Violating Equidispersion: A Comparative Analysis of Poisson and Negative Binomial Regression

by
Hunter Evans
Advisor: Dr. Samantha Seals, PhD

An Undergraduate Proseminar
In Partial Fulfillment of the Degree of
Bachelor of Science in Mathematics
The University of West Florida
November 26, 2024

The Proseminar of Student's Name is approved:

Samantha Seals, PhD, Proseminar Advisor	Date
Samantha Seals, PhD, Committee Chair	Date

Accepted for the Department/Division:

Jia Liu, PhD, Chair	Date

Table of Contents

Abstract

Objectives: This paper aims to explore the implications of analyzing overdispersed count data under the Poisson distribution, which breaks the equidispersion assumption. Using negative binomial as a comparison, our goal is to understand the potential consequences of model misspecification and to guide researchers toward more robust statistical practices.

Methods: We conducted a simulation study to compare Poisson and negative binomial. We wrote a function that would repeatedly generate datasets while also allowing a way to change the relationship between μ and σ^2. We did this by simulating our outcome variable as $y \sim \text{NegBin}(\mu, \theta)$, where θ is the dispersion parameter that changes the relationship between μ and σ^2. After the data was simulated, each dataset was used to construct a model under both Poisson and negative binomial distributions. The simulated raw data and, separately, the analysis results from both Poisson and negative binomial regressions were stored for further analysis. The regression analysis results were examined with respect to bias, mean square error (MSE), type I error, and model selection using AIC.

Results: In our simulation study, our results agreed with statistical theory. The bias and MSE results showed that there wasn't a major difference between the two models, which was expected as the slopes themselves aren't affected when the data is overdispersed. In the standard error results, we saw that negative binomial had a higher SE than Poisson for $\theta \in \{1, 10, 50\}$, where the data was overdispersed. This visualizes the underestimation of the standard error. We saw similar results in the rejection rate graphs as well, which showed when the data was overdispersed, the Poisson model would reject the null, $H_0 : \beta = 0$, 30-40% of the time. In the vast majority of the simulated datasets, the AIC suggested that negative binomial was the better fit when the data was overdispersed.

Conclusions: Based on these results, we can conclude that when the data is overdispersed and Poisson is used, our conclusions may not be accurate. In the $\theta = 500$ case, where $\mu \approx \sigma^2$, Poisson and negative binomial are behaving similarly. Because there is not a clear advantage in estimating an additional parameter with the negative binomial, we recommend using Poisson regression for simplicity.

Chapter 1

Introduction

1.1 Statement of Problem

Checking assumptions is a critical aspect of statistical analysis. Certain tests or models can have multiple assumptions that need to be met, otherwise, the conclusions may not be valid. For example the normal distribution assumes random sampling, normality of data distribution, homogeneity of variances, a reasonably large sample size, and interval or ratio scaled measurements. In particular, we are interested in examining the equidispersion assumption of Poisson regression.

This paper aims to explore the implications of analyzing overdispersed count data under the Poisson distribution, which breaks the equidispersion assumption. Our goal is to understand the potential consequences of model misspecification and to guide researchers toward more robust statistical practices.

1.2 Relevance of Problem

Statistical analysis can be difficult to navigate due to the variety of methods to choose from and the corresponding assumptions that must be satisfied. With real-world problems, meeting the assumptions for these models can be very strenuous and sometimes even impossible. Poisson regression is a good example of this, due to the assumption that $\mu = \sigma^2$, which is called equidispersion. The question then becomes, what are the ramifications of applying this model when the mean and variance differ? This project compares Poisson regression to negative binomial, a model that accounts for overdispersion, we hope to answer this question.

1.3 Literature Review

The Poisson distribution upon which Poisson regression is based, was first derived in the early 1800's by Siméon Poisson [1]. Poisson regression quickly became the starting point of the analysis of count data due to how it can incorporate common issues like clustering data, repeated measurements, overdispersion, and excess zeros [2]. One of the main issues, overdispersion, which is when the variance is greater than the mean, can be dealt with using models such as the negative binomial [3]. When overdispersion is ignored, the Poisson regression model underestimates standard errors of regression coefficients, in turn inflating type I error rates [4] [5]. Another issue could also be loss of efficiency in using statistics appropriate for the assumed distribution [4].

A simulation study performed by Owusu (2020) to examine the effects of overdispersion. Simulated data

was analyzed under both Poisson and negative binomial regressions to examine the effects of overdispersion. In this study, the Poisson regression standard errors were much lower than those of negative binomial, which caused the Poisson model to incorrectly identify some covariates as significant. [6]

Chapter 2

Methods

2.1 Statistical Modeling

2.1.1 Analysis of Count Data

Count data consists of non-negative integers, such as 0, 1, 2, 3, etc. It represents countable quantities, such as the number of occurrences or events. Common examples of count data used are the number of people who check out at a grocery store or the number of births in a certain county.

By definition, the normal distribution is a continuous probability distribution that is symmetric about the mean. Because count data is discrete, the normal distribution is not appropriate as it assumes $y \in \mathbb{R}$. Applying the normal distribution to count data may yield predictions that would not make sense, such as $\hat{y} < 0$. To apply this to our examples from earlier, we might predict -4 people going to the grocery store.

Because we are analyzing count data and the normal distribution is not appropriate, we now move from general linear models to generalized linear models. In particular, we now specify the underlying distribution to be either the Poisson or negative binomial.

Poisson Distribution

The Poisson distribution is a discrete probability distribution, and is used to model count data, specifically the number of occurrences within a certain time period. The Poisson distribution uses one parameter, λ, and assumes that the data is count, equidispersed, and the observations are independent of each other. The way we often see Poisson is as the Probability Mass Function (PMF), the formula is below:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \ldots$$

The graph of the PMF of Poisson is below:

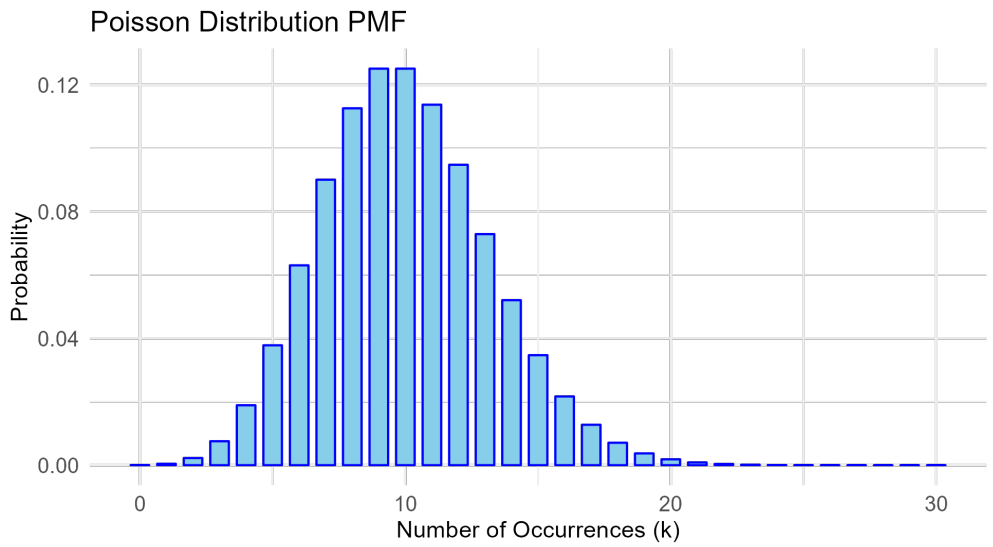

Figure 2.1: PMF of the Poisson distribution with $\lambda = 10$

The moment generating function of Poisson is below, which we will use in a proof later:

$$M_{\text{Poi}}(t) = E[e^{tY}] = e^{\lambda(e^t - 1)}$$

Negative Binomial Distribution

The negative binomial distribution is a discrete probability distribution that measures the number of trials required to achieve a certain number of successes. The negative binomial distribution is a generalized form of the Poisson distribution that adds an extra parameter to account for dispersion. This extra parameter allows for modeling of overdispersed data, making it a good alternative of Poisson when the data is overdispersed. The PMF of negative binomial is below:

$$P(X = x) = \binom{x + r - 1}{x} p^r (1 - p)^x, \quad x = 0, 1, 2, \ldots$$

In this study we will be using an algebraically equivalent form so we can have one of our parameters be the dispersion parameter.

$$P(X = x) = \frac{\Gamma(x + r)}{\Gamma(r)x!} \left(\frac{\mu}{\mu + r}\right)^x \left(\frac{r}{\mu + r}\right)^r$$

where:

- μ is the mean of the distribution.

- r is the dispersion parameter, also called θ in our paper.
- $\Gamma(x)$ is the Gamma function.

The graph of the PMF is below:

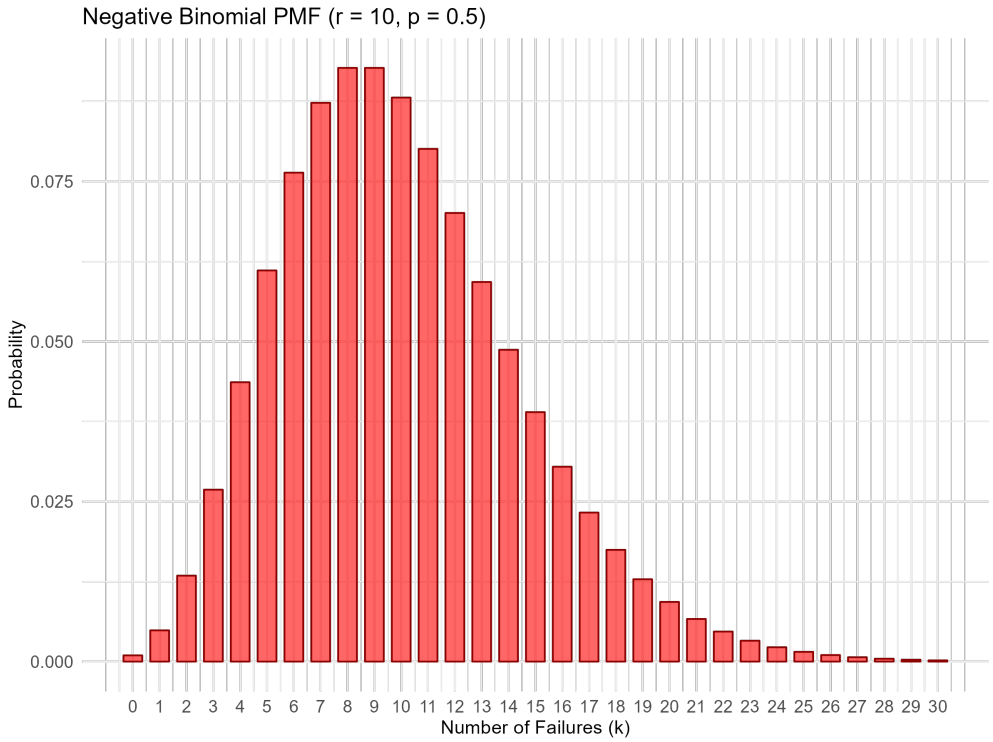

Figure 2.2: PMF of the Negative Binomial

The moment generating function of negative binomial is below:

$$M_{NB}(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$$

Comparing the Poisson and Negative Binomial Distributions

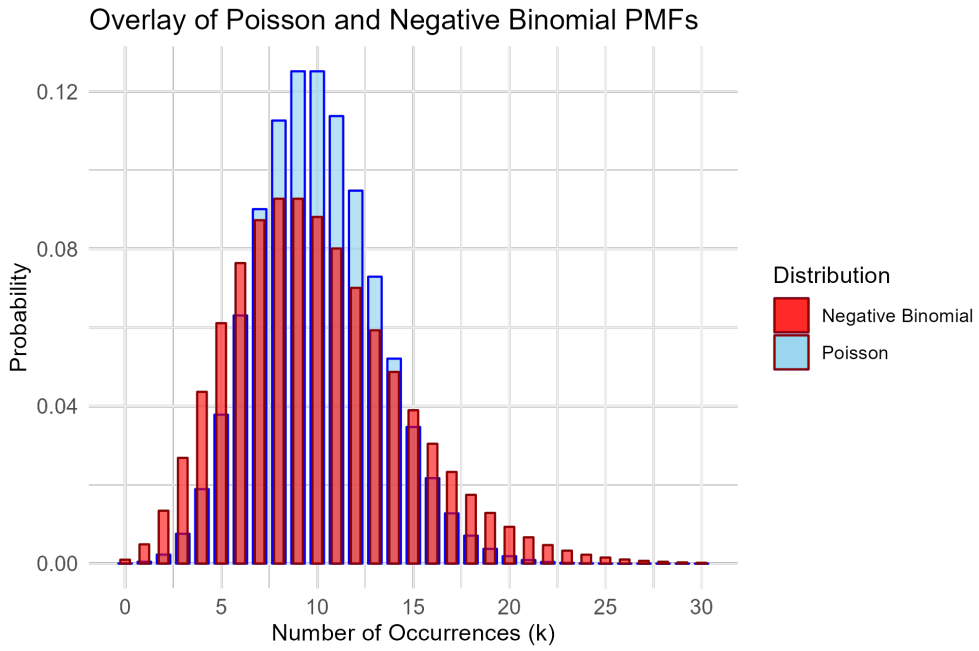

Figure 2.3: Overlay of PMF: Poisson and negative binomial

Figure 2.1 shows the PMF of the Poisson distribution with $\lambda = 10$, Figure 2.2 shows the PMF of the negative binomial distribution with $r = 10$ and $p = 0.5$, and Figure 2.3 shows the two PMFs overlaid one another.

From Figure 2.3 we see that the negative binomial has higher variability than the Poisson. This increase in variability can be seen when comparing $\text{var}[y]_{\text{Poi}}$ and $\text{var}[y]_{\text{negbin}}$.

$$\text{var}[y]_{\text{Poi}} = \lambda$$

$$\text{var}[y]_{\text{negbin}} = \frac{\mu + r}{\mu^2}$$

This extra term in the variance under the negative binomial addresses overdispersion. Because the negative binomial handles overdispersion, it is a good alternative to the Poisson when $\sigma^2 > \mu$. In fact, as the dispersion parameter increases (meaning overdispersion decreases), the negative binomial converges in distribution to the Poisson.

One useful lemma for this proof comes from *Statistical Inference* (Casella and Berger). The lemma states as $\lim_{n\to\infty} a_n = a$ [7],

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n = e^a$$

Proof that negative binomial $\xrightarrow{\text{d}}$ Poisson [7]:

$$\begin{aligned}
\lim_{r\to\infty} M_{\text{NB}}(t) &= \lim_{r\to\infty}\left(\frac{p}{1-(1-p)e^t}\right)^r \\
&= \lim_{r\to\infty}\left(\frac{1-(1-p)}{1-(1-p)e^t}\right)^r \\
\text{If } r(1-p)=\lambda \to 1-p &= \frac{\lambda}{r}, \\
&= \lim_{r\to\infty}\left(\frac{1-(1-p)}{1-(1-p)e^t}\right)^r \\
&= \lim_{r\to\infty}\frac{\left(1+\frac{1}{r}\left(-\lambda\right)\right)^r}{\left(1+\frac{1}{r}e^t\left(-\lambda\right)\right)^r} \\
&= \lim_{r\to\infty}\frac{e^{-\lambda}}{e^{-\lambda e^t}} \\
&= e^{\lambda(e^t-1)} \\
&= M_{\text{Poi}}(t)
\end{aligned}$$

2.1.2 Generalized Linear Models

Generalized linear models (GzLM) relax the assumption of a normally distribution random error term. Instead of assuming the normal distribution as the underlying distribution, we may also consider other distributions such as the Poisson and negative binomial.

Poisson Regression

The first model we tested is Poisson regression, which is a linear model that is used to analyze count data. Poisson regression models the relationship between a count variable and one or more predictor variables. Poisson regression assumes the outcome variable follows the Poisson distribution, which makes it take on the same assumptions as the Poisson distribution, which are the data must be a count and equidispersed. Even though it is highly restrictive with the equidispersion assumption, it is the go to model for count data due to how it can adapt with the common issues that count data has such as clustering, zero-inflation, and overdispersion [2]. The regression formula for Poisson is below:

$$\ln(y) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

Negative Binomial Regression

Negative binomial regression is a linear model that also analyzes count data. Similarly to Poisson regression, it models the relationship between a count variable, which follows the negative binomial distribution, and one or more predictor variables. However, unlike Poisson regression that assumes equidispersion, negative binomial regression adds an extra variable so it can handle overdispersed data. This commonly makes negative binomial regression a better choice when the data exhibits the variance being greater than the mean. The formula for negative binomial regression is below:

$$\ln(y) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$$

2.1.3 Simulation Study

To explore the differences in these models, a simulation study was conducted. A simulation study is the process of creating synthetic, or made up, data. This was done using R [8] with the packages tidyverse [9], MASS [10], and broom[11].

Our goal was to make a function that would repeatedly generate datasets while also allowing a way to change the relationship between μ and σ^2. The predictor, x, was simulated first, with $x \sim N(0, 1)$. Then the mean was defined as $\mu = \exp\{\beta_0 + \beta_1 x, \}$, leading to simulating the count outcome, $y \sim \text{NegBin}(\mu, \theta)$. Note that θ is the dispersion parameter and allows us to change the relationship between μ and σ^2. After the data was simulated, each dataset was used to construct a model under both the Poisson and negative binomial distributions. Regardless of Poisson or negative binomial, we modeled

$$\ln(y) = \beta_0 + \beta_1 x.$$

The simulated raw data and, separately, the analysis results from both Poisson and negative binomial regressions were stored for further analysis. The regression analysis results were examined with respect to bias, mean square error (MSE), type I error, and model selection using AIC.

Bias

Bias is a statistical measurement that measures the systematic deviation of the value that is being tested from the true value. Bias is used to tell the difference between the slopes of models to see how off the estimated value was from the actual value. When performing a simulation study, bias can be accurately measured because we know the true value of the population parameters. In this context, bias is calculated

as

$$\text{bias}(\hat{\beta}) = E[\hat{\beta}] - \beta,$$

where:

- $\hat{\beta}$ is the estimator of the parameter β,
- $E[\hat{\beta}]$ is the expected value of the estimator $\hat{\beta}$,
  - This represents the average value of $\hat{\beta}$ over many samples from the population.
- β is the true value of the parameter being estimated.

In our simulation study we want to see a low bias from both of the models. Since bias takes the estimated value subtracted from the true value, the lower the bias is, the closer the estimated value is to the true value. A low bias means that the models can capture the underlying data, allowing a good fit. A high bias would mean that our estimated value is consistently off from the true value, which can lead to the model underfitting the data. Underfitting the data will cause a poor performance in the model because it fails to capture the complexity of this data [12].

MSE

The Mean Squared Error (MSE) quantifies the the error in the model. In particular, MSE is the average squared difference between the true values and the predicted values of the model. The MSE can be computed as follows,

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2,$$

where:

- y_i is the true value for the i-th observation,
- $\hat{y}_i$ is the predicted value for the i-th observation, and
- n is the number of observations.

MSE can also be examined using the bias,

$$\text{MSE}(\hat{\beta}) = \text{Var}(\hat{\beta}) + \Big(\text{Bias}(\hat{\beta})\Big)^2,$$

where:

- $\text{var}(\hat{\beta})$ is the variance of the estimator and
- $\text{bias}(\hat{\beta})$ is the bias of the estimator.

Within our study we hope to see a low MSE for both of our models. Low MSE shows that the model is accurate in predicting the observed data. On the contrary, a high MSE shows that the model is not accurate in predicting the data. Similar to a high bias, a high MSE can show that the model is not a good fit for the underlying data [13].

Type I Error

Type I error is a false positive in hypothesis testing. That means we reject the null hypothesis when it was, in fact, true. This is typically compared to finding an innocent person guilty. The rate of false positives is referred to as α. It is typical to use $\alpha = 0.05$ in hypothesis testing. This means that we are willing to withstand drawing an incorrect conclusion 5% of the time.

In this project, we are interested in estimating β, so the corresponding hypotheses are,

$$H_0 : \ \beta = 0$$
$$H_1 : \ \beta \neq 0$$

Because type I error is based on incorrect rejection of the null hypothesis, we should consider what causes the type I error to increase: the p-value determines rejection and the test statistic determines the p-value. In the case of Poisson and negative binomial regressions, Wald's z is used to test the slope, β. Wald's z is as follows,

$$z_0 = \frac{\hat{\beta}}{\text{SE}_{\hat{\beta}}}$$

Based on the formula, it can be observed that as standard error decreases, the z statistic increases. As the z statistic increases, the p-value decreases, inflating the type I error rate. Thus, if the standard error is underestimated, we are necessarily increasing the type I error rate.

To examine type I error, we found the rejection rate of both models under each simulated scenario. Each estimated model was used to test $H_0 : \ \beta = 0$ at $\alpha = 0.05$. Then, the proportion of models that rejected was calculated. When comparing the Poisson and negative binomial regressions, it is known that in overdispersed data, the Poisson estimates a smaller standard error than with the negative binomial. That implies that there should be an increase in the type I error when applying the Poisson distribution if the data is overdispersed.

Akaike Information Criterion

The Akaike Information Crierion (AIC) measures "how good" a model fits. It takes into account both goodness of fit and parsimony, or the complexity of the model. One thing to note is that the actual value for AIC is meaningless on its own, but is rather used to compare the fit of different models. The lowest value of the AIC signifies the "best" fitting model [14]. AIC is calculated as follows, The Akaike Information Criterion (AIC) is given by the formula:

$$\text{AIC} = 2k - 2\ln(L(\hat{\theta})),$$

where k is the number of parameters in the model, and $L(\hat{\theta})$ is the likelihood function evaluated at the estimated parameters $\hat{\theta}$.

2.1.4 Definition of Parameters

For the simulation, we held the intercept and slope constant at $\beta_0 = 1.5$ and $\beta_1 = 0.25$, respectfully. Then, we considered three sample sizes, $n \in \{25, 100, 500\}$, and five dispersion parameters, $\theta \in \{1, 10, 50, 500, 2000\}$. The dispersion parameters correspond to the following scenarios:

- $\theta = 2000$: $\mu = \sigma^2$
- $\theta = 500$: $\mu \approx \sigma^2$
- $\theta = 50$: $\mu < \sigma^2$
- $\theta = 10$: $\mu << \sigma^2$
- $\theta = 1$: $\mu <<< \sigma^2$

2.2 Results

2.2.1 Expected Results

The purpose of this simulation study was to verify what is suggested by statistical theory and determine "how bad" the results are as a function of the overdispersion. What we expect to see is at high θ values, or low dispersion, the results from Poisson and negative binomial are similar to each other, and at low θ values, or high dispersion, we expect to see more differences between the two methods. Generally, the slopes of the two models are similar, even at high levels of overdispersion. Given this fact, we expect to see the results

of bias and MSE for both models to be similar to each other. We also expect to see the two models differ when it comes to the standard errors. At high levels of overdispersion, we expect to see the Poisson model underestimate standard errors, which will lead drawing incorrect conclusions. These incorrect conclusions can increase the risk of type I error, which leads to invalid results. When comparing these models under AIC, we expect to see negative binomial showing better results at high overdispersion due to the extra dispersion parameter it offers. As θ gets higher, and the data becomes equidispersed expect to see the Poisson model getting better results due to the simplicity of the model.

2.2.2 Bias Results

Figure 2.4: Bias Results: Poisson and Negative Binomial

Figure 2.4 displays the average bias across the sample sizes. The lines on the graph represent the different values of θ. We can see that the Poisson does have a lower bias at every θ, however, this analysis uses count data, so a 0.04 difference will not greatly affect the models.

2.2.3 MSE Results

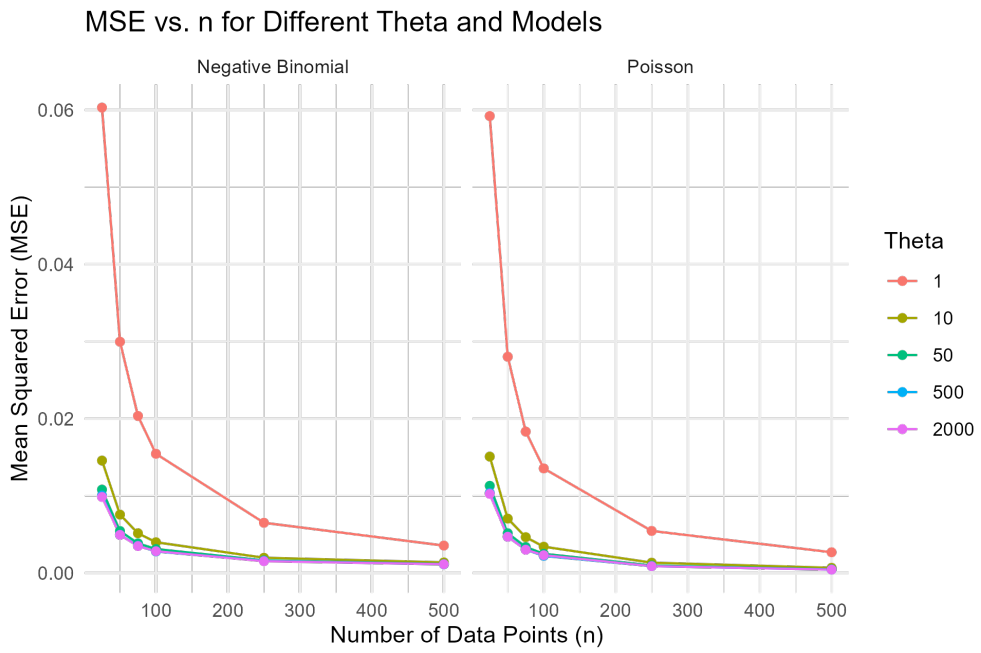

Figure 2.5: MSE Results: Poisson and Negative Binomial

Figure 2.5 displays the average MSE across the sample sizes. The lines on the graph represent the different values of θ. These results are similar to the bias results: the differences in MSE are very small and represent count data. That is, the differences are so small that it doesn't affect the overall model efficiency. Unsurprisingly, the MSE for the Poisson is slightly less than the MSE for the negative binomial across all values of θ. This is unsurprising because we know that the Poisson regression model is underestimating the standard errors.

2.2.4 Standard Error Results

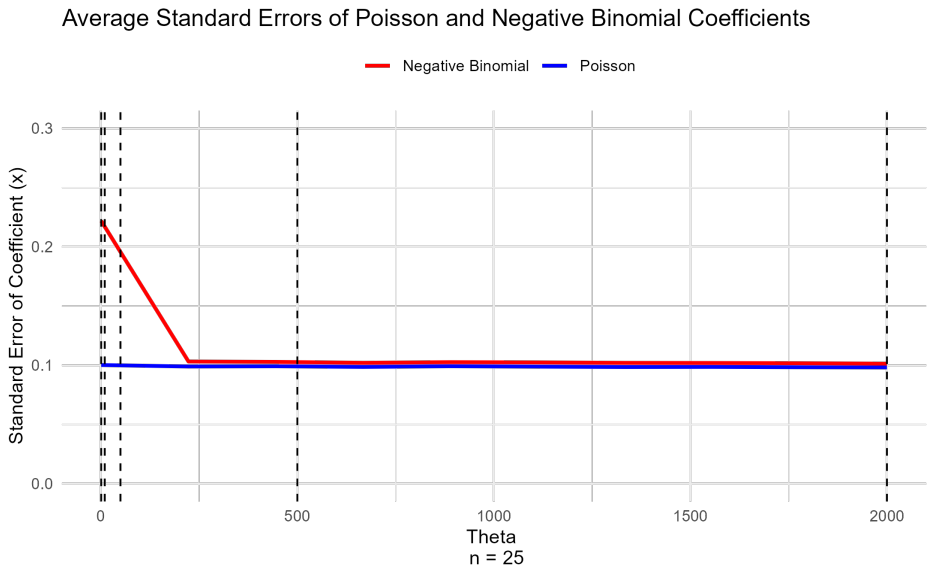

Figure 2.6: Standard Error Results: n = 25

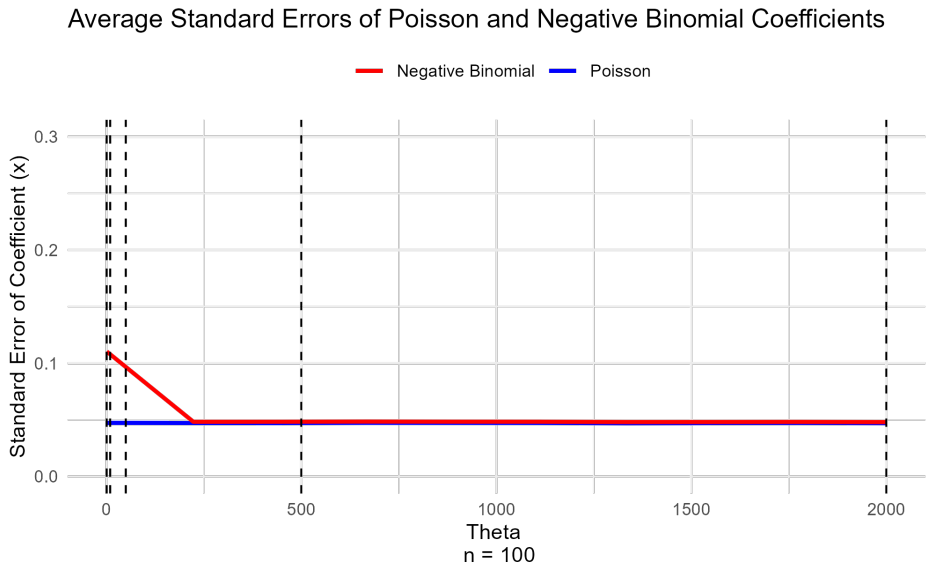

Figure 2.7: Standard Error Results: n = 100

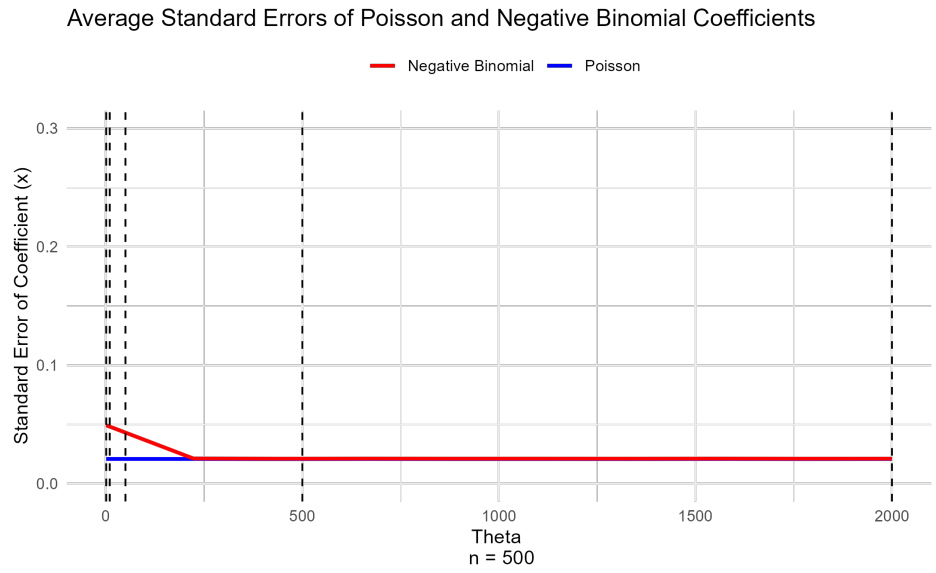

Figure 2.8: Standard Error Results: n = 500

Figures 2.6, 2.7, and 2.8 show the average $SE_{\hat{\beta}}$ across the values of θ. The lines are determined by which distribution was applied. The vertical black dashed lines indicate the values of θ under consideration.

On all three sample sizes, we see that negative binomial has higher standard errors than Poisson at low θ values, or high overdispersion. The difference in standard errors difference between the two models begins to decrease as sample size increases. This makes sense because we know that as sample size increases, error in estimation decreases.

From these graphs we can conclude that Poisson underestimates standard errors at low θ values, leading to high levels of type I error. Of note, around $\theta = 250$, we see the models begin to converge, shedding light on the range where it might be possible to use Poisson over negative binomial, even if some overdispersion is present.

2.2.5 Rejection Rate Results

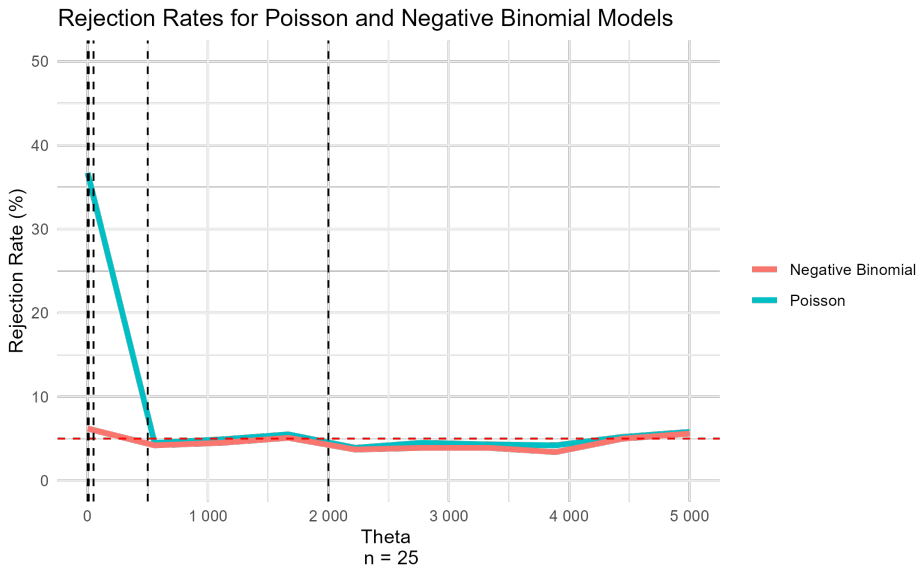

Figure 2.9: Rejection Rate Results: n = 25

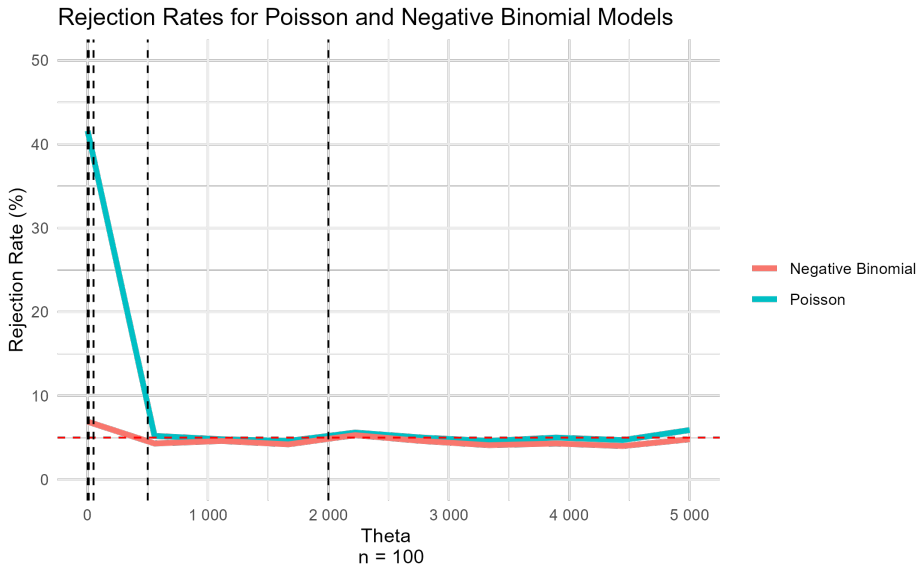

Figure 2.10: Rejection Rate Results: n = 100

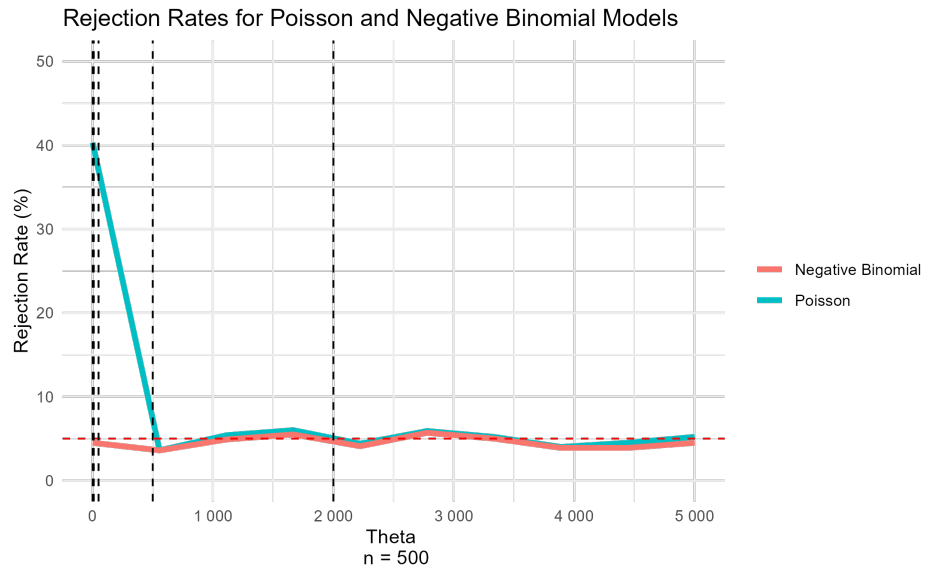

Figure 2.11: Rejection Rate Results: n = 500

Figures 2.9, 2.10, and 2.11 display the rejection rates across θ for the three sample sizes. The black dashed lines represent the θ values used in the simulation study. The red dashed line to represents the 5% rejection rate that is expected.

Across all sample sizes, we see a very high rejection rate for Poisson at high overdispersion ($\theta = 1$). Using the Poisson when the outcome is highly overdispersed, the Poisson will incorrectly determine significance 30-40% of the time. This shows that the type I error is inflated under the Poisson, invalidating results. The negative binomial hovers around 5% rejection line, which was initially expected. Finally, we see the models start to converge at $\theta = 500$, which is the $\mu \approx \sigma^2$ case. At this value of θ the models are providing similar estimates and Poisson can be used.

2.2.6 AIC results

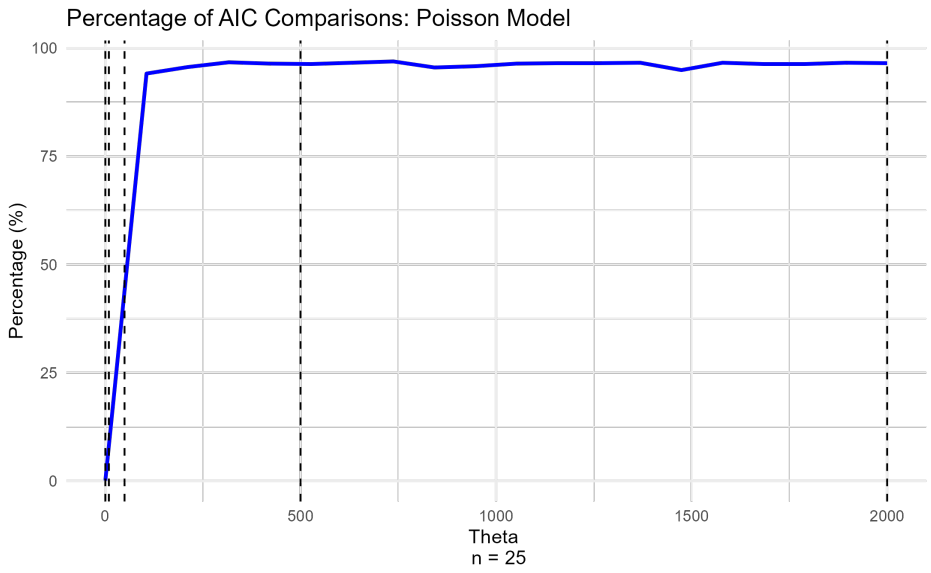

Figure 2.12: AIC Results: n = 25

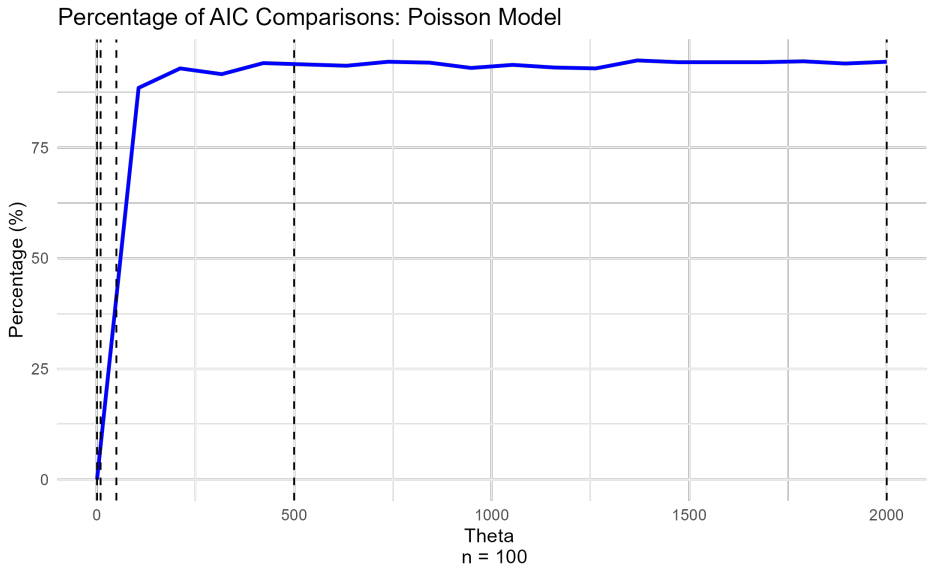

Figure 2.13: AIC Results: n = 100

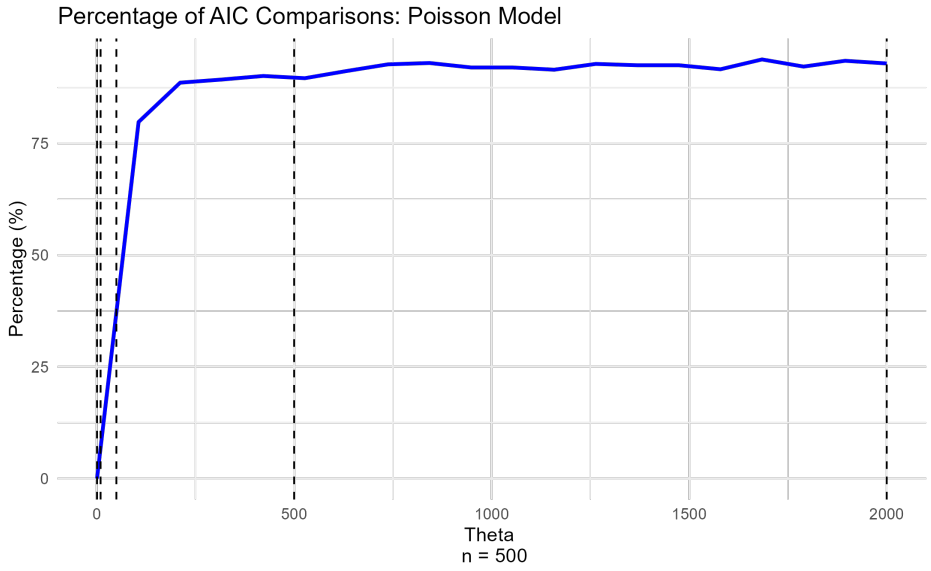

Figure 2.14: AIC Results: n = 500

Figures 2.12, 2.13, and 2.14 show the percentage of the time that the AIC pointed to the Poisson distribution. That is,

$$\text{AIC}_{\text{Poi}} < \text{AIC}_{\text{NB}}.$$

Examining the graphs, we see that at low θ values (high overdispersion), the negative binomial is picked 75-100% of the time. As θ increases, the AIC begins to point towards the Poisson more. Even when $\theta = 500$ ($\mu \approx \sigma^2$), the Poisson is picked more than 90% of the time. Looking closer at the AIC values, we see that as θ increases the Poisson and negative binomial models converge and the values of AIC are very close to each other. Poisson is picked more often due to estimating fewer parameters than in the negative binomial.

Chapter 3

Conclusions

3.1 Summary of Key Findings

In our simulation study, our results agreed with statistical theory. The bias and MSE results showed that there wasn't a major difference between the two models, which was expected as the slopes themselves aren't affected when the data is overdispersed. In the standard error results, we saw that negative binomial had a higher SE than Poisson for $\theta \in \{1, 10, 50\}$, where the data was overdispersed. This visualizes the underestimation of the standard error.

We saw similar results in the rejection rate graphs as well, which showed when the data was overdispersed, the Poisson model would reject the null, $H_0 : \beta = 0$, 30-40% of the time. In the vast majority of the simulated datasets, the AIC suggested that negative binomial was the better fit when the data was overdispersed. Based on these results, we can conclude that when the data is overdispersed and Poisson is used, our conclusions may not be accurate.

One thing to highlight is that in the $\theta = 500$ case, where $\mu \approx \sigma^2$, Poisson and negative binomial are behaving similarly. Because there is not a clear advantage in estimating an additional parameter with the negative binomial, we recommend using Poisson regression for simplicity.

3.2 Suggestions for Further Study

To further this research, this study can be tailored to a specific scientific application using real world data and parameters. To better examine what is happening between $\theta = 50$ and $\theta = 500$, additional simulations can be performed with $50 \leq \theta \leq 500$. We can also look into overdispersion in other models, such as the quasi-Poisson, and zero-inflated models.

Bibliography

[1] J. Hilbe, *Negative Binomial Regression*. Cambridge University Press, 2011.

[2] J. Hilbe, *Modeling Count Data*. Cambridge books online, Cambridge University Press, 2014.

[3] N. E. Breslow, "Extra-poisson variation in log-linear models," *J. R. Stat. Soc. Ser. C. Appl. Stat.*, vol. 33, no. 1, p. 38, 1984.

[4] D. R. Cox, "Some remarks on overdispersion," *Biometrika*, vol. 70, no. 1, pp. 269–274, 1983.

[5] A. Palmer, J. M. Losilla, J. Vives, and R. Jiménez, "Overdispersion in the poisson regression model," *Methodology (Gott.)*, vol. 3, pp. 89–99, jan 2007.

[6] S. Owusu, "Analysis of the effects of overdispersion in population dynamics," Master's thesis, Jomo Kenyatta University of Agriculture and Technology, 2020.

[7] G. Casella and R. Berger, *Statistical Inference*. Duxbury Resource Center, 2001.

[8] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.

[9] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019.

[10] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*. New York: Springer, fourth ed., 2002. ISBN 0-387-95457-0.

[11] D. Robinson, A. Hayes, and S. Couch, *broom: Convert Statistical Objects into Tidy Tibbles*, 2023. R package version 1.0.5.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2nd ed., 2009.

[13] A. Zeileis and C. Kleiber, "A short introduction to the poisson and negative binomial models," *Statistical Software and Computational Biology*, vol. 2, no. 1, pp. 1–5, 2007.

[14] D. J. Beal, "Information criteria methods in sas for multiple linear regression models," in *Proceedings of the Annual Conference of the SouthEast SAS Users Group*, 2007.