2001

Using generalized linear models with a mixed random component to analyze count data

Jungah Jung

USING GENERALIZED LINEAR MODELS WITH A MIXED RANDOM COMPONENT TO ANALYZE COUNT DATA

By

Jungah Jung

B.S. Kyungpook National University, 1999

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of *Arts*

(in Mathematics)

The Graduate School

The University of Maine

August, 2001

Advisory Committee:

William **A.** Halteman, Associate Professor of Mathematics and Statistics, Advisor

Robert D. Franzosa, Professor of Mathematics and Statistics

Sundar Subramanian, Assistant Professor of Mathematics and Statistics

USING GENERALIZED LINEAR MODELS WITH A MIXED RANDOM COMPONENT TO ANALYZE COUNT DATA

By

Jungah Jung

Thesis Advisor: Dr. William A. Halteman

An Abstract of the Thesis Presented
In Partial Fulfillment of the Requirements for the
Degree of Master of Arts
(in Mathematics)
August, 2001

Many discrete response variables have counts as possible outcomes. Poisson regression has been recognized as an important tool for analyzing count data. This technique includes the simple Poisson generalized linear model and mixtures of independent Poisson models as special cases. Generalized linear models have been found useful in many statistical analysis.

Count data analyzed under such models often exhibit overdispersion. In many practical circumstances the restriction that the mean and variance are equal is not realistic. Especially, when there is overdispersion in the data, a conditional negative binomial mixed model, given some random effects, could be an attractive alternative.

This paper focuses on the data analysis using mixed Poisson regressions and mixed Negative Binomial regressions.

The motivation comes from attempts to analyze habitat use from the snow tracking data.

ACKNOWLEDGMENTS

To my Father and Mother

TABLE OF CONTENTS

ACKNOWLEDGMENTS.. ii

LIST OF TABLES.. v

LIST OF FIGURES.. .vi

Chapter

1. GENERALIZED LINEXR MODEL.. 1

1.1. Birthweight Example... 2

1.2. Horseshoe Crabs and Satellite Example................................ .4

1.3. Space Shuttle Challenger Accident Example.................................... 7

1.4. Parameter Estimation..12

1.5. Generalized Linear Model(GLM).. .13

2. MIXED POISSON REGRESSIONS USING GENERALIZED

LINEAR MODELS.. 16

2.1. Poisson Generalized Linear Model.. 17

2.2. Mixed Poisson Regression Models.. ..23

2.3. Parameter Estimation...25

2.4. Implementation Issues...31

2.4.1. Model Selection..31

2.4.2. Residual Analysis and Goodness of fit.....................................33

2.5. Seizure Frequency Data Analysis.. .34

3. THE **ANALYSIS** OF A NEW DATA SET USING GENERALIZED LINEAR MIXTURE MODELS.. .38

3.1. Data Analysis Using Mixed Poisson Regressions.............................. .40

3.2 Overdispersion... .44

3.3. Negative Binomial Model.. 47

3.4 Data Analysis Using Mixed Negative Binomial Models...................... .51

3.5. Conclusion... 64

REFERENCES.. 66

APPENDICES... .... .68

Appendix A. Mixed Poisson Regression Program................................ 69

Appendix B. Mixed Negative Binomial Regression Program......................... .75

BIOGRAPHY OF THE AUTHOR....................................... 81

LIST OF TABLES

Table 1.1. Birthweight and gestational age for 24 babies.................................. 2

Table 2.1. Data analysis with mixed Poisson regressions.................................. 35

Table 3.1. The estimation results of mixed Poisson regressions........................... 40

Table 3.2 The estimation results for overdispersion tests
for mixed Poisson regressions.. 46

Table 3.3. The results of the mixed negative binomial regressions...................... 55

Table 3.4. The results of estimation of the best appropriate model...................... 61

LIST OF FIGURES

Figure 1.1. Birthweight and gestational ages for 24 babies.. 3

Figure 1.2. **A** line using mean values of birthweight....................................... **.3**

Figure 1.3. Number of satellites by width of female crab................................. ***.5***

Figure 1.4. Smoothing of horseshoe crab counts.. ***.6***

Figure 1.5. The estimated mean number of satellites

for log and identity links.. **.7**

Figure 1.6. Temperatures versus the number of O-ring failures

with incidents(**1**).................................. 8

Figure 1.7. Temperatures versus the number of O-ring failures

with incidents(2)......................... 9

Figure 1.8. Decreasing tendency between incidents and temperatures................. **10**

Figure 2.1. Daily seizure counts... **18**

Figure 2.2. The fitted values of the Poisson generalized linear model..................20

Figure 2.3. The residual and the QQ plot for the Poisson

generalized linear model.......................... ....**.22**

Figure 2.4. The residual and the QQ plot of the two-component

mixture Poisson model.. **.36**

Figure 2.5. The fitted values of the two-component

mixture Poisson model .... **.37**

Figure 3.1. Snow tracking data .... .39

Figure 3.2. The residual plot and the ***QQ*** plot for the mixed Poisson regression model with two-component mixture .... .41

Figure 3.3. The fitted values **of** number of tracks per segment for the mixed Poisson regression model with two-component mixture.................**.42**

Figure 3.4. The residual plot and the ***QQ*** plot for the mixed Poisson regression model with three-component mixture.................................... .42

Figure 3.5. The fitted values of number of tracks per segment for the mixed Poisson regression model with three-component mixture.................43

Figure 3.6. The residual plot and the ***QQ*** plot for the mixed Poisson regression model with four-component mixture .... .43

Figure 3.7. The fitted values of number of tracks per segment for the mixed Poisson regression model with four-component mixture 44

Figure 3.8. The residual plot and the ***QQ*** plot for the negative binomial generalized linear model... .53

Figure 3.9. The fitted values of number of tracks per segment from the negative binomial generalized linear model .... .54

Figure 3.10. The residual plot and the ***QQ*** plot for the two-component mixture of negative binomial model.. .57

Figure 3.11. The fitted values of number of tracks per segment for the two-component mixture of negative binomial model...............58

Figure 3.12. The residual plot and the ***QQ*** plot for the three-component mixture of negative binomial model.. .59

Figure 3.13. The fitted values of number of tracks per segment for the three-component mixture of negative binomial model..............***.60***

Figure 3.14. The residual plot and the ***QQ*** plot for the best appropriate model in two-component mixture... ***.62***

Figure 3.15. The fitted values of number **of** tracks per segment of the best appropriate model in two-component mixture............................ ***.63***

Chapter 1

GENERALIZED LINEAR MODEL

This chapter presents the concepts for a generalized linear model. These models provide a unified theoretical and conceptual framework for many of the most commonly used statistical methods. The class of generalized linear models is a natural generalization of classical linear models. We introduce the concept of generalized linear models with three examples, sections 1.1 to **1.3.** Section **1.4** gives the parameter estimation, which is maximum likelihood estimation, and Section 1.5 discusses the definition of generalized linear models.

1.1. Birthweight Example

The data in Table1. 1 are the birthweights (g) and estimated gestational ages (weeks) of 24 babies born in a certain hospital. The data are shown in the scatter plot in Figure 1.1. The question of interest is how to model the apparent linear trend of birthweight increasing with gestational age.

Age (Weeks)	40	38	40	35	36	37	41	40
Birth Weight (g)	2968	2795	3163	2925	2625	2847	3295	3473
Age (Weeks)	40	36	40	38	42	39	40	37
Birth Weight (g)	3317	2729	2935	2754	3210	2817	3126	2539
Age (Weeks)	37	38	40	38	36	38	39	40
Birth Weight (g)	2628	3176	3421	2975	2412	2991	2875	3231

Table 1.1. Bithweight and gestational age for 24 babies.

Figure 1.1 shows one or more observations for each gestational age. In order to construct a model, we use the sample mean of birthweights for each gestational age. Figure 1.2 shows a straight line placed to approximate the upward trend of these birthweight means. Neither the mean nor the individual data points lie exactly on this line.

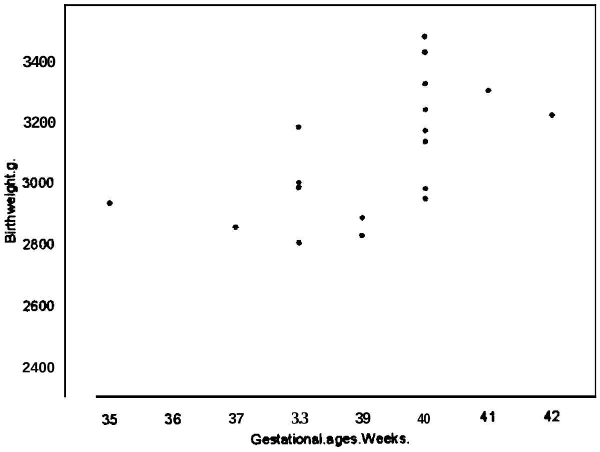


Figure 1.1. Bithweight and gestational ages for 24 babies.

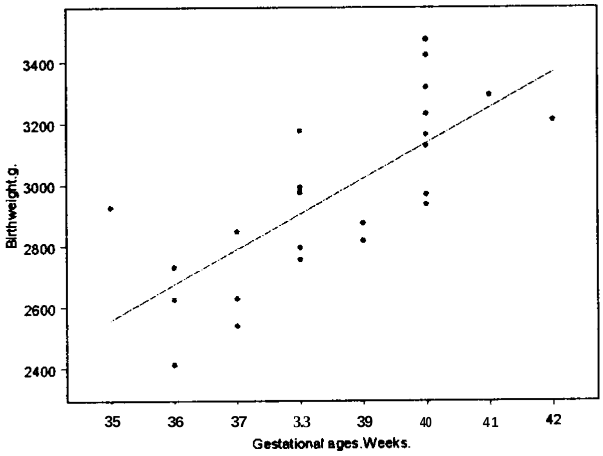


Figure 1.2. A line using mean values of birthweight.

The distance from data point to the line is denoted as ε_k for k=1,. ..,24 and assume that the ε_k's are statistically independent and all have the same probability distribution, Gaussian with mean 0 and constant variance σ^2, this is denoted by ε_k~N(0, σ^2).

A general statistical model for these data may be given by

$Y_k = \alpha + \beta x_k + \varepsilon_k$ where the response Y_k is the birthweight for the k-th baby (k=1,…,24), the parameter a represents the intercept of the line, the parameter β represents the slope or rate of increase of average birthweight with age, and the independent variable x_k is the age for the k-th baby.

We might consider birthweight to be a normal random variable, Y_k **,** because it is continuous, and E(ε_k) = 0, so we have E(Y_k) = $a + \beta x_k$, then it follows that Y_k is N(E(Y_k), σ^2).

1.2. Horseshoe Crabs and Satellite Example

These data are from a study of nesting horseshoe crabs. Each female horseshoe crab in the study had some number of male crabs, called ***satellites,*** residing nearby her. Satellite males form large groups around female horseshoe crabs. This results in a nonrandom distribution that cannot be explained by local environmental conditions or habitat selection. A. Agresti (1996) presented a data analysis of the habitat of horseshoe crabs. The study investigated factors that affect how many male crabs each female crab had.

Explanatory variables that might affect the study include the female crab's color, spine condition, weight, and carapace width. The response outcome for each female crab is her number of satellites. For now we use width alone as a predictor of the response. This variable is measured in centimeters.

Figure **1.3** plots the response counts against width. There are many different observations for each width, and the substantial variability in counts makes it difficult to discern a clear pattern. To obtain a clearer picture of overall trend, we group the female crabs into a set of width categories,(**≤ 23.25,23.25-24.25, 24.25-25.25,25.25-26.25, 26.25-27.25,27.25-28.25,28.25-29,25, > 29.25)** and calculate the sample mean number of satellites for female crabs in each width category.

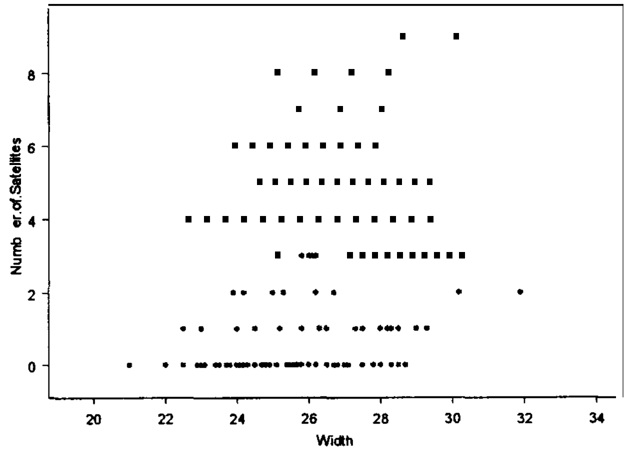


Figure **1.3.** Number of satellites by width of female crab.

Figure **1.4** plots these sample means against the sample mean width for crabs in each category. The sample means show a strong increasing trend with width. The trend seems to be approximately linear, or a smooth curve.

We discuss models for which the mean or the log of the mean is linear in width. Let μ denote the expected number of satellites for a female crab, and let x denote her width. **A** statistical model that is often used for count data **is** the Poisson distribution. Using this distribution leads to a Poisson regression model with identity link, $\mu = a, + \beta_1 x$ or the Poisson loglinear model with log link, $\log \mu = a, + \beta_2 x$.

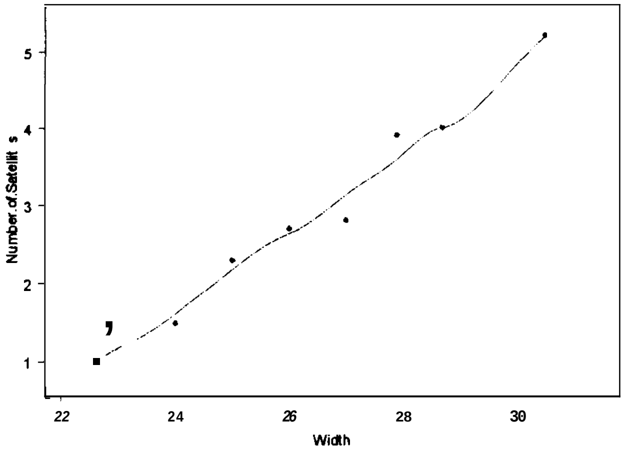
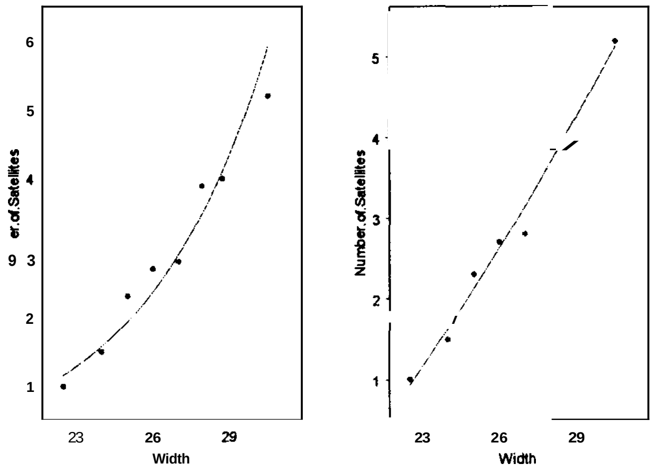


Figure **1.4.** Smoothing of horseshoe crab counts.

Figure **1.5** plots the fitted number of satellites against width, for models with log link and with identity link,



Log link **Identity** link

Figure 1.5. Estimated mean number of satellites for log and identity links.

1.3. Space Shuttle *Challenger* Accident Example

These data are from the space shuttle Challenger accident in 1986 (Dalal, Fowlkes and Hoadley 'Risk analysis of the space shuttle: Pre-Challenger Prediction of Failure', in JASA, 1989). On January 28, 1986 America was shocked by the destruction of the space shuttle Challenger, and the death **of** its seven crew members.

The investigation concluded that the accident was caused by a combustion gas leak in a joint, which resulted from the failure of a device called an ***O-ring. An*** O-ring does not work properly at low temperatures. The temperature of the O-rings at the time of the Challenger launch was **3**1°F. The data are from the **23** preaccident launches of the space shuttle and were used to predict O-ring performance under the Challenger launch conditions. There were 6 O-rings in the shuttle. On the night of January **27,** the night before the accident, there was a teleconference among the engineers. The discussion focused on the forecast of **3** 1"F temperature at launch time the next morning, and the effect of low temperatures on O-ring performance. The data used by them are plotted in Figure 1.6. Each plotted point represents a shuttle flight that experienced thermal distress on the O-rings; the **X** axis shows the temperatures at launch and the **Y** axis shows the number of O-ring failures, Based on the U-shaped configuration of points, it was concluded that there was no evidence from the historical data for a temperature effect.

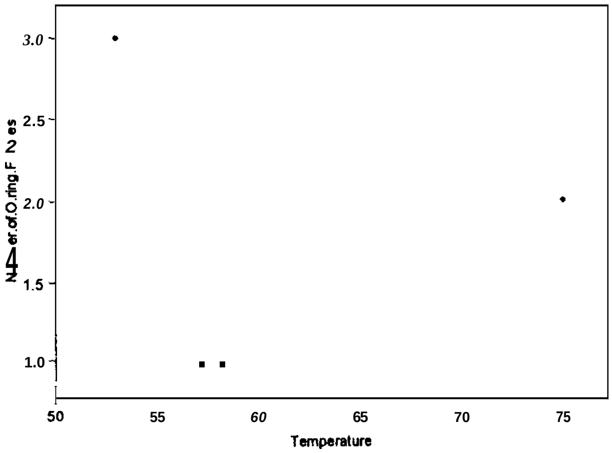


Figure 1.6. Temperatures versus the number of O-ring failures with incidents (1).

After this accident, the engineers noted that a mistake made in the analysis of these data (Figure 1.6) was that the flights with zero number of O-ring of failures were left off the plot because it was felt that these flights did not contribute any information about the temperature effect. After the accident, they reanalyzed using all of the data.

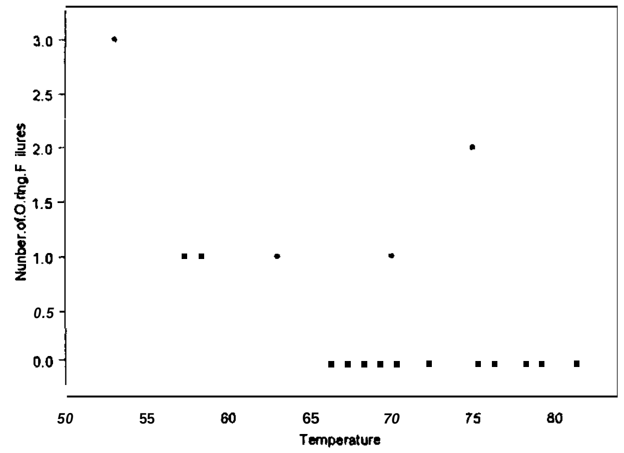


Figure 1.7. Temperatures versus the number of O-ring failures with incidents **(2).**

Figure 1.7 shows a plot of the number of O-ring failures versus temperature for 23 shuttle flights. This is the same plot as Figure 1.6 with the flights with zero incidents. This suggests that aside from one point (75,2), there is a tendency for the number of O-ring failures to decrease with increasing temperature as depicted in Figure 1.8.

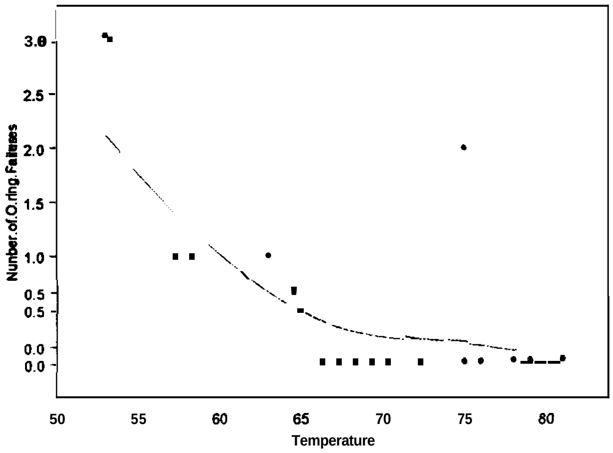


Figure **1.8.** Decreasing tendency between incidents and temperatures.

A statistical model appropriate for these data follows.

If $p(t)$ denotes the probability of a O-ring failure for a given temperature, t, $p(t)$ is a decreasing function with increasing temperature. We can consider $p(t) = a + \beta t$. There are two possible approaches to a model for these data. One is using the Binomial probability distribution and the other is using the Bernoulli probability distribution.

If X is the number **of** O-ring failures, then X has a binomial distribution with n=6 (total number of O-ring in the shuttle). The probability function for the number of failures is given by

$P(X = x) = \binom{n}{x} p(t)^{x} (1 - p(t))^{n-x}$ where $p(t) = a + \beta t$. The expected value of X is $E(X) = np(t) = n(a+\beta t)$.

This model has a weakness. There would be values t for which $p(t) < 0$ or $p(t) > 1$. Relationships between $p(t)$ and t are better modeled nonlinearly rather than linearly.

A fixed change in t may have less impact when $p(t)$ is near 0 or 1 than when $p(t)$ is near the middle of its range. In practice, nonlinear relationships between $p(t)$ and t are often monotonic, with $p(t)$ decreasing continuously as t increases. For this we turn to a logistic regression model.

The logistic regression model is $\log\left[\frac{p(t)}{1-p(t)}\right] = a + \beta t$.

An alternative approach is to look at the probability of any O-ring damage. Denote **Y** as follows:

$$Y = \begin{cases} 1 & \text{if there was one or more O-ring failures.} \\ 0 & \text{otherwise} \end{cases}$$

Y is a binary random variable with the probability $p^*(t)$ of at least one **O**-ring incident. Note that $Y=0$ iff $X=0$, and p and p^* can be compared with $p^*(f) = 1-(1-p(t))^n$ where $P(Y=1) = p^*(t)$. The logistic regression model for this approach is $\log\left[\frac{p^*(t)}{1-p^*(t)}\right] = \alpha^* + \beta^* t$. The expected value of Y is

$$E(Y) = p^*(t) = \frac{e^{\alpha^*+\beta^* t}}{1 + e^{\alpha^*+\beta^* t}} \; .$$

For each of these situations, the data are a realization of a random process, which means that we must use the probability model functions to relate the data to the parameters of the models.

1.4. Parameter Estimation

Generally the parameters of the model are estimated using the method of maximum likelihood. We describe this approach using an example below.

[Maximum Likelihood Estimation]

For the Gaussian distribution with meanp , and standard deviation σ, the probability model for one data point **is**

$$f(y;\mu,\sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{1}{2}} \exp[-(y-\mu)^2 / 2\sigma^2] \, .$$

And for the model with N data points, it is

$$f(y;\mu,\sigma) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}} \exp[-\sum_{i=1}^{N}(y_i-\mu)^2 / 2\sigma^2] \, .$$

Using a model $\mu_i = \boldsymbol{a} + \beta x_i$ to relate an explanatory variable, X , to the mean of Y , the probability model becomes

$$f(y;\alpha,\beta,\sigma)=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N}{2}}\exp[-\sum_{i=1}^{N}(y_i-(\alpha+\beta x_i))^2/2\sigma^2].$$

Estimates of α and β are found by maximizing **f** .

1.5. Generalized Linear Model (GLM)

Generalized linear models (GLMs) extend linear models to accommodate both non-linear response distributions and transformations to linearity. All generalized linear models have three components. These are the random component, the systematic component and the link function.

1.The ***random component:*** For a sample size N, let $Y_1,Y_2,...,Y_N$ denote the observations on the response variable Y . The GLM treats $(Y_1,Y_2,...,Y_N)$ as sequence of independent observations. The random component of a GLM consists of identifying the response variable **Y** and selecting a probability distribution to describe it.

In section 1.1, we assumed a Normal error regression for birthweight. In section 1.2, we used a Poisson distribution for ***satellites.*** In section 1.3, we presented two possible models. **A** model for O-ring failures using a Binomial distribution, and a model for the damage of any O-ring using a Bernoulli distribution.

2. The ***systematic component:*** The systematic component of a GLM specifies the $x_1, x_2, \ldots, x_p$ variables. These enter linearly as predictors on right hand side of the model equation. That is, the systematic component specifies the variables that play the roles of $x_1, x_2, \ldots, x_p$ in the formula $\sum_{i=l}^{P} \alpha_i x_i$. This linear combination of the covariates is called the ***linear predictor*** η given by

$$\eta = \sum_{i=1}^{P} \alpha_i x_i$$

In each example, the linear predictor involves a simple model with one covariate. The linear predictor is $a + \beta x$.

3. The ***link function***: the link function relates the linear predictor η to the expected value, μ , of a datum Y. So we write a link function as this form

$\eta = g(\mu)$.

In the birthweight example, the link function is identity link, i.e. $E(Y_k) = \mu = \alpha + \beta x_k$. For the horseshoe crabs and satellite example, two link functions were used, identity and log link, i.e. $\mu = a, + \beta_1 x$ or $\log \mu = a, + \beta_2 x$. In the Challenger accident example, the link function is the logit link, i.e. $\log\left[\frac{p(t)}{1-p(t)}\right] = \mathbf{a} + \beta t$ but this function is a function of μ . For the Binomial $\mu = E(X) = np(t)$ **was** used.

Chapter 2

MIXED POISSON REGRESSIONS USING GENERALIZED LINEAR MODELS

This chapter discusses a specific type of generalized linear model. The model uses Poisson regression and a mixture of independent Poisson regressions as special cases.

Section **2.1** introduces data and analyzes them with a Poisson generalized linear model. Section **2.2** discusses mixed Poisson regression models. Section **2.3** and section **2.4** describe the parameter estimation, model selection, residual analysis, and goodness-of-fit. Section 2.5 shows data analysis with mixed Poisson regressions.

2.1. Poisson Generalized Linear Model

In this section we analyze data with a generalized linear model. The data are from a clinical trial carried out at British Columbia's Children's Hospital which investigated the effect of intravenous gammaglobulin (IVIG) on suppression of epileptic seizures (Wang, Puterman, Cockburn and Le, **1996).** Subjects were randomized into two groups. After 28 days of baseline observation the treatment group received monthly infusions of M G . The primary end point of the trial was the daily seizure frequency. The principal data source was a daily seizure diary that contained the number of hours of parental observation and the number of seizures of each type during the observation period. The number of seizures depends on how long parents observed their children during the trial. The more time they see their children, the larger number of seizures they can count.

We use Poisson regression to analyze the seizure counts from a single subject receiving IVIG. The data extracted from the seizure dairy were the daily counts, y_i and the hours of parental observation, t_i, for the i-th day (Figure **2.1).** *As* covariates we use treatment (x_{i1}), trend (x_{i2}) and treatment-trend interaction (x_{i3}), where

$$x_{i1} = \begin{cases} 1 & \text{if there is a treatment } (i > 28) \\ 0 & \text{otherwise, } (i \leq 28) \end{cases} \tag{2.1}$$

$$x_{i2} = \log(i) \tag{2.2}$$

$$x_{i3} = x_{i1}\, x_{i2}\,. \tag{2.3}$$

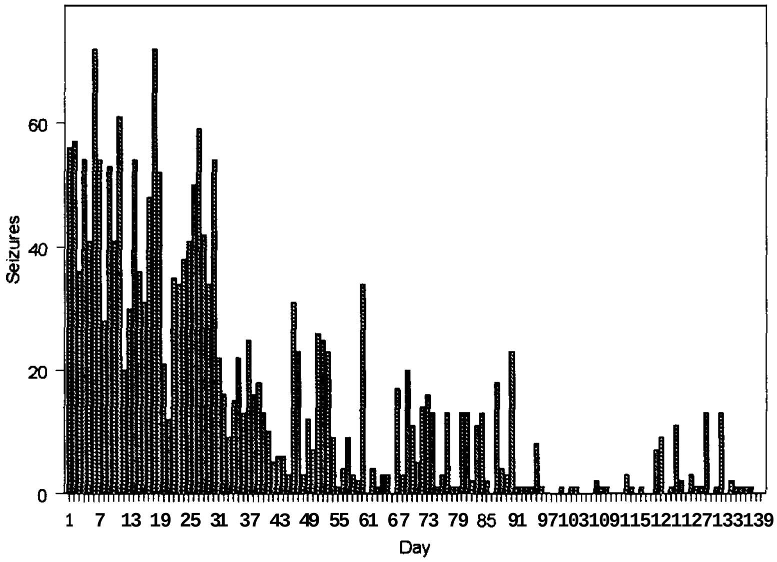


Figure 2.1. Daily seizure counts.

We have a generalized linear model using Poisson regression with covariates (2.1), (2.2), (2.3) and a log link function. We apply the generalized linear model assuming that:

(1) Each daily observed seizure count, y_i , is associated with time exposure (observation hours), t_i , and covariates $\underline{x_i} = (x_{i1}, x_{i2}, x_{i3})$;

(2) Daily seizure counts are independent and follow a generalized linear model with means equal to the product of observation time (t_i) and the Poisson rate (number of seizures per hour). Rates are specified by the log link function, which are $\log\lambda(\underline{x_i},\underline{\alpha}) = \exp(\alpha_0 + \alpha_1 x_{il} + \alpha_2 x_{i2} + \alpha_3 x_{i3})$ where i=1,...,140.

Recall that the Poisson density function of Y_i is

$$f(y_i \mid \lambda_i) \equiv p(y_i \mid \lambda_i) = \frac{1}{y_i!}\lambda_i^{y_i}\exp(-\lambda_i).$$

The mean is $\lambda_i \equiv t_i\lambda(\underline{x_i},\underline{\alpha}) \equiv t_i \exp(\underline{\alpha}' \cdot \underline{x}_i)$

$$So\ f(y_i \mid \underline{x_i}, t_i, \underline{\alpha}) = \frac{1}{Yi!}(t_i \exp(\underline{\alpha}' \cdot \underline{x}_i))^{y_i} \exp(-t_i \exp(\underline{\alpha}' \cdot \underline{x}_i)).$$

The maximum likelihood parameter estimates obtained for this model are $\hat{\alpha}$ = (-2.9484, -2.1525, -1.8768, 0.6551).

Figure 2.2 compares the model fit to the data. The right-hand side shows the fitted values of this Poisson regression and the left-hand side shows the original data. The two plots do not look similar, and we may conclude the Poisson generalized model does not fit the data well.

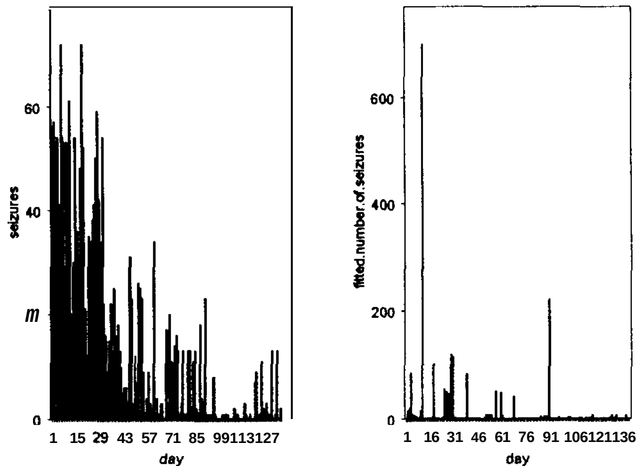


Figure 2.2. The fitted values of the Poisson generalized linear model.

Since the data are being modeled, each response value, y, , is not exactly equal to the model's parameters (called a fitted value and denoted $\hat{\mu}_i$). The question then arises of how discrepant they are, because while a small discrepancy may be tolerable a large discrepancy is not. **A** measure of discrepancy is called ***goodness*** *of fit.* It may be formed in various ways, but we will consider only the Pearson residual. The Pearson residual (residual) has this form: $\dfrac{y_i - \hat{\mu}_i}{\sqrt{\operatorname{var}(\hat{\mu}_i)}}$

There are three tools to access goodness of fit; (1) **A** scattered plot of the residuals versus the fitted values. (2) **A** Normal ***QQ*** plot of the residuals.

(3) The calculation of a goodness of fit statistic.

If data are fit well, these residuals are randomly scattered around 0 on the Y-axis and QQ plot shows a straight line.

The goodness of fit statistic is computed by summating the squares of Pearson residuals. If the data is fit, this result will follow a probability distribution called Chi-squared (Mood, A. M, Grabill, F.A., Boes, D.C. **1973).** Small values relative to the parameter of Chi-squared distribution indicate a goodness of fit to the data.

Figure **2.3** presents both residual plot (top) and **QQ** plot (bottom). The residuals are not scattered randomly around 0 on the Y-axis nor is the ***QQ*** plot straight. This indicates these seizure data are not well fit by a Poisson regression model.

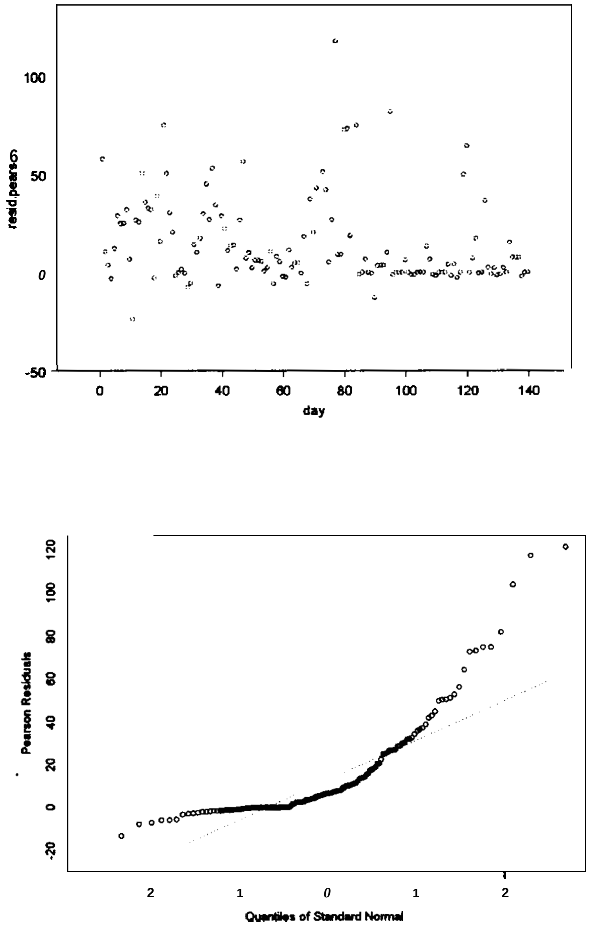


Figure **2.3.** The Residual plot and the QQ plot for the Poisson generalized linear model.

For this model, the goodness of fit statistic is **8162.072** on **136** degrees of freedom. This value exceeds the upper 95% critical point of the χ^2_{136} distribution, suggesting that there is an evidence of a lack of fit.

Now, we proceed to analyze this data with a mixed model.

2.2. Mixed Poisson Regression Models

Wang, Puterman, Cockburn and Le (1996) presented a mixed Poisson regression model analysis of the seizure data. This is their approach.

Let the random variable Y_i denote the z-th response variable and let $\{(y_i, t_i, \underline{x}_i), i=1,\ldots,n\}$ denote observations where y_i is the observed value of Y_i, t_i a non-negative value representing the time or extent of exposure and $\underline{x}_i$ a k-dimensional covariate vector corresponding to the linear predictor part of the model. Usually the first element of $\underline{x}_i$ is a 1 corresponding to an intercept.

Our mixed Poisson regression model assumes:

(1) The unobserved mixing process can occupy any one of c states where c is finite and unknown;

(2) For each observed count, y_i, there is an unobserved random variable, Λ_i, representing the component at which y_i is generated. Further, the (Y_i, Λ_i) are pairwise independent;

(3) Λ_i follows a discrete distribution with c points of support, and $\Pr(\Lambda_i = j) = p_j$ where $\sum_{j=1}^{c} p_j = 1$;

(4) Conditional on Λ_i=j, Y_i follows a Poisson distribution which we denote by

$$Y_i \sim f_j(y_i \mid \underline{x}_i, t_i, \underline{\alpha}_j) \equiv P_0(y_i \mid \underline{x}_i, t_i, \underline{\alpha}_j) = \frac{1}{y_i!}\lambda_{ij}^{y_i}\exp(-\lambda_{ij}) \qquad (2.4)$$

where

$$\lambda_{ij} \equiv t_i\lambda_j(\underline{x}_i, \underline{\alpha}_j) \equiv t_i \exp(\underline{\alpha}_j' \underline{x}_i), \qquad \text{for } j = 1,\ldots,c,$$

with $\underline{a} \equiv (\underline{\alpha}_1,\ldots,\underline{\alpha}_c)'$ denoting unknown parameters, and $\underline{\alpha}_j = (\underline{\alpha_{j1}},\ldots,\underline{\alpha_{jk}})'$,

j $=1,\ldots,$c.

Note that we could also choose another positive link function.

The above assumptions define the unconditional distribution of the observations, y_i , as a finite Poisson mixture in which the mixing probabilities, $\boldsymbol{p}$, , are constant and the component distributions are Poisson distributions with means, λ_{ij} , which is determined by the exposure, t_i , and by the Poisson rate, $\lambda_j(\underline{x}_i, \underline{\alpha}_j)$, which is related to covariates $\underline{x}_i$ through a log link function.

Under the above assumptions the probability function of Y_i satisfies

$$f(y_i \mid \underline{x}_i, t_i, \underline{\alpha}, \underline{p}) = \sum_{j=1}^{c} p_j P_0(y_i \mid \lambda_{ij}) \qquad (2.5)$$

where $P_0(y_i \mid A_j)$ is given by (4), and $p = (p_1,\ldots,p_c)'$ is an unknown parameter vector.

We may equivalently view the model as arising from the following sampling scheme. Observations are independent. For observation i, component j is chosen according to a multinomial distribution with probability p_j. Subsequently, y_i is generated from a Poisson distribution with mean λ_{ij}. When the data are observed, the source (i.e. component) of the observation is unknown.

For the above model, the unconditional mean and variance of Y_i are, respectively,

$$E(Y_i) = E(E(Y_i \mid \Lambda_i)) = t_i \sum_{j=1}^{c} p_j \lambda_{ij} \qquad \text{and}$$

$$Var(Y_i) = E(Var(Y_i \mid \Lambda_i)) + Var(E(Y_i \mid \Lambda_i))$$

$$= t_i \sum_{j=1}^{c} p_j \lambda_{ij} + t_i^2 \left\{ \sum_{j=1}^{c} p_j \lambda_{ij}^2 - \left\{ \sum_{j=1}^{c} p_j \lambda_{ij} \right\}^2 \right\}$$

Obviously, $Var(E(Y_i \mid \Lambda_i)) = 0$ if and only if $\lambda_{il} = \lambda_{i2} = \ldots = \lambda_{ic}$.

2.3. Parameter Estimation

For a fixed number of components c, we obtain maximum likelihood estimates of the parameters in the above model using the EM algorithm as first suggested by Dempster et al. **(1977).** They described a general method for computing maximum likelihood estimates when observations are missing. For the mixture model estimation, we implement the EM algorithm by treating unobservable component membership of the observations as missing data.

We discuss the choice of the number of components later.

Suppose that $(Y,X,T) \equiv \{(y_i,t_i,x_i)$, i=1,...,n} are the observed data generated by the above mixture model. Let $(Y,Z,X,T) \equiv \{(y_i,z_i,t_i,x_i)$, i=1,...,n} denote the complete data for the mixture, where the observed quantity $z_i = (z_{i1},...,z_{ic})'$ satisfies

$$z_{ij} = \begin{cases} 1 & \text{if } \Lambda_i = j \\ 0 & \text{otherwise.} \end{cases}$$

The log of the probability function Y for the complete data is

$$f(Y,Z,\alpha,p,X,T) = \sum_{i=1}^{n}\sum_{j=1}^{c} z_{ij}\log(p_j) + \sum_{i=1}^{n}\sum_{j=1}^{c} z_{ij}\log P_0(y_i \mid \lambda_{ij}).$$

The EM approach finds the maximum likelihood estimates using an iterative procedure consisting of two steps: an E-step and an M-step. The E-step replaces the missing data by its expectation conditional on the observed data. The M-step finds the parameter estimates that maximize the expected log probability function for the complete data, conditional on the expected values of the missing data. **In** our case, this procedure can be stated as follows.

E-step. Given $\alpha^{(0)}$ and $p^{(0)}$, replace the missing data Z by its expectation conditioned on these initial values of the parameters and the observed data. $(Y, \underline{X}, T)$. In this case, the conditional expectation of the j-th component of $\underline{z}_i$ equals the probability that the observation y_i was generated by the j-th component of the mixture distribution, conditional on the parameters, the data, and the covariates. Denote the conditional expectation of the j-th component of $\underline{z}_i$ by $\tilde{z}_{i,j}(\underline{\alpha}^{(0)}, \underline{p}^{(0)})$. Then

$$\tilde{z}_{i,j}(\underline{\alpha}^{(0)}, \underline{p}^{(0)}) = \frac{p_j f_j(y_i \mid \underline{x}_i, t_i, \underline{\alpha}_j^{(0)})}{\sum_{l=1}^{c} p_l f_l(y_i \mid \underline{x}_i, t_i, \underline{\alpha}_l^{(0)})}, \qquad \text{j=1,..,c} \qquad (2.6)$$

M-step. Given conditional probabilities $\{ \tilde{z}_{i,j}(\underline{\alpha}^{(0)}, \underline{p}^{(0)}) = (\tilde{z}_{i,1}, \ldots, \tilde{z}_{i,c})'$; i=1,…,n}, obtain estimates of the parameters by maximizing, with respect to $\underline{\alpha}$ and $\underline{p}$,

$$Q(\alpha, p \mid \underline{\alpha}^{(0)}, \underline{p}^{(0)}) = E\{f(Y, Z, \alpha, p, X, T) \mid Y, \alpha^{(0)}, p^{(0)}, X, T\}$$

$$= Q_1 + Q_2$$

where $Q_1 = \sum_{i=1}^{n} \sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) \log(p_j)$ and

$$Q_2 = \sum_{i=1}^{n} \sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) \log P_0(y_i \mid \lambda_{ij}) .$$

Then

$$Q, = \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{,j}(\alpha^{(0)}, p^{(0)})\log(p_j)$$

$$= \sum_{i=1}^{n}[\tilde{z}_{i,1}(\alpha^{(0)}, p^{(0)})\, \boldsymbol{log}\, p_1 + \tilde{z}_{,2}(\alpha^{(0)}, p^{(0)})\, \boldsymbol{log}\, p_2 + \cdots + z_{i,c}(\alpha^{(0)}, p^{(0)})\log(p_c)]$$

$$= \sum_{i=1}^{n}[\tilde{z}_{i,1}(\alpha^{(0)}, p^{(0)})\, \boldsymbol{log}\, p_1 + \tilde{z}_{1,2}(\alpha^{(0)}, p^{(0)})\, \boldsymbol{log}\, p_2 + \cdots + z_{i,c}(\alpha^{(0)}, p^{(0)})\log(1 - \sum_{j=1}^{c-1} p_j)]$$

since $\sum_{j=1}^{c} p_j = 1$. i.e. $p_c = 1 - \sum_{j=1}^{c-1} p_j$.

The estimated parameters, $\hat{\alpha}$ and $\hat{p}$ **,** satisfy the following M- step equation:

$$\frac{\partial Q_1}{\partial p_j}|_{\hat{p}_j} = \sum_{i=1}^{n}\left| \frac{\tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)})}{\hat{p}_j} - \frac{\tilde{z}_{i,c}(\alpha^{(0)}, p^{(0)})}{1 - \sum_{j=1}^{c-1}\hat{p}_j} \right|$$

$$= \sum_{i=1}^{n}\left(\frac{\tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)})}{\hat{p}_j} - \frac{\tilde{z}_{i,c}(\alpha^{(0)}, p^{(0)})}{\hat{p}_c} \right)$$

So we have

$$\frac{}{\partial p_j} = \sum_{i=1}\left(\frac{\tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)})}{\hat{p}_j} - \frac{\tilde{z}_{i,c}(\alpha^{(0)}, p^{(0)})}{\hat{p}_c} \right) = 0, \quad \text{for } j=1,\ldots,c\text{-}1. \qquad \textbf{\textit{(2.7)}}$$

The above result yields c-1 simultaneous equations with c-1 unknowns (the $\hat{p}_j$).

Solving the system (2.7) yields

$$\hat{p}_j = \frac{1}{n}\sum_{i=1}^{n} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) \qquad \text{for j} = 1,\ldots,\text{c-1}. \tag{2.8}$$

and

$$Q_2 = \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) \log P_0(y_i \mid \lambda_{ij})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) \log \frac{1}{y_i!} \lambda_{ij}^{\ y_i} \exp(-\lambda_{ij})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)})[\log \frac{1}{y_i!} + y_i \log \lambda_{ij} - \lambda_{ij}]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)})[-\log y_i! + y_i(\log t_i + \mathrm{a}'x_,) - t_i \exp(\alpha_j' x_i)]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{c} - \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) \log y_i! + \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) y_i \log t_i$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) \alpha_j' x_i y_i \text{ - } \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)}) t_i \exp(\alpha_j' x_i) .$$

Thus,

$$\frac{\partial Q_2}{\partial \alpha}\Big|_{\hat{\alpha}} = \sum_{i=1}^{n}\sum_{j=1}^{c} \tilde{z}_{i,j} \frac{\partial}{\partial \alpha} \log P_0(y_i \mid \lambda_{ij})$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{c}\tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)})x_i y_i - \sum_{i=1}^{n}\sum_{j=1}^{c}\tilde{z}_{i,j}(\alpha^{(0)}, p^{(0)})t_i x_i \exp(\alpha_j' x_i) = 0. \qquad (2.9)$$

Since a closed form solution of equation **(2.9)** is unavailable, we use an iterative method such as Newton's method to obtain the estimates. Hence we implement the E- and M- steps in the following way to obtain parameter estimates.

Step 0: Specify starting values $\alpha^{(0)} = (\alpha_1^{(0)},\ldots,\alpha_c^{(0)})$ and $P^{(0)} = (p_1^{(0)},\ldots,p_c^{(0)})$,

and two tolerances ε_0 and ε ;

Step 1: (E-step) Compute $\tilde{z}_i = (\tilde{z}_{i,1},\ldots,\tilde{z}_{i,c})'$ (i= 1,...,n) ,using **(2.6).**

Step **2:** (M-step)

(a) Find values of $\hat{p}$ using **(2.9);**

(b) Find values of $\hat{\alpha}$ by solving **(2.8)** using Newton's method;

Step **3:** **(a)** If at least one of the following conditions **is** true, set $\underline{\alpha}^{(0)} = \hat{\alpha}$ and $\underline{p}^{(0)} = \hat{p}$ **,** and go to Step 1; otherwise, go to (b).

(1) $\left\|\hat{\alpha} - \alpha^{(0)}\right\| \equiv \sum_{j=1}^{c}\sum_{l=1}^{k}\left|\hat{\alpha}_{j,l} - \alpha_{j,l}^{(0)}\right| \geq \varepsilon$;

(2) $\left\|\hat{p} - p^{(0)}\right\| \equiv \sum_{j=1}^{c}\left|\hat{p}_j - p_j^{(0)}\right| \geq \varepsilon$;

(3) $\left|f(Y,\hat{\alpha},\hat{p},X,T) - f(Y,\alpha^{(0)},p^{(0)},X,T)\right| \geq \varepsilon_0$.

(b) Maximize the observed probability function $f(Y,\hat{\alpha},\hat{p},X,T)$ using an iterative approach with $\hat{\alpha}$ and $\hat{p}$ as initial values. Then stop.

When the number of components, c, is known, the asymptotic normality of $\sqrt{n}((\hat{\alpha},\hat{p}) - (ap))$ can be shown under standard regularity conditions (Lehmann,1983). To approximate the standard error, we compute $\hat{\sigma}(\hat{\alpha}_{j,l})$ and $\hat{\sigma}(\hat{p}_j)$ from the diagonal elements of the inverse of the (c*k + (c-1))-dimensional observed information matrix with c fixed at $\hat{c}$ which is defined as

$$\frac{\partial^2 f(Y,\alpha,p,X,T)}{\partial(\alpha,p)^2} = \begin{pmatrix} \frac{\partial^2 f}{\partial\alpha^2} & \frac{\partial^2 f}{\partial\alpha\partial p} \\ \frac{\partial^2 f}{\partial\alpha\partial p} & \frac{\partial^2 f}{\partial p^2} \end{pmatrix}$$

2.4. Implementation Issues

2.4.1. Model Selection

To apply the mixed Poisson regression model we must know the number of components, c, and we require a method for inference about the model parameters.

When c is known, inference for the parameters can be carried out using by likelihood ratio tests. In practice, this is rarely the case. When c is unknown, we use the following approach for model selection. This is based on maximum likelihood estimation.

Two widely used model selection criteria are Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). McLachlan and Basford(1998) and Leroux and Puterman(1992) discussed the use of AIC and BIC to determine the number of components in a finite mixture model without covariates. Leroux (1992) established consistency of parameter estimates under the following penalized likelihood criteria.

AIC: Choose the model for which $\hat{f}_c(X) - a_c(X)$ is largest;

BIC: Choose the model for which $\hat{f}_c(X) - (\frac{1}{2})(\log(n))a_c(X)$ is largest.

where $\hat{f}_c(X)$ is the probability function of the mixture with c components and covariate **X,** $a_c(X)$ = c*k + (c-1) where k is the dimension of α_j and n is the total number of observations.

A good model is one that fits the data very well. By including enough parameters in the model we can make the fit as close as we please, and indeed by having as many parameters as observations we can make the fit perfect.

However, simplicity, represented by parsimony of parameters, is also a desirable feature of a model; we do not include parameters that we do not need. Not only does a parsimonious model enable the analyst to think about the data, but one that is substantially correct gives better predictions than one that includes unnecessary parameters.

The model which maximizes AIC and BIC, also minimizes $a,(X)$where $a,(X)$ is a function of c, the number of components. So we can choose the model which maximizes the log-likelihood with the smallest number of components.

Using the BIC (AIC), our selection approach consists of two stages. At the first stage, we determine c to maximize BIC (AIC) values for the saturated mixture models that contain all possible covariates. At the second stage, our goal is choosing an appropriate model to fit the data, by finding the combination of covariates of a model that maximizes BIC (AIC) values for the selected c-component mixture model.

2.4.2. Residual Analysis and Goodness of fit

Once a mixed Poisson regression model has been fit to a set of observations, it is essential to check the quality of fit. For this purpose, we consider the

Pearson residuals for mixed Poisson regression models. The ***Pearson residual*** satisfies

$$r = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}},$$

where

$$\hat{\mu}_i = t_i \sum_{j=1}^{c} p_j \hat{\lambda}_{ij}$$

$$V(\hat{\mu}_i) = t_i \sum_{j=1}^{c} p_j \hat{\lambda}_{ij} + t_i^{\,2} \left\{ \sum_{j=1}^{c} p_j \hat{\lambda}_{ij}^{\,2} - \left\{ \sum_{j=1}^{c} p_{j_{ij}} \hat{\lambda}_{ij} \right\}^2 \right\}$$

Note that the sum of the squared Pearson residuals, $\sum_{i=I}^{n} r_i^{\,2}$, gives the goodness-of-fit statistic for the mixed Poisson regression model.

2.5. Seizure Frequency Data Analysis

In this section, we apply the mixture models to our data. Table 2.1 shows the estimation results of mixed Poisson regressions. We choose the two-component mixture model because its AIC and BIC are larger than those of three-component mixture model. ***So*** this is the good fit of the data.

		Poisson rate				Log-likeliHood	AIC	BIC
	Pj	αj1	**aj2**	**aj3**	αj4			
1	0.7	1.9870	7.2759	-.2470	-2.2487	-297.78	-304.78	-320.02
2	0.3	2.4156	1.5015	-.2455	-.5406			
1	0.01	-20.00	**26.1732**	**6.1329**	**-7.9335**	-347.91	-357.91	-382.50
2	**0.86**	**2.289**	**48.2688**	**-19.987**	**-14.504**			
3	**0.13**	**-12.07**	**26.1731**	-.1980	**-7.9335**			

Table **2.1.** Data analysis with mixed Poisson regressions.

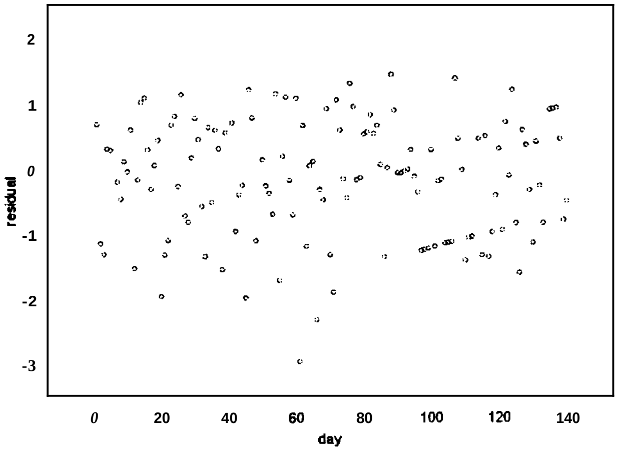
In the two component mixture model, the mixing probabilities equal **0.7** and **0.3** and the respective conditional rate functions are

$$\lambda_1(\underline{x}_i, \underline{\alpha}_1) = \exp(1.9870 + 7.2759 x_{i1} - .2470 x_{i2} - 2.2487 x_{i3})$$

and $$\lambda_2(\underline{x}_i, \underline{\alpha}_2) = \exp(2.4156 + 1.5015 x_{i1} - .2455 x_{i2} - .5406 x_{i3}).$$

The sum of Pearson residual, r^2, is **115.5128** with **131** degrees of freedom and p-value is **0.83**. Thus, there is strong evidence of a good fit because the value does not exceed the upper **95%** critical point of the χ^2_{131} .

AIC is **-304,7776** and BIC is **-320.015.**



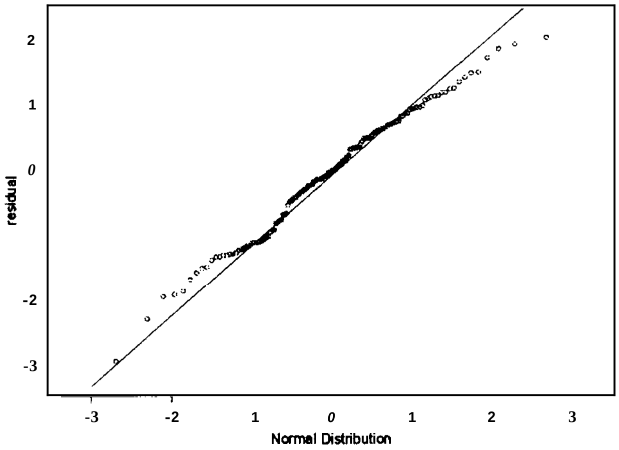


Figure 2.4. The residua plot and the ***QQ*** plot for the two-component mixture Poisson model.

In Figure 2.4, residuals are randomly distributed around 0 of the Y-axis, the curve follows the straight line in ***QQ*** plot. The fitted values of the two-component mixture model are displayed in Figure 2.5.

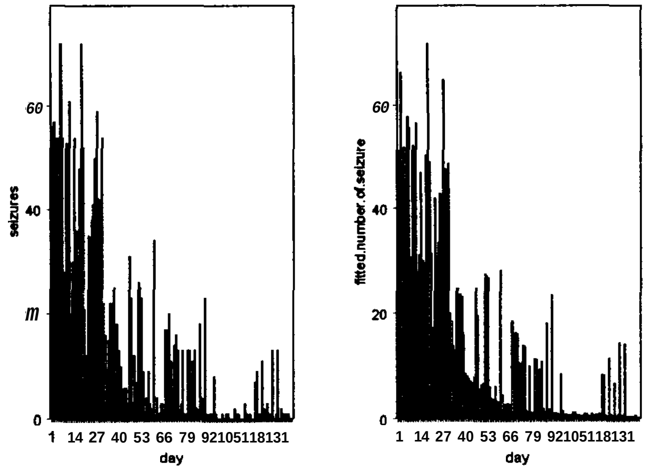


Figure 2.5. The fitted values of the two-component mixture Poisson model.

The right-hand side of Figure *2.5* shows the fitted values of this Poisson regression model. We put the original data (Figure **2.1)** on the left-hand side to compare with these fitted values. The two plots look almost the same, it means the fitted values are very close to the data. ***So*** this is a good model that fits the data well.

We conclude that the two-component mixture model describes seizure frequency data well.

Chapter 3

THE ANALYSIS OF A NEW DATA SET USING GENERALIZED LINEAR MIXTURE MODELS

Wildlife ecologists want to know if snowshoe hare use habitat depending upon which vegetation types are around. In order to answer this question, a technique called snow tracking is used.

Lines called transects are randomly placed through the area to be studied. After a snowfall, the lines are examined in 100M(meter) sections and the number of hare tracks in each section are counted. The goal of the model is to determine whether the average number of tracks differs among various types of vegetation.

If the habitat use depends upon the vegetation types, one would expect to see a higher average number of tracks in more frequently used vegetation types.

The response outcome for each 100M section is the number of hare tracks.

The snow tracking data set contains several covariates: the number of days since the last snowfall and an indication of the various vegetation types. The number of days since the last snowfall plays an important role in this data set. The larger the interval between a snow and counting of the tracks, the more tracks there will be to count. There are 10 vegetation types such as White Pine Forest, Hemlock Forest, Mixed Center Forest, Spruce Fir Forest, Northern White Cedar Forest, Birch-Aspen Forest, Northern Hardwood Forest, Mixed

Hardwood-Conifer Forest, Pitch Pine Forest.

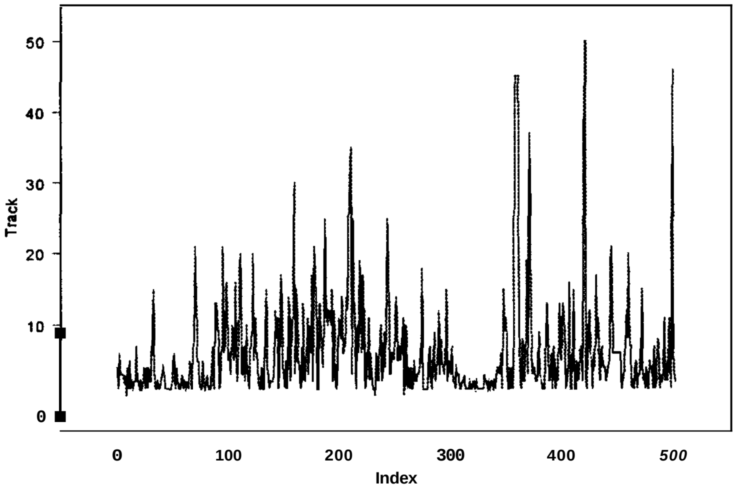


Figure 3.1. The snow tracking data

We apply mixed Poisson regressions to this data, but it is not easy to find a good fit of the data using these models because of overdispersion. They suggest that there might be different models for describing the data set where overdispersion is a prominent feature. In this case, certain types of negative binomial regression models are perhaps the most convenient to deal with, and have been studied by various authors.(Lawless, J.F. 1987.b)

Section 3.1 analyzes snow tracking data with mixed Poisson regressions. Section 3.2 discusses overdispersion and shows the result of overdispersion in mixed Poisson regressions. Section 3.3 describes the idea of negative binomial regressions and section 3.4 analyzes the data with negative binomial regressions.

3.1. Data Analysis Using Mixed Poisson Regressions

Table 3.1 presents the estimation results of mixed Poisson regressions. The goodness of fit statistic reveals that these models are inappropriate for the snow tracking data.

Mixing prob-ability	Goodness of fit			Log-like-lihood	AIC	BIC
	value	df	P-value			
.5956	1178.23	480	0	-1299.75	-1320.75	-1371.264
.4044						
Three component mixture						
.4203	1958.84	472	0	-1601.45	-1632.448	-1710.274
.3625						
.2172						
Four component mixture						
.0498	3738.95	468	0	-2164.96	-2205.955	-2311.092
.4243						
.3187						
.2182						

Table 3.1. The estimation results of mixed Poisson regressions

Figure 3.2, 3.4, and 3.6 also support that there is evidence of lack of fit in the mixed Poisson regression model. Residual plots show that almost every count centers for small number of fitted values and there are outliers. ***QQ*** plots show that residuals diverge somewhat for relatively small and relatively large Normal values.

Figure 3.3,3.5, and **3.7** compare the fit of each model to the data. The right-hand side shows the fitted values of each mixed Poisson regression and the left-hand side shows the original data (Figure **3.1).** Since these two plots do not look very similar, we conclude that the data are not fit well using mixed Poisson regressions.

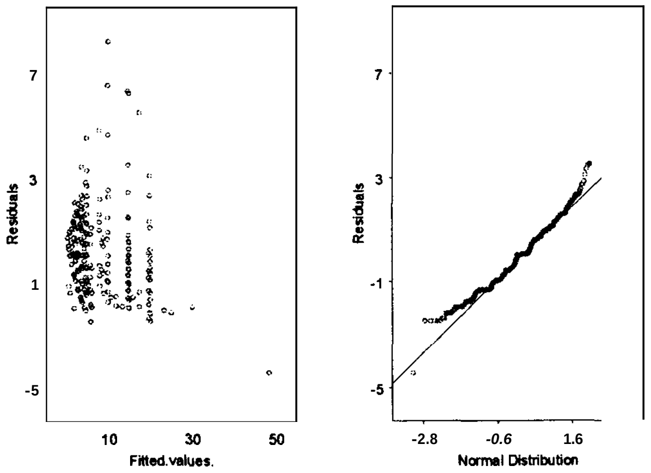


Figure **3.2.** The residual plot and the ***QQ*** plot for the mixed Poisson regression model with two-component mixture.

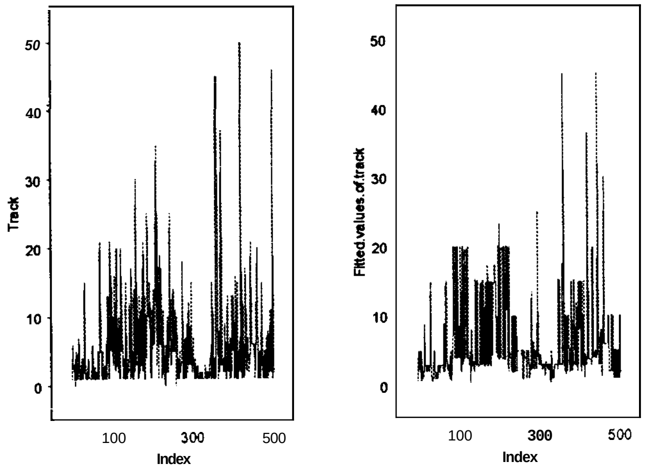


Figure **3.3.** The fitted values of number of tracks per segment for the mixed Poisson regression model with two-component mixture.

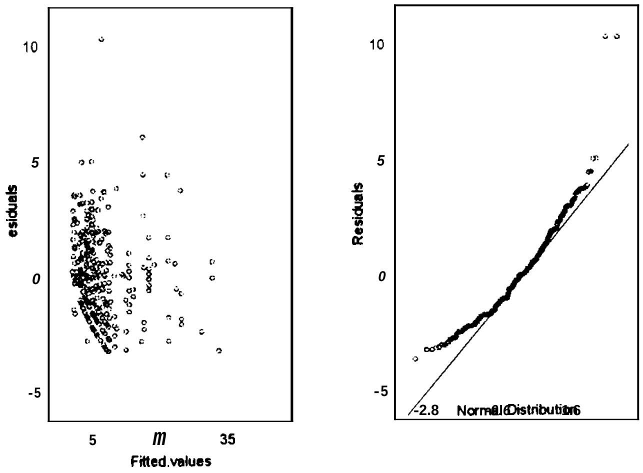


Figure **3.4.** The residual plot and the QQ plot for the mixed Poisson regression model with three-component mixture.

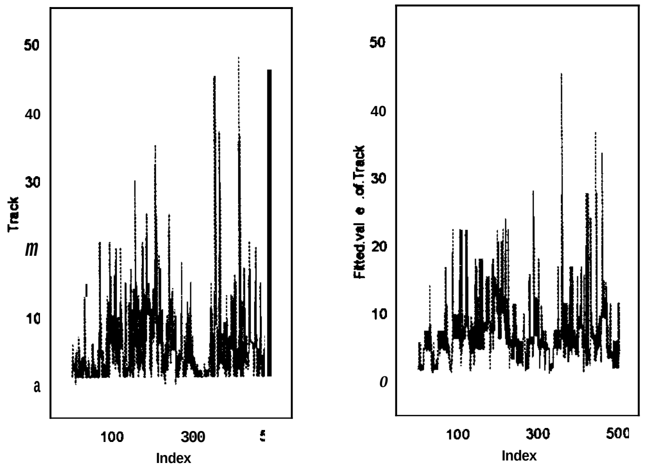


Figure 3.5. The fitted values of number of tracks per segment

for the mixed Poisson regression model with three-component mixture.

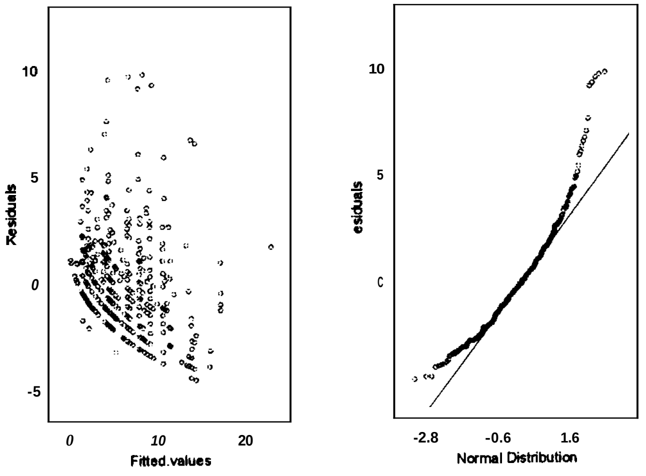


Figure *3.6.* The residual plot and the ***QQ*** plot for the mixed Poisson regression model with four-component mixture.

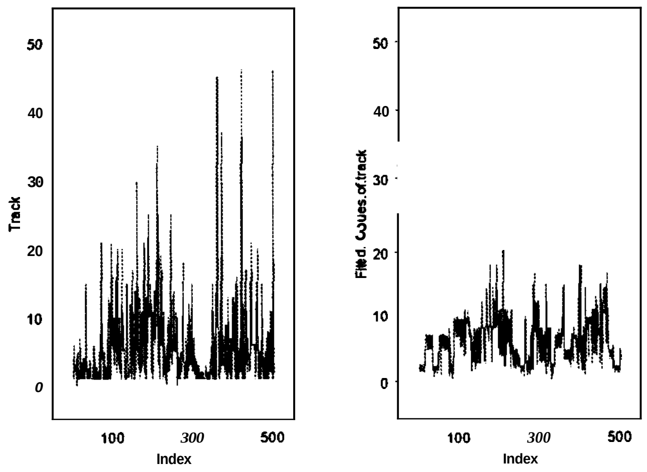


Figure 3.7. The fitted values of number of tracks per segment for the mixed Poisson regression model with four-component mixture.

Why can we not get a goodness-of-fit in this data set using mixed Poisson regressions?

Because there is overdispersion in the data set. We will discuss overdisversion in the next section.

3.2. Overdispersion

Count data often show greater variability in the response counts than one would expect if the response distribution truly were Poisson. The variances in these count data are much larger than the means, whereas Poisson distributions have identical mean and variance. The phenomenon of a generalized linear model having greater variability than predicted by the random component of the model is called ***Overdispersion.*** **A** common cause of overdispersion is heterogeneity among responses.

To determine whether the data are overdispersed with respect to the Poisson distribution in a Poisson regression model, we use three score test statistics proposed by Dean (1992). He tested the hypothesis of no overdispersion against alternatives representing different forms of overdispersion.

The test statistics are

$$P_a = \frac{\sum((y_i - \hat{\mu}_i)^2 - \hat{\mu}_i)}{\sqrt{2\sum \hat{\mu}_i^2}},$$

$$P_b = \frac{\sum((y_i - \hat{\mu}_i)^2 - y_i)}{\sqrt{2\sum \hat{\mu}_i^2}},$$

and $$P_c = \frac{1}{\sqrt{2n}} \sum \frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i}$$

Corresponding to the following specifications of overdispersion:

(a) $E(y_i) \cong \mu_i$, $Var(y_i) \cong \mu_i (1 + \tau\mu_i)$ for τ small;

(b) $E(y_i) = \mu_i$, $Var(y_i) = \mu_i (1 + \tau\mu_i)$;

(c) $E(y_i) = \mu_i$, $Var(y_i) = \mu_i (1 + \tau)$.

In these formulas μ is the estimated mean value for the case of independent and identically distributed observations in a Poisson regression model. Under $H_0 : \tau = 0$, each test statistic asymptotically follows a standard normal distribution.

Table **3.2** shows the estimated values of the overdispersion test for these data. The more components a model has, the higher overdispersion there is. Values less than **3** would indicate no overdispersion.

	Pa	Pb	Pc
Two-component mixture	**30.4125**	**30.4125**	**20.3396**
Three-component mixture	**41.3946**	**41.3946**	**45.0201**
Four-component mixture	**98.2676**	**98.2676**	**100.8950**

Table **3.2.** The estimation results for overdispersion tests for mixed Poisson regressions.

We conclude that mixed Poisson regressions are not appropriate for describing the snow tracking data well because of overdispersion .

We need to think of better models to analyze these data, we consider the negative binomial model.

3.3. Negative Binomial Model

An unpublished Ph.D. dissertation (Plassmann, F., 1997) describes the negative binomial distribution from a Poisson distribution that is mixed with a gamma distribution. The derivation is as follows:

Let Y_i follow a Poisson distribution with parameter λ_i . Assume that this parameter follows a two-parameter gamma distribution $f(\lambda_i;\theta_i,\phi_i)$, whose density fbnction is given by

$$f(\lambda_i;\theta_i,\phi_i) = \frac{\lambda_i^{\theta_i-1}e^{-\frac{\lambda_i}{\phi_i}}}{\phi_i^{\theta_i}\Gamma(\theta_i)}$$

For the purpose of finding an interpretation of the parameters of the negative binomial distribution, it is common to redefine the second parameter as

$\phi_i = \frac{\mu_i}{\theta_i}$, which results in the density fbnction

$$f(\lambda_i)=\frac{\lambda_i^{\theta_i-1}e^{-\frac{\lambda_i\theta_i}{\mu_i}}\theta_i^{\theta_i}}{\mu_i^{\theta_i}\Gamma(\theta_i)}$$

with mean equal to μ_i and Var(λ_i) = $\frac{\mu_i^2}{\theta_i}$.

On the other hand, the basic Poisson model can be generalized by relaxing the assumption that A, is a deterministic function, and by replacing it with the assumption A, is generated by A, = $\lambda(\cancel{x_i}, \cancel{a_i})$ .The resulting mixed distribution is described by $E[f(Y_i \mid$ A,)] , that is, the expectation taken with respect to the distribution of A, . If $f(\lambda_i)$ is the density function of the random parameter λ_i , the distribution of each Y_i is obtained by integrating over A,, which results in

$$P(Y_i = y_i) = \int_0^\infty P(Y_i = y_i \mid \lambda_i) f(\lambda_i) d\lambda_i .$$

So the marginal density $P(Y_i = y_i)$ can now be calculated as

$$P(Y_i = y_i) = \int_0^\infty \frac{\lambda_i^{y_i} e^{-\lambda_i} \lambda_i^{\theta -1} \theta_i^{\theta_i}}{y_i! \mu_i^{\theta_i} \Gamma(\theta_i)} e^{-\frac{\lambda_i \theta_i}{\mu_i}} d\lambda_i$$

$$= \frac{\theta_i^{\theta_i}}{y_i! \mu_i^{\theta_i} \Gamma(\theta_i)} \int_0^\infty \lambda_i^{y_i + \theta_i - 1} e^{-\lambda_i\left(1+\frac{\theta_i}{\mu_i}\right)} d\lambda_i$$

$$= \begin{pmatrix} \theta_i + y_i - 1 \\ \theta_i - 1 \end{pmatrix} \left(\frac{\mu_i}{\mu_i + \theta_i} \right)^{y_i} \left(\frac{\theta_i}{\mu_i + \theta_i} \right)^{\theta_i}$$

This density is called a negative binomial distribution with the parameters $\theta_i >$ Oand $\mu_i > 0$.

As a result of the index-parameterization of the gamma distribution, the mean of the negative binomial distribution is equal to the parameter μ_i , and the variance is given by $\mu_i + \frac{\mu_i^2}{\theta_i}$. The parameter θ_i determines the degree of dispersion, that is, the degree which the variance differs from the mean. For $\theta_i \to \infty$ the distribution converges to the Poisson distribution which implies the variance equals the mean. **As** both parameters are positive, the variance of the negative binomial distribution is larger than the mean and the distribution can be used to model data with overdispersion.

As θ_i can be any positive rational number, it is necessary to calculate the factorial in the binomial coefficient by using the relationship between factorials and the gamma function $\Gamma(x) = (\mathbf{x} - 1)!$ for the integer **x** . The probability $P(Y_i = y_i)$ can then be calculated as

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \theta_i)}{y_i!\Gamma(\theta_i)} \left(\frac{\mu_i}{\mu_i + \theta_i}\right)^{y_i} \left(\frac{\theta_i}{\mu_i + \theta_i}\right)^{\theta_i}$$

The most widely used estimation technique to estimate the negative binomial model is the maximum likelihood method. If n is the number of independent observations, then the likelihood function of the negative binomial distribution can be determined according to

$$L(\mu_i, \theta_i \mid y_i) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \theta_i)}{\Gamma(y_i + 1)\Gamma(\theta_i)} \left(\frac{\mu_i}{\mu_i + \theta_i}\right)^{y_i} \left(\frac{\theta_i}{\mu_i + \theta_i}\right)^{\theta_i}$$

For any nonnegative integer y and any θ_i>0, it is possible to write $\frac{\Gamma(y_i + \theta_i)}{\Gamma(\theta_i)} =$

$\theta_i(\theta_i + 1)\cdots(\theta_i + y_i - 1)$, so that the loglikelihood function can be written without using the gamma function as

$$\ln L(\mu_i, \theta_i \mid y_i)$$

$$= \sum_{i=1}^{n}[\{\sum_{k=0}^{y_i=1} \ln(\theta_i + k)\} - \ln y_i! + \theta_i(\ln\theta_i - \ln(\mu_i + \theta_i)) - y_i(\ln\mu_i - \ln(\mu_i + \theta_i))].$$

Now we want to find the estimates, μ_i, θ_i, that maximize the loglikelihood function.

$$\frac{\partial \ln L}{d\mu_i} = \frac{\partial}{\partial \mu_i}\left[\sum_{i=1}^{n} -\theta_i \ln(\mu_i + \theta_i) - y_i \ln\mu_i + y_i \ln(\mu_i + \theta_i)\right]$$

$$= \sum_{i=1}^{n} -\frac{\theta_i}{\mu_i + \theta_i} - \frac{y_i}{\mu_i} + \frac{y_i}{\mu_i + \theta_i}$$

$$= \sum_{i=1}^{n} \frac{y_i - \theta_i}{\mu_i + \theta_i} - \frac{y_i}{\mu_i}$$

So we have $\dfrac{\partial \ln L}{\partial \mu_i} = \displaystyle\sum_{i=1}^{n} \frac{y_i - \theta_i}{\mu_i + \theta_i} - \frac{y_i}{\mu_i} = 0$ (3.1)

And $\dfrac{\partial \ln L}{\partial \theta_i} = \dfrac{\partial}{\partial \theta_i}\left[\displaystyle\sum_{i=1}^{n}\left\{\sum_{k=0}^{y_i-1} \ln(\theta_i + k)\right\} + \theta_i \ln\theta_i - \theta_i \ln(\mu_i + \theta_i) + y_i \ln(\mu_i + \theta_i)\right]$

$$= \left[\sum_{i=1}^{n}\left\{\sum_{k=0}^{y_i-1} \frac{1}{\theta_i + k}\right\} + \ln\theta_i + 1 - \ln(\mu_i + \theta_i) - \frac{\theta_i}{\mu_i + \theta_i} + \frac{y_i}{\mu_i + \theta_i}\right]$$

so we have $\dfrac{\partial \ln L}{\partial \theta_i} = \left[\sum_{i=1}^{n}\left\{\sum_{k=0}^{y_i-1}\frac{1}{\theta_i+k}\right\} + \ln\theta_i + 1 - \ln(\mu_i+\theta_i) + \frac{y_i-\theta_i}{\mu_i+\theta_i}\right]$=0 (3.2)

Since closed form solutions of equation (3.1), (3.2) are unavailable, we use an iterative method as was done in Chapter 2, to estimate μ_i, θ_i .

3.4. Data Analysis Using Mixed Negative Binomial Models

In this section the analysis of the snow tracking data is repeated with the negative binomial distribution.

In the mixed Poisson regressions, the parameter A is equal to the expected value of the Poisson distribution, and the independent variables are introduced into the model by expressing A as a deterministic function of these variables. In order to guarantee a positive expected A value, the functional form estimated is A $=\exp(\underline{\alpha}'\underline{x})$ which is equal to μ in this case ; i.e. $\mu = \exp(\underline{\alpha}'\underline{x})$ as discussed in mixed Poisson regressions in Chapter 2.

Now we apply a generalized linear negative binomial model for the snow tracking data set. We use the same link function

$\mu(\underline{x}_i, \mathbf{a}) = \exp(\alpha_0 + \alpha_1 x_{i1} + \cdots + \alpha_{10} x_{i10})$ as with mixed Poisson regressions

for i=1,. ..,502 (the number of data point) where

$$f(y_i \mid x_i, t_i, \alpha, \theta_i) = \frac{\Gamma(y_i+\theta_i)}{\Gamma(y_i+1)\Gamma(\theta_i)}\left(\frac{\mu_i}{\mu_i+\theta_i}\right)^{y_i}\left(\frac{\theta_i}{\mu_i+\theta_i}\right)^{\theta_i}$$

and $\mu_i = t_i \mu(\underline{x}_i, \underline{\alpha}) = t_i \exp(\underline{\alpha}' \underline{x}_i)$ .We can estimate $\underline{\alpha}$ by replacing the μ_i in **(3.1)** and **(3.2)** with the link function $\mu(\underline{x}_i, \underline{\alpha})$.

The parameter estimates become

$\hat{\underline{\alpha}}$ = **(-.8995, .2488, -.8923, -1.8086, -.3414, -.4380, .2285, .6680, .4160, .1778, .3978)** and $\hat{\theta}$=**1.2886**.

For this model, the residual deviance is **518.8409** on **491** degrees of freedom. It does not exceed the upper **95%** critical point of the χ^2_{491} distribution and the p-value is .**18**, suggesting that there is an evidence of goodness of fit.

But the residual plot and the ***QQ*** plot of this model reveal that there is something insufficient to choose this model as good of fit, and these plots are displayed in Figure **3.8.** The residual plot shows some pattern of counts and the QQ plot does not show the straight line.

Figure **3.9** compares the fit to original data. The right-hand side shows the fitted values of number of track per segment and the left-hand side shows the original data. Since the two plots do not look similar we conclude that the generalized linear negative binomial model does not fit well to the data.

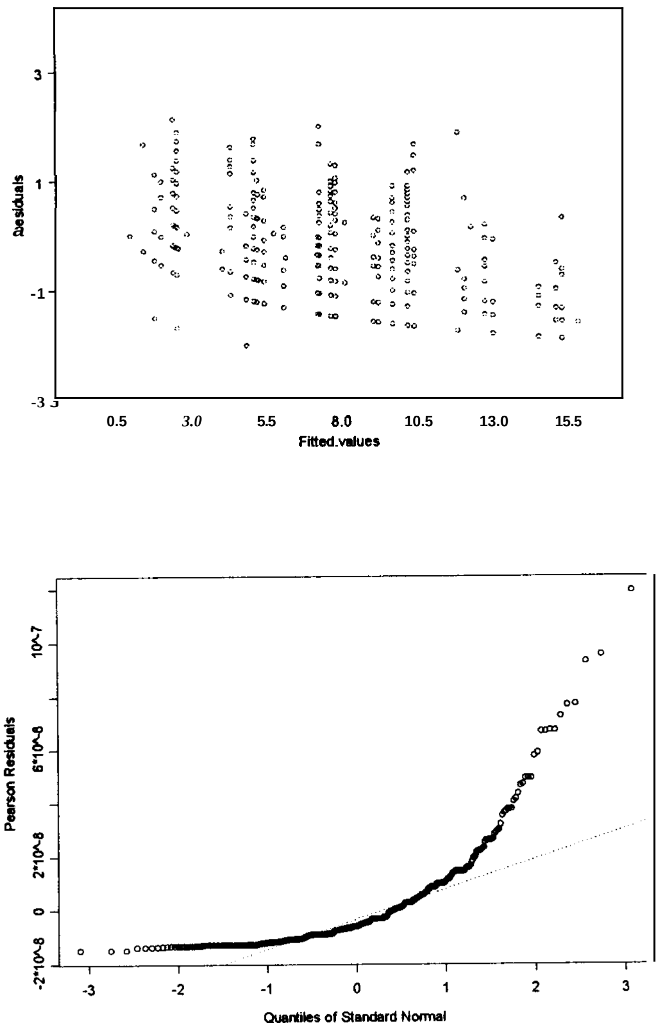


Figure **3.8.** The residual plot and the QQ plot for the negative binomial generalized linear model.

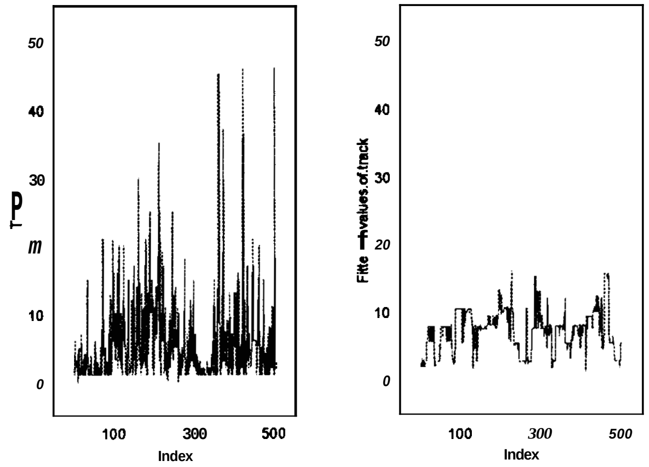


Figure **3.9.** The fitted values of number of tracks per segment from the negative binomial generalized linear model.

So we continue the data analysis using mixed negative binomial regressions. We use the same method with mixed Poisson regressions to estimate parameters $\mu, \underline{\theta}$ in negative binomial models. This includes EM algorithm, iterative steps, their properties, model selection using AIC and **BIC,** residual analysis and goodness of fit test.

pro-bability		Value	df	p-value	like-lihood		
Two-component mixture							
.3307	**6.0617**	**.8930**	**477**	**1**	**-1572.02**	**-1593.02**	**-1643.53**
.6693	**6.8496**						
Three-component mixture							
.1195	3.079	9.772 *e-014	465	1	-1722.36	1-1753.36	-1831.19
.5140	1.9741						
.3665	1.9186						

Table 3.3. The results of the mixed negative binomial regressions.

Chi-square tests give that both negative binomial regressions are appropriate for describing data because the p-value is equal to 1. Between these two models, we choose the two-component mixture model because it has the larger AIC and BIC values than the other. The residual plot and the ***QQ*** plot of the negative binomial regressions follow in Figure 3.10 and 3.12.

Both residual plots in Figure 3.10 and 3.12 are randomly placed around the 0-axis though they center at some small numbers of fitted values, but residuals in two-component mixture model is better randomness than three-component mixture model. Both ***QQ*** plots in Figure 3.10 and 3.12 diverge somewhat for relatively small and relatively large Normal values, but the ***QQ*** plot **of** the two-component mixture model is better than the three-component mixture model because it shows the divergence for relatively large Normal values while the other does for both relatively large Normal values.

The right-hand side of Figure 3.11 and 3.13 shows the fitted values of these negative binomial regression models. We put the original data (Figure 3.1) to compare with these fitted values on the left-hand side. The two plots look similar, meaning the model fits the data well.

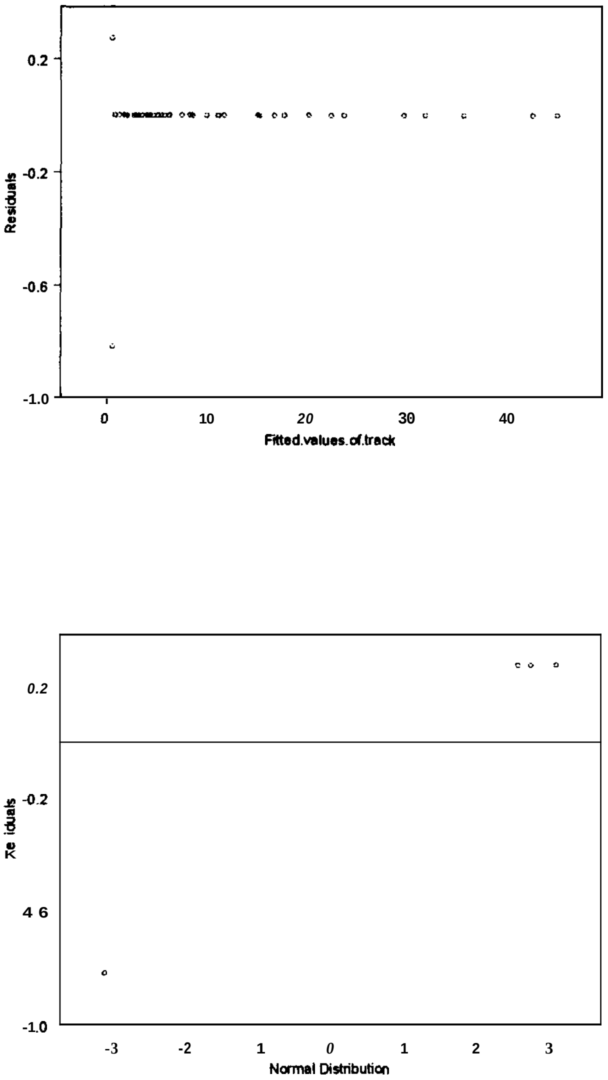


Figure 3.10. The residual plot and the QQ plot for **the two-component mixture** of **negative binomial model.**

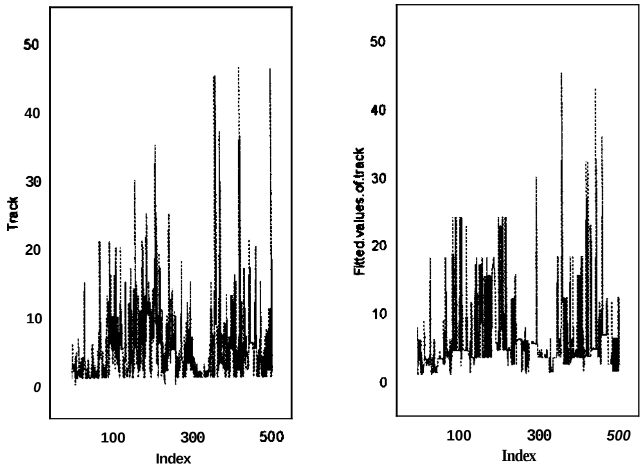


Figure 3 11. The fitted values of number of tracks per segment for the two-component mixture of negative binomial model.

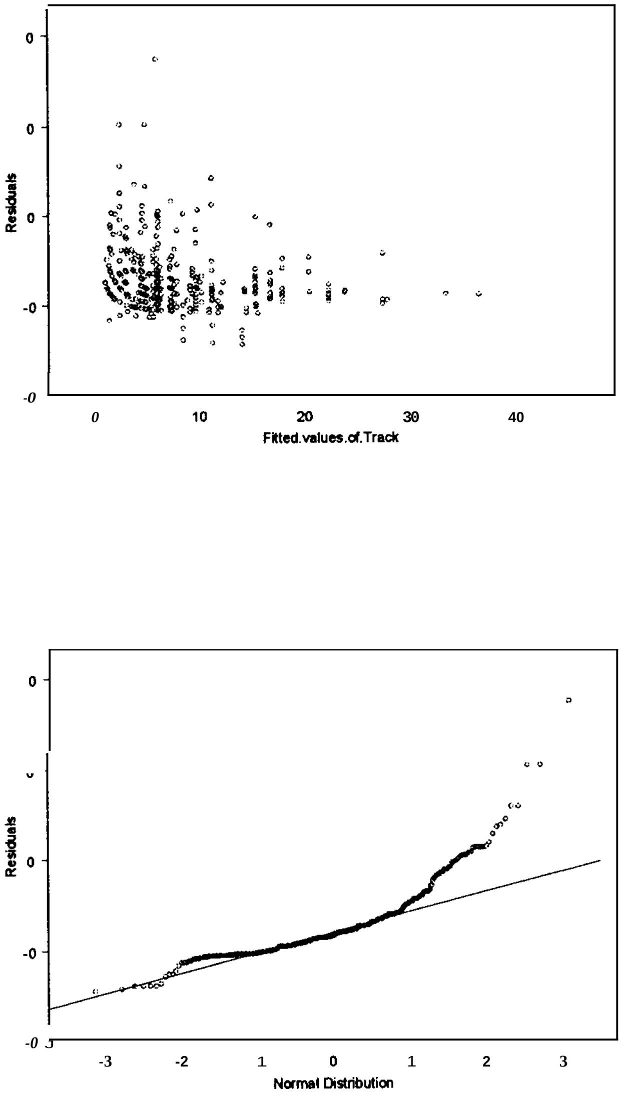


Figure 3.12. The residual plot and the QQ plot for the three-component mixture of **negative binomial model.**

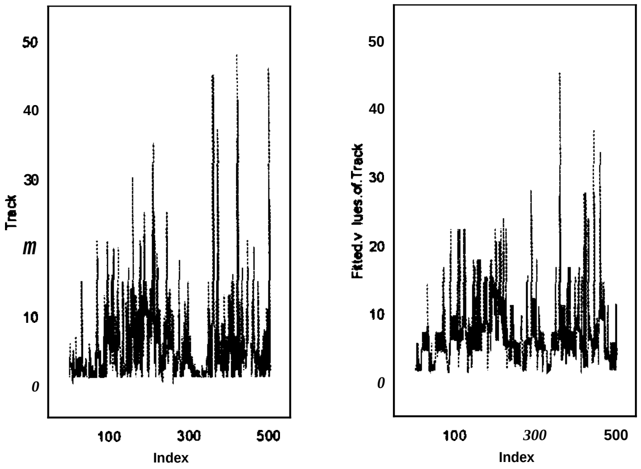


Figure **3.13.** The fitted values of number of tracks per segment for the three-component mixture of negative binomial model.

We consider the two-component mixture model as good fit of data, we want to reduce the number of covariates. Recall that the goal is choosing an appropriate model to fit data, we decide the best model by finding the model that has the largest AIC and **BIC** values among the two-component mixture models.

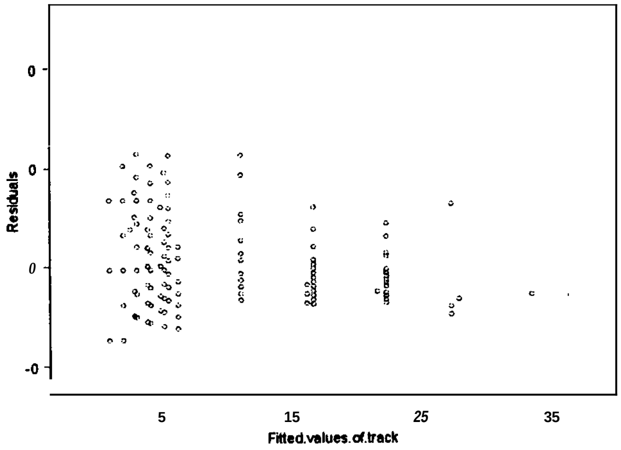
Mixing probability	θ	Negative binomial rate μ		
		Intercept	Northern Hardwood Forest	Pitch Pine Forest
.3486	**5.7087**	.1109	**-1.4326**	**-1.7225**
.6514	**5.6450**	**-1.557**	**1.6370**	**2.1577**

Log likelihood	AIC	BIC	Goodness of fit		
			Value	df	p-value
-1569.46	-1574.46	-1591.23	4.3748*e-014	493	1

Table **3.4.** The results of estimation of the best appropriate model

The $\hat{\theta}_1$ and $\hat{\theta}_2$ are equal to **5.7087** and **5.6450** respectively, the Pearson residual, X^2 , is **4.3748**e^{-14} with **493** degrees of freedom and the p-value is 1. AIC and **BIC** are **−1574.46** and **−1591.23** respectively. Thus, this model fits the data well.

The residuals and ***QQ*** plot of this model are displayed in Figure **3.14.** We can see randomness in the residual plot and check a straight line in the ***QQ*** plot. We put the original data (Figure 3.1) on the left-hand side in Figure **3.15** to compare with these fitted values. The fitted values of the model are shown on the right-hand side in Figure **3.15.** The two plots look similar, meaning the data is fit well using this model.



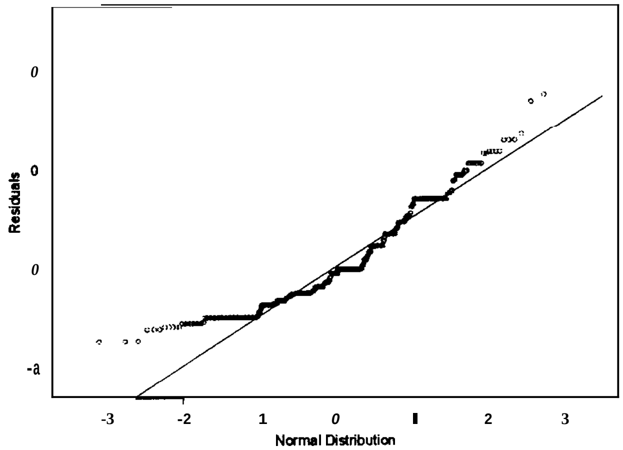


Figure 3.14. The residual and the QQ plot for the best appropriate model in two-component mixture.

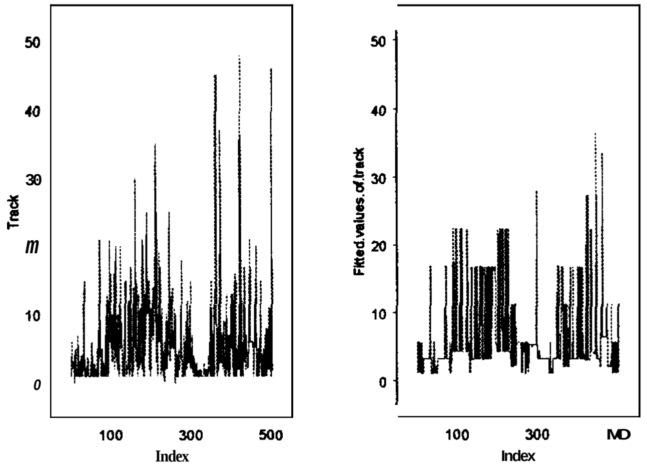


Figure 3.15. The fitted values of number of tracks per segment of the best appropriate model in two-component mixture.

We interpret this fitted model. The mixing probabilities are .3486 and .6514 and the respective rates are

$$\mu_1(x_i, \alpha_1) = \exp(.1109 - 1.4326 * NorthernHardwoodForest_{i1}$$

$$- 1.7225 * PitchPineForest_{i2})$$

$$\mu_2(x_i, a,) = \exp(-1.557 + 1.6370 * NorthernHardwoodForest_{i1}$$

$$+ 2.1577 * PitchPineForest_{i2})$$

for i=1,...,502.

For instance, $\hat{\alpha}_{11} = -1.4326$ is the estimated *NorthernHardwoodForest* effect when the data come from component one. While $\hat{\alpha}_{21} = 1.6370$ is the estimated ***NorthernHardwoodForest*** effect when the data come from component two.

Recall that our goal of the model is to determine whether the average number of tracks differ among various types of vegetation. This model suggests that the average number of track differ among response values which have the two types of vegetation, which are ***NorthernHardwoodForest*** and ***PitchPineForest.***

Since we used the indication of various vegetation type as covariates $\mu_1(x_i, a,)$ has only three values, **1.1173,** .2667 and .1996: the average number of tracks is **1.1173** when there are no effect of these two vegetation type. The average number of tracks is .2667 when there is the only effect of ***NorthernHardwoodForest*** while the average number of tracks is .1996 when there are the only effect of ***PitchPineForest.*** There is no case with both effect of these two vegetation types at the same time. $\mu_2(x_i, a,)$ has also three values, which are .2108, **1,0833** and **1.8234** respectively.

3.5. Conclusion

This paper provides a mixed generalized linear Poisson regression in which the rates of the component distributions depend on covariates. This model can be used to explain overdispersion in Poisson regression models. The negative binomial regression is derived as a mixed Poisson distribution and can deal with overdispersion in Poisson regression models.

Two examples illustrate the use of these models and provide results. In the first application, we analyze the data using mixed Poisson regressions and in the second example, we examine the data using mixed negative binomial regressions.

REFERENCES

Agresti, A. **(1996).** ***An*** Introduction to Categorical Data Analysis. New York Wiley Series in Probability and Statistics.

Dalal S.R., Fowlkes, E.B. and Hoadley, B. **(1989).** Risk Analysis of the Space Shuttle: Pre-Challenger Prediction of Failure. Journal of American Statistical Association, **84, 945-957.**

Dean, C.B. **(1992).** Testing for Overdispersion in Poisson and Binomial Regression Models. Journal of American Statistical Association, **87, 45 1-457.**

Dempster, A.P., Laird, N.M. and Rubin, D.B. **(1976).** Maximum Likelihood from Incomplete data via the EM Algorithm, Journal of Royal Statistical Society, Series B, **39, 1-38.**

Dobson, A.J. **(1990).** ***An*** Introduction to Generalized Linear Models. London: Chapman and Hall.

Lawless, J.F. **(1987).** Regression Method for Poisson Process Data. The Canadian Journal of Statistics **15, 209-225**.

Lehmann, E.L. **(1983).** Theory of Point Estimation. New **York:** Wiley.

Leroux, **B.G. (1992).** Consistent Estimation of a Mixing Distribution. Annals of Statistics 20, 1350-1360.

Leroux, B.G. and Puterman, M. L. (1992). Maximum Penalized Likelihood Estimation for Independent and Markov Dependent Mixture Models. Biometrics 48, 545-558.

McCullagh, **P.** and Nelder, J.A.(1989). Generalized Linear Models, **2nd** edition. Boca Raton: Chapman and Hall.

McLachlan, G.J. and Basford, K.E. (1988) Mixture Models. New York: Maecel Dekker, Inc.

Mood, A.M., Graybill, F.A., Boes, D.C. (1973). Introduction to the Theory of Statistics. New York: McGraw-Hill.

Plassmann, F.(1997). The Impact of Two-Rate Taxes on Construction in Pennsylvania. VT ETD [Online]. Available: http://scholar.lib.vt.edu/theses/available/etd-61097-3834/unrestricted/Ch3.pdf, 44-48.

Venables, W.N., Ripley, B.D. (1999). Modern Applied Statistics with S-plus, 3rd edition. New York: Springer.

Wang, P., Puterman, M.L., Cockburn, I. and Le, N.(1996). Mixed Poisson Models with Covariate Dependent Rates. Biometrics 52, 381-400.

APPENDICES

Appendix A. Mixed Poisson regression program

Appendix B. Mixed negative binomial regression program

Appendix A. Mixed Poisson Regression Program

```
"mixpoisson"<-function(data.frame, vars, offset = T, comp)

{

# Comp is the number of componentsto be examined

# If used (i.e. offset = T) the offset variable comes first

# The response variable is first with all covariatesfollowing

# Use numerical indices in Vars to identify which variablesto use

# Initialize script

    trms <- length(vars) - 1

    data.mod <- as.matrix(data.frame[vars])

    if(!offset) {

        trms <- trms + 1

        data.mod <- cbind(1, data.mod)

    I

    if(trms > 1)

        dimnames(data.mod) <- list(NULL, c("t", "Y", paste("X", 1:(trms -1), sep = "")))

    else dimnames(data.mod) <- list(NULL, c("t", "Y"))

    n <- nrow(data.mod)

    k <- trms * comp # Build formula for model

    zmod <- paste("z", 2:comp, sep = "")

    if(trms > 1) {

        xmod <- paste("X", 1:(trms - 1), sep = "")

        intmod <- outer(xmod, zmod, paste, sep = ":")

        dim(intmod) <- c(1, (trms - 1) * (comp - 1))

    I
```

```
  else {
      xmod <- null()

  intmod <- null()
  }
  model <- paste(c("Y~offset(log(t))", xmod, zmod, intmod), collapse="+")   # Assign
  uninformativeprior mixingprobs to components
  pj.old <- rep((1/comp), comp)
  pj.new <- rep(0, comp)      # Setup vector to receive parameter
             # estimates
  a.old <- matrix(0, comp, trms)
  # Build matrix to compute component parameters from regession
# parameters
  parm.bld <- diag(comp)
  parm.bld[, 1] <- 1
  # Build indicator of component and randomly assign each obs
# to a component
  rints <- matrix(c(1:n, floor(runif(n, 1, (comp + 0.999)))), nrow = n)
  z <- matrix(0, NOW = n, ncol = comp, dimnames = list(NULL, paste("z", 1:comp, sep =
  "")))
  z[rints] <- 1
  data.mod <- cbind(data.mod, z)      # initializethe likelihood
                # keeper
  p <- dpois(data.mod[, "Y"], data.mod[, "Y"])
  p[is.na(p)] <- 1
```

```
   loglik <- sum(log(p))
   loglike.old <- 0   # i will keep track of number of iterations
   i <- 0    # Start the process
   repeat {
# run Poisson regression -- z's are indicator of components.
       out.glm <- glm(formula(model), family = poisson, link = log,data =
   as.data.frame(data.mod), control = glm.control(
   maxit = 25)) # save parameter estimates from model
   out.glm[["coefficients"]][is.na(out.glm[["coefficients"]])] <-   0
       a.new <- parm.bld %*% matrix(out.glm[["coefficients"]], ncol = trms, byrow = T)
       loglike.new <- loglik - out.glm[["deviance"]]/2
   # compute estimates of new lambdas
       if(trms > 1) {
       lambda <- data.mod[, "t"] * exp(cbind(1, data.mod[, 3:(
               1 + trms)])%*% t(a.new))
       }
       else {
           lambda <- data.mod[, "t"] * exp(matrix(1, nrow = n,
               ncol = 1) %*% t(a.new))
       }
       p <- pj.old * matrix(dpois(data.mod[, "Y"], lambda), ncol = comp)
       p[is.na(p)] <- 1
   # Rank conditional probs from smallest to largest
       p.max <- t(apply(p, 1, order))
   # Assign component membership based upon size of conditional prob.
```

```
        data.mod[, dimnames(z)[[2]]] <- ifelse(p.max == comp, 1,0)
    # Compute new mixing probabilities
        pj.new <- colMeans(data.mod[, dimnames(z)[[2]]])
    # Check to see if a's converged
        a.diff <- sum(abs(a.new - a.old))
        pj.diff <- sum(abs(pj.new - pj.old))
        loglike.diff <- abs(loglike.new - loglike.old)
    # get ready to accept next round parameter estimates
        a.old <- a.new
        pj.old <- pj.new
        loglike.old <- loglike.new # count iterations
        i <- i + 1# exit if a.estimates converge or i exceeds 10
        if((i > 30) || ((a.diff < 1e-007) && (pj.diff < 1e-007) && (
            loglike.diff < 0.0001)))
            break



# Compute analysis results
# Standard errors (from inverse of information matrix)
    z <- colSums(data.mod[, (2 + trms):(1 + trms + comp)] %*% diag(1/pj.new^2))
    pinfo <- matrix(z[comp], nrow = (comp - 1), ncol = (comp - 1))
    pinfo <- sqrt(diag(ginverse(pinfo + diag(z)[1:(comp - 1), 1:(comp - 1)] )))
    x <- kronecker(parm.bld, diag(trms))
    subs <- (1:k)[out.glm$coefficients != 0]
    se <- matrix(0, k, k)
```

```
  se[subs, subs] <- summary.lm(out.glm)$cov.unscaled

  se <- matrix(summary.lm(out.glm)$sigma * sqrt(diag(x %*% se %*% t(x))),ncol = trms,

  byrow = T)   # Parameters se's and Wald stats.

  probs <- c(pj.old, se[trms * comp + 1:(comp - 1)])

  names(probs) <- c(paste("comp", 1:comp, sep = ""), paste("se", 1:(comp - 1), sep = ""))

  parms <- matrix(t(cbind(a.old, se, a.old/se)), ncol = trms, byrow = T)

  dimnames(parms) <- list(rep(c("comp", "se", "Wald stat"), times = comp),   c(paste("a",

  0:(trms - 1), sep = "")))

  # Goodness of fit Statistics

  chistat <- sum(residuals.glm(out.glm, type = "pearson")^2)

  chistat <- c(Chistat = chistat, df = (out.glm[["df.residual"]] - (comp - 1)), pvalue = (1 –


pchisq(chistat, out.glm[["df.residual"]] - ( comp - 1), ncp = 0)))

  AIC <- loglike.old - k + (comp - 1)

  BIC <- loglike.old - ((k + (comp - 1)) * log(n))/2

  Fit <- c(AIC = AIC, BIC = BIC)     # Overdispersion measures

  Pa <- sum((data.mod[, "Y"] - fitted(out.glm))^2 –

        fitted(out.glm))/sqrt(2 * sum(fitted(out.glm)^2))

  Pb <- sum((data.mod[, "Y"] - fitted(out.glm))^2 - data.mod[, "Y"])/sqrt(2 *

  sum(fitted(out.glm)^2))

  Pc <- (1/sqrt(2 * n)) * sum(((data.mod[, "Y"] - fitted(out.glm))^2 - data.mod[,

  "Y"])/fitted(out.glm))

  OverDisp <- c(Pa = Pa, Pb = Pb, Pc = Pc)

  finaldata.mod <<- data.mod

  poissonglm.out <<- out.glm
```

```
# Show final parameter estimates and log likelihood

list(Reps = i, "Component Weights w/SE" = probs, "Comp Parameters" = parms,

Loglikelihood = loglike.old, "Chi-square Fit" = chistat,

    "AIC and BIC Fit" = Fit, "OverDispersion Meas." = OverDisp) }
```

Appendix B. Mixed Negative Binomial Regression Program

```
"negbi.prob"<-
function(y, mu, theta)
{
    exp((lgamma(y + theta) + y * log(mu) + theta * log(theta)) - (lgamma(
        theta) + lgamma(y + 1) + (theta + y) * log(mu + theta)))
}


"mixnegb2"<-function(data.frame, vars, offset = "T", comp)
{
# Comp is the number of components to be examined
# If used (i.e. offset = T) the offset variable comes first
# The response variable is first with all covariates following
# Use numerical indices in Vars to identify which variables to
# use
# Initialize script
    library(Mass)     # Need this lib to do Neg Bin Glm
    offset <- as.logical(offset)
    trms <- length(vars) - 1
    data.mod <- as.matrix(data.frame[vars])
    if(!offset) {
        trms <- trms + 1
        data.mod <- cbind(1, data.mod)
}
if(trms > 1)
    dimnames(data.mod) <- list(NULL, c("t", "Y", paste("X", 1:(trms -1), sep = "")))
else dimnames(data.mod) <- list(NULL, c("t", "Y"))
n <- nrow(data.mod)
k <- trms * comp # Build formula for model
if(trms > 1) {
    xmod <- paste("X", 1:(trms - 1), sep = "")
}
else {
    xmod <- null()
}
model <- paste(c("Y~offset(log(t))", xmod), collapse = "+")
# Assign uninformative prior mixing probs to components
pj.old <- rep((1/comp), comp)
pj.new <- rep(0, comp)     # Setup vector to receive parameter
                # estimates
a.new <- a.old <- matrix(0, comp, trms)
se.parms <- matrix(0, comp, trms)
theta <- rep( 1, comp)
se.theta <- rep(0, comp)
# Build indicator of component and randomly assign each obs to a component
rints <- matrix(c(1:n, floor(runif(n, 1, (comp + 0.999)))), nrow = n)
z <- matrix(0, nrow = n, ncol = comp, dimnames = list(NULL, paste("z", 1:comp, sep =
""))) 
```

```
    z[rints] <- 1

    data.mod <- cbind(data.mod, z)

    # add columns to receive fitted and residuals

    data.mod <- cbind(data.mod, matrix(0, n, 2))

    dimnames(data.mod)[[2]][1+trms+comp + 1:2]<- c("fitted",

        "residual")   # i will keep track of number of iterations

    i <- 0

    loglike.old <- 0    # Start the process

    repeat {

# run Neg Bin regression -- z's are indicator of components.

        for(j in 1:comp) {

            pick.rows <- data.mod[, paste("z", j, sep = "")]

            out.glm <- glm.nb(formula(model), link = log, data =

                as.data.frame(data.mod[pick.rows == 1, ]),

                control = glm.control(maxit = 25))

    # save parameter estimates from model

            out.glm[["coefficients"]][is.na(out.glm[["coefficients"

                ]])] <- 0

            a.new[j, ] <- out.glm[["coefficients"]]

            1<- length(summary.lm(out.glm)$sigma * sqrt(diag(

                summary.lm(out.glm)$cov.unscaled)))

            se.parms[j, ] <- c(summary.lm(out.glm)$sigma * sqrt(


    diag(summary.lm(out.glm)$cov.unscaled)), rep(0,

                trms - 1))
            theta[j] <- out.glm$theta
            se.theta[j] <- out.glm$SE.theta
            data.mod[pick.rows == 1, "fitted"] <- fitted(out.glm)
            data.mod[pick.rows == 1, "residual"]<- residuals.glm(
                out.glm, type = "pearson")
        }
# compute estimates of new means
        if(trms > 1) {
            mu <- data.mod[, "t"] * exp(cbind(1, data.mod[, 3:(1 +
                trms)]) %*% t(a.new))
        }
        else {
            mu <- data.mod[, "t"] * exp(matrix(1, NOW = n, ncol =
                1)%*% t(a.new))
        }
        p <- pj.old * matrix(negbi.prob(data.mod[, "Y"], mu, theta), ncol = comp)
        p[is.na(p)] <- 1
    # Rank conditional probs from smallest to largest
        loglike.new <- sum(log(apply(p, 1, max)))
        p.max <- t(apply(p, 1, order))
    # Assign component membership based upon size of conditional prob. data.mod[,
    dimnames(z)[[2]]] <- ifelse(p.max == comp, 1, 0)

    # Compute new mixing probabilities
        pj.new <- colMeans(data.mod[, dimnames(z)[[2]]])
```

```
      # Check to see if a's converged
          a.diff <- sum(abs(a.new - a.old))
          pj.diff <- sum(abs(pj.new - pj.old))
          loglike.diff <- abs(loglike.new - loglike.old)
      # get ready to accept next round parameter estimates
          a.old <- a.new
        pj.old <- pj.new
          loglike.old <- loglike.new # count iterations
          i <- i + 1# exit if a.estimates converge or i exceeds 30
          if((i > 30) || ((a.diff < 0.0001) && (pj.diff < 0.0001) && (loglike.diff < 0.01)))
              break
      }
  # Compute analysis results
  # Standard errors (from inverse of information matrix)
      z <- colSums(data.mod[, (2 + trms):(1 + trms + comp)] %*% diag(1/pj.new^2))
      pinfo <- matrix(z[comp], nrow = (comp - 1), ncol = (comp - 1))
      pinfo <- sqrt(diag(ginverse(pinfo + diag(z)[1:(comp - 1), 1:(comp - 1)]  )))  # Parameter
      se's and Wald stats.
      probs <- c(pj.old, pinfo)
      names(probs) <- c(paste("comp", 1:comp, sep = ""), paste("se", 1:(comp - 1), sep = ""))
       parms <- matrix(t(cbind(a.old, se.parms, a.old/se.parms)), ncol = trms, byrow = T)
      dimnames(parms) <- list(rep(c("comp", "se", "Wald stat"), times = comp),   c(paste("a",
      0:(trms - 1), sep = "")))
      # Goodness of fit Statistics
      chistat <- sum(data.mod[, "residual"]^2)
```

```
chistat <- c(Chistat = chistat, df = n - (trms + 1)* comp - 1, pvalue = (1 - pchisq(chistat, n
- (trms + 1)* comp - 1,
    ncp = 0)))
AIC <- loglike.old - k + (comp - 1)
BIC <- loglike.old - ((k + (comp - 1))* log(n))/2
Fit <- c(AIC = AIC, BIC = BIC)     # Overdispersion measures
Theta <- rbind(Theta = theta, SE = se.theta)
finaldata.mod <<- data.mod
# Show final parameter estimates and log likelihood
list(Reps = i, "Component Weights w/SE" = probs, "Comp Parameters" = parms,
Loglikelihood = loglike.old, "Chi-square Fit" = chistat,
  "AIC and BIC Fit" = Fit, "Ests. of Theta" = Theta)
```

BIOGRAPHY OF THE AUTHOR

Jungah Jung was born in Seoul, Korea on May 12,1974. She was raised in Seoul and Taegu and graduated Kyung-I1 Women's High School in 1993. She attended the Kyungpook National University and graduated in 1999 with a Bachelor's degree in Statistics. She also attended Graduate School of the Kyunpook National University majoring in Statistics in 1999. After one semester, she came to United States and entered the Mathematics and Statistics graduate program at the University of Maine in the fall 1999.

During her university years she has been to the English Institute at The University of Oregon to learn speaking English in 1996.

She likes to travel, taking photography, reading books, watching movies and playing outdoor sports such as bowling, skiing, running and tennis.

After receiving her degree Jungah plans to enter Ph.D program in Statistics. Jungah is a candidate for the Master of *Arts* degree in Mathematics from The University of Maine in August, 2001.