KU LEUVEN

FACULTY OF SCIENCES

An alternative Conway-Maxwell-Poisson model to handle dispersion settings

Josiah MCMILLAN

Supervisor: Prof. Martial Luyts
Leuven Statistics Research Centre (LStat)

Master thesis submitted in fulfillment
of the requirements for the degree in
Master of Science in Statistics and Data Science

Academic year 2021-2022

Preface

I would like to thank my supervisor Prof. Martial Luyts for his guidance, tremendous help in the writing process, and the provision of data and code. His comments and checks on this process were invaluable.

The thesis aims to explore a new family of probability functions in order to determine if there are any specifications that lead to better modelling outcomes in comparison to the state of the art count models.

Summary

Count models are important to modelling real world processes, these models are varied in their characteristics and necessary assumptions. Since count data comes in many forms a variety of distributions are needed to model these data. While the familiar Poisson and Negative Binomial models are able to model count data well, they are unable to model under-dispersion. The Conway-Maxwell-Poisson (CMP) distribution can flexibly model count data that is over and under dispersed however it has issues with coefficient interpretation and run time.

A new family of distributions for modeling count data has its statistical properties explored. The family of distributions is an extension of the Poisson distribution, with a key change. The exponent of the θ parameter in the Poisson distribution is allowed to take on varied functional forms rather than being restricted to the identity. Due to this added flexibility, a normalizing constant is added to ensure the specification remains a Probability Mass Function (PMF). We analyze some of the analytic and numerical properties of the normalizing constant and explore the numerical limitations on the parameter space for some specifications.

Three specifications are chosen from this new family of distributions which are shown to be able to model either under- and over- dispersion. Using Maximum Likelihood Estimation (MLE) to fit our chosen specifications to five simulated datasets with varying levels of dispersion, and one with zero-inflation we show that two of our specifications are able to perform well on under-dispersed data with some limitations.

With limitations on two of our specifications we choose one specification and fit a model to two real world data sets where the outcome variable exhibits either over- or under- dispersion and compare our model to the standard count models in terms of Akaike Information Criterion (AIC) and run time. Coefficients are found to have the correct sign and significance, but interpretation remains difficult.

Ultimately we find that the specification used does a better job in terms of fit than the Poisson but does worse than other models, even underperforming the CMP-μ model in terms of run time.

Glossary

λ location parameter for the CMP.

c parameter for $f(n) = c \cdot n$ in the alternative distribution.

μ mean in reference to the CMP-μ model developed by Huang (2017).

θ location parameter for the alternative distribution.

v Dispersion parameter for the CMP.

Acronyms

AIC Akaike Information Criterion.

BMI Body Mass Index.

CDC Centers for Disease Control and Prevention.

CMP Conway-Maxwell-Poisson.

GLM Generalized Linear Model.

MGF Moment Generating Function.

MLE Maximum Likelihood Estimation.

PMF Probability Mass Function.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

When dealing with count data, GLM frameworks like Poisson and Negative Binomial models are often used as a standard approach due to their simplicity in use and easy interpretation. The Negative-Binomial is able to model data that is over-dispersed while the Poisson is able to model data that is equi-dispersed. This limitation of the Poisson means that for over-dispersed data the Negative-Binomial is superior, however when it comes to under-dispersion both do not provide a good fit. The purpose of this thesis is to investigate a new model approach that may be able to deal with count data and see how it performs in comparison to the standard models.

Each distribution has its use in modeling count data with Negative-Binomial and the CMP functions being able to handle data that has variance that is larger than the mean and the CMP having the ability to model data with variation in the dispersion around the mean. Each of these functions provides value to those who model counts with different levels of flexibility and different types of assumptions used.

The CMP model was first explored by: Shmueli et al. (2005) which led to the popularization of the model and subsequent exploration over the years. One of the primary advantages that was explored was the great flexibility of the CMP model over the poisson and the Negative-Binomial models as explored in the paper by Shmueli (2005) but this came at the cost of greater computation times. The authors also detailed three methods of dealing with estimating the parameters of the CMP with the infinite sum $Z(\lambda, v)$ being truncated for all practical purposes.

Interpretation of the coefficients in the CMP model is done with an alternative parameterization of the CMP regression model with the approximate $\mu = \lambda^{\frac{1}{v}}$ but this leads to difficulty with interpretation of the covariates.

This alternative formulation was first proposed by Lord et al. (2008) and allowed the distribution to be more easily interpretable as the effect of covariates now directly influenced the centering of the model.

The CMP model had an interpretation issue resolved with the paper by Huang (2017) which outlined the properties of the CMP-μ model. In particular, the CMP-μ model overcomes the issue of the lambda parameterization of the CMP model by using an alternative formulation of the probability mass function of the Conway-Maxwell distribution. Using the mean rather than the lambda parameterization allows covariates to be more directly interpretable in a standard Generalized Linear Model (GLM) format and the model was found to be faster to converge. Beyond truncating the normalizing constant which has limitations on the bounds of μ and v, other methods of approximating the normalizing constant have been taken.

Additional extensions of the CMP model are those that have applied it to zero-inflated data such as the model explored by Choo-Wosoba et al. (2015). Other models have sought to extend the CMP such that it unifies both the Gamma and the Negative Binomial distributions as well, what has been called the Extended CMP model by Chakraborty and Imoto (2016). Beyond these extensions in the flexibility of the model, the CMP has been augmented to fit time series data as well in the form of an Autoregressive Moving Average model as seen in the paper by Melo and Alencar (2020).

One of the main issues of the CMP model is the presence of a normalizing constant in the distribution, making the model often slow in convergence. The CMP distribution suffers because of the difficulty in explaining its covariates when applied in a Generalized Linear Model context which is only overcome by the CMP-μ model. The aim of my thesis is to find a family of functions that are able to model the count data more quickly and provide a better way of interpreting said count data. Additionally we aim to see if our alternative can be fit to both real world and simulated data and perform at similar levels to the CMP or to the Negative Binomial.

Our research aims to explore a family of distributions, investigate its characteristics, to check for model fit, and compare with commonly used models. This family of distributions can serve as an interesting alternative to the standard distributions and the CMP distribution when it comes to modeling count data in some contexts.

We aim to show the explore the qualities of our distribution by a number of both analytical and numerical methods. The alternative distribution is applied to benchmark count data as well as simulated data in order to see the difference in performance. Data simulation shall be used in order to see how the model deals with different levels of dispersion among different parameters and how it deals with a low (high) global dispersion with low (high) mean counts.

We proceed with the thesis as follows: The real and simulated datasets that are used for the analysis are described in Section 2. In Section 3 the CMP, the Poisson, and our own alternative distribution are described. Next, we provide a proof for the convergence of the normalizing constant as well as show convergence in numerical simulations. Findings about the numerical simulations for different function specifications and θ values are discussed as well. Then in Section 4 we delve into proving that the alternative is in fact a probability mass function, then we seek to determine the key properties of our distribution such as the moment generating function along with the first and second moments. The application of our method to simulated count data using MLE is discussed in Section 5. The results of the application of the PMF to real data is in Section 6. We end with an evaluation of our method and our conclusions in Section 7.

Chapter 2

Dataset Descriptions

2.1 Simulated Data

We generated six datasets using the R programming language. These seven data sets have characteristics of being over-, under- and equi- dispersed as well as different μ values. The different choices of μ are explored to leverage the increasing dispersion of the Alternative distribution with specification $f(n) = n^{\frac{13}{12}}$. These values were generated using base R functions such as rnbinom and rpois as well as the function simcmp from the "degreenet" library.

The simcmp function allows us to sample the CMP distribution with different values of λ and v respectively. We leverage this fact to create a datasets with both moderate and strong under dispersion.

We also generate zero-inflated data to compare our methods to the standard methods. We simulate a Bernoulli trial with the rbinom function in R with a probability of 0.80 which controls the proportion of zeros in our dataset. For the zero-inflated dataset, we used a Poisson distribution with $\lambda = 10$ for the non-zero portion of the distribution.

Summary Statistics for simulated datasets

Dataset	n	mean	sd	median	min	max
equi-dispersed	500	4.10	2.08	4	0	12
over-dispersed, $\mu = 4$	500	4.28	4.75	3	0	29
over-dispersed, $\mu = 10$	500	9.74	4.58	9	0	28
under-dispersed	500	4.14	1.56	4	1	9
strongly under-dispersed	500	4.01	0.96	4	2	7
zero-inflated	500	1.62	3.79	0	0	18

2.2 Sleep Duration Data

The sleep time dataset comes from an annual telephone survey of United States residents conducted by the Centers for Disease Control and Prevention (CDC) for the year 2020 as a part of the Behavioral Risk Factor Surveillance System. This dataset is composed of over 400,000 respondents from all 50 states, the District of Columbia and three U.S. territories. The dataset includes health status questions and demographic data of respondents. The variable of interest is the hours of sleep each night, the variance of sleep time is equal to 2.1 and the mean is equal to 7.1 which means it is underdispersed. The full dataset of 300 variables was reduced down to 20 variables of interest and a set of summary statistics can be found in the tables below. All variables that are not starred take numerical values. This dataset is the primary dataset of interest for the regression modelling we will be doing in subsequent sections.

Summary statistics, Sleep Duration data

Variable	n	mean	sd	medin	min	max
SleepTime	319795	7.10	1.44	7.00	1.00	24.00
BMI	319795	28.33	6.36	27.34	12.02	94.85
MentalHealth	319795	3.90	7.96	0.00	0.00	30.00
PhysicalHealth	319795	3.37	7.95	0.00	0.00	30.00

2.3 Publications Data

The dataset contains a count for the number of articles published by a sample of 126 lecturers at Covenant University in Nigeria from 2013 to 2015. The variable of interest is the number of articles published by the lecturer and the covariates that were kept from the original dataset are: Number of children, years of teaching experience, and number of undergraduate courses taught within the period of interest. This data was chosen due to the over-dispersed nature of the response variable and it was sourced from Mendeley Data (Sunday, 2021).

Summary statistics, Publication data

Variable	n	mean	sd	median	min	max
Publications	126	6.44	5.57	6	0	32
Children	126	1.92	1.20	2	0	4
Undergrad Courses Taught	126	7.87	3.91	6	3	27
Years of Experience	126	10.38	7.41	9	1	41

Chapter 3

Methods

3.1 The Standard Count Models

Data that comes in counts are only defined for numbers in $\mathbb{N}$ whereas other models may be able to $\mathbb{R}$, this lends count data to have their own class of models. Before endeavoring to explore the alternative proposition we must first discuss the state of the art. The models used range from the relatively simple but restrictive Poisson distribution to the flexible but computationally complex Conway-Maxwell-Poisson Distribution.

Poisson Model

The most recognizable count model in use is the Poisson model characterized below. This model importantly assumes equi-dispersion which limits its flexibility.

$$P(N = n) = \frac{\lambda^n}{(n!)} e^{-\lambda}$$

The expected value and the variance are given by:

$$E(N) = \lambda$$

$$Var(N) = \lambda$$

Equi-dispersion is an assumption that means that the mean of the data is the same as the variance of the data. This restriction is often violated when dealing with data in the real world that is either over- or under- dispersed. The Poisson gets equi-dispersion from the formulation of its mean and variance which are equivalent and are both equal to λ. Fitting a Poisson regression on over- or under- dispersed data doesn't influence the mean of regression coefficients but it does influence the standard errors meaning that the confidence intervals of the coefficients will be larger than they ought to be. Larger confidence intervals from the Poisson regression could in some cases lead to the confidence intervals containing 0 which would lead to an increase in Type I error in this circumstance.

Negative-Binomial Model

he Negative binomial model is able to model both equi- and over-dispersion. It is related to the Poisson distribution by an alternative parameterization.

$$P(N=n) = \left(\frac{\Gamma(n+\alpha^{-1})}{\Gamma(n+1)\Gamma(\alpha^{-1})}\right)(\lambda\alpha)^n\,(1+\alpha\lambda)^{(-n+\alpha^{-1})}$$

The expected value and the variance are given by:

$$E(N) = \lambda$$

$$Var(N) = \lambda + \alpha\lambda^2$$

The Negative Binomial is a generalization of the Poisson distribution, having the same mean structure but with an additional parameter for over-dispersion. Because the model is more dispersed than the Poisson it is employed in a regression context whenever the conditional distribution of the response variable is over-dispersed. A limitation of the Negative-Binomial Model is that while it can model over-dispersed data it is unable to handle data that is under dispersed much like the Poisson.

The Conway-Maxwell-Poisson Model

For count data the most flexible model that currently exists is the Conway-Maxwell-Poisson, it is able to model both over- and under-dispersion for within some bounds on the parameters: λ, v

$$P(N) = \frac{\lambda^{\,n}}{(n!)^v}e^{-\lambda}\,\frac{1}{Z\left(f\left(n\right),\,\lambda\right)}$$

where:

$$Z\left(s,\,\lambda\right) = \sum_{s=0}^{\infty}\frac{\lambda^{\,s}}{(s!)^v}e^{-\lambda}$$

The expected value and the variance are approximated by:

$$E(N) \approx \lambda^{\frac{1}{v}} - \frac{v-1}{2v} \tag{3.1}$$

$$Var(N) \approx \frac{1}{v}\left(\lambda^{\frac{1}{v}} - \frac{v-1}{2v}\right) \approx \frac{1}{v}\left(E(N)\right)$$

The CMP was rediscovered by Shmueli et al. (2005) after having been initially introduced by it's namesakes Maxwell and Conway (1962). Subsequently, the CMP had it's attractive features discovered, namely the flexibility when modelling different types of dispersion. Extensions to the CMP have also allowed for the model to take into account different dispersion levels across different groups in the data as shown in Sellers and Shmueli (2010). An issue with the CMP is that it has difficult to interpret coefficients when used in a regression context; the approximation used for generating regression coefficients was outlined by Sellers and Shmueli (2008) however the point estimates are often quite different to those of the Poisson and Negative-Binomial. Numerical methods and approximations must be made for the CMP to work properly, numerical methods are necessary to do MLE and the normalizing constant must be computed directly which increases the time required to employ this method.

The Conway-Maxwell-Poisson-μ Model

Due to these limitations the CMP-μ model was developed by Huang (2017) which reparameterizes the distribution in the following way.

$$P(N = n|\mu, v) = \frac{\lambda(\mu, v)^{\,n}}{(n!)^{v}} \frac{1}{Z\left(\lambda(\mu, v), v\,\right)}$$

where $\lambda(\mu, v)$ is given by the solution to:

$$0 = \sum_{n=0}^{\infty}(n - \mu)\frac{\lambda^{n}}{(n!)^{v}}$$

Nearly all of the same benefits and negatives as the CMP exist for the CMP-μ in terms of the flexibility of modelling dispersion and the difficulty evaluating the infinite sum for the normalizing constant. The benefit of the CMP-μ model is that the rate parameter corresponds to the mean of the distribution and therefore covariates become directly comparable to those produced by the Negative-Binomial and the Poisson.

3.2 Alternative Proposition

Thus the limitations that exist in the models, be they restrictions on dispersion or slowness in generating a normalizing constant, are what motivates the exploration of our family of distributions. The topic of this thesis is to investigate a new model family to check for its flexibility under different function specifications. Let $N \in \mathbb{N}$, i.i.d. with $\theta > 0$, the proposed PMF that I explored for this thesis is of the form:

$$P(N) = \frac{\theta^{\,f(n)}}{(n)!} e^{-\theta}\, \frac{1}{C\left(f\left(n\right),\, \theta\right)} \tag{3.2}$$

where:

$$C\left(f\left(s\right),\, \theta\right) = \sum_{s=0}^{\infty}\frac{\theta^{\,f(s)}}{(s)!}e^{-\theta} \tag{3.3}$$

Our model is of a similar to that of the CMP in that it is a part of the exponential family, and we hope to explore different types of dispersion given the functional forms that are given. We show the relationship of our distribution to the CMP when the $v = 1$ in the CMP it converges to the Alternative proposition. The proofs that our alternative method is a Probability Mass Function and that it is a special case of the CMP are in Appendix 1.

One choice that must be made for our family of distributions is choosing the functional specification in the exponent of θ. The specification of the functional form determines the shape of the distribution and thus whether it is heavy-tailed, over- or under- dispersed, or zero-inflated.

$$\frac{P(N = n - 1)}{P(N = n)} = \frac{n \cdot \theta^{f(n-1)}}{\theta^{f(n)}} \tag{3.4}$$

The change in ratios of successive probabilities can be non-linear in n which allows for differences in shapes. For example, modelling $f(n) = \sqrt{n}$ and $f(n) = n^2$ lead to two very

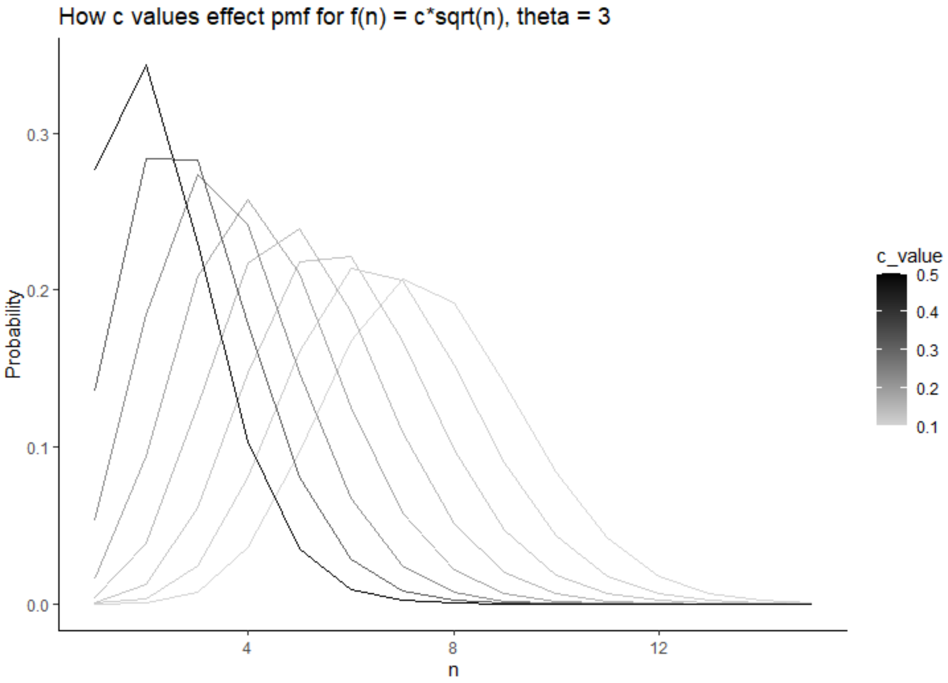


Figure 3.1: The effect of parameter c on pmf with $f(n) = c \cdot sqrt(n)$

different ratios of successive probabilities which in turn leads to two differently shaped PMFs that have tails of different thicknesses.

An extra parameter can be added to the specification to change the location of the mass of the function as shown in Figure 3.1. This can be shown in the ratios of successive probabilities equation 3.5. This is also why the specification $f(n) = c \cdot n$ is in essence a Poisson model with $\lambda = \theta^{\frac{1}{c}}$ which precludes it from further study.

$$\frac{P(N = n-1)}{P(N = n)} = \frac{n \cdot \theta^{c \cdot g(n-1)}}{\theta^{c \cdot g(n)}} = \frac{n \cdot \theta^{g(n-1)}}{\theta^{g(n)}} \tag{3.5}$$

Since closed forms for the expectation and variance don't exist for some specifications we must compute them directly by their infinite series which imposes a burden in terms of computation time. However this allows us to find the indices necessary for exploring different functions as seen in Section 4.

3.3 Normalizing Constant Convergence

Analytic Tests for Convergence

One issue of dealing with the alternative distribution that is true of the CMP is that if the normalizing constant does not converge then it becomes numerically difficult to deal with, growing faster too large to store in computer memory. This presents a limitation on the parameters and functional forms we can choose in practice. In this section we will be exploring these limitations in some depth.

For the functional forms it was necessary to search for those that would converge, thus I analyzed a select number of functional forms $f(n)$ for our distribution: n, n^k, $log(n)$, $a*n+b$, etc. In practice n is the same as picking Poisson which we are trying to compare our family of functions to. It is trivial to show that alternative specifications such as n^k are divergent and therefore cause numerical issues.

We use d'Alembert's test in order to determine convergence of the normalizing constant.

Specific Functional Forms

Using d'Alembert's test from Appendix B, we can be sure our series does not converge to anything greater than or equal to 1 if we find some functions that satisfy the following:

$$\lim_{s\to\infty} \left|\theta^{\,f(s+1)-f(s)} < (s+1)\right|$$

Adjusting terms, since $s \in \mathbb{N}$ and $\theta^{f(s)} > 0 \forall\, s$

$$\lim_{s\to\infty} \left[log(\theta) * (f(s+1) - f(s)) < log(s+1)\right]$$

Thus the difference between the terms in our sum must follow this pattern:

$$\lim_{s\to\infty} \left[f(s+1) - f(s) < log(s+1) - log(\theta)\right]$$

Specification: $f(s) = log(s+1)$

Using the rule we developed above:

$$\lim_{s\to\infty} \left[log(s+2) - log(s+1) < log(s+1) - log(\theta)\right]$$

Rearranging and cancelling terms we find

$$\lim_{s\to\infty} \left[-log(s) < -log(\theta)\right]$$

Since $\lim_{s\to\infty} -log(s) = -\infty$ we know that the above relation holds and therefore $log(s)$ is a viable function for use in our PMF.

Specification: $f(s) = \sqrt{s}$

Using the rule we developed above:

$$\lim_{s\to\infty} \left[\sqrt{s+1} - \sqrt{s} < log(s+1) - log(\theta)\right]$$

Rearranging and cancelling terms we find:

$$\lim_{s\to\infty} \left[\sqrt{s+1} - \sqrt{s} - log(s+1) < -log(\theta)\right]$$

Since $\lim_{s\to\infty} \sqrt{s+1} - \sqrt{s} = 0$, our relation reduces to the familiar:

$$\lim_{s\to\infty} \left[-log(s+1) < -log(\theta)\right]$$

We know that the above relation holds and therefore $\sqrt{s}$ is a viable function for use in our PMF.

These are two specific forms that are consider for our analysis in the realm of the specific rather than the general.

Numeric Convergence and Considerations

This numerical method of convergence is important for the computation of the values of the likelihood estimate deviance. If the sum was divergent we would be unable to compute any of the values of the probability mass function even though it may be a valid PMF.

One issue with computing a normalizing constant is the limitations of the machine to compute the factorial in the denominator. The factorial function is limited to taking an input value of 170 at the maximum due to overflow, with that being said with simulation we found that this limitation is not typically an issue for the specifications used in later sections.

The numerical issues crop up in the evaluation of the normalizing constant, as θ increases for the specifications of $\sqrt{n}$ and $log(n+1)$ the normalizing constant approaches zero at which point the log-likelihood of the normalizing constant becomes undefined. These numerical issues put a hard cap on the maximum parameter value θ of approximately 744 at which point the normalizing constant for $f(n) = log(n+1)$ is 3.02×10^{-320} and for $f(n) = \sqrt{n}$ is 4.69×10^{-319}.

Tested functional forms include $n, n^2, log(n), n^k$ the results from these tests confirmed our results from the analytic work we did: functional forms that grow too quickly diverge and make our PMF unusable. With additional parameters such as in the case of n^k the convergence of the normalizing constant was sensitive to changes in the parameter values especially for $k > 1$. For $f(n) = an + c$, there exists some parameter value d such that $an + c = dn$ and therefore specifications of this form reduce to the Poisson.

For specifications that don't pass the test in , numerical approximations may still exist for the normalizing constant. In the case of the specification $f(n) = n^{\frac{13}{12}}$ it was observed that the values in the normalizing constant series decreased toward 0 and thus the sum can be truncated and the error can be bound.

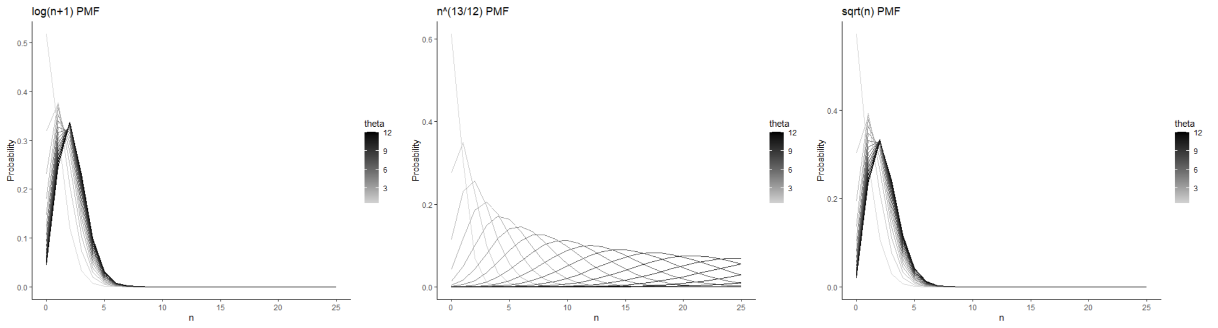
For some specifications, values of θ, and n can lead $\theta^{f(n)}$ to evaluate larger than memory, specifically if $\theta^{f(n)} > 1.798 \times 10^{308}$ then the value will be unable to be held in memory using R or any language storing numbers in the double-precision floating point format.

Chapter 4

Characterizing Our PMF

4.1 Choosing $f(n)$

Of interest when describing the shape of this PMF is the differences there are between the different specifications $f(n)$. The specifications have influence on the types of data that can be modeled by our family of distributions. For example, in 4.1 we can see that both $f(n) = sqrt(n)$ and $f(n) = log(n+1)$ have some similarities in their shapes – both have the majority of their mass nearer the origin for varying levels of theta and being quite peaked for those values. The other specification $f(n) = n^{\frac{13}{12}}$ is quite different, it has a shape that more closely resembles a Poisson however it appears to become flatter as θ increases. Additionally, a functional form $f(n) = c \cdot n$ was considered but was found to be equivalent to the Poisson as shown in the Appendix 5.



(a) $f(n) = \log(n+1)$

(b) $f(n) = n^{\frac{13}{12}}$

(c) $f(n) = \sqrt{n}$

Figure 4.1: Comparison of PMF for family of functions with specifications of the form $f(n) = \log(n+1)$, $f(n) = n^{\frac{13}{12}}$, and $f(n) = \sqrt{n}$, θ ranges from 0 to 12

$$E(N) \approx \theta + (0.07)\theta^2 + (0.0093)\theta^3 \tag{4.1}$$

One benefit of the specification $f(n) = n^{\frac{13}{12}}$ is that it has an approximation to the expected value of the PMF for any given θ given by 4.1. The approximation was taken by fitting a least squares of the expected value of the distribution for values of $\theta \leq 15$ which resulted in $\mu \leq 60$. The adjusted R-squared of this approximation is reported to be 1 and additional information can be found in the Appendix A.1.

Moments of PMF & Closed Forms

The moments of specification $f(n) = c \cdot n$ were found by using a closed form of the Moment Generating Function (MGF) which can be found in the Appendix 4. This representation allowed for computing the moments by plugging into a formula rather than by evaluation of a summation. For the specifications other than $f(n) = c \cdot n$ that we're interested in, finding closed forms proved to be impractical and thus we evaluated two summations $E[N]$ and $E[N^2]$ for different values of θ and c to derive the expected value and the variance for these specifications.

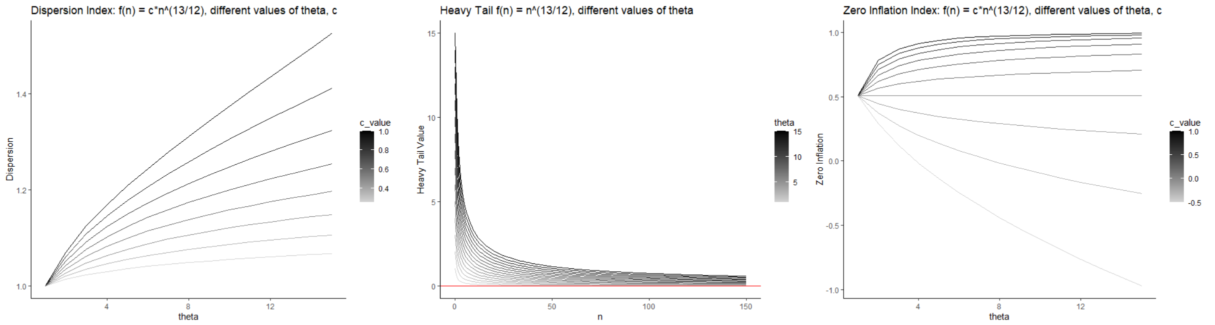
4.2 Heavy Tails, Dispersion Index, and Zero Inflation

To understand how our model behaves, I compute indices for dispersion, heavy-tails, and zero-inflation taken from (Luyts et al., 2018). The indices are given below:

$$DI(N_i) = \frac{Var(N_i)}{E(N_i)} \qquad HT(N_i) = \frac{P(N_i = n_i + 1)}{P(N_i = n_i)}, n_i \to \infty \qquad ZI(N_i) = 1 + \frac{log(P(N_i = 0)}{E(N_i)} \tag{4.2}$$

We characterize a PMF as being over-, under- or equidisperse for values of the dispersion index $DI > 1$, $DI < 1$ and $DI = 1$ respectively. The dispersion index is defined relative to the Poisson distribution such that for a given specification if our index indicates that it is characterized by over-dispersion then it is characterized by over-dispersion relative to the Poisson distribution. The zero-inflation index was introduced by (Puig and Valero, 2006) if the value of $ZI = 0$ then our distribution follows a Poisson distribution, otherwise if $ZI < 0$ or $ZI > 0$ then our distribution is zero-deflated or zero inflated. The heavy tail index indicates that the distribution is heavy-tailed if $HT \to \infty$ when $n \to \infty$ and not heavy-tailed otherwise.

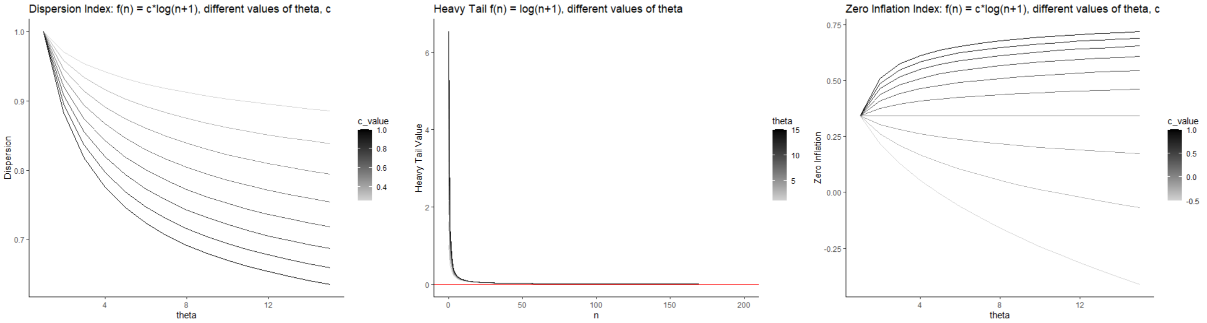
For the functional specification $f(n) = n^{\frac{13}{12}}$ we observe the indices in the 4.2. We can see from the dispersion index that for all values of θ this specification of the probability mass function exhibits over dispersion. The specification is able to exhibit both zero-deflation, for most θ values, however it approaches the same ratio of zeroes as the Poisson as θ increases. If we take a negative c value then we get zero deflation for all theta values. This specification is the only one of the three under study that exhibits heavy tails for $\theta > 1$ which lends it additional flexibility. The proof of heavy-tails for this specification can be found in the Appendix 6.



(a) Dispersion Index (b) Heavy-Tail Index (c) Zero Inflation Index

Figure 4.2: Indices for specification $f(n) = n^{\frac{13}{12}}$

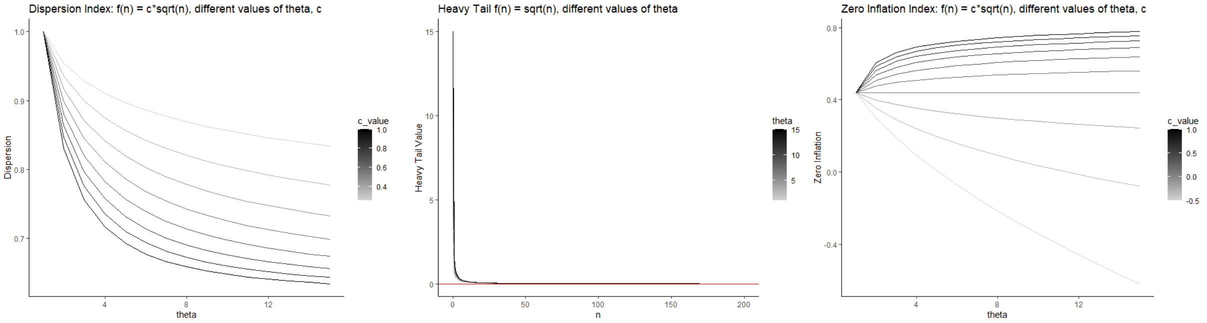
The specification $f(n) = \log(n+1)$ exhibits under- dispersion for all values of theta. This specification does not exhibit heavy tails, as can be seen in the figure below 4.3b. This specification has zero-deflation for values of θ that are close to zero. All indices can be found in 4.3.



(a) Dispersion Index (b) Heavy-Tail Index (c) Zero Inflation Index

Figure 4.3: Indices for specification $f(n) = \log(n+1)$

Figure 4.4 shows that the $f(n) = \sqrt{n}$ specification behaves quite similarly to the specification $f(n) = \log(n+1)$. The indices for dispersion, zero-inflation, and heavy-tailedness have similar bounds. This specification is under-dispersed, zero-deflated, and does not have heavy tails for the theta values under study.



(a) Dispersion Index (b) Heavy-Tail Index (c) Zero Inflation Index

Figure 4.4: Indices for specification $f(n) = \sqrt{n}$

Chapter 5

Application of Model to Simulated Data

Since our simulated data comes in a univariate flavor, we can use the log-likelihood from 5.1 to find an estimate of θ for each of our simulated data sets and compare the AIC with those of the standard models.

5.1 Maximum Likelihood Estimation

In order to generate estimates of our parameter from a dataset we can use Maximum Likelihood Estimation. The likelihood for any particular observation i can be written as:

$$L_i(\theta|n_i) = \frac{\theta_i^{f(n_i)}}{n_i!}e^{-\theta_i}\frac{1}{C(f(n_i,\theta_i)}$$

The log-likelihood for observation i can be written as:

$$logL_i(\theta|n_i) = f(n_i)log(\theta_i) - log(n_i!) - \theta_i - log(C(f(n_i,\theta_i))$$

Summing over N observations

$$logL = \sum_{i=0}^{N} f(n_i)log(\theta_i) - \sum_{i=0}^{N} log(n_i!) - \sum_{i=0}^{N} \theta_i - \sum_{i=0}^{N} log(C(f(n_i,\theta_i)) \tag{5.1}$$

In order to make our maximum likelihood estimable we use the log-likelihood and apply the Nelder-Mead optimization which is the default optimization method of the optim package in R. Other methods can be done to achieve similar results as has been shown with the quasi-Newton implementation. Using Maximum Likelihood estimation allows us to compare models by the AIC and therefore check the fit of our model in future sections. Importantly, using MLE allows us to compute standard errors of our regression coefficients which will allow us to compare with those of the standard count models.

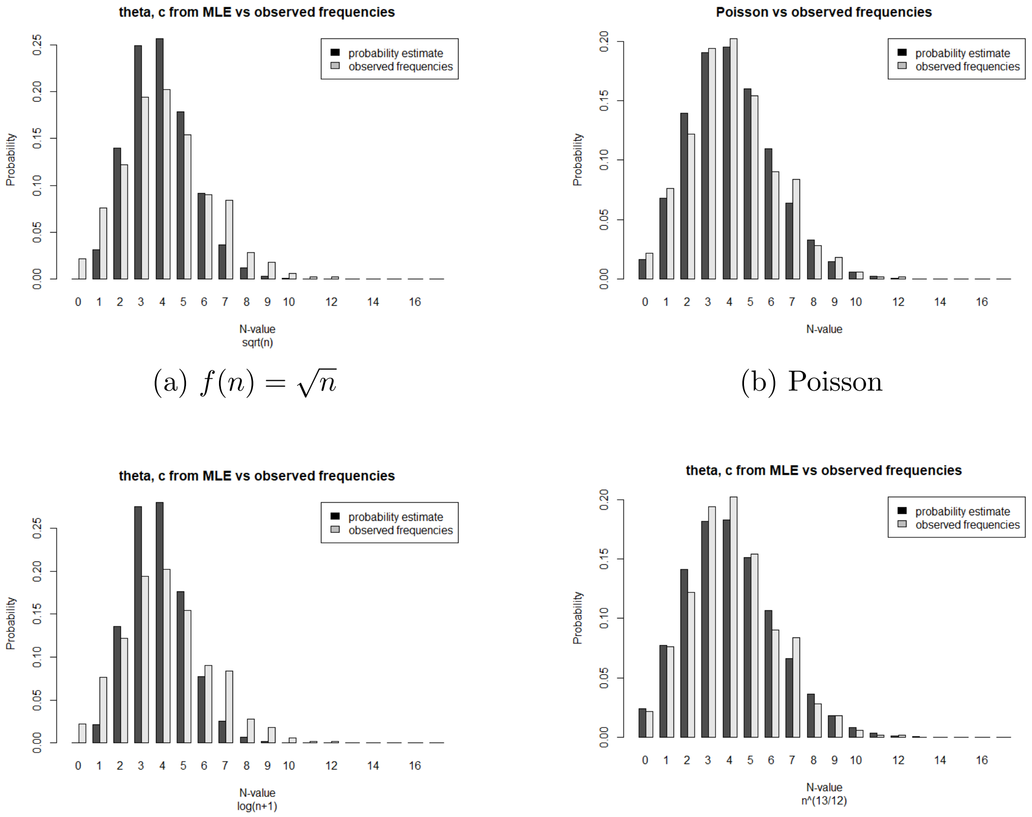
In practice, the model specification we choose is important as can be seen in Figure 4.1 the specification $f(n) = n^{\frac{13}{12}}$ has flexibility in where it is located versus the other two specifications. This flexibility is important to modelling counts with $E(n) > 5$ without running into issues of scaling the data.

5.2 Equi- dispersed

The AIC values for the simulated equi-disperse data can be found in Table 5.1. The lowest AIC is given by the Poisson model, followed by the alternative with specification $f(n) = n^{\frac{13}{12}}$. The worst performing model was the alternative with specification $f(n) = \log(n+1)$. Figures showing the PMF for the Poisson and the Alternative model for all specifications can be found in Figure 5.1.

Model	**AIC**
Poisson	**2129.57**
Negative Binomial	2130.88
CMP	2130.63
Alternative $f(n) = n^{\frac{13}{12}}$	2129.7
Alternative $f(n) = c \cdot \sqrt{n}$	2307.46
Alternative $f(n) = c \cdot log(n+1)$	2384

Table 5.1: AIC for equi-dispersed data, Poisson, NB, CMP, and Alternative



(a) $f(n) = \sqrt{n}$

(b) Poisson

(c) $f(n) = \log(n)$

(d) $f(n) = n^{\frac{13}{12}}$

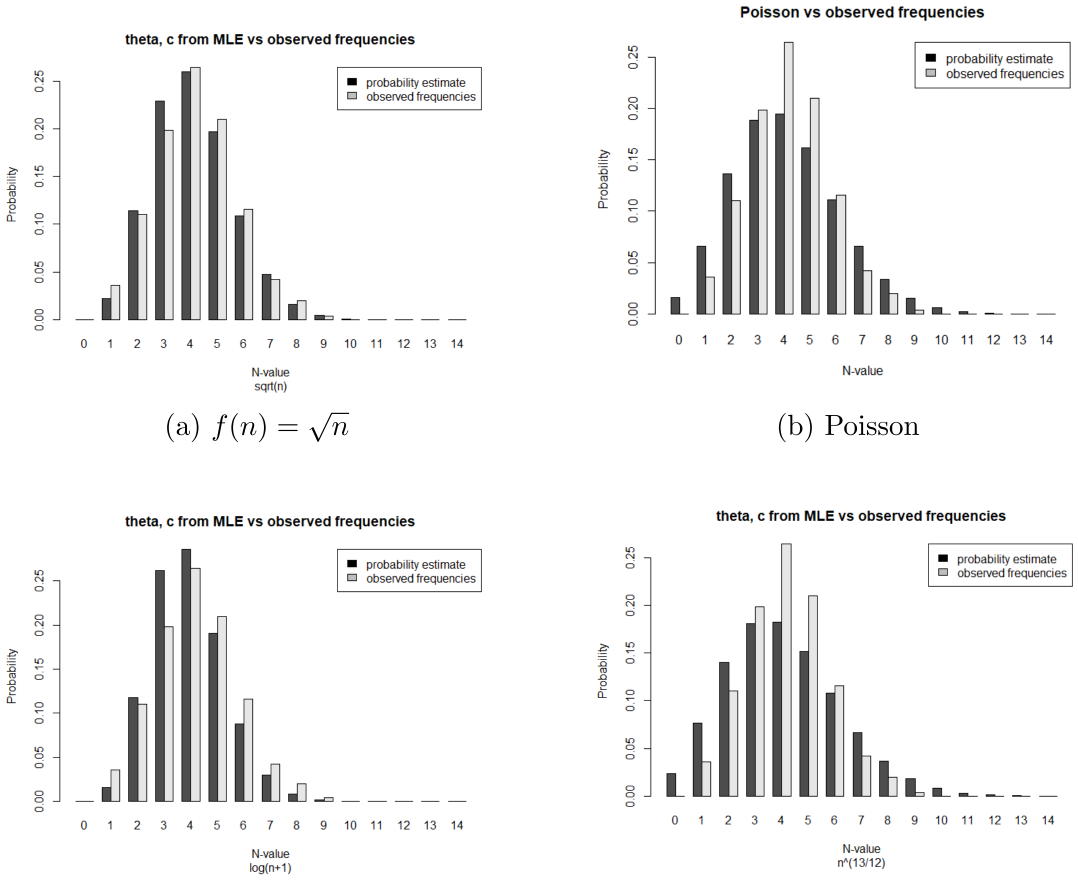
Figure 5.1: Fit comparison, different specifications of Alternative vs Poisson on equi-dispersed data

5.3 Under- dispersed

The results for model comparison are found in Table 5.2. Based on these results, for under-dispersed data, the CMP performs the best, followed by the two alternatives $f(n) = log(n+1)$ and $f(n) = \sqrt{n}$. The Negative-Binomial model performs the worst. It is clear from Figure 5.2 that the Poisson and Alternative model with $f(n) = n^{\frac{13}{12}}$ are more disperse relative to the other models.

Model	AIC
Poisson	1917.4
Negative Binomial	1972.2
CMP	**1856.47**
Alternative $f(n) = n^{\frac{13}{12}}$	1947.89
Alternative $f(n) = c \cdot \sqrt{n}$	1856.78
Alternative $f(n) = c \cdot log(n+1)$	1878.89

Table 5.2: AIC for under-dispersed data, Poisson, NB, Alternative



(a) $f(n) = \sqrt{n}$

(b) Poisson

(c) $f(n) = \log(n)$

(d) $f(n) = n^{\frac{13}{12}}$

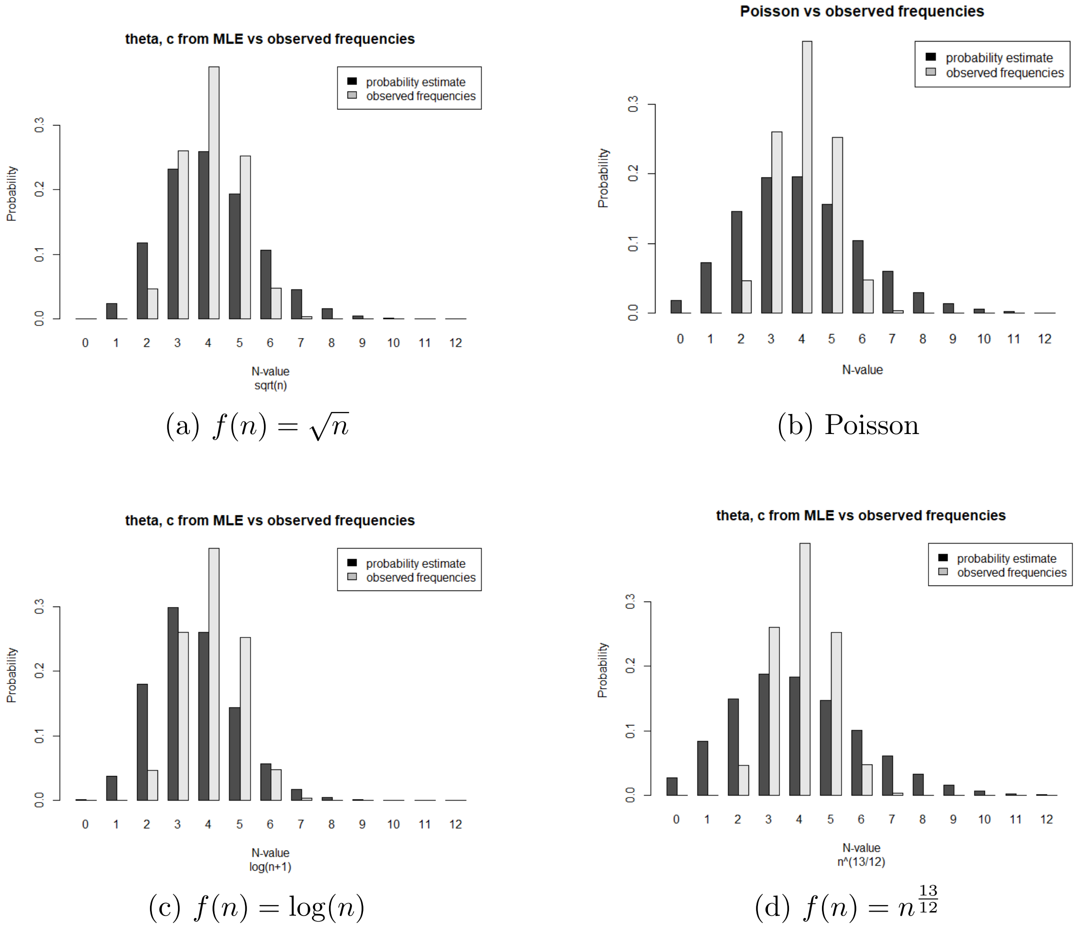
Figure 5.2: Fit comparison, different specifications of Alternative vs Poisson on under-dispersed data

5.4 Strongly Under- dispersed

The results for model comparison are found in Table 5.3, strongly under- dispersed data, the CMP performs the best, followed by the two alternatives $f(n) = log(n+1)$ and $f(n) = \sqrt{n}$. The alternative with specification $f(n) = n^{\frac{13}{12}}$ performs the worst as it is only able to model over-dispersion.

Model	AIC
Poisson	1739.26
Negative Binomial	1741.27
CMP	**1382.31**
Alternative $f(n) = n^{\frac{13}{12}}$	1790.26
Alternative $f(n) = c \cdot \sqrt{n}$	1540.208
Alternative $f(n) = c \cdot log(n+1)$	1562.895

Table 5.3: AIC for very under-dispersed data, Poisson, NB, Alternative



(a) $f(n) = \sqrt{n}$

(b) Poisson

(c) $f(n) = \log(n)$

(d) $f(n) = n^{\frac{13}{12}}$

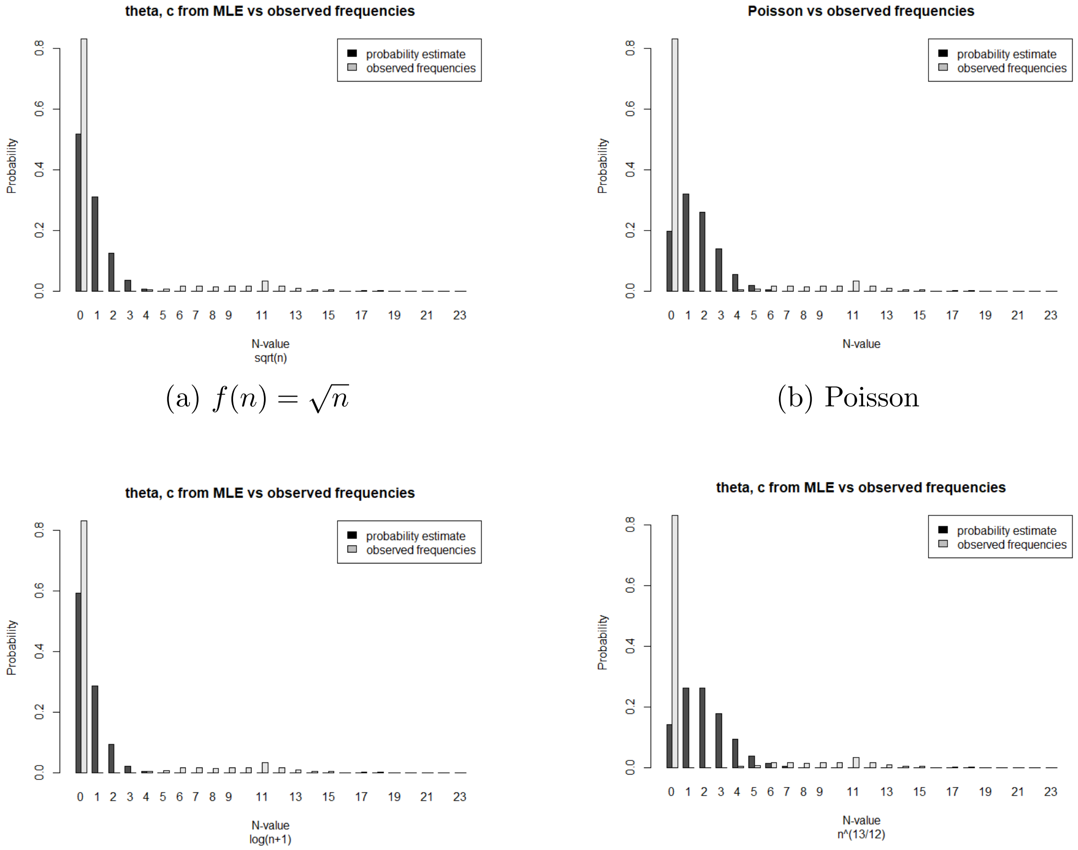
Figure 5.3: Fit comparison, different specifications of Alternative vs Poisson on very under-dispersed data

5.5 Zero-inflated

The performance of the models can be found in Figure 5.4 and Table 5.4. The CMP had an AIC of 1745.99, however the λ parameter was negative violating a key assumption of the CMP. Restricting our comparison to the models besides the CMP, the Alternative with specification $f(n) = n^{\frac{13}{12}}$ performed the best, followed by the Poisson and the Negative Binomial.

Model	**AIC**
Poisson	3297.22
Negative Binomial	3299.11
CMP	1745.99
Alternative $f(n) = n^{\frac{13}{12}}$	**3190.75**
Alternative $f(n) = c \cdot \sqrt{n}$	3377.29
Alternative $f(n) = c \cdot log(n+1)$	3388.79

Table 5.4: AIC for zero-inflated data, Poisson, NB, Alternative



(a) $f(n) = \sqrt{n}$

(b) Poisson

(c) $f(n) = \log(n)$

(d) $f(n) = n^{\frac{13}{12}}$

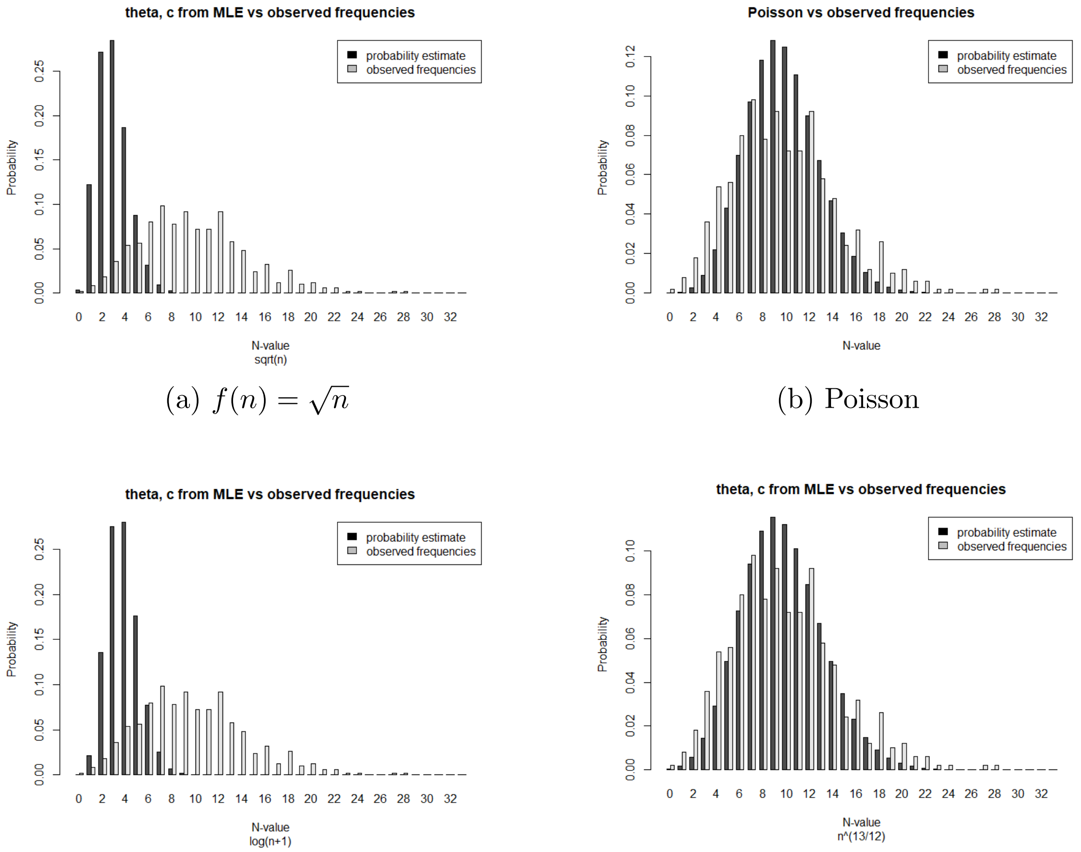
Figure 5.4: Fit comparison, different specifications of Alternative vs Poisson on zero-inflated data

5.6 Over- dispersed $\mu = 10$

For over-dispersed data with $\mu = 10$ our model performances are marked in Table 5.5. We can also observe the performance of the Poisson vs the Alternative Proposition in Figure 5.5. Our model with specification $f(n) = n^{\frac{13}{12}}$ fits the distribution better than the Poisson but under performs the CMP and Negative Binomial fits. Note however, that the $f(n) = log(n)$ and $f(n) = \sqrt{n}$ specifications have poorer fits to the distribution than expected due to numerical issues in calculating the normalizing constant as it approached zero value as explained in Section 3.3.

Model	**AIC**
Poisson	3091.1
Negative Binomial	2904.7
CMP	**2906.03**
Alternative $f(n) = c \cdot \sqrt{n}$	10318.52
Alternative $f(n) = c \cdot log(n+1)$	9449.817
Alternative $f(n) = c \cdot n^{\frac{13}{12}}$	2993.21

Table 5.5: AIC for over-dispersed non zero-inflated data, Poisson, NB, Alternative



(a) $f(n) = \sqrt{n}$

(b) Poisson

(c) $f(n) = \log(n)$

(d) $f(n) = n^{\frac{13}{12}}$

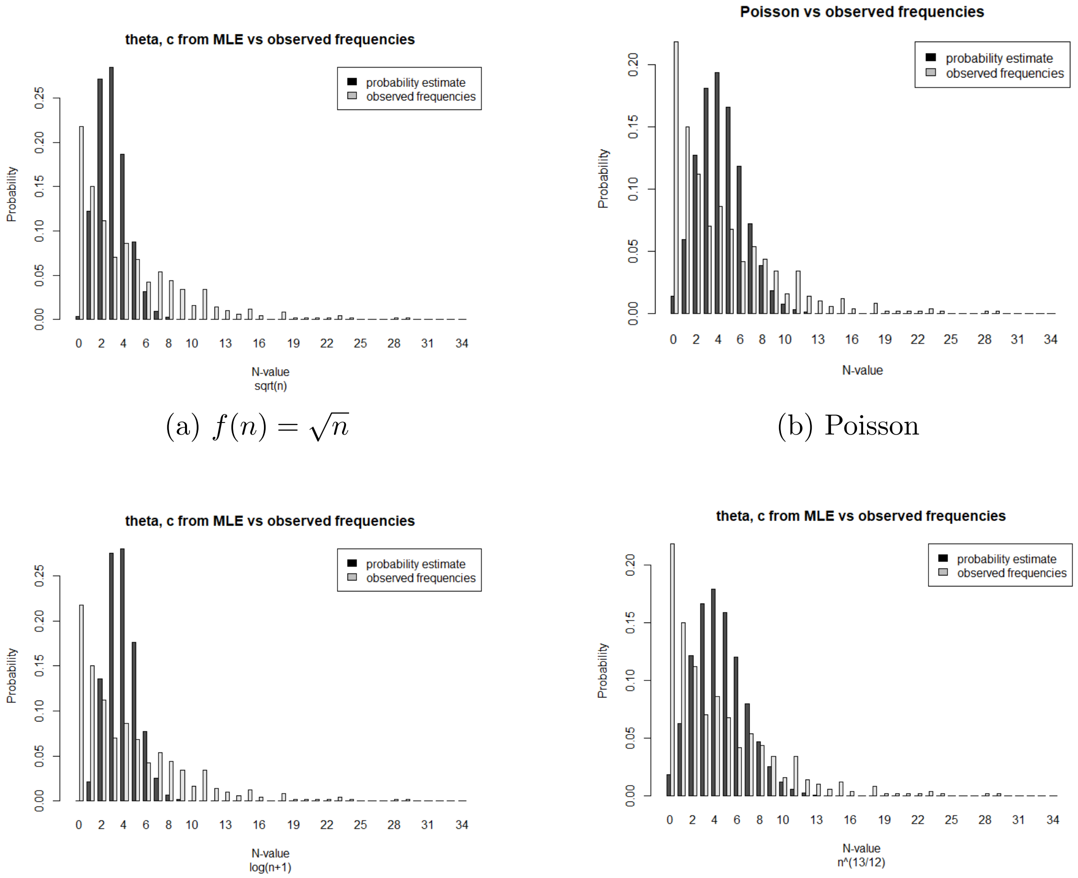
Figure 5.5: Fit comparison, different specifications of Alternative vs Poisson on over-dispersed and zero-inflated data

5.7 Over- dispersed $\mu = 4$

Figure 5.6 shows the Poisson and the $f(n) = n^{\frac{13}{12}}$ have similar shapes to their distributions, modelling some of the over-dispersion of the data. Both, the $f(n) = log(n)$ and the $f(n) = \sqrt{n}$ specifications are fit more tightly around $\mu = 4$. The results in Table 5.6 show the alternative with specification $f(n) = n^{\frac{13}{12}}$ outperforming the Poisson, but underperforming the Negative-Binomial and the CMP fits.

Model	AIC
Poisson	3689.68
Negative Binomial	**2564.47**
CMP	2566.19
Alternative $f(n) = c \cdot n^{\frac{13}{12}}$	3466.71
Alternative $f(n) = c \cdot \sqrt{n}$	7300.279
Alternative $f(n) = c \cdot log(n+1)$	5457.83

Table 5.6: AIC for over-dispersed and zero-inflated data, Poisson, NB, Alternative



(a) $f(n) = \sqrt{n}$

(b) Poisson

(c) $f(n) = \log(n)$

(d) $f(n) = n^{\frac{13}{12}}$

Figure 5.6: Fit comparison, different specifications of Alternative vs Poisson on over-dispersed and zero-inflated data

Chapter 6

Application of Model to Real Data

Using the two datasets in Section 2, we explore how well our model is able to fit to real world data and if it can recover the direction and statistical significance of the covariates. Based on the performance of the specifications $f(n) = \log(n)$ and $f(n) = \sqrt{n}$, the need to scale the data in order use them, and the lack of a good approximation for the expected values, we forgo their use in favor of $f(n) = n^{\frac{13}{12}}$ for this section.

6.1 Regression formulation

Importantly, we can apply our distribution in a regression scenario with covariates. We can do this by employing the same method outlined in Sellers and Shmueli (2010). We can model our distribution with the following parameterization:

$$log(\theta_i) = \beta_0 + \beta_1 n_{i1} + \beta_2 n_{i2} + ... + \beta_p n_{ip} \tag{6.1}$$

Our θ parameter will maximize the log-likelihood of the form:

$$logL_i(\theta|y_i) = y_i log(\theta_i) - log(y_i!) - \theta_i - log(C\left(f\left(n\right),\, \theta_i\right))$$

implying that

$$logL = \sum_{i=0}^{N} y_i log(\theta_i) - \sum_{i=0}^{N} log(y_i!) - \sum_{i=0}^{N} \theta_i - \sum_{i=0}^{N} log(C\left(f\left(n\right),\, \theta_i\right))$$

With θ that maximizes the log-likelihood, we can then use equation 4.1 in order to get coefficients that give us the direction and significance of the effect of each covariate. However even with this approximation we will be unable to interpret our values in terms of mean contrasts.

6.2 Publishing Data

For the Publishing dataset, the $f(n) = n^{\frac{13}{12}}$ parameterization was fit and compared to the standard models detailed in Section 3. For the Poisson, Negative-Binomial, and CMP-μ models the means is specified as:

$$\log(E(Y|X)) = \beta_0 + \beta_1 \cdot Experience + \beta_2 \cdot Children + \beta_3 \cdot Undergraduate\_Courses \tag{6.2}$$

For both the CMP and Alternative models the rate parameter θ is specified as:

$$\log(\theta) = \beta_0 + \beta_1 * Experience + \beta_2 * Children + \beta_3 * Undergraduate\_Courses \quad (6.3)$$

Table 6.1: Publishing Dataset: Estimated coefficients, standard errors (se), dispersion, AIC for competing models

	Number of Articles Published				
	Poisson	Neg-Bin	Alt	CMP	CMP-mu
Intercept	1.408^{***}	1.321^{***}	1.0159^{***}	7.57	1.358^{***}
	(0.097)	(0.224)	(0.0611)	(0.0202)	(0.209)
Experience	-0.002	0.0017	-0.0018	−0.0047	−0.0001
	(0.005)	(0.012)	(0.0033)	(0.0002)	(0.0112)
Children	0.138^{***}	0.153^{*}	0.1412^{***}	−0.01	0.1472^{*}
	(0.032)	(0.075)	(0.0207)	(0.0002)	(0.0694)
UndergraduateCourses	0.025^{**}	0.026	0.0252^{***}	−0.0023	0.0256
	(0.008)	(0.021)	(0.0053)	(0.0002)	(0.0189)
AIC	989.48	740.44	907.73	737.33	**736.71**
Dispersion	1	1	NA	0.0922	0.09174
Run Time (seconds)	<1	<1	106	74	<1

standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Strangely enough, the CMP-mu and the Negative Binomial models both identify the intercept and the children covariate as being statistically significant, however the CMP finds no significant covariates. Just as well the Alternative and the Poisson identify the intercept, number of children and number of Undergraduate courses taught during the period of study as important. The Alternative also agrees directionally with the Poisson, Neg-Binomial, and the CMP-mu models for the its significant covariates, however while it has a lower AIC than the Poisson, the difficulties interpretating the covariates and the larger run time make this model less useful than the Negative Binomial model.

6.3 Sleep Duration Data

For the sleep duration dataset, the $f(n) = n^{\frac{13}{12}}$ parameterization was fit and compared to the standard models detailed in Section 3. For the Poisson, Negative-Binomial, and CMP-μ models the means is specified as in equation 6.5; this parameterizaton allows for transparent interpretation of the coefficients in a standard log-linear form.

$$\log(E(Y|X)) = \log(\mu) = \beta_0 + \beta_1 \cdot BMI + \beta_2 \cdot MentalHealth + \beta_3 \cdot PhysicalHealth \quad (6.4)$$

For both the CMP and Alternative models the rate parameter θ is specified as:

$$\log(\theta) == \beta_0 + \beta_1 \cdot BMI + \beta_2 \cdot MentalHealth + \beta_3 \cdot PhysicalHealth \quad (6.5)$$

Table 6.2: Sleep Dataset: Estimated coefficients, standard errors (se), dispersion, AIC for competing models

	SleepTime				
	Poisson	Neg-Bin	Alt-$n^{\frac{13}{12}}$	CMP	CMP-mu
Intercept	2.011***	2.011***	1.587***	7.57***	2.063***
	(3.053e-03)	(3.053e-03)	(1.968e-03)	(0.0202)	(0.027)
BMI	-1.35e-3***	-1.35e-3***	-0.146e-2***	−0.0047***	−0.0034***
	(1.06e-04)	(1.06e-04)	(6.89e-5)	(0.0002)	(0.0097)
MentalHealth	-2.87e-3***	-2.87e-3***	-0.305e-2***	−0.01***	−0.0042***
	(8.96e-05)	(8.96e-05)	(5.84e-5)	(0.0002)	(0.0008)
PhysicalHealth	-6.55e-4***	-6.55e-4***	-0.635e-3***	−0.0023***	−0.0002
	(8.87e-05)	(8.87e-05)	(5.78e-5)	(0.0002)	(0.0008)
AIC	1307741	1307747	1350301	**1132176**	3436.742
Dispersion	1	1	NA	3.68	4.09
Run Time (seconds)	1	5	66	1008	4

Note: CMP-mu model was run on a smaller sample of the dataset the AIC value is non-comparable.
standard errors in parentheses
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

The AIC values for these different models in Table 6.2 show that the alternative model does the worst among all the models on this under-dispersed data set, followed by the Poisson, Negative Binomial, and finally the CMP. This is to be expected as the outcome variable Sleep Time is under dispersed as discussed in Section 2 and the Poisson, Negative-Binomial, and the Alternative are incapable of modelling under-dispersion. Note, the CMP-μ model was used on a random sample of 1000 rows in the Sleep Time dataset due to issues with memory allocation using the full dataset. From this table we can see that while the CMP has the best fit it requires much more time than the other specifications. The sign and significance for the Alternative model agree with those of the other models, however due to the under-dispersion of the response variable, the Alternative model had the highest AIC indicating a worse fit.

For both the Alternative and the CMP we can get the direction of the direction and significance for each covariate. However using the approximations given by equations 3.1 and 4.1 does not give simple interpretations of these covariates. Knowing the direction and significance of a parameter does provide some information in our case, for example, patients with higher Body Mass Index (BMI) tend to get fewer hours of sleep each night.

Chapter 7

Conclusions and Further Research Questions

This thesis explored a novel family of distributions for count data we were able to explore the characteristics of this family and show that different specifications of this family are able to model over- and under- dispersion. This alternative family of distributions has many desirable qualities that make it useful in practice such as the link between itself and the Poisson, the ease of computing the normalizing constant, and the broad family of functional forms that can be used.

The results from Table 6.2 showed that some specifications for this family of functions have normalizing constants that take less time to compute than those of the CMP model given a large sample size. The normalizing constant was shown to converge for the three specifications explored. One drawback of the normalizing constant is some functional specifications are infeasible due divergence. Additionally for two specifications we explored $f(n) = \sqrt{n}$ and $f(n) = \log(n+1)$ if θ was too large then the terms in the normalizing constant are evaluated as zero and the log-likelihood becomes undefined. This limitation means that the expected value of the PMFs given by these specifications is restricted to being less than 5.

The characteristics of the selected functional forms were determined in 4, we found that for $f(n) = \sqrt{n}$ and for $f(n) = \log(n)$ the PMFs exhibited under-dispersion and didn't have heavy-tails. For $f(n) = n^{\frac{13}{12}}$ the PMF exhibited over-dispersion. All specifications exhibited zero-deflation.

In Section5 we further tested how well the distributions fit simulated data to compare how these characteristics allow the distributions to perform relative to a set of standard models. In most cases, the CMP outperformed all competing models, however for under- dispersed data, two specifications from our family of functions performed better than the Poisson and the Negative-Binomial. For over-dispersed data, the specification $f(n) = n^{\frac{13}{12}}$ –which was the only of the three chosen specifications which could model over-dispersion was unable– to fit the data as well as the standard models, beating only the Poisson as the μ of the data increased.

In Section 6 we applied these functions to two real-world datasets, finding that while the specifications did not perform well in terms of fit, they were able to recover the sign and significance of coefficients in agreement with the standard models. Additionally, while the outcome variable for the sleep time dataset was under-dispersed, the mean put a limitation on the fit of the two specifications which are able to handle under-dispersion.

A key drawback of our model is the inability to interpret coefficients in terms of the effect on the mean of the outcome variable, further exploration of a mean-centered regression model such as that of Huang (2017) would provide some additional clarity for users of this model. Further specifications should be explored in order to determine if other unseen characteristics can be had from this family of functions.

Bibliography

Chakraborty, S. and Imoto, T. (2016). Extended conway-maxwell-poisson distribution and its properties and applications. *Journal of Statistical Distributions and Applications*, 3(1).

Choo-Wosoba, H., Levy, S. M., and Datta, S. (2015). Marginal regression models for clustered count data based on zero-inflated conway-maxwell-poisson distribution with applications. *Biometrics*, 72(2):606–618.

Huang, A. (2017). Mean-parametrized conway–maxwell–poisson regression models for dispersed counts. *Statistical Modelling*, 17(6):359–380.

Lord, D., Guikema, S. D., and Geedipally, S. R. (2008). Application of the conway–maxwell–poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis amp; Prevention*, 40(3):1123–1134.

Luyts, M., Molenberghs, G., Verbeke, G., Matthijs, K., Ribeiro Jr, E. E., Demétrio, C. G., and Hinde, J. (2018). A weibull-count approach for handling under- and overdispersed longitudinal/clustered data structures. *Statistical Modelling*, 19(5):569–589.

Maxwell, W. L. and Conway, R. W. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, 12:132–136.

Melo, M. and Alencar, A. (2020). Conway–maxwell–poisson autoregressive moving average model for equidispersed, underdispersed, and overdispersed count data. *Journal of Time Series Analysis*, 41(6):830–857.

Puig, P. and Valero, J. (2006). Count data distributions. *Journal of the American Statistical Association*, 101(473):332–340.

Sellers, K. F. and Shmueli, G. (2008). A flexible regression model for count data. *SSRN Electronic Journal*.

Sellers, K. F. and Shmueli, G. (2010). Data dispersion: Now you see it... now you don't. *SSRN Electronic Journal*.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., and Boatwright, P. (2005). A useful distribution for fitting discrete data: Revival of the conway-maxwell-poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):127–142.

Sunday, A. O. (2021). Over-dispersed count data for number of journal articles published by university lecturers.

Appendices

Appendix A

First appendix

1. Proposal with General form: $\frac{1}{e^{-\theta}\sum_{n=0}^{\infty}\frac{\theta^{f(n)}}{n!}} * \frac{\theta^{f(s)} * e^{-\theta}}{s!}$ is a PMF.

Proof. Goal: To show that sum of proposal w.r.t s is equal to 1
Simplify terms:
$\frac{1}{e^{-\theta}\sum_{n=0}^{\infty}\frac{\theta^{f(n)}}{n!}} * \frac{\theta^{f(s)} * e^{-\theta}}{s!}$
Introduce summation in terms of s
$\frac{1}{\sum_{n=0}^{\infty}\frac{\theta^{f(n)}}{n!}} * \sum_{s=0}^{\infty}\frac{\theta^{f(s)} *}{s!}$
Since each term of the summations is equal:
$\frac{1}{\sum_{n=0}^{\infty}\frac{\theta^{f(n)}}{n!}} * \sum_{s=0}^{\infty}\frac{\theta^{f(s)} *}{s!} = \frac{\frac{\theta^{f(1)}}{1!}+\frac{\theta^{f(2)}}{2!}+\frac{\theta^{f(3)}}{3!}+...}{\frac{\theta^{f(1)}}{1!}+\frac{\theta^{f(2)}}{2!}+\frac{\theta^{f(3)}}{3!}+...} = 1$

□

Proof. For any function $f(n)$ that is defined for all n
$\forall\, f(n)$: $e^{-\theta} * \theta^{f(n)} > 0$
$\forall\, n$: $n! > 0$
Implies:
$\forall\, n$: $\frac{1}{n!} > 0$
Implies:
$\forall\, n$: $e^{-\theta} * \theta^{f(n)} * \frac{1}{n!} > 0$
$C(f(n), \theta) = \sum_{n=0}^{\infty}\frac{\theta^{f(n)}}{n!} * e^{-\theta}$
Since $\forall\, n$ $\theta^{f(n)} * \frac{1}{n!} > 0$, $\forall\, n$: $\frac{1}{\sum_{n=0}^{\infty}\frac{\theta^{f(n)}}{n!}} > 0$
Implies: $C(f(n), \theta) > 0$ and $\frac{1}{C(f(n),\theta)} > 0$
Therefore: $\forall\, n$: $\frac{1}{C(f(n),\theta)} * e^{-\theta} * \theta^{f(n)} * \frac{1}{n!} > 0$

□

2. It can be shown that $f(n) = n$ is a special case for this PMF which is equivalent to that of the Poisson distribution.

$$P(N) = \frac{\theta^{\,n}}{(n)!} e^{-\theta} \ \frac{1}{C\left(f\left(n\right),\, \theta\right)}$$

Where:

$$C\left(f\left(s\right),\, \theta\right) = \sum_{s=1}^{\infty} \frac{\theta^{\,s}}{(s)!} e^{-\theta}$$

Pulling the constant out of the sum

$$C\left(f\left(s\right),\, \theta\right) = e^{-\theta} \ \sum_{s=1}^{\infty} \frac{\theta^{\,s}}{(s)!}$$

Where the term in the summation converges to e^{θ} and the normalizing constant therefore is equal to 1.

$$C\left(f\left(s\right),\, \theta\right) = e^{-\theta} \ * e^{\theta} \ = 1$$

Thus we can rewrite the proposal PMF as

$$P(N) = \frac{\theta^{\,n}}{(n)!} e^{-\theta}$$

3. Relationship of our proposal to the Conway-Maxwell Poisson Model
Conway-Maxwell Poisson:
$P(X = x) = \frac{\lambda^x}{(x!)^v} * \frac{1}{Z(\lambda,v)}$ where $Z(\lambda, v) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^v}$
Alternative Proposal:
$P(N = n) = \frac{1}{C(f(n),\theta)} * e^{-\theta} * \theta^{f(n)} * \frac{1}{n!}$ where $C(f(n), \theta) = \sum_{s=0}^{\infty} \frac{\theta^{f(s)}}{s!}$
Take $f(n) = n$ for Alternative Proposal
$P(N = n) = \frac{1}{C(f(n),\theta)} * e^{-\theta} * \theta^{n} * \frac{1}{n!}$ where $C(f(n), \theta) = \sum_{s=0}^{\infty} \frac{\theta^{s}}{s!} * e^{-\theta}$
Take $v = 1$ for Conway-Maxwell Poisson
$P(X = x) = \frac{\lambda^x}{(x!)} * \frac{1}{Z(\lambda,v)}$ where $Z(\lambda, v) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)}$
Expand CMP out:
$P(X = x) = \frac{\lambda^x}{(x!)} * \frac{1}{\sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)}}$
Expand AltProp out:
$P(N = n) = \frac{1}{\sum_{s=0}^{\infty} \frac{\theta^s}{s!} * e^{-\theta}} * e^{-\theta} * \frac{\theta^n}{n!}$
Which becomes:
$P(N = n) = \frac{1}{\sum_{s=0}^{\infty} \frac{\theta^s}{s!} * e^{-\theta}} * e^{-\theta} * \frac{\theta^n}{n!}$
$P(N = n) = \frac{\theta^n}{n!} * \frac{1}{\sum_{s=0}^{\infty} \frac{\theta^s}{s!}}$
Thus by change of parameters we find both CMP and AltProp to be equivalent when $v = 1$ and $f(n) = n$

4. Moment Generating Function $f(n) = c \cdot n$

$$M_n(t) = \sum_{n=0}^{\infty} e^{tn} e^{-\theta} \frac{\theta^{cn}}{n!} \frac{1}{C(f(n),\theta)} \tag{A.1}$$

$$e^a = \sum_{n=0}^{\infty} a^n/n! \tag{A.2}$$

thus,

$$M_n(t) = \frac{e^{-\theta} e^{\theta^c e^t}}{C(f(n),\theta)} \tag{A.3}$$

$$M'_n(t) = \frac{\theta^c e^{-\theta+\theta^c e^t + t}}{C(f(n),\theta)} \tag{A.4}$$

$$M'_n(0) = \frac{\theta^c e^{-\theta+\theta^c}}{C(f(n),\theta)} \tag{A.5}$$

$$M''_n(t) = \frac{\theta^c e^{-\theta+\theta^c e^t + t}(\theta^c e^t + 1)}{C(f(n),\theta)} \tag{A.6}$$

$$M''_n(0) = \frac{\theta^c e^{-\theta+\theta^c}(\theta^c + 1)}{C(f(n),\theta)} \tag{A.7}$$

5. Expected value and Variance of specification $f(n) = c \cdot n$

$$E(n) = M'_n(0) = \frac{\theta^c e^{-\theta+\theta^c}}{C(f(n),\theta)} \tag{A.8}$$

$$Var(N) = \frac{\theta^c e^{\theta^c - \theta}(\theta^c + 1)}{C(f(n),\theta)} - \left(\frac{\theta^c e^{-\theta+\theta^c}}{C(f(n),\theta)}\right)^2 \tag{A.9}$$

Both A.8 and A.9 are equal to θ for $c = 1$, showing that our distribution converges to the Poisson as expected.

6. Analytic proof of Heavy-Tails for $f(n) = n^{\frac{13}{12}}$

$$\frac{P(N_i = n_i + 1)}{P(N_i = n_i)} = \frac{\theta^{(n+1)^{\frac{13}{12}}}}{(n+1)!} \cdot \frac{n!}{\theta^{(n)^{\frac{13}{12}}}} \cdot \frac{C(f(n) = n^{\frac{13}{12}})}{C(f(n) = (n+1)^{\frac{13}{12}})} \tag{A.10}$$

$$\frac{P(N_i = n_i + 1)}{P(N_i = n_i)} = \frac{\theta^{(n+1)^{\frac{13}{12}} - n^{\frac{13}{12}}}}{(n+1)} \cdot \frac{C(f(s) = s^{\frac{13}{12}})}{C(f(s) = (s+1)^{\frac{13}{12}})} \tag{A.11}$$

$$\frac{P(N_i = n_i + 1)}{P(N_i = n_i)} = \frac{\theta^{(n+1)^{\frac{13}{12}} - n^{\frac{13}{12}}}}{(n+1)} \cdot C_{normalizing-ratio} \tag{A.12}$$

$$\lim_{n\to\infty} \frac{P(N_i = n_i + 1)}{P(N_i = n_i) \cdot C_{normalizing-ratio}} = \lim_{n\to\infty} \frac{\theta^{(n+1)^{\frac{13}{12}} - n^{\frac{13}{12}}}}{(n+1)} \cdot C_{normalizing-ratio} = \infty \tag{A.13}$$

Thus it has been shown that $f(n) = n^{\frac{13}{12}}$ exhibits heavy-tails by the definition given in 4.2

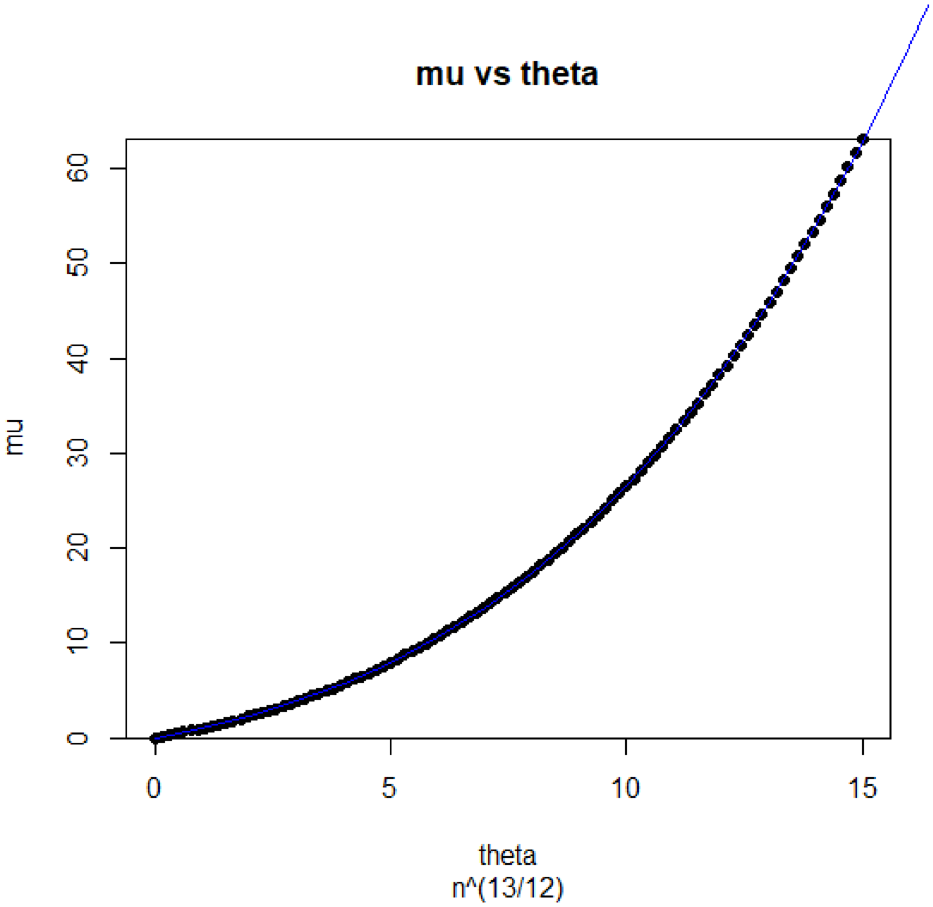


Figure A.1: Relationship between $\mu = E(N)$ and θ

	μ
β_{θ}	0.98^{***}
	(0.01)
β_{θ^2}	0.07^{***}
	(0.00)
β_{θ^3}	0.01^{***}
	(0.00)
R^2	1.00
Adj. R^2	1.00
Num. obs.	100

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.1: Estimates of coefficients for fit μ vs θ

Appendix B

Second appendix

Akaike Information Criterion (AIC) defined by the following:

$$AIC = 2k - 2log(\hat{L})$$

D'alembert's test for convergence

$$L = \lim_{n\to\infty} |\frac{a_{n+1}}{a_n}| \tag{B.1}$$

Where if $L < 1$ then the summation converges absolutely, if $L > 1$ then the summation diverges, and if $L = 1$ then the test is inconclusive. For our purposes we use the generalized form of the PMF, find limits on the parameters that can be used, and then we will explore the specific functional forms used later. Our normalizing constant takes the form:

The subsequent term is:

$$a_n = e^{-\theta}\ \frac{\theta^{\ f(s)}}{(s)!} \qquad a_{n+1} = e^{-\theta}\ \frac{\theta^{\ f(s+1})}{(s+1)!} \tag{B.2}$$

Thus given B.2 and B d'Alembert's test for our series looks like:

$$L = \lim_{s\to\infty} \left| \frac{e^{-\theta}\ \frac{\theta^{\ f(s+1)}}{(s+1)!}}{e^{-\theta}\ \frac{\theta^{\ f(s)}}{(s)!}} \right|$$

Which simplifies to:

$$L = \lim_{s\to\infty} \left| \frac{\theta^{\ f(s+1)}}{\theta^{\ f(s)}} \frac{1}{(s+1)} \right|$$

Further:

$$L = \lim_{s\to\infty} \left| \frac{\theta^{\ f(s+1)-f(s)}}{(s+1)} \right|$$

To make sure our series does not converge to anything greater than or equal to 1 we must find some family of functions that satisfy the following:

$$\lim_{s\to\infty} \left| \theta^{\ f(s+1)-f(s)} < (s+1) \right|$$

Adjusting terms, since $s \in \mathbb{N}$ and $\theta^{f(s)} > 0 \forall\, s$

$$\lim_{s \to \infty} [log(\theta) * (f(s+1) - f(s)) < log(s+1)]$$

Thus the difference between the terms in our sum must follow this pattern:

$$\lim_{s \to \infty} [f(s+1) - f(s) < log(s+1) - log(\theta)]$$

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be