) + r\ln\left[\frac{r}{r+\mu(1-p_i)}\right]\right. \\
&\quad \left. + y_{ij}\ln\left[\frac{\mu(1-p_i)}{r+\mu(1-p_i)}\right]\right\} .
\end{aligned}$$

7.4.2 Score Functions

The derivation of the score functions for the negative binomial model with parameters r and $\mu(1-p)$ involves differentiating expressions of the form $\ln\Gamma(y+r) - \ln\Gamma(r)$ with respect to r, which introduces the digamma function into the problem. In order to simplify this task, Lawless (1987) considered expanding the ratio $\dfrac{\Gamma(y+r)}{\Gamma(r)}$ in the following manner

$$\frac{\Gamma(y+r)}{\Gamma(r)} = r(r+1)\ldots(r+y-1) , \tag{7.2}$$

an expansion which holds for any $r > 0$ and $y \geq 1$. Therefore the log-likelihood for the jth observation from the control group can be rewritten as

$$\ell_{cj} = \sum_{s=0}^{y_{cj}-1}\ln(r+s) - \ln y_{cj}! + y_{cj}\ln\mu - (y_{cj}+r)\ln(r+\mu) + r\ln r$$

and does not depend on the parameters α and β. The score functions for a single observation from this treatment group are therefore

$$\begin{aligned}
\frac{\partial\ell_{cj}}{\partial r} &= \sum_{s=0}^{y_{cj}-1}\frac{1}{r+s} + \ln\left(\frac{r}{r+\mu}\right) + \frac{\mu - y_{cj}}{r+\mu} \\
\frac{\partial\ell_{cj}}{\partial\mu} &= \frac{r(y_{cj}-\mu)}{\mu(\mu+r)} \\
\frac{\partial\ell_{cj}}{\partial\alpha} &= 0 \\
\frac{\partial\ell_{cj}}{\partial\beta} &= 0
\end{aligned}$$

Lawless(1987).

Using the expansion given by equation (7.2), the log-likelihood for a single observation corresponding to the administration of a non-zero dose of the drug, $d_i,\ i = 1, \ldots, D$, is given by

$$\begin{aligned} \ell_{ij} &= \sum_{s=0}^{y_{ij}-1} \ln(r+s) - \ln y_{ij}! + y_{ij} \ln \mu + y_{ij} \ln(1-p_i) \\ &\quad -(y_{ij}+r) \ln[r + \mu(1-p_i)] + r \ln r \ . \end{aligned}$$

Therefore the scores for a single observation treated with a non-zero dose of the drug are

$$\begin{aligned} \frac{\partial \ell_{ij}}{\partial r} &= \sum_{s=0}^{y_{ij}-1} \frac{1}{r+s} + \ln\left[\frac{r}{r+\mu(1-p_i)}\right] + \frac{\mu(1-p_i) - y_{ij}}{r+\mu(1-p_i)} \\ \frac{\partial \ell_{ij}}{\partial \mu} &= \frac{r[y_{ij} - \mu(1-p_i)]}{\mu[\mu(1-p_i)+r]} \\ \frac{\partial \ell_{ij}}{\partial \alpha} &= \frac{-p_i r(y_{ij} - \mu(1-p_i))}{r+\mu(1-p_i)} \\ \frac{\partial \ell_{ij}}{\partial \beta} &= \frac{-p_i r x_i[y_{ij} - \mu(1-p_i)]}{r+\mu(1-p_i)} \ . \end{aligned}$$

The score functions for the data can then be found by summing the score functions for each observation over all observations. Note that setting these score functions to zero and solving them simultaneously for the parameters in the model gives the maximum likelihood estimates. However, the solutions cannot be obtained explicitly and numerical methods are required for parameter estimation.

7.4.3 Information Matrix

The information matrix of the negative binomial distribution is minus the matrix of expected values of the second-order derivatives of the log-likelihood with respect to the parameters $r,\ \mu,\ \alpha$ and β.

Consider the joint information for the parameter r and the remaining parameters in the model by first observing that the associated second-order derivatives

of the log-likelihood are

$$\begin{aligned}
\frac{\partial^2 \ell_{ij}}{\partial r \partial \mu} &= \frac{(1-p_i)[y_{ij} - \mu(1-p_i)]}{[\mu(1-p_i)+r]^2} \\
\frac{\partial^2 \ell_{ij}}{\partial r \partial \alpha} &= \frac{-\mu p_i (1-p_i)[\mu(1-p_i) - y_{ij}]}{[\mu(1-p_i)+r]^2} \\
\frac{\partial^2 \ell_{ij}}{\partial r \partial \beta} &= \frac{-\mu x_i p_i (1-p_i)[\mu(1-p_i) - y_{ij}]}{[\mu(1-p_i)+r]^2} \; .
\end{aligned}$$

The expected values of these derivatives are all clearly zero, indicating that the information between the parameter r and the remaining parameters in the model is zero. Therefore consider partitioning the information matrix as

$$\mathfrak{I} = \begin{bmatrix} \mathfrak{I}_{11} & \underline{0}^T \\ \underline{0} & \mathfrak{I}_{22} \end{bmatrix}$$

where $\mathfrak{I}_{11}$ contains the information on r and $\mathfrak{I}_{22}$ is the information for the parameters $\mu,\ \alpha$ and β. Thus

$$\mathfrak{I} = -E \begin{bmatrix} \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial r^2} & 0 & 0 & 0 \\ 0 & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \mu^2} & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \mu \partial \alpha} & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \mu \partial \beta} \\ 0 & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \mu \partial \alpha} & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \alpha^2} & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \alpha \partial \beta} \\ 0 & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \mu \partial \beta} & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \alpha \partial \beta} & \frac{\partial^2 \ell(r,\mu,p;Y)}{\partial \beta^2} \end{bmatrix} .$$

For a single response from the control treatment group, the diagonal sub-matrices of the information matrix are

$$\mathfrak{I}_{11}^{(c)} = E\left\{ \sum_{s=0}^{Y_{cj}-1} \frac{1}{(r+s)^2} \right\} + \frac{\mu}{r(\mu+r)}$$

and

$$\mathfrak{I}_{22}^{(c)} = \begin{bmatrix} \frac{r}{\mu(\mu+r)} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

An observation arising from the administration of a non-zero dose of the drug has an information matrix with the following sub-matrices

$$\mathfrak{I}_{11}^{(d)} = E\left\{\sum_{s=0}^{Y_{ij}-1}\frac{1}{(r+s)^2}\right\} + \frac{\mu(1-p_i)}{r[\mu(1-p_i)+r]}$$

and

$$\mathfrak{I}_{22}^{(d)} = \frac{(1-p_i)r}{\mu(1-p_i)+r}\begin{bmatrix} \frac{1}{\mu} & -p_i & -p_i x_i \\ -p_i & \mu p_i^2 & \mu p_i^2 x_i \\ -p_i x_i & \mu p_i^2 x_i & \mu p_i^2 x_i^2 \end{bmatrix}.$$

Due to the independence of the observations, the information matrix for the data is obtained by summing the information matrices for individual observations over all of the observations.

7.5 Models for Separate Doses

For comparative purposes, two models that consider each dose of the drug separately are introduced here. In particular, the first model, which is referred to as the $\{r_i, \mu_i\}$ model, fits a separate negative binomial distribution to each dose of the drug, i.e. $Y_{ij} \sim \text{NegBin}(r_i, \mu_i)$, and the second model, which is denoted the $\{r, \mu_i\}$ model, is a nested version of the first which fits a common parameter r to each dose, i.e. $Y_{ij} \sim \text{NegBin}(r, \mu_i),\ j = 1, \ldots, n_i,\ i = 1, \ldots, D$. Although these are not maximal models, they can be used as base models for comparison because they have large numbers of parameters.

7.6 Estimation

7.6.1 Parameter Estimates

Three methods of obtaining the maximum likelihood estimates of the parameters $r,\ \mu,\ \alpha$ and β for the negative binomial model with a logit tolerance distribution are considered.

Anscombe's Method

Anscombe (1949) suggested methods of approximating the maximum likelihood estimates of the parameters that were appropriate for computation by hand and avoided the complicated calculations required for maximum likelihood estimation. Since direct calculation of maximum likelihood estimates is simplified by the software that is currently available, these methods of approximation are not considered in this study.

Genstat Procedure

The WADLEY procedure in Genstat has already been described in Chapter 2 and contains an option for specifying the distribution of the unknown number of parasites, including the negative binomial. However, attempts to use this procedure for the latter distribution proved troublesome. In particular certain results appeared spurious and the output did not include an estimate of r, making the identification of inherent problems difficult. Further, the WADLEY procedure requires more than one observation per dose in order to fit the model to a data set. As a result of these complications, the WADLEY procedure was not used and no Genstat output is therefore included here.

Constrained Nonlinear Optimization

The nonlinear log-likelihood function specified in Section 7.4.1 can be maximized using a nonlinear optimization routine. In the present study, a Gauss program incorporating the procedure OPTMUM was written for this purpose. The starting values were chosen in the same manner as that used for the Poisson model of Chapter 2 and, since the negative binomial distribution becomes more Poisson-like as r tends to infinity, large starting values of r were used in order to avoid convergence problems.

A complication in the computations was caused by the presence of the gamma function in the likelihood equation. Specifically, functions of the form $\ln \Gamma(y + r)$

must necessarily be evaluated. In certain cases, such as the malaria data, values of y are very large, causing overflow problems in the direct calculation of $\Gamma(y+r)$. In the present study, when y was sufficiently small, $\ln\Gamma(y+r)$ was calculated directly and for large values of y, the command `lnfact` in Gauss was used. The constraints implicit in the model are $r > 0$ and $\beta > 0$ and these were incorporated into the optimization procedure.

7.6.2 Standard Errors of Parameter Estimates

The standard errors of the parameter estimates $\hat{r},\ \hat{\mu},\ \hat{\alpha}$ and $\hat{\beta}$ are approximated by substituting these estimates into the information matrix and taking the square root of the diagonal elements of the inverse of that matrix.

Computation of the standard error for $\hat{r}$ involves calculating the expected value of the sum

$$\sum_{s=0}^{Y-1} \frac{1}{(r+s)^2} \tag{7.3}$$

with respect to the random variable Y. Terms of this form are contained in $\mathfrak{I}_{11}$, which was defined in Section 7.4.3. The calculation of the expectation of (7.3) is equivalent to evaluating the infinite sum

$$\sum_{k=0}^{\infty} \frac{1}{(r+k)^2} \left\{1 - \Pr(Y \leq k)\right\}, \tag{7.4}$$

(Lawless, 1987). Now $1 - \Pr(Y \leq k)$ is a monotone decreasing function in k and thus, for the purposes of calculation, the infinite sum (7.4) can be approximated by the finite sum

$$\sum_{k=0}^{k_{\text{cut}}} \frac{1}{(r+k)^2} \left\{1 - \Pr(Y \leq k)\right\}$$

where k_{cut} is chosen so that $1 - \Pr(Y \leq k_{\text{cut}})$ is suitably close to 0. In the present case the value

$$\begin{aligned} k_{\text{cut}} &= \left\lfloor \mathrm{E}(Y) + 10\ \sqrt{\mathrm{Var}(Y)}\ \right\rfloor \\ &= \left\lfloor \mu(1-p) + 10\ \sqrt{\mu(1-p) + \frac{[\mu(1-p)]^2}{r}}\ \right\rfloor \end{aligned}$$

was used and was found to give a satisfactory approximation due to the fact that $1 - \Pr(Y \leq k_{\text{cut}})$ was less than 10^{-6}. The first component of the information for r is then found by evaluating (7.4) for each observation and summing over all observations.

95% Wald intervals can be computed for the parameters, r, μ, α and β of the negative binomial distribution. For example, the 95% Wald interval for r is given by $\hat{r} \pm 1.96\ se(\hat{r})$ where $se(\hat{r})$ is the standard error of the estimate of r (Kutner, Nachtsheim, Neter and Li, 2005, p.579). Profile likelihood plots can also be produced for r, μ, α and β and these plots can be used to obtain approximate 95% confidence intervals for the parameters. Specifically, the 95% profile likelihood confidence interval for the parameter r is given by $[r_1, r_2]$, with r_1 and r_2 satisfying

$$\ell_P(r_1) = \ell_P(r_2) = \ell(\hat{r}, \hat{\mu}, \hat{p}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05},$$

where $\ell_P(r)$ is the profile likelihood for r evaluated at r and $\ell(\hat{r}, \hat{\mu}, \hat{p}; \underline{y})$ is the value of the maximum log-likelihood of the NegBin$[r, \mu(1-p)]$ distribution (Azzalini, 1996, p.146).

7.6.3 Parameter Estimates for the Separate Dose Models

The base models, $\{r_i, \mu_i\}$ and $\{r, \mu_i\}$, from Section 7.5 were fitted using maximum likelihood estimation. It can easily be seen that the maximum likelihood estimate of μ_i is the mean of the observations corresponding to that treatment group, i.e. $\hat{\mu}_i = \overline{y}_i$ where

$$\overline{y}_i \quad = \quad \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i} \ i = 1, \ldots, D, \ j = 1, \ldots, n_i \ .$$

A concentrated log-likelihood, which is a function of r_i only, can then be obtained by substituting $\overline{y}_i$ for μ_i (Jones, 1993, p.16) and for the ijth observation is given as

$$\ell_{ij} \quad = \quad \ln\Gamma(y_{ij} + r_i) - \ln y_{ij}! - \ln\Gamma(r_i) + r_i \ln\left(\frac{r_i}{r_i + \overline{y}_i}\right) + y_{ij} \ln\left(\frac{\overline{y}_i}{r_i + \overline{y}_i}\right) \ .$$

This single-variable function can then be maximized in order to estimate r_i by using a nonlinear optimization routine such as the OPTMUM procedure in Gauss. The same idea can be used when a common parameter r is adopted for all doses of the drug.

7.6.4 Goodness of Fit

The calculation of the deviance involves the specification of a maximal model, which has one parameter for each observation. This is problematic when the parameter, r, in the negative binomial model is not regarded as a nuisance parameter. As a result AIC can be used to compare the fit of the NegBin$[r, \mu(1-p)]$ model with those of the $\{r_i, \mu_i\}$ and $\{r, \mu_i\}$ models. The model yielding the smallest AIC should be selected as the best fitting model. The AIC for the NegBin$[r, \mu(1-p)]$ model is computed as AIC=-$2\ell(\hat{r}, \hat{\mu}, \hat{p}; \underline{y}) + 8$, where $q = 4$.

7.7 Examples

The negative binomial model, with a logit function modelling the probability of the death of an organism, was fitted to the algae and malaria data sets using a Gauss program which incorporates the OPTMUM routine.

7.7.1 Algae Data

The NegBin$[r, \mu(1-p)]$ model was fitted to the algae data using the Gauss program developed in this study. The parameter estimates, standard errors and 95% Wald intervals are summarized in Table 7.1. The estimate of the parameter r is extremely large and thus the model being fitted to this data is essentially the Poisson model of Chapter 2. Therefore the negative binomial model does not improve in any way on the fit of the Poisson model to this data and as a result profile likelihood plots were not obtained and no further analyses were performed.

Table 7.1: Results from fitting the NegBin$[r, \mu(1-p)]$ model to the algae data.

Parameter	Estimate	Standard Error	95% Wald Interval
$\hat{r}$	293728.802	58770.450	(178538.720;408918.884)
$\hat{\mu}$	224.833	6.595	(211.907;237.759)
$\hat{\alpha}$	-0.902	0.126	(-1.149;-0.655)
$\hat{\beta}$	1.415	0.090	(1.239;1.591)

7.7.2 Malaria Data

The NegBin$[r, \mu(1-p)]$ model was then fitted to the malaria data and since the estimate of r was not very large, profile likelihood plots were produced for each of the parameters. The profile likelihood plots are presented in Figures 7.1 (a), (b), (c) and (d). The parameter estimates together with their standard errors, 95% Wald intervals and 95% profile likelihood intervals are summarized in Table 7.2.

Table 7.2: Results from fitting the NegBin$[r, \mu(1-p)]$ model to the malaria data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
$\hat{r}$	1.505	0.401	(0.719;2.291)	*
$\hat{\mu}$	6581.768	2496.454	(1688.718;11474.818)	(3500;16500)
$\hat{\alpha}$	-1.388	0.997	(-3.342;0.566)	(-3.200;0.250)
$\hat{\beta}$	3.752	0.558	(2.658;4.846)	(3.050;4.800)

* A profile likelihood interval for r could not be computed due to convergence problems.

Note in particular that the standard error of the mean is very large.

The fit of this model was compared with those of the $\{r_i, \mu_i\}$ and $\{r, \mu_i\}$ models that consider each dose separately and the results of the fits are presented in Table 7.3. The AIC values for the $\{r_i, \mu_i\}$ and $\{r, \mu_i\}$ models differ substantially,

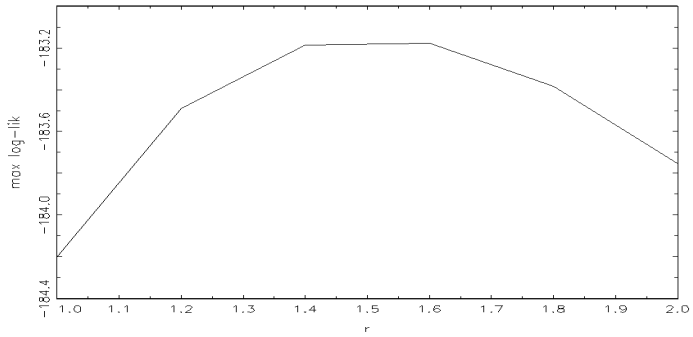


Figure 7.1: (a)

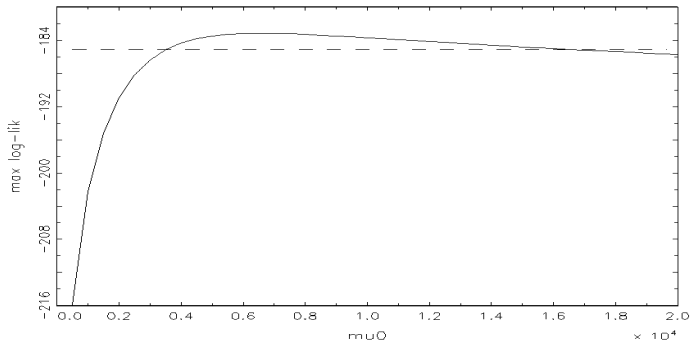


Figure 7.1: (b)

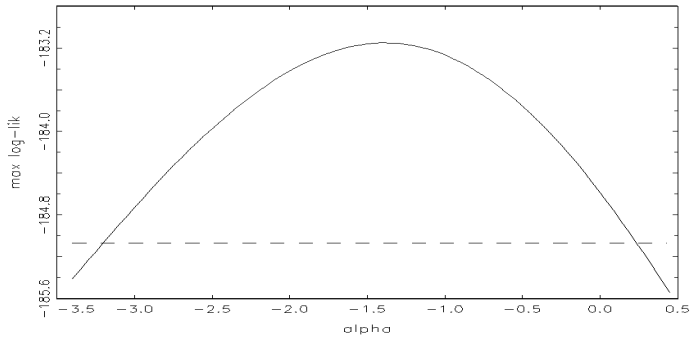


Figure 7.1: (c)

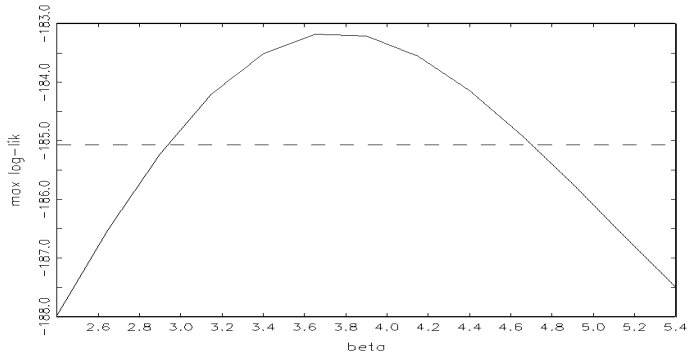


Figure 7.1: (d)

Figure 7.1: Profile likelihood plots for the parameters (a) r, (b) μ, (c) α and (d) β of the negative binomial model with a logit tolerance distribution fitted to the malaria data.

Table 7.3: A comparison of the various models fitted to the malaria data.

Model	Maximum log-likelihood	Number of parameters	AIC
$\{r_i, \mu_i\}$	-128.753	16	289.506
$\{r, \mu_i\}$	-151.253	9	320.506
μ, r, α, β	-183.152	4	374.304

indicating that a common value for r cannot be adopted for all doses of the drug. The NegBin$[r, \mu(1-p)]$ model of interest in this study has an AIC value which is much higher than the AIC values of the base models and can thus be regarded as a very poor model for this data. In fact the base model $\{r_i, \mu_i\}$ provides the best fit for the malaria data.

7.8 Conclusions

The results from fitting the negative binomial distribution to the algae data indicated that the model simply reduced to the Poisson model, which was seen in Chapter 2, to provide a poor fit to the data. When fitted to the malaria data, the NegBin$[r, \mu(1-p)]$ distribution did not adequately account for the overdispersion. Therefore the negative binomial distribution was not an effective tool for modelling overdispersion in the data considered in this thesis. These results inspired Chapter 8 which considers an alternative distribution for the unknown number of organisms initially treated.

Chapter 8

The Binomial-EWP Model

8.1 Introduction

This chapter follows the rationale behind modelling Wadley's problem with the negative binomial distribution by using a weighted Poisson distribution to model the binomial parameter n. Ridout and Besbeas (2004) investigated a number of weighted Poisson distributions that can be used to model over- and under-dispersed count data. In particular, they introduced the exponentially weighted Poisson distribution for underdispersed count data, of which there are two types: the three-parameter exponentially weighted Poisson distribution, denoted EWP_3, and the two-parameter exponentially weighted Poisson distribution, which is denoted EWP_2. These exponentially weighted Poisson distributions are also suitable for modelling overdispersed count data (Kokonendji, Mizère and Balakrishnan, 2008). In the present study, each of the weighted Poisson distributions described by Ridout and Besbeas (2004) was considered for modelling the unknown number of organisms initially treated with a drug and the EWP_2 seemed to be the most attractive and most algebraically tractable. Furthermore, Ridout and Besbeas (2004) did not report much work on this distribution and there are very few references to it in the literature. For these reasons the two-parameter exponentially weighted Poisson distribution was selected for further study and for convenience

it is referred to here as the EWP distribution. Various properties of the EWP distribution are explored and it is considered for modelling the binomial number of trials, thereby giving rise to a distribution aptly termed the binomial-EWP distribution.

In Section 8.2 the EWP distribution itself is explored in further detail. The p.m.f. of the binomial-EWP distribution is derived in Section 8.3 and Section 8.4 includes the moments of the binomial-EWP distribution. Some examples of the binomial-EWP distribution are presented in Section 8.5. The model is then considered for dose-response data in the Wadley problem setting in Section 8.6 and estimation and inference in this context are outlined in Section 8.7. Examples are discussed in Section 8.8 and Section 8.9 contains some concluding comments.

8.2 The EWP Distribution

8.2.1 Distribution

Suppose that the random variable Y follows an EWP distribution. Then the p.m.f. of Y is that of a weighted Poisson distribution and is given by

$$\Pr(Y = y) = \frac{e^{-\tau}\tau^{y}}{y!}\frac{e^{-\theta|y-\tau|}}{W}, n = 0, 1, 2, \ldots, \tag{8.1}$$

where τ is the Poisson parameter, θ is a parameter that incorporates over- or underdispersion and W is a normalizing constant which is written as

$$W = \sum_{y=0}^{\infty} \frac{e^{-\tau}\tau^{y}e^{-\theta|y-\tau|}}{y!}$$

and which ensures that the probabilities sum to one. This distribution is denoted as $Y \sim$EWP(τ, θ) for the remainder of this thesis. The distribution of Y reduces to a Poisson distribution when $\theta = 0$ and is suitable for modelling over- and underdispersed count data when $\theta < 0$ and $\theta > 0$ respectively (Ridout and Besbeas, 2004).

Ridout and Besbeas (2004) stated that the normalizing constant W can be approximated by a finite sum with the "upper summation limit replaced by a suitably large value". However, suppose that $\lfloor \tau \rfloor$ denotes the integer part of the value of τ. Then the normalizing constant W can be calculated as

$$\begin{aligned}
W &= \sum_{y=0}^{\infty} \frac{e^{-\tau}\tau^y e^{-\theta|y-\tau|}}{y!} \\
&= \sum_{y=0}^{\lfloor\tau\rfloor} \frac{e^{-\tau}\tau^y e^{\theta(y-\tau)}}{y!} + \sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{e^{-\tau}\tau^y e^{-\theta(y-\tau)}}{y!} \\
&= \sum_{y=0}^{\lfloor\tau\rfloor} \frac{e^{-\tau}\tau^y e^{\theta(y-\tau)}}{y!} + e^{\tau(\theta-1)} \sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{(\tau e^{-\theta})^y}{y!} \\
&= \sum_{y=0}^{\lfloor\tau\rfloor} \frac{e^{-\tau}\tau^y e^{\theta(y-\tau)}}{y!} + e^{\tau(\theta-1)} \left[e^{\tau e^{-\theta}} - \sum_{y=0}^{\lfloor\tau\rfloor} \frac{(\tau e^{-\theta})^y}{y!} \right] \\
&= \sum_{y=0}^{\lfloor\tau\rfloor} \frac{e^{-\tau}\tau^y e^{\theta(y-\tau)}}{y!} - \sum_{y=0}^{\lfloor\tau\rfloor} \frac{e^{-\tau}\tau^y e^{-\theta(y-\tau)}}{y!} + e^{\tau(\theta-1+e^{-\theta})} \\
&= e^{-\tau} \sum_{y=0}^{\lfloor\tau\rfloor} \frac{\tau^y}{y!} \left[e^{\theta(y-\tau)} - e^{-\theta(y-\tau)} \right] + e^{\tau(\theta-1+e^{-\theta})} \\
&= 2e^{-\tau} \sum_{y=0}^{\lfloor\tau\rfloor} \frac{\tau^y}{y!} \sinh[\theta(y-\tau)] + e^{\tau(\theta-1+e^{-\theta})} \; . \qquad (8.2)
\end{aligned}$$

Since the sum in expression (8.2) is finite, the normalizing constant of the EWP distribution can be calculated explicitly, which negates the statement made by Ridout and Besbeas (2004).

8.2.2 Moments of the EWP Distribution

According to Ridout and Besbeas (2004), "explicit formulae for the moments of the EWP distributions are not available" and they suggest approximating the moments of these EWP distributions by using "a finite upper limit to the summation".

Consider, however, the first two factorial moments of the EWP distribution, which are defined as the expectations of Y and $Y(Y-1)$ respectively. The first

factorial moment can be obtained in the following way:

$$\begin{aligned}
\mathrm{E}(Y) &= \sum_{y=0}^{\infty} y\Pr(Y=y) \\
&= \sum_{y=0}^{\lfloor\tau\rfloor} \frac{ye^{-\tau}\tau^y}{y!}\frac{e^{\theta(y-\tau)}}{W} + \sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{ye^{-\tau}\tau^y}{y!}\frac{e^{-\theta(y-\tau)}}{W} \\
&= \frac{\tau e^{-\tau(1+\theta)+\theta}}{W}\sum_{y=1}^{\lfloor\tau\rfloor}\frac{(\tau e^{\theta})^{y-1}}{(y-1)!} + \frac{\tau e^{\tau(\theta-1)-\theta}}{W}\sum_{y=\lfloor\tau\rfloor+1}^{\infty}\frac{(\tau e^{-\theta})^{y-1}}{(y-1)!} \\
&= \frac{\tau e^{-\tau(1+\theta)+\theta}}{W}\sum_{y=0}^{\lfloor\tau\rfloor-1}\frac{(\tau e^{\theta})^{y}}{y!} + \frac{\tau e^{\tau(\theta-1)-\theta}}{W}\sum_{y=\lfloor\tau\rfloor}^{\infty}\frac{(\tau e^{-\theta})^{y}}{y!} \\
&= \frac{\tau e^{-\tau(1+\theta)+\theta}}{W}\sum_{y=0}^{\lfloor\tau\rfloor-1}\frac{(\tau e^{\theta})^{y}}{y!} + \frac{\tau e^{\tau(\theta-1)-\theta}}{W}\left[e^{\tau e^{-\theta}} - \sum_{y=0}^{\lfloor\tau\rfloor-1}\frac{(\tau e^{-\theta})^{y}}{y!}\right] . \qquad (8.3)
\end{aligned}$$

The second factorial moment of the EWP distribution can be calculated as

$$\begin{aligned}
\mathrm{E}[Y(Y-1)] &= \sum_{y=0}^{\infty} y(y-1)\Pr(Y=y) \\
&= \sum_{y=0}^{\lfloor\tau\rfloor} \frac{y(y-1)e^{-\tau}\tau^y}{y!}\frac{e^{\theta(y-\tau)}}{W} + \sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{y(y-1)e^{-\tau}\tau^y}{y!}\frac{e^{-\theta(y-\tau)}}{W} \\
&= \frac{e^{-\tau(1+\theta)}(\tau e^{\theta})^2}{W}\sum_{y=2}^{\lfloor\tau\rfloor}\frac{(\tau e^{\theta})^{y-2}}{(y-2)!} + \frac{e^{\tau(\theta-1)}(\tau e^{-\theta})^2}{W}\sum_{y=\lfloor\tau\rfloor+1}^{\infty}\frac{(\tau e^{-\theta})^{y-2}}{(y-2)!} \\
&= \frac{e^{-\tau(1+\theta)}(\tau e^{\theta})^2}{W}\sum_{y=0}^{\lfloor\tau\rfloor-2}\frac{(\tau e^{\theta})^{y}}{y!} + \frac{e^{\tau(\theta-1)}(\tau e^{-\theta})^2}{W}\sum_{y=\lfloor\tau\rfloor-1}^{\infty}\frac{(\tau e^{-\theta})^{y}}{(y)!} \\
&= \frac{e^{-\tau(1+\theta)}(\tau e^{\theta})^2}{W}\sum_{y=0}^{\lfloor\tau\rfloor-2}\frac{(\tau e^{\theta})^{y}}{y!} + \frac{e^{\tau(\theta-1)}(\tau e^{-\theta})^2}{W}\left[e^{\tau e^{-\theta}} - \sum_{y=0}^{\lfloor\tau\rfloor-2}\frac{(\tau e^{-\theta})^{y}}{y!}\right] . \qquad (8.4)
\end{aligned}$$

Since the summations in expressions (8.3) and (8.4) are finite the first two factorial moments of the EWP distribution can be calculated explicitly, which again negates the claim made by Ridout and Besbeas (2004). The variance of the EWP distribution can then be obtained from these two moments in the usual way as

$$\mathrm{Var}(Y) = \mathrm{E}[Y(Y-1)] + \mathrm{E}(Y) - [\mathrm{E}(Y)]^2 .$$

8.2.3 Examples

Some illustrative examples of the p.m.f. of the EWP distribution are presented, together with their mean and variance, in Figure 8.1. The graphs indicate that the EWP distribution can have more than one mode. Ridout and Besbeas (2004) stated that the variance of the EWP distribution decreases as θ increases, which appears to be the case in the examples presented here.

8.2.4 Likelihood, Score Functions and Information Matrix

Consider a random sample of n observations $y_1, \ldots, y_n$ from an EWP distribution with parameters τ and θ. The likelihood function is the joint p.m.f. of these observations and is thus written as

$$L(\tau, \theta; \underline{y}) \;=\; \prod_{i=1}^{n} \left[\frac{e^{-\tau} \tau^{y_i}}{y_i!} \frac{e^{-\theta|y_i - \tau|}}{W} \right] .$$

As a result the log-likelihood function is given by

$$\ell(\tau, \theta; \underline{y}) \;=\; \sum_{i=1}^{n} y_i \ln \tau - n\tau - \sum_{i=1}^{n} \ln y_i! - \sum_{i=1}^{n} \theta |y_i - \tau| - n \ln W \; . \qquad (8.5)$$

Consider the score functions of τ and θ. The absolute value $|y - \tau|$ in expression (8.5) leads to discontinuities in the log-likelihood function with respect to τ. As a result this function is not differentiable with respect to τ and the score function for τ cannot be derived. The score function for θ can be obtained but its expression is long and complicated and is therefore not included here. Since the log-likelihood function cannot be differentiated with respect to τ, the information matrix of the EWP distribution cannot be derived.

8.2.5 Estimation and Inference

Recall that when W, the normalizing constant of the EWP distribution, was expressed as a finite sum, the integer part of the value of τ, $\lfloor \tau \rfloor$, was introduced in the limits of the summations. This can be seen in expression (8.2) and is as a

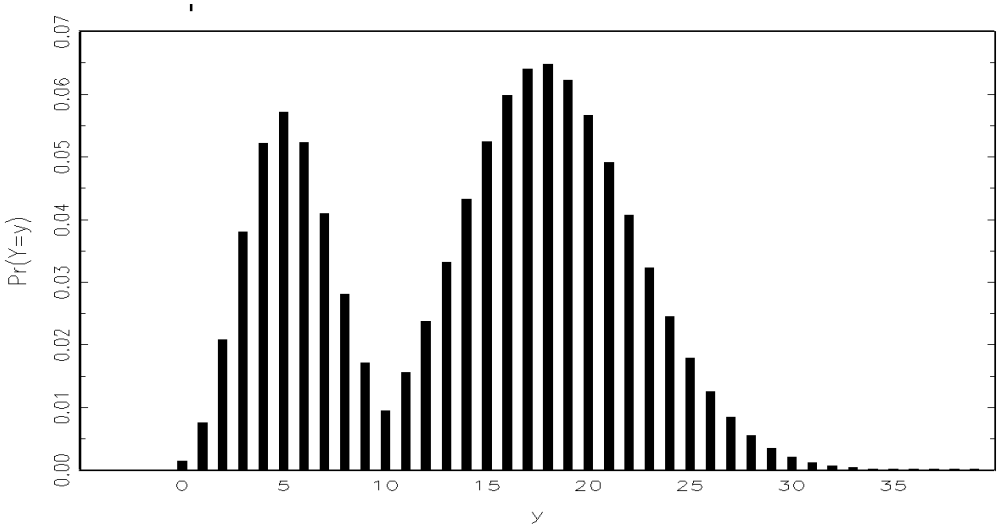


Figure 8.1: (a) $\tau = 10,\ \theta = -0.6$, E($Y$)=14.202, Var($Y$)=50.423

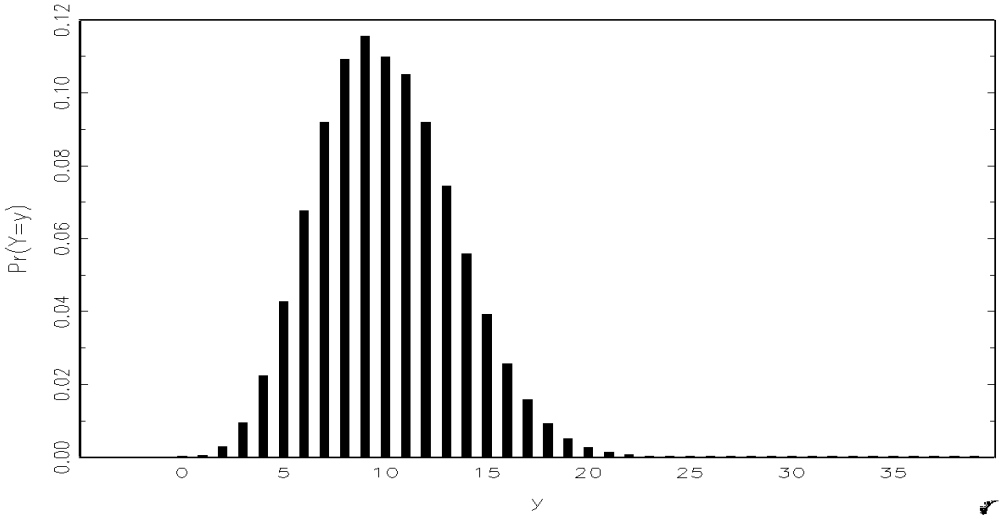


Figure 8.1: (b) $\tau = 10,\ \theta = -0.05$, E($Y$)=10.050, Var($Y$)=11.398

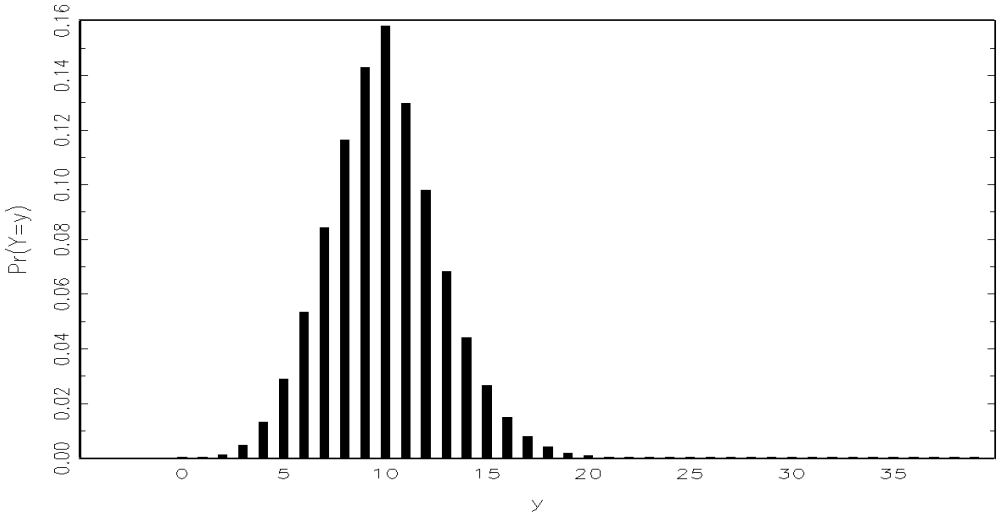


Figure 8.1: (c) $\tau = 10,\ \theta = 0.1$, E(Y)=9.940, Var(Y)=7.768

Figure 8.1: Plots showing the shapes of various EWP distributions.

result of the absolute value of $y - \tau$ in the log-likelihood function. These limits of the summations in the expression for W cause discontinuities in the log-likelihood function to occur at integer values of τ.

In order to illustrate these discontinuities in the log-likelihood function, data were simulated from an EWP distribution with $\tau = 6$ and $\theta = -0.5$ and a profile likelihood plot for the parameter τ was constructed. This plot is presented in Figure 8.2 and indicates that the discontinuities in the log-likelihood function are indeed at integer values of τ.

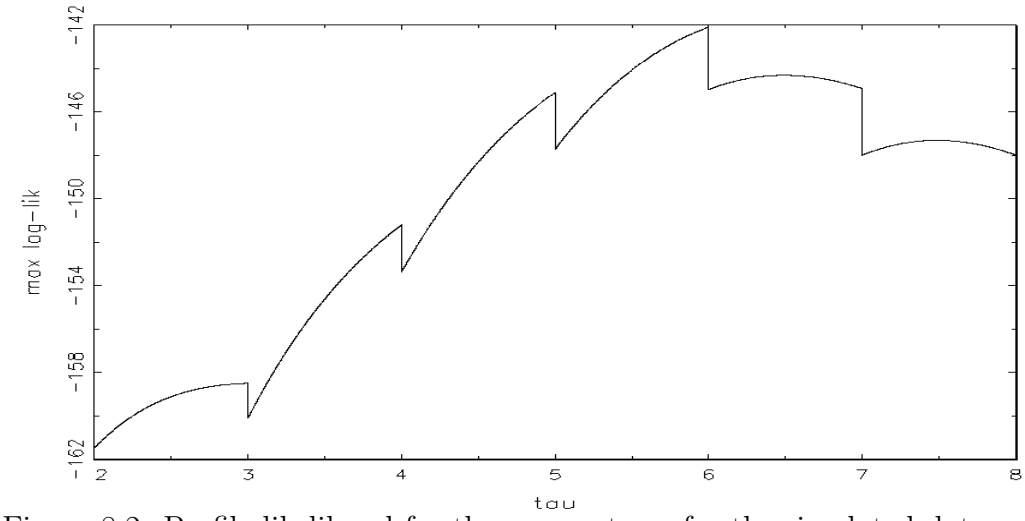


Figure 8.2: Profile likelihood for the parameter τ for the simulated data.

It is usual to estimate the parameters of a model by maximizing the log-likelihood function with respect to these parameters. The Gauss optimization routine, OPTMUM, uses derivative-based algorithms to numerically optimize a function and since the log-likelihood function is not differentiable with respect to τ this procedure cannot be used to estimate the parameters of the EWP distribution. Ridout and Besbeas (2004) used the Nelder-Mead method of optimization, which is a derivative-free algorithm, to estimate the parameters τ and θ but an alternative routine is proposed in the present study.

The parameters of the EWP distribution can be estimated by selecting a range of possible values for τ. Since the log-likelihood function is discontinuous at integer values of τ, it is necessary to fix τ at an integer value and to select a grid of integer

values for τ. In effect a profile likelihood is developed for τ by fixing τ at each value within the selected range and then estimating θ for that particular value of τ. Since the log-likelihood function is differentiable with respect to θ, the OPTMUM procedure in Gauss can then be used to estimate this parameter for a specified value of τ.

The information matrix of the EWP distribution cannot be estimated by the observed Fisher information matrix since the latter cannot be calculated. The standard errors of the parameter estimates can however be estimated by a parametric bootstrap (Efron and Tibshirani, 1993, Chapter 6). It is assumed that the data under consideration form a random sample from a probability distribution and that the parameter estimates $\hat{\tau}$ and $\hat{\theta}$ obtained from the data are the true parameter values of the probability distribution from which the data are sampled. A number of samples, say B, of size n can then be drawn from this probability distribution using the table look-up method (Ross, 1998, p.462) and the expression for the probability that Y is equal to y from expression (8.1). The parameters τ and θ are then estimated for each of the simulated samples and the standard deviations of the B parameter estimates can be used to provide the approximate standard errors of the parameter estimates. A value between 25 and 200 is an appropriate choice for B (Efron and Tibshirani, 1993, Chapter 6). These standard errors can be used to construct 95% Wald intervals for the parameters. In particular, the 95% Wald interval for θ is given by $\hat{\theta} \pm 1.96 se(\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimate of θ and $se(\hat{\theta})$ is the standard error of $\hat{\theta}$ (Kutner, Nachtsheim, Neter and Li, 2005, p.579).

The profile likelihood plot for τ is created when the parameters of the EWP distribution are estimated. However, in order to estimate the standard error of τ, a range of non-integer values can be specified for this parameter. A profile likelihood for θ can also be constructed by specifying a range of values for θ and for each value of θ a range of integer values for τ can be considered and the log-likelihood function computed. The maximum values of the log-likelihood for each

value of θ are then used to construct the profile likelihood plot for this parameter. These profile likelihood plots can be used to obtain approximate 95% confidence intervals for τ and θ. Specifically, the 95% confidence interval for τ from the profile likelihood plot is given by $[\tau_1; \tau_2]$, where τ_1 and τ_2 satisfy the condition

$$\ell_P(\tau_1) = \ell_P(\tau_2) = \ell(\hat{\tau}, \hat{\theta}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05} \ ,$$

and where $\ell_P(\tau)$ is the profile likelihood for τ evaluated at τ and $\ell(\hat{\tau}, \hat{\theta}; \underline{y})$ is the value of the maximum log-likelihood of the EWP distribution (Azzalini, 1996, p.146).

8.2.6 Goodness of Fit

A chi-squared goodness of fit test can be performed to determine whether the EWP distribution is suitable for modelling a particular count data set. A significant χ^2 statistic indicates that the model of interest is not suitable for the data.

The fit of the EWP distribution to count data can also be compared with that of the Poisson, CMP, beta-Poisson, Altham-Poisson and negative binomial distributions using AIC. In particular, the model with the smallest AIC value is the best fitting model for the data and the AIC of the EWP distribution, which has two parameters, is given by $AIC = -2\ell(\hat{\tau}, \hat{\theta}; \underline{y}) + 4$.

8.2.7 An Example

The EWP distribution was fitted to the sales data presented in Section 4.7 using the approach described in Section 8.2.5. The standard errors of the parameter estimates were approximated using a parametric bootstrap. Since the sales data set includes 3168 observations, 150 samples of size 3168 were drawn from the EWP distribution and the standard deviations of the parameter estimates obtained were used to approximate the standard errors of $\hat{\tau}$ and $\hat{\theta}$. 95% Wald intervals were then computed for the parameters τ and θ. Profile likelihood plots were produced for the parameters τ and θ of the EWP distribution and are presented in Figures 8.3

(a) and (b) respectively. These plots were used to obtain 95% confidence intervals for τ and θ. The parameter estimates, together with their standard errors and 95% Wald and profile likelihood intervals are recorded in Table 8.1.

Table 8.1: Results from fitting the EWP model to the sales data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	3	0.082	(2.839;3.161)	(2.370;3.260)
θ	-0.597	0.023	(-0.642;-0.552)	(-0.608;-0.558)

The adequacy of the EWP distribution for modelling the sales data was assessed using a chi-squared goodness of fit test as discussed in Section 8.2.6. AIC values were used to compare the fits of the Poisson, CMP, beta-Poisson, Altham-Poisson and negative binomial models with that of the EWP distribution. The chi-squared test results and AIC values are presented in Table 8.2. The p-values from the

Table 8.2: A comparison of the various models fitted to the sales data.

Model	χ^2 p-value	AIC
Poisson	0.000	17 921.622
CMP	0.242	15 056.712
Beta-Poisson	0.002	15 057.934
Altham-Poisson	0.005	15 083.308
Negative Binomial	0.636	15 059.220
EWP	0.000	15 806.331

chi-squared goodness of fit test indicate that the CMP and negative binomial distributions are suitable for the sales data. A comparison of the AIC values suggests that the fit of the EWP model is a great improvement on that of the Poisson model but that the EWP model does not model the overdispersion in the sales data as well as the CMP, beta-Poisson, Altham-Poisson and negative binomial models.

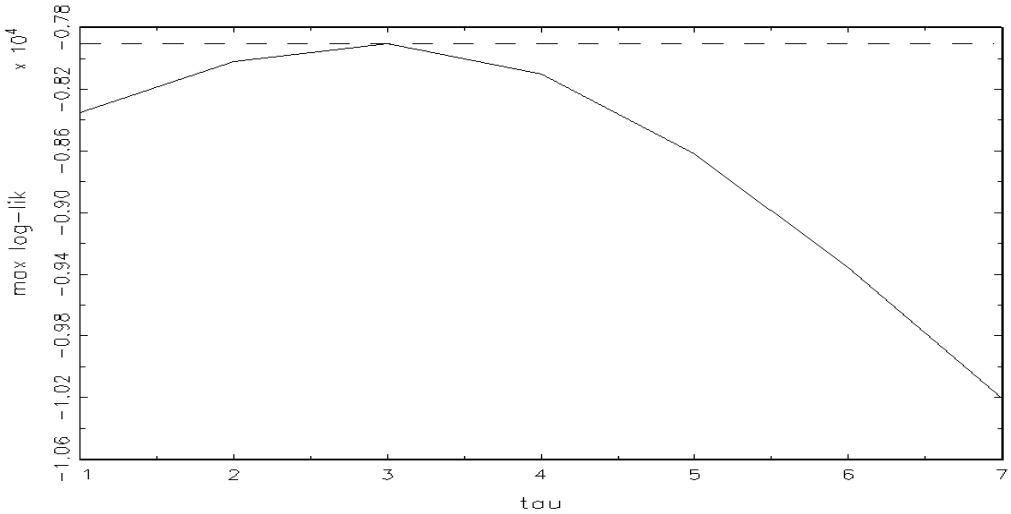


Figure 8.3: (a)

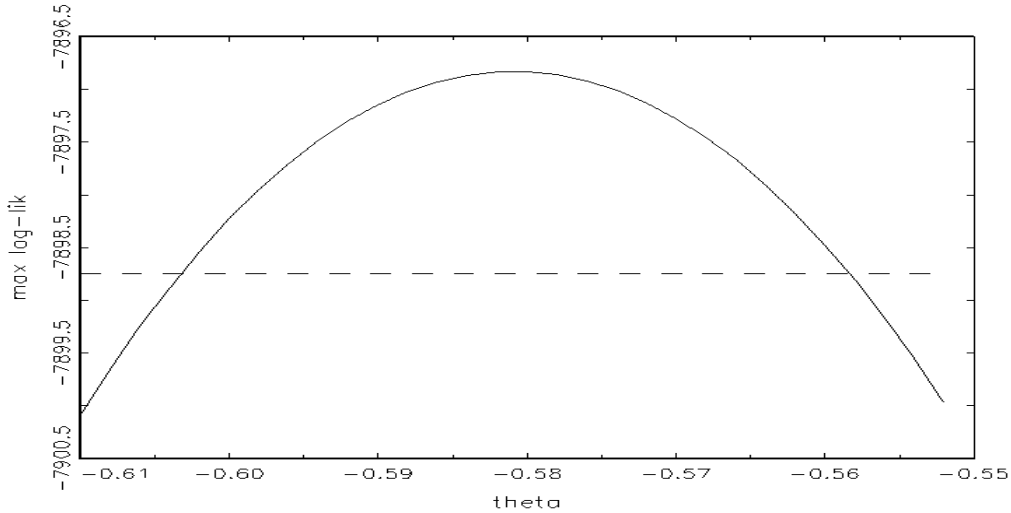


Figure 8.3: (b)

Figure 8.3: Profile likelihood plots for the parameters (a) τ and (b) θ of the EWP model fitted to the sales data.

8.3 Derivation of the Binomial-EWP p.m.f.

In the context of a dose-response study, suppose that Y, the number of organisms that survive exposure to varying doses of a drug, is observed and that the number initially treated, N, is unknown. Then Y given that $N = n$ follows a binomial distribution and the binomial parameter p can be modelled with an appropriate tolerance distribution. If an EWP distribution is used to model the random variable N the marginal p.m.f. of Y can be derived from first principles. Consider

$$\begin{aligned}\Pr(Y=y) &= \sum_{n=y}^{\infty}\Pr(Y=y|N=n)\Pr(N=n)\\ &= \sum_{n=y}^{\infty}\binom{n}{y}p^y(1-p)^{n-y}\frac{e^{-\tau}\tau^n e^{-\theta|n-\tau|}}{n!W}\end{aligned}$$

and the two scenarios $y > \tau$ and $y \leq \tau$. When $y > \tau$

$$\begin{aligned}\Pr(Y=y) &= \frac{e^{\tau(\theta-1)}(\tau p e^{-\theta})^y}{y!W}\sum_{n=y}^{\infty}\frac{[\tau(1-p)e^{-\theta}]^{n-y}}{(n-y)!}\\ &= \frac{e^{\tau(\theta-1)}(\tau p e^{-\theta})^y}{y!W}\sum_{n=0}^{\infty}\frac{[\tau(1-p)e^{-\theta}]^{n}}{n!}\\ &= \frac{(\tau p e^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W}\end{aligned}$$

and when $y \leq \tau$

$$\begin{aligned}\Pr(Y=y) &= \frac{\left(\frac{p}{1-p}\right)^y e^{-\tau}}{y!W}\left\{\sum_{n=y}^{\lfloor\tau\rfloor}\frac{[\tau(1-p)]^n e^{-\theta(\tau-n)}}{(n-y)!}+\sum_{n=\lfloor\tau\rfloor+1}^{\infty}\frac{[\tau(1-p)]^n e^{-\theta(n-\tau)}}{(n-y)!}\right\}\\ &= \frac{\left(\frac{p}{1-p}\right)^y e^{-\tau}}{y!W}\left\{\sum_{n=y}^{\lfloor\tau\rfloor}\frac{[\tau(1-p)]^n e^{-\theta(\tau-n)}}{(n-y)!}+\sum_{n=y}^{\infty}\frac{[\tau(1-p)]^n e^{-\theta(n-\tau)}}{(n-y)!}\right.\\ &\quad\left.-\sum_{n=y}^{\lfloor\tau\rfloor}\frac{[\tau(1-p)]^n e^{-\theta(n-\tau)}}{(n-y)!}\right\}\\ &= \frac{\left(\frac{p}{1-p}\right)^y e^{-\tau}}{y!W}\left\{\sum_{n=y}^{\lfloor\tau\rfloor}\frac{[\tau(1-p)]^n}{(n-y)!}\left[e^{-\theta(\tau-n)}-e^{\theta(\tau-n)}\right]+\sum_{n=y}^{\infty}\frac{[\tau(1-p)]^n e^{-\theta(n-\tau)}}{(n-y)!}\right\}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\left(\frac{p}{1-p}\right)^y e^{-\tau}}{y!W}\left\{\sum_{k=0}^{\lfloor\tau\rfloor-y}\frac{[\tau(1-p)]^{k+y}}{k!}\left[e^{\theta(k+y-\tau)}-e^{-\theta(k+y-\tau)}\right]\right.\\
&\quad\left.+\sum_{k=0}^{\infty}\frac{[\tau(1-p)]^{k+y}e^{-\theta(k+y-\tau)}}{k!}\right\}\\
&= \frac{(\tau p)^y e^{-\tau}}{y!W}\sum_{k=0}^{\lfloor\tau\rfloor-y}\frac{[\tau(1-p)]^k}{k!}\left[e^{\theta(k+y-\tau)}-e^{-\theta(k+y-\tau)}\right]+\frac{(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W}\\
&= \frac{2(\tau p)^y e^{-\tau}}{y!W}\sum_{k=0}^{\lfloor\tau\rfloor-y}\frac{[\tau(1-p)]^k}{k!}\sinh[\theta(k+y-\tau)]+\frac{(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W}\,.
\end{aligned}
$$

Thus the p.m.f. of Y can be written concisely as

$$
\Pr(Y=y) = \begin{cases}
\dfrac{2(\tau p)^y e^{-\tau}}{y!W}\displaystyle\sum_{k=0}^{\lfloor\tau\rfloor-y}\frac{[\tau(1-p)]^k}{k!}\sinh[\theta(k+y-\tau)]+\dfrac{(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W}\,, & y\le\tau\\[2ex]
\dfrac{(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W}\,, & y>\tau
\end{cases}
\tag{8.6}
$$

which is the p.m.f. of the binomial-EWP distribution with parameters τ, p and θ and is denoted as $Y\sim\text{BinEWP}(\tau,p,\theta)$ throughout this thesis. Observe that the p.m.f. of the binomial-EWP distribution includes finite sums and can therefore be calculated explicitly.

8.4 Moments of the Binomial-EWP Distribution

The mean of the binomial-EWP distribution can be written as

$$
\begin{aligned}
\text{E}(Y) &= \sum_{y=0}^{\infty} y\Pr(Y=y)\\
&= \sum_{y=0}^{\lfloor\tau\rfloor}\left\{\frac{2y(\tau p)^y e^{-\tau}}{y!W}\sum_{k=0}^{\lfloor\tau\rfloor-y}\frac{[\tau(1-p)]^k}{k!}\sinh[\theta(k+y-\tau)]+\frac{y(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W}\right\}\\
&\quad+\sum_{y=\lfloor\tau\rfloor+1}^{\infty}\frac{y(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W}\,.
\end{aligned}
$$

Consider simplifying the infinite sum in this expression in the following way:

$$\sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{y(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W} = \frac{\tau pe^{-\theta} e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{(\tau pe^{-\theta})^{y-1}}{(y-1)!}$$

$$= \frac{\tau pe^{-\theta} e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \sum_{y=\lfloor\tau\rfloor}^{\infty} \frac{(\tau pe^{-\theta})^y}{y!}$$

$$= \frac{\tau pe^{-\theta} e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \left[e^{\tau pe^{-\theta}} - \sum_{y=0}^{\lfloor\tau\rfloor-1} \frac{(\tau pe^{-\theta})^y}{y!} \right] .$$

Thus the mean of the binomial-EWP distribution can be written as

$$\begin{aligned} \mathrm{E}(Y) &= \sum_{y=0}^{\lfloor\tau\rfloor} \left\{ \frac{2y(\tau p)^y e^{-\tau}}{y!W} \sum_{k=0}^{\lfloor\tau\rfloor-y} \frac{[\tau(1-p)]^k}{k!} \sinh[\theta(k+y-\tau)] + \frac{y(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W} \right\} \\ &\quad + \frac{\tau pe^{-\theta} e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \left[e^{\tau pe^{-\theta}} - \sum_{y=0}^{\lfloor\tau\rfloor-1} \frac{(\tau pe^{-\theta})^y}{y!} \right] . \end{aligned} \tag{8.7}$$

Consider the second factorial moment of the binomial-EWP distribution, which is given by

$$\mathrm{E}(Y(Y-1)) = \sum_{y=0}^{\infty} y(y-1)\Pr(Y=y)$$

$$= \sum_{y=0}^{\lfloor\tau\rfloor} \left\{ \frac{2y(y-1)(\tau p)^y e^{-\tau}}{y!W} \sum_{k=0}^{\lfloor\tau\rfloor-y} \frac{[\tau(1-p)]^k}{k!} \sinh[\theta(k+y-\tau)] \right.$$

$$\left. + \frac{y(y-1)(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W} \right\} + \sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{y(y-1)(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W} .$$

The infinite sum in this expression can be simplified as

$$\sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{y(y-1)(\tau pe^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W} = \frac{(\tau pe^{-\theta})^2 e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \sum_{y=\lfloor\tau\rfloor+1}^{\infty} \frac{(\tau pe^{-\theta})^{y-2}}{(y-2)!}$$

$$= \frac{(\tau pe^{-\theta})^2 e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \sum_{y=\lfloor\tau\rfloor-1}^{\infty} \frac{(\tau pe^{-\theta})^y}{y!}$$

$$= \frac{(\tau pe^{-\theta})^2 e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \left[e^{\tau pe^{-\theta}} - \sum_{y=0}^{\lfloor\tau\rfloor-2} \frac{(\tau pe^{-\theta})^y}{y!} \right] .$$

Therefore the expected value of $Y(Y-1)$, where $Y \sim$BinEWP(τ, p, θ) can be written as

$$\mathrm{E}[Y(Y-1)] = \sum_{y=0}^{\lfloor\tau\rfloor} \left\{ \frac{2y(y-1)(\tau p)^y e^{-\tau}}{y!W} \sum_{k=0}^{\lfloor\tau\rfloor - y} \frac{[\tau(1-p)]^k}{k!} \sinh[\theta(k+y-\tau)] \right.$$
$$\left. + \frac{y(y-1)(\tau p e^{-\theta})^y e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y!W} \right\} + \frac{(\tau p e^{-\theta})^2 e^{\tau(\theta-1+(1-p)e^{-\theta})}}{W} \left[e^{\tau p e^{-\theta}} - \sum_{y=0}^{\lfloor\tau\rfloor - 2} \frac{(\tau p e^{-\theta})^y}{y!} \right]. \tag{8.8}$$

The expectations in expressions (8.7) and (8.8) can then be used to compute the variance of the binomial-EWP distribution in the usual way. The expressions for the moments of the binomial-EWP distribution are complicated and cumbersome. However, since the sums included in these expressions are finite, the moments can be readily computed for given τ, p and θ. The moments in equations (8.7) and (8.8) include the term $y!$ which can be computationally awkward when y is large. In order to accommodate large values of y, the logarithm of $y!$ can be approximated using commands such as `lnfact` in the programming language Gauss and the logarithm of the terms in the sum can be computed. By doing this, the impact of the large value of $y!$ is accommodated and the moment under consideration can then be estimated by exponentiating the resulting terms and calculating their sum.

An alternative method for estimating the mean and variance of the binomial-EWP distribution is to simulate observations from the distribution and then to compute the sample mean and variance of the simulated observations. Two methods of simulation can be used. The observations can be simulated using the table look-up method (Ross, 1998, p.462) and the expressions for the probability that Y is equal to y from (8.6). Alternatively, observations from the binomial-EWP distribution can be simulated by first obtaining a value for N from the EWP distribution using the table look-up method described in Section 8.2.5 and then simulating a value for y given $N = n$ from the binomial distribution. Either method can be used with the number of simulated observations selected by observing the moments for a range of sample sizes. When the simulated sample size is sufficiently large the

estimates of the moments become stable.

8.5 Examples

Illustrative examples of the p.m.f. of the binomial-EWP distribution, together with the mean and variance, are presented in Figure 8.4. Observe that the distribution can have more than one mode. The mean and variance were estimated by simulating observations from the distribution according to the second method described in Section 8.4. In each case a sample size of 10^6 observations was sufficient for calculating the moments. The examples illustrate that the p.m.f. of the binomial-EWP distribution can take different forms and is therefore flexible in terms of modelling.

8.6 Modelling Dose-Response Data

Consider a dose-response setting and let p_i denote the probability of the death of an organism exposed to a dose d_i of a drug, with $x_i = \log d_i$ and $i = 1, \ldots, D$. When a logit tolerance distribution is used, p_i can be written in terms of x_i and the parameters α and β as

$$p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}, i = 1, \ldots, D . \tag{8.9}$$

8.6.1 Likelihood Function

In accordance with the notation introduced in Chapter 2, let the random variable Y_{cj} refer to the number of survivors in the control group, where $j = 1, \ldots, n_c$. Denote the random variable representing the number of organisms surviving exposure to a non-zero dose d_i of the drug by Y_{ij}, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, D$.

Since the increase in probability of death for organisms in the control group is zero, these observations simply constitute count data. Thus consider modelling the observations in the control group with an EWP distribution. Following Ridout

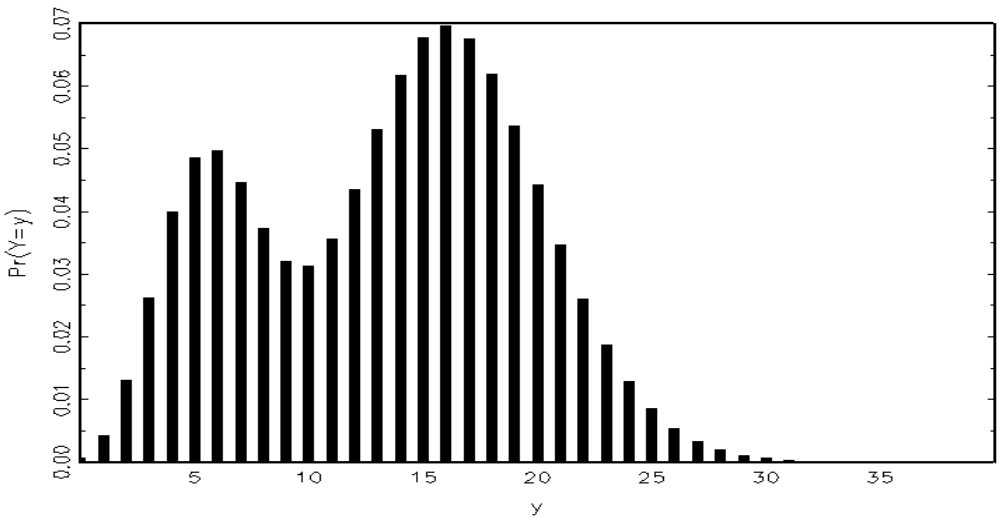


Figure 8.4: (a) $\tau = 20,\ p = 0.5,\ \theta = -0.5$, E($Y$)=13.3870, Var($Y$)=36.4910

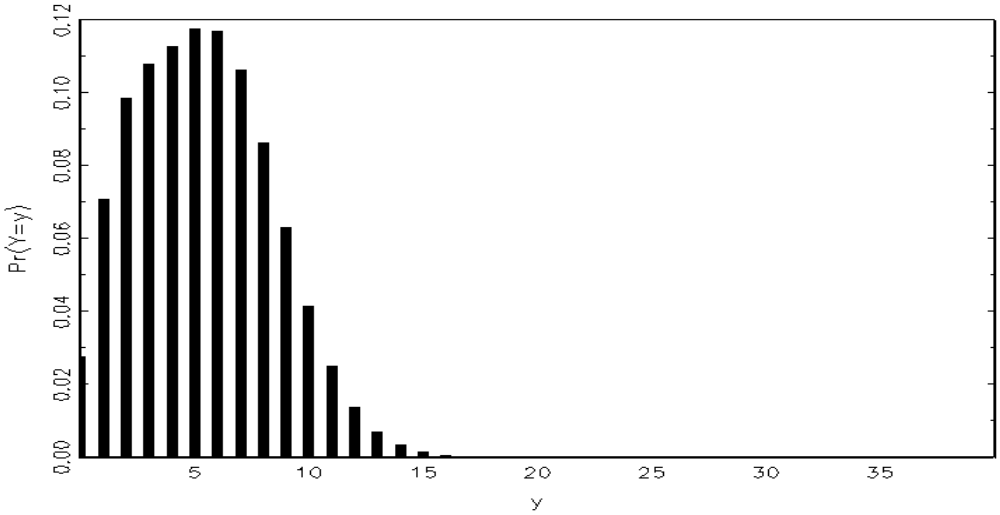


Figure 8.4: (b) $\tau = 20,\ p = 0.2,\ \theta = -0.5$, E($Y$)=5.3570, Var($Y$)=9.0502

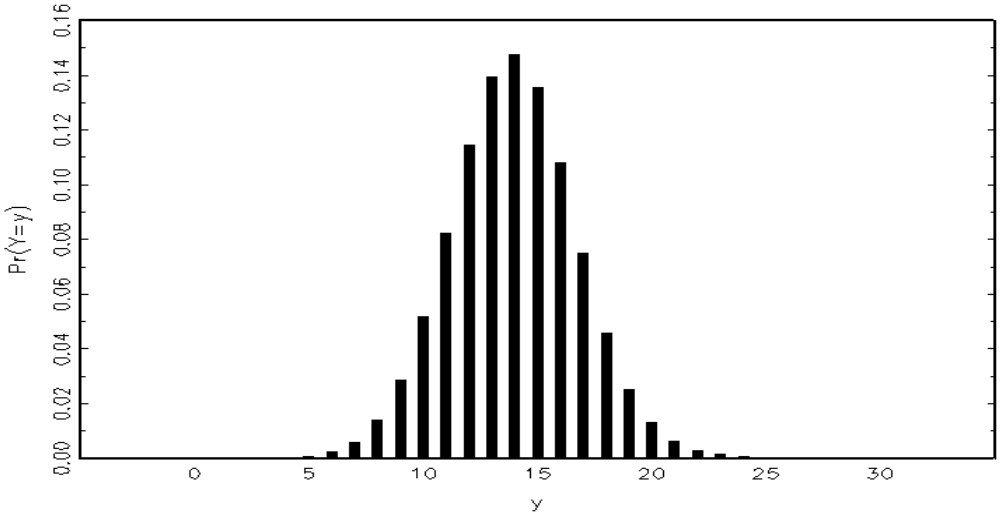


Figure 8.4: (c) $\tau = 20,\ p = 0.7,\ \theta = 0.32$, E($Y$)=13.9359, Var($Y$)=9.1410

Figure 8.4: Plots showing the shapes of various binomial-EWP distributions.

and Besbeas (2004), the likelihood function for an observation from the control group can therefore be written as

$$L_{cj} \;=\; \frac{e^{-\tau}\tau^{y_{cj}}}{y_{cj}!}\frac{e^{-\theta|y_{cj}-\tau|}}{W}$$

and the log-likelihood function as

$$\ell_{cj} \;=\; y_{cj}\ln\tau - \tau - \ln y_{cj}! - \theta|y_{cj}-\tau| - \ln W \;,$$

where W is the normalizing constant of the distribution derived previously and is given by

$$W \;=\; 2e^{-\tau}\sum_{y_{cj}=0}^{\lfloor\tau\rfloor}\frac{\tau^{y_{cj}}}{y_{cj}!}\sinh[\theta(y_{cj}-\tau)] + e^{\tau(\theta-1+e^{-\theta})} \;.$$

The observed number of organisms y_{ij} surviving exposure to a non-zero dose d_i of the drug can be modelled with a binomial-EWP distribution, where $j = 1,\ldots,n_i$ and $i = 1,\ldots,D$. Then the likelihood function for an observation less than or equal to τ is given by

$$L^{(1)}_{d(ij)} \;=\; \frac{2(\tau p)^{y_{ij}}e^{-\tau}}{y_{ij}!W}\sum_{k=0}^{\lfloor\tau\rfloor-y_{ij}}\frac{[\tau(1-p)]^k}{k!}\sinh[\theta(k+y_{ij}-\tau)] + \frac{(\tau p e^{-\theta})^{y_{ij}}e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y_{ij}!W}$$

and for an observation exceeding τ the likelihood function is

$$L^{(2)}_{d(ij)} \;=\; \frac{(\tau p e^{-\theta})^{y_{ij}}e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y_{ij}!W} \;.$$

The log-likelihood functions for an observed response less than or equal to τ is therefore given by

$$\ell^{(1)}_{d(ij)} \;=\; \ln\left[\frac{2(\tau p)^{y_{ij}}e^{-\tau}}{y_{ij}!W}\sum_{k=0}^{\lfloor\tau\rfloor-y_{ij}}\frac{[\tau(1-p)]^k}{k!}\sinh[\theta(k+y_{ij}-\tau)] + \frac{(\tau p e^{-\theta})^{y_{ij}}e^{\tau(\theta-1+(1-p)e^{-\theta})}}{y_{ij}!W}\right]$$

and the log-likelihood function for an observation greater than τ is therefore

$$\ell^{(2)}_{d(ij)} \;=\; y_{ij}\ln(\tau p e^{-\theta}) + \tau[\theta - 1 + (1-p)e^{-\theta}] - \ln(y_{ij}!) - \ln W \;.$$

The log-likelihood function for the data, denoted $\ell(\tau, p, \theta; \underline{y})$, is then found by summing the log-likelihood for each observation over all of the observed responses.

8.6.2 Score Functions and the Information Matrix

The log-likelihood function of the binomial-EWP distribution is discontinuous at integer values of τ and is therefore not differentiable with respect to τ. The score functions for α, β and θ can be derived but these expressions are awkward and cumbersome and as a result they are not included here.

Since the derivation of the information matrix entails differentiating the log-likelihood function with respect to the parameters, the information matrix for the binomial EWP distribution cannot be derived.

8.7 Model Fitting and Checking

8.7.1 Estimation

The log-likelihood function of the binomial-EWP distribution is discontinuous in τ and thus it cannot be optimized numerically using the Gauss OPTMUM procedure. However, the parameters τ, α, β and θ can be estimated using the method of estimation of the parameters for the EWP model outlined in Section 8.2.5. In particular, a range of integer values of τ can be specified and a profile likelihood plot can be produced for this range of values by estimating the remaining parameters in the model using the OPTMUM routine in Gauss. This method of estimation requires starting values for the parameters τ, α, β and θ. Ridout and Besbeas (2004) recommended using a starting value of zero for θ and then obtaining starting values for the remaining parameters by fitting a Poisson distribution to the data. Following their recommendation, the classic Poisson$[\tau(1-p)]$ model can be fitted to the data under investigation and the estimates of α and β can then be used as starting values for these parameters. The estimate of τ from the Poisson$[\tau(1-p)]$ model can be used to give an indication of the range of values to be selected for this parameter.

8.7.2 Inference

Since the information matrix of the binomial-EWP distribution cannot be derived, the standard errors of the estimates of τ, α, β and θ can be approximated using the parametric bootstrap described in Section 8.2.5. Data can be simulated from the binomial-EWP distribution by first simulating a value for N from the EWP distribution using the table look-up method. A y value, given that $N = n$, can then be simulated from the binomial distribution with p depending on dose through α and β. This method of simulation was described in Section 8.4 but without the dependence of the probability on the dose of the drug. 95% Wald intervals can then be computed for each of the parameters. For the parameter τ, this interval is calculated as $\hat{\tau} \pm 1.96\ se(\hat{\tau})$, where $\hat{\tau}$ is the maximum likelihood estimate of τ and $se(\hat{\tau})$ is the standard error of the estimate of τ (Kutner, Nachtsheim, Neter and Li, 2005, p.579).

A profile likelihood plot for the parameter τ is generated by the method used for estimating the parameters of the binomial-EWP distribution. Once again, non-integer values of τ can be included in the range of τ values when approximating the standard error of this parameter estimate. Profile likelihood plots can also be produced for the remaining parameters α, β and θ. Consider the profile likelihood plot for θ. This can be obtained by specifying a range of values for θ and for each value of θ a grid of integer values for τ can be considered. The log-likelihood function can then be maximized with respect to α and β for each combination of θ and τ and the profile likelihood plot constructed accordingly. The profile likelihood plots for α and β can be constructed in a similar way. The method proposed here will give an approximate profile likelihood and will capture any asymmetry in the distribution of the parameters. The profile likelihood plots can then be used to obtain approximate 95% confidence intervals for the parameters. For example, the 95% confidence interval for the parameter τ is given by $[\tau_1,\tau_2]$ where

$$\ell_P(\tau_1) = \ell_P(\tau_2) = \ell(\hat{\tau}, \hat{\alpha}, \hat{\beta}, \hat{\theta}; \underline{y}) - \frac{1}{2}\chi^2_{1,0.05}\ ,$$

and where $\ell_P(\tau)$ is the profile likelihood for the parameter τ evaluated at τ and $\ell(\hat{\tau}, \hat{\alpha}, \hat{\beta}, \hat{\theta}; \underline{y})$ is the maximum log-likelihood of the binomial-EWP distribution (Azzalini, 1996, p.146).

8.7.3 Goodness of Fit

The fit of the binomial-EWP distribution to dose-response data can be compared with that of the Poisson, beta-Poisson, Altham-Poisson and negative binomial distributions using AIC. For the binomial-EWP distribution with a logit tolerance distribution this statistic is calculated as $-2\ell(\hat{\tau}, \hat{\alpha}, \hat{\beta}, \hat{\theta}; \underline{y}) + 8$, where $q = 4$. The model with the lowest AIC value provides the best fit to the data.

8.8 Examples

8.8.1 Algae Data

The binomial-EWP distribution for dose-response data was fitted to the algae data using the method of estimation described in Section 8.7.1. The standard errors of the parameter estimates were obtained by generating 150 parametric bootstrap samples of size 25, 5 observations for each dose, and then calculating the standard errors of the parameter estimates obtained from these samples. The resulting standard errors were used to construct 95% confidence intervals for the parameters. Profile likelihood plots were also obtained for the parameters τ, α, β and θ and used to compute 95% confidence intervals for the parameter estimates. The profile likelihood plots are time consuming to construct and are presented in Figures 8.5 (a), (b), (c) and (d).

The estimates of τ, α, β and θ, together with their standard errors, 95% Wald intervals and 95% profile likelihood intervals are recorded in Table 8.3. Recall that when θ is zero the EWP distribution reduces to a Poisson distribution. The estimate of θ for the algae data is very close to zero and in fact zero lies within the

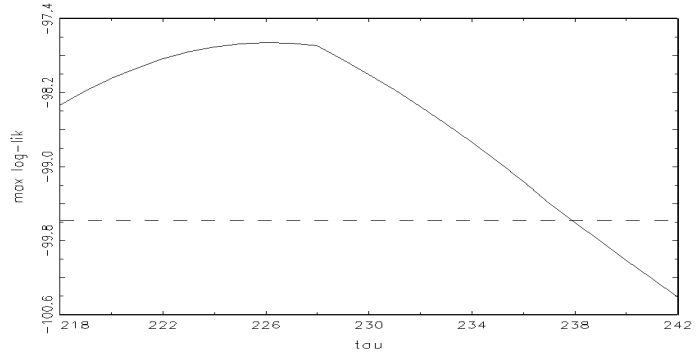


Figure 8.5: (a)

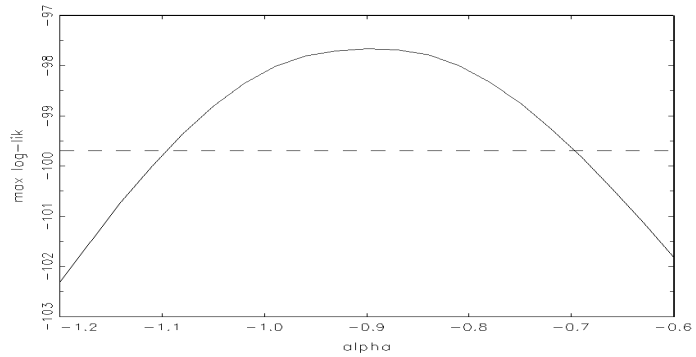


Figure 8.5: (b)

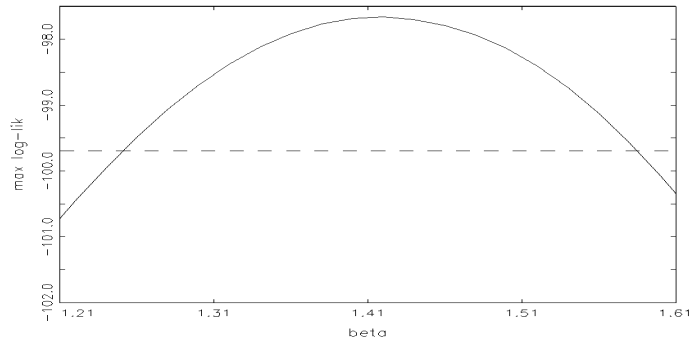


Figure 8.5: (c)

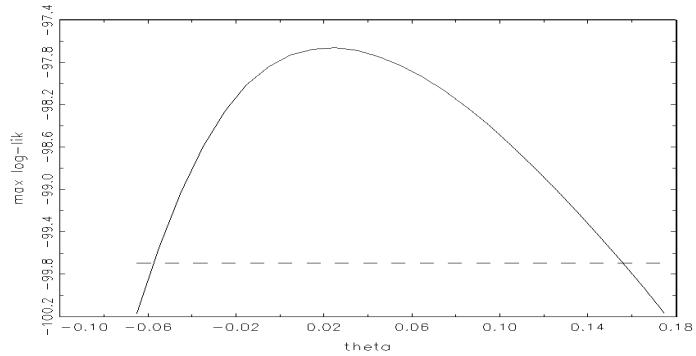


Figure 8.5: (d)

Figure 8.5: Profile likelihood plots for the parameters (a) τ, (b) α, (c) β and (d) θ of the binomial-EWP model with a logit tolerance distribution fitted to the algae data.

Table 8.3: Results from fitting the binomial-EWP model to the algae data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	226	1.325	(223.403;228.597)	(*;238)
α	-0.898	0.104	(-1.102;-0.694)	(-1.110;-0.690)
β	1.419	0.089	(1.245;1.593)	(1.261;1.583)
θ	0.022	0.042	(-0.060;0.104)	(-0.055;0.155)

*This limit of the confidence interval cannot be computed

confidence intervals for θ. As a result the initial Poisson distribution in a Wadley context is essentially being fitted to the algae data here.

A comparison of the AIC values for the different models fitted to the algae data is presented in Table 8.4. The AIC values indicate that the negative binomial and

Table 8.4: A comparison of the various models fitted to the algae data.

Model	Maximum log-likelihood	Number of parameters	AIC
Poisson	-97.776	3	201.552
Beta-Poisson	-96.251	4	200.501
Altham-Poisson	-96.314	4	200.628
Negative Binomial	-97.778	4	203.556
Binomial-EWP	-97.661	4	203.322

binomial-EWP distributions did not accommodate the overdispersion in the algae data at all. This is because the models were essentially reduced to Poisson models. The beta-Poisson and Altham-Poisson distributions are better at accommodating the overdispersion in the algae data. The AIC values of these two models were very similar, with the AIC of the beta-Poisson distribution being slightly lower than that of the Altham-Poisson distribution. It can therefore be deduced that the source of extra variability in the algae data is implicit in the probability of

death rather than the unobserved number of organisms initially treated.

8.8.2 Malaria Data

The binomial-EWP distribution was fitted to the malaria data described in Section 2.2.2 by considering a range of values of τ and optimizing the likelihood function with respect to α, β and θ. 150 parametric bootstrap samples of size 24 were generated and used to approximate the standard errors of the parameter estimates. A profile likelihood plot was obtained for the parameter τ when the parameters were estimated but profile likelihood plots could not be produced for the remaining parameters in the model due to convergence problems. The profile likelihood plot for the parameter τ is included in Figure 8.6. This plot was used to construct

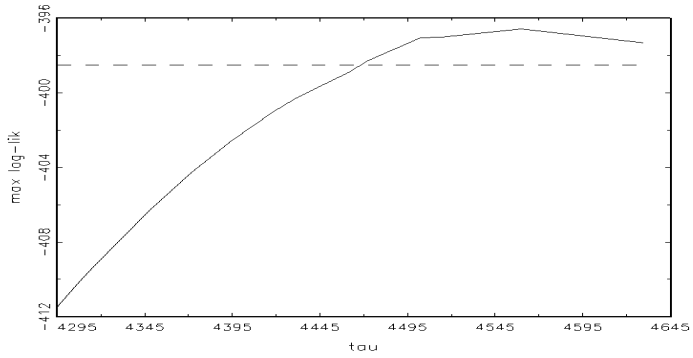


Figure 8.6: Profile likelihood plot for the parameter τ of the binomial-EWP model with a logit tolerance distribution fitted to the malaria data.

a 95% confidence interval for τ but the upper limit of the interval could not be obtained because of problems with convergence. The estimates of τ, α, β and θ, together with their standard errors, 95% Wald intervals and the 95% profile likelihood interval for τ are presented in Table 8.5.

The fit of the binomial-EWP distribution was compared with that of the Poisson, beta-Poisson and negative binomial distributions using AIC. Recall that the Altham-Poisson distribution could not be fitted to the malaria data. Table 8.6 contains a comparison of the AIC values for the various models and from these values it can be deduced that by providing a better fit than the Poisson distribu-

Table 8.5: Results from fitting the binomial-EWP model to the malaria data.

Parameter	Estimate	Standard Error	95% Wald Interval	95% Profile Interval
τ	4560	1.560	(4556.942;4563.058)	(4465;*)
α	-2.805	0.745	(-4.265;-1.345)	-
β	2.246	0.959	(0.366;4.126)	-
θ	-0.105	0.038	(-0.179;-0.031)	-

*This limit of the confidence interval cannot be computed

tion, the binomial-EWP distribution accounts for some of the overdispersion in the malaria data. However, the AIC values of the beta-Poisson and negative binomial

Table 8.6: A comparison of the various models fitted to the malaria data.

Model	Maximum log-likelihood	Number of parameters	AIC
Poisson	-870.023	3	1746.046
Beta-Poisson	-174.670	4	357.340
Negative Binomial	-183.152	4	374.304
Binomial-EWP	-396.581	4	747.162

distributions are much lower than the AIC value of the binomial-EWP distribution and thus these models seem better able to accommodate the overdispersion in the malaria data.

8.9 Conclusions

Properties of the EWP distribution are explored in this chapter and its moments are derived. Ridout and Besbeas (2004) did not explore the algebraic tractability of the EWP distribution but it is demonstrated in the present study. In fact, the algebraic tractability of the EWP distribution carries through to the binomial-

EWP distribution. The p.m.f. is discontinuous in τ and hence problems arise with parameter estimation and inference. These problems were addressed here by using profile likelihood and simulation techniques.

The binomial-EWP distribution is introduced when the EWP distribution is considered for modelling data arising from a Wadley's problem setting. The moments of the binomial-EWP distribution are derived and while the expressions for the moments are explicit, they are somewhat cumbersome. As a result, simulation can be used as a means of estimating these moments. As with the EWP distribution, the p.m.f. of the binomial-EWP distribution is discontinuous in the parameter τ. The problems of estimation and inference were addressed in the same way as for the EWP distribution.

The binomial-EWP model was fitted to the algae data and the fit was extremely poor. This model was also fitted to the malaria data and, while it was observed that the distribution accounted for some of the overdispersion in the data, it was not the best fitting model. Profile likelihood plots could not be obtained for all of the parameters when fitting the binomial-EWP distribution to the malaria data. It can therefore be concluded that the binomial-EWP model can accommodate overdispersion in some cases but that the beta-Poisson and Altham-Poisson distributions seem better able to model the overdispersed data from a Wadley's problem setting considered here. It can be surmised that the overdispersion in these data sets is implicit in the probability of death rather than the unknown number of organisms initially treated, N.

Chapter 9

Conclusions

Various models for overdispersion have been used for Wadley-type data in this study. The well-known beta-binomial model was considered and crudely adapted to a Wadley's problem setting. This distribution was then extended in a more elegant way to a Wadley problem setting, resulting in the beta-Poisson model. Although some work has been done on the beta-Poisson distribution, properties of the distribution and parameter estimation were explored in detail and the model was used in a dose-response context. Very little work has been done on the multiplicative binomial distribution which was introduced by Altham in 1978. This model was considered in the present study and some of its properties, such as modality and the information matrix, were explored and derived. An extension of the multiplicative binomial distribution to a Wadley's problem setting yielded the Altham-Poisson model. This model has not been derived or studied before and was not particularly tractable algebraically. Numerical techniques were therefore employed to accommodate the algebraic intractability of this model.

The negative binomial is frequently used in a Wadley's problem setting and was therefore considered in the present study. Finally, the EWP distribution was investigated. This model has not been examined in detail in the literature and thus its knowledge base was developed in this thesis. The EWP model was then used for Wadley's problem and the resulting binomial-EWP distribution, which has not

been considered in previous studies, was derived. Properties of the binomial-EWP distribution were investigated.

The beta-Poisson model, which is suitable for overdispersion in Wadley's problem setting, proved to be effective for modelling the overdispersion in the algae data. The Altham-Poisson distribution adequately models overdispersed count data and overdispersed dose-response data from a Wadley's problem setting. It has limited usefulness, however, in that it cannot easily model large observations. Finally, distributions for overdispersed count data were considered for modelling Wadley's problem and the negative binomial distribution fitted the algae and malaria data sets poorly. The binomial-EWP distribution provided a very poor fit to the algae data and accounted for only some of the overdispersion in the malaria data. The apparent cause of overdispersion in the two primary data sets considered in this thesis was clearly the probability of death rather than the unknown number of organisms initially treated.

There is scope for future research in overdispersion within the Wadley context. In particular, random effects models and the double binomial model (Efron, 1986), which are suitable for modelling overdispersion in the binomial probability, can be adapted to a Wadley's problem setting. The CMP distribution, which was mentioned in this thesis, has recently been reparametrized and used in a logistic regression setting by Guikema and Goffelt (2008). The reparametrized model can be explored for modelling Wadley's problem, thereby modelling overdispersion in the unknown number of organisms initially treated. Methods proposed by Faddy and Smith (2005), which consider a bivariate generalization of the Poisson process for the binomial parameter n, can also be considered.

The Wadley-type distributions considered in this study can be investigated in a broader context. In particular, mixtures of these distributions and their applications to zero-inflated data can be explored.

Software

Gauss. Aptech Systems Inc., 23804 SE Kent-Kangley Road, Maple Valley, WA 98038, United States of America. http://www.aptech.com.

Genstat. VSN International Ltd, 5 The Waterhouse, Waterhouse Street, Hemel Hempstead, HP1 1ES, United Kingdom. http://www.vsni.co.uk.

Mathematica. Wolfram Research Inc., 100 Trade Center Drive, Champaign, IL 61820-7237, United States of America. http://www.wolfram.com.

References

Altham, P. M. E. (1978). Two generalizations of the binomial distribution. *Applied Statistics* **27**, 162-167.

Anscombe, F. J. (1949). The statistical analysis of insect counts based on the negative binomial distribution. *Biometrics* **5**, 165-173.

Anscombe, F. J. (1949). Note on a problem in probit analysis. *Annals of Applied Biology* **36**, 203-205.

Arfken, G. B., Weber, H. J. (1995). *Mathematical Methods for Physicists, International Edition*, 4th edition. San Diego: Academic Press.

Azzalini, A. (1996). *Statistical Inference Based on the Likelihood*. London: Chapman and Hall.

Baker, R. J., Pierce, C. B., Pierce, J. M. (1980). Wadley's problem with controls. *GLIM Newsletter* No. 3, 32-35.

Bhattacharya, S. K., Holla, M. S. (1965). On a discrete distribution with special reference to the theory of accident proneness. *Journal of the American Statistical Association* **60**, 1060-1066.

Collett, D. (2003). *Modelling Binary Data*. London: Chapman and Hall.

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied Statistics*, **21**, 113-120.

Darroch, J. N. (1974). Multiplicative and additive interaction in contingency tables. *Biometrika*, **61**, 207-214.

Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*, 2nd edition. Boca Raton: Chapman and Hall.

Efron, B. (1986). Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association* **81**, 709-721.

Efron, B., Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Eickhoff, J. C. (2006). Personal communication.

Engel, B., te Brake, J. (1993). Analysis of embryonic development with a model for under- or overdispersion relative to binomial variation. *Biometrics*, **49**, 269-279.

Erdélyi, A., Magnus, W., Oberhettinger, F., Tricomi, F. G. (1981). *Higher Transcendental Functions, 3 vols*. New York: McGraw Hill.

Faddy, M. J., Smith, D. M. (2005). Modeling the dependence between the number of trials and the success probability in binary trials. *Biometrics*, **61**, 1112-1114.

Finney, D. J. (1971). *Probit Analysis*, 3rd edition. Cambridge: University Press.

Gouws, E. (1995). Drug resistance in malaria research: the statistical approach. MSc thesis, University of Natal.

Grattan-Guinness, I. (1970). *The development of the foundations of mathematical analysis from Euler to Rieman*. Massachusetts: The Colonial Press Inc.

Griffiths, D. A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics* **29**, 637-648.

Grimmett, G., Welsh, D. (1986). *Probability an Introduction.* New York: Oxford University Press.

Guikema, S. D., Goffelt, J. P. (2008). A flexible count data regression model for risk analysis. *Risk Analysis*, **28**, 213-223.

Haines, L. M., Clarke, G. P. Y., Gouws, E., Rosenberger, W. F. (2001). Optimal design for the testing of anti-malarial drugs. In *mODa 6 - Advances in Model-Oriented Design and Analysis*, Atkinson, A. C., Hackl, P., Müller, W. G. (eds), 119-126. New York: Physica-Verlag.

Hogg, R. V., McKean, J. W., Craig, A. T. (2005). *Introduction to Mathematical Statistics*, 6th edition. New Jersey: Prentice-Hall.

Holgate, P. (1970). The modality of some compound Poisson distributions. *Biometrika*, **57**, 666-667.

Holla, M. S., Bhattacharya, S. K. (1965). On a discrete compound distribution. *Annals of the Institute of Statistical Mathematics, Tokyo* **17**, 377-384.

Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. (2005). *Applied Linear Statistical Models*, 5th edition. Boston: Prentice-Hall.

Lawless, J. F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics* **15**, 209-225.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Volume 2, 2nd edition. New York: Wiley.

Johnson, N. L., Kotz, S., Kemp, A. W. (1992). *Univariate Discrete Distributions*, 2nd edition. New York: Wiley.

Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach*. London: Chapman and Hall.

Kokonendji, C. C, Mizère, D., Balakrishnan, N. (2008). Connections of the Poisson weight function to overdispersion and underdispersion. *Journal of Statistical Planning and Inference*, **138**, 1287-1296.

Kotz, S., Johnson, N. L. (1986). Poisson-Beta Distribution. *Encyclopedia of Statistical Sciences* **7**, 19. New York: Wiley.

Lang, S. (1974) *Analysis I*, Reading: Addison-Wesley Publishing Company.

Lindsey, J. K., Altham, P. M. E. (1998). Analysis of the human sex ratio by using overdispersion models. *Applied Statistics* **47**, 149-157.

Lindsey, J. K., Mersch, G. (1992). Fitting and comparing probability distributions and log linear models. *Computational Statistics and Data Analysis* **13**, 373-384.

Lindsey, J. K. (1995). *Modelling Frequency and Count Data*, New York: Oxford University Press.

Little, R. J. A. (Discussant) (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 25.

Lovison, G. (1998). An alternative representation of Altham's multiplicative binomial distribution. *Statistics and Probability Letters*, **36**, 415-420.

Madden, L. V., Hughes, G. (1995). Plant disease incidence: distributions, heterogeneity, and temporal analysis. *Annual Review of Phytopathology*, **33**, 529-564.

McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models*, 2nd edition. Cambridge: Chapman and Hall.

Molenberghs, G., Ryan, L. M. (1999). An exponential family model for clustered multivariate binary data. *Environmetrics*, **10**, 279-300.

Morgan, B. J. T. (1992). *Analysis of Quantal Response Data*. London: Chapman and Hall.

Nelder, J. A. (Discussant) (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 23-25.

Olkin, I., Petkau, A. J., Zidek, J. V. (1981). A comparison of n estimators for the binomial distribution. *Journal of the American Statistical Association* **76**, 637-642.

Pack, S. E. (1986). The analysis of proportions from toxicological experiments. Unpublished Ph.D. thesis, University of Kent, Canterbury.

Ridout, M. S., Besbeas, P. (2004). An empirical model for underdispersed count data. *Statistical Modelling*, **4**, 77-89.

Ross, S. M. (1998). *A First Course in Probability*, 5th edition. New Jersey: Prentice Hall.

Ross, S. M. (2002). *Introduction to Probability Models*, 9th edition. Orlando: Academic Press.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the Conway-Maxwell-Poisson distribution. *Applied Statistics*, **54**, 127-142.

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between sets of trials. *Journal of the Royal Statistical Society, Series B* **10**, 257-261.

Smith, D. M., Morgan, B. J. T. (1989). Extended models for Wadley's problem. *GLIM's Newsletter* No. 18, 21-28.

Tarone, R. E. (1979). Testing the goodness of fit of the binomial distribution. *Biometrika*, **66**, 585-590.

Trajstman, A. C. (1989). Indices for comparing decontaminants when data come from dose-response survival and contamination experiments. *Applied Statistics*, **38**, 481-494.

Wadley, F. M. (1949). Dosage-mortality correlation with number treated estimated from a parallel sample. *Annals of Applied Biology* **36**, 196-202.

Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31**, 949-952.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics* **31**, 144-148.

Young, L. J., Campbell, N. L., Capuano, G. A. (1999). Analysis of overdispersed count data from single-factor experiments: a comparative study. *Journal of Agriculture, Biological, and Environmental Statistics*, **4**, 258-275.

Zelterman, D., (2004). *Discrete Distributions: Applications in the Health Sciences*, Sussex: Wiley.

Zhu, J., Eickhoff, J. C., Kaiser, M. S. (2003). Modeling the dependence between number of trials and success probability in beta-binomial-Poisson mixture distributions. *Biometrics*, **59**, 955-961.