BACHELOR

Estimating quantiles of the negative binomial distribution

Jacobs, Luuk C.P.

Award date:
2016

Estimating Quantiles of the Negative Binomial Distribution

Eindhoven University of Technology

Bachelor Final Project

2WH40

Author:
Luuk Jacobs, ID 0853299

Supervisor:
Dr. A. Di Bucchianico

July 8, 2016

Abstract

In laboratoria drugs are produced. These drugs should be produced in an environment that contains a limited amount of bacteria. The amount of bacteria should be monitored continuously such that it does not increase over time. A method to do this is to describe the sample with a probability distribution. The Poisson distribution is a frequently used probability distribution. But this distribution has the disadvantage that the mean is equal to the variance. In reality this is not true in most cases. The negative binomial distribution is also a distribution that can be used for counting, and for this distribution the mean is not necessarily equal to the variance. However estimating an upper bound for the amount of bacteria becomes much more difficult with the negative binomial distribution. There are multiple parametrizations for the negative binomial distribution, and in literature there is no comparison between the different parametrizations. In this report we derive two forms of the negative binomial distribution and analyze the parametrizations that are used in the literature. For each parametrization we estimate the parameters with the method of moments and the method of maximum likelihood. For two parameterizations there is also the quasi maximum likelihood estimator, but it will not be used later in the report for presenting results. After estimating the parameters we describe two methods for estimating quantiles for the negative binomial distribution. The first method is transforming the data so that the data is approximately normal. Quantiles can then be calculated by using quantiles of the normal distribution. The second method is to make use of a chi-squared distribution. The built-in function of R can also give an exact quantile for the estimator. By comparing the estimators for parameters and quantiles with true values we will see that the maximum likelihood estimator gives the best results for each parametrization. The best results are in general given by parametrization 2. So the best way to estimate quantiles is to use parametrization 2 in combination with the method of maximum likelihood.

Contents

Chapter 1

Introduction

In laboratoria many drugs are produced to cure patients. These patients are often very ill and have a poor resistance against harmful organisms. Therefore it is necessary that these drugs are produced in a sterile environment, or in an environment that has not more than a certain amount of bacteria. It is also not allowed that the amount of bacteria increases in time. So there is a need to monitor how many bacteria there are in a room, and to estimate an upper bound. By continuously monitoring how many bacteria there are in the room, one can ring an alarm when there is a shift in the data. In this case the amount of bacteria exceeds the estimated upper bound too frequently. The amount of bacteria should not exceed the upper bound more than for example in 1 of the 1000 samples. If this happens more frequently, than this means that the amount of bacteria is increasing over time and action should be taken.

A method for modeling counts of micro organisms like bacteria is to make use of a probability distribution. A well-known distribution for counting is the Poisson distribution. The Poisson distribution is a discrete distribution that has a simple form, and therefore calculations may be easy. However, the Poisson distribution has a disadvantage. When using the Poisson distribution then the data should satisfy that the expected value is equal to the variance. In practice this is often not true. In many situations there is a higher variance than the mean. Therefore the Poisson distribution may not be the most suitable distribution for counting bacteria.

A probability distribution that can have the property of having larger variance than the mean, is the negative binomial distribution. The advantage of the negative binomial distribution is that this distribution takes all kinds of overdispersion into account. The disadvantage might be that there exist multiple parametrizations. This means that one has to be careful using literature, because every author may use another parametrization. Also estimating parameters for the negative binomial distribution differs between the multiple forms. It might be that those differences in parameters may cause big differences in estimated quantiles. These quantiles are needed to be estimateed if we want an upperbound for the amount of bacteria.

In literature there are several articles about the negative binomial distribution. Pieters et al. (1977) and Clark and Perry (1989) wrote about parameter estimation, while Anscombe (1948) and Hoffman (2003) wrote about estimating quantiles. They all used their own parametrization. The methods that they discuss are not discussed by them for other parametrizations. For that reason there is no comparison between the different parametrizations.

In this report we will answer the following questions:

- Which paramerizations of the negative binomial distribution are known and often used?
- How should the parameters of each form of the negative binomial distributed be estimated?
- Which estimating methods are available to estimate quantiles of the negative binomial distribution?
- Which parametrization of the negative binomial distribution is most suitable to estimate the quantiles?
- Which method of estimating parameters gives the best quantiles within each parametrization?

Chapter 2

The Negative Binomial Distribution

In this chapter we explain in detail the required background for the negative binomial distribution. We will derive two ways in which the negative binomial distribution may appear. The first variation will be based on counting succes and failures with a certain probability. The second method will describe the Poisson-gamma mixture model. Although there are many more ways to derive the negative binomial distribution, we derive now only two of them. The two forms that will be discussed in this chapter are probably the easiest forms to understand, because they make use of other distributions that are well-known.

2.1 Negative binomial based on successes and failures

The first interpretation of the negative binomial distribution is the probability that there are r failures before the k^{th} success occurs in independent Bernoulli trials. The negative binomial distribution has some properties of the binomial distribution and the geometric distribution. The binomial distribution is a series of Bernoulli experiments which counts the number of successes given that there are n trials. The geometric distribution is also connected to a series of Bernoulli experiments, and it counts the number of failures before the first succes. So the description of the negative binomial distribution stated above seems very similar to those two distributions. This is because the counting until success part is very similar to the geometric distribution part. The part of the given number of successes is similar to the binomial distribution. Now we will derive the probability mass function of this version of the negative binomial distribution.

The negative binomial distribution has always a succes in the last attempt (the last Bernoulli trial), this is the x^{th} trial, and the r^{th} succes. This means that there are $r-1$ successes in the previous $(x-1)$ trials. Then the number of possible outcomes for the previous trials is $\binom{x-1}{r-1}$. We denote p as the succes probability of one single Bernoulli experiment. Then each combination that is included in the $\binom{x-1}{r-1}$ possible outcomes has probability $p^r(1-p)^{x-r}$. Now we have that x is the number of trials and r is the number of successes. We can also define y as the number of failures, so $y = x - r$. We will verify that this y is the same as the y in the Poisson-gamma mixture model that we are going to describe in the next section.

For the probability mass function of the negative binomial we now have:

$$f(y; p, r) = \binom{y+r-1}{r-1} p^r (1-p)^y \tag{2.1}$$

The parametrization of (2.1) also shows the similarities with the binomial distribution. The structure is similar: Binomial coefficient, success probability to the power the number of successes, failure probability to the power the number of failures. The differences are that the number of trials is not fixed for the negative binomial, and therefore the binomial coefficient differs.

2.2 Negative binomial Poisson-gamma mixture model

Another parametrization of the negative binomial distribution is the Poisson-gamma mixture model. For this parametrization of the negative binomial distribution, we start with a simple Poisson distribution. The probability mass function of a Poisson distribution is given bij:

$$f(y; \lambda) = \frac{e^{-\lambda}\lambda^y}{y!}, \quad y = 0, 1, 2, ... \tag{2.2}$$

For a Poisson distribution we know that the expected value and variance are equal. In practice, count data has larger variance than the expected value. This phenomenon is called overdispersion. This overdispersion can be taken into account by considering a gamma distribution. In (2.2) the expected value of the Poisson distribution was known, namely λ. This means that the total distribution was fixed by the parameter λ. But now we want to drop this property and construct a new distribution. In this distribution we take λ randomly from a gamma distribution (G) with parameters α and β. The result is that we have a series of Poisson distributions with all different parameters that are coming from the gamma distribution. Each Poisson distribution occurs with probability $\mathbb{P}(G = \lambda)$. So we get an integral of Poisson distributions that sums up to 1. In this way we combined a Poisson distribution with a gamma distribution to a new distribution. The gamma distribution is given by:

$$g(\lambda; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \, \lambda^{\alpha-1} \, e^{-\frac{\lambda}{\beta}}, \quad \lambda \geq 0 \tag{2.3}$$

Formula (2.3) contains the gamma function. The gamma function ($\Gamma(x)$) is defined for $x > 0$ as follows:

$$\Gamma(x) = \int_0^\infty e^{-t} \, t^{x-1} \, dt$$

Some properties of the gamma function are:

- $\Gamma(n+1) = n!$ if n is a non-negative integer
- $\Gamma(x+1) = x \, \Gamma(x)$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

The second property also defines the gamma function for negative, non-integer x. The compound distribution can now be constructed as follows:

$$f(y; \alpha, \beta) = \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \, g(\lambda; \alpha, \beta) \, d\lambda \tag{2.4}$$

If we substitute (2.3) into (2.4) we get that:

$$f(y; \alpha, \beta) = \int_0^\infty \frac{e^{-\lambda}\lambda^y}{y!} \, \frac{1}{\Gamma(\alpha)\beta^\alpha} \, \lambda^{\alpha-1} \, e^{-\frac{\lambda}{\beta}} \, d\lambda \tag{2.5}$$

The next thing to do is to evaluate the integral in (2.5). We use the relation $x = \lambda \left(1 + \frac{1}{\beta}\right)$.

$$\begin{aligned}
f(y; \alpha, \beta) &= \int_0^\infty \frac{e^{-\lambda} \, \lambda^y}{y!} \frac{1}{\Gamma(\alpha) \, \beta^\alpha} \, \lambda^{\alpha-1} \, e^{-\frac{\lambda}{\beta}} \, d\lambda \\
&= \frac{1}{\Gamma(\alpha) \, \beta^\alpha \, y!} \int_0^\infty e^{-\lambda(1+\frac{1}{\beta})} \, \lambda^{y+\alpha-1} \, d\lambda \\
&= \frac{1}{\Gamma(\alpha) \, \beta^\alpha \, y!} \frac{1}{1+\frac{1}{\beta}} \int_0^\infty e^{-x} \left(\frac{x}{1+\frac{1}{\beta}}\right)^{(y+\alpha)-1} dx \\
&= \frac{1}{\Gamma(\alpha) \, \beta^\alpha \, y!} \frac{1}{1+\frac{1}{\beta}} \left(\frac{1}{1+\frac{1}{\beta}}\right)^{y+\alpha-1} \int_0^\infty e^{-x} \, x^{(y+\alpha)-1} dx \\
&= \frac{1}{\Gamma(\alpha) \, \beta^\alpha \, y!} \left(\frac{\beta}{\beta+1}\right)^{y+\alpha} \Gamma(y+\alpha) \\
&= \frac{\Gamma(y+\alpha)}{\Gamma(\alpha) \, y!} \left(\frac{\beta}{\beta+1}\right)^{y} \left(\frac{1}{\beta+1}\right)^{\alpha}
\end{aligned}$$

Instead of using the scale parameter β we could also use the rate parameter $\gamma = \frac{1}{\beta}$. Then (2.6) gives an alternative parametrization of the negative binomial distribution based on the mixed Poisson gamma model.

$$\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!} \left(\frac{1}{\gamma+1}\right)^{y} \left(\frac{\gamma}{\gamma+1}\right)^{\alpha} \tag{2.6}$$

Finally, we can also simplify the gamma function terms since y is a nonnegative integer. This will be done as follows.

$$
\begin{aligned}
f(y;\alpha,\beta) &= \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!}\left(\frac{\beta}{\beta+1}\right)^{y}\left(\frac{1}{\beta+1}\right)^{\alpha} \\
&= \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!}\left(\frac{\beta}{\beta+1}\right)^{y}\left(\frac{1}{\beta+1}\right)^{\alpha} \\
&= \frac{(y+\alpha-1)!}{y!(\alpha-1)!}\left(\frac{\beta}{\beta+1}\right)^{y}\left(\frac{1}{\beta+1}\right)^{\alpha} \\
&= \binom{y+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{y}\left(\frac{1}{\beta+1}\right)^{\alpha}
\end{aligned}
\tag{2.7}
$$

So eventually, the probability mass function for the negative binomial distribution becomes:

$$
f(y;\alpha,\beta) = \binom{y+\alpha-1}{\alpha-1}\left(\frac{\beta}{\beta+1}\right)^{y}\left(\frac{1}{\beta+1}\right)^{\alpha} \tag{2.8}
$$

If we take $\alpha = r$ and $\frac{1}{\beta+1} = p$, then we get the same formula as in (2.1). This means that we have derived the negative binomial distribution in two different ways. However, there are many more ways to parametrize the negative binomial distribution. We will not discuss these parametrizations here, but we make use of these parametrizations to estimate parameters and quantiles. An overview of parametrizations used is given in chapter 3.

Chapter 3

Overview of Parametrizations

There are several articles about the negative binomial distribution. For example Clark and Perry (1989) and Pieters et al. (1977) wrote about parameter estimation. However, they used a different parametrization. It is not possible to directly compare the results of those two articles. A translation is needed to compare the articles. It is therefore very helpful to make a translation table. This avoids problems in the future like using the wrong results that are obained by parametrization. It is not only necessary to make this table to prevent using wrong results, but we also need to switch many times between the multiple parametrizations. So calculations will be easier when the translations between the parametrizations are known. In this chapter we will give an overview of the parametrizations that can occur in the negative binomial distribution.

First it is helpful to see how the different probability mass functions look like. Table 3.1 describes the different parametrizations that occur in the articles in Table 3.2, with their probability mass function, expected value and variance.

Parametrization	**PMF**	**Expected Value**	**Variance**
Parametrization 1	$\frac{\Gamma(y+\kappa)}{\Gamma(\kappa)\ y!}\left(\frac{\mu}{\mu+\kappa}\right)^y\left(\frac{\kappa}{\mu+\kappa}\right)^\kappa$	$\mathbb{E}[Y]=\mu$	$\text{Var}(Y)=\mu(1+\frac{\mu}{\kappa})$
Parametrization 2	$\frac{\Gamma(y+\frac{1}{a})}{\Gamma(\frac{1}{a})\ y!}\left(\frac{u}{u+\frac{1}{a}}\right)^y\left(\frac{1}{1+au}\right)^{\frac{1}{a}}$	$\mathbb{E}[Y]=u$	$\text{Var}(Y)=u(1+au)$
Parametrization 3	$\frac{\Gamma(y+k)}{\Gamma(k)\ y!}\ (1-p)^y\ p^k$	$\mathbb{E}[Y]=\frac{k(1-p)}{p}$	$\text{Var}(Y)=\frac{k(1-p)}{p^2}$
Parametrization 4	$\frac{\Gamma(y+\alpha)}{\Gamma(\alpha)\ y!}\left(\frac{\beta}{\beta+1}\right)^y\left(\frac{1}{\beta+1}\right)^\alpha$	$\mathbb{E}[Y]=\alpha\beta$	$\text{Var}(Y)=\alpha\beta+\alpha\beta^2$
Parametrization 5	$\frac{\Gamma(y+\frac{\gamma\delta}{\delta-1})}{\Gamma(\frac{\gamma\delta}{\delta-1})y!}\left(1-\frac{1}{\delta}\right)^y\left(\frac{1}{\delta}\right)^{\frac{\gamma\delta}{\delta-1}}$	$\mathbb{E}[Y]=\gamma\delta$	$\text{Var}(Y)=\gamma\delta^2$

Table 3.1: Categorization of parametrizations

In Table 3.1 all parameterizations can be found for the literature of Table 3.2 that wrote about the negative binomial distribution. From Table 3.1 we also see that the Poisson-Gamma mixture model is the same as Parametrization 4, and the negative binomial model based on successes and failures is the same as Parametrization 3. Although there are differences in notation in the literature it can be noticed that some parametrizations are the same. For the sake of clarity, we make some categories so we know which articles use the same parametrization. The articles in Table 3.2 wrote about estimating parameters or estimating quantiles of the negative distribution. So these might be helpful for comparing the different parametrizations of the negative binomial distribution. And if one wants to use one of these articles it takes less effort to understand the distribution.

Parametrization 1	Parametrization 2	Parametrization 3
Piegorsch (1990)	Piegorsch (1990)	Guenther (1972)
Anscombe (1948)	Clark and Perry (1989)	Best and Gipps (1974)
Clark and Perry (1989)	Hoffman (2003)	
Willson et al. (1984)	Anraku and Yanagimoto (1990)	

Parametrization 4	Parametrization 5
Scheaffer (1976)	Guenther (1972)
Pieters et al. (1977)	
Blom (1954)	

Table 3.2: Categorization of parametrizations

In Table 3.2 we see some articles twice. The reason for this is that these articles start with a certain parametrization, but then use another parametrization to avoid problems. It could also be that another parametrization is easier to calculate with or that it has a special form. For example Piegorsch (1990) and Clark and Perry (1989) start both with Parametrization 1, but switch then to Parametrization 2. It is the question whether it is indeed a good idea to do this switch. In Table 3.3 we present an overview of all the translations. The horizontal parametrizations are the parametrizations in the articles, and the vertical parametrizations are the parametrizations that one wishes to translate to. This tabel is very helpful to do all the translations automatically, because this table is easy to implement in software like R. So after implementing this table, it is easy translating parameters.

	Param 1	**Param 2**	**Param 3**	**Param 4**	**Param 5**
Param1		$a = \frac{1}{\kappa}$, $u = \mu$	$k = \kappa$, $p = \frac{\kappa}{\kappa+\mu}$	$\alpha = \kappa$, $\beta = \frac{\mu}{\kappa}$	$\gamma = \frac{\kappa\mu}{\mu+\kappa}$, $\delta = \frac{\mu}{\kappa} + 1$
Param2	$\kappa = \frac{1}{a}$, $\mu = u$		$k = \frac{1}{a}$, $p = \frac{1}{1+au}$	$\alpha = \frac{1}{a}$, $\beta = ua$	$\gamma = \frac{u}{ua+1}$, $\delta = ua + 1$
Param3	$\kappa = k$, $\mu = \frac{k}{p} - k$	$a = \frac{1}{k}$, $u = \left(\frac{1}{p} - 1\right) k$		$\alpha = k$, $\beta = \frac{1}{p} - 1$	$\gamma = (1-p)k$, $\delta = \frac{1}{p}$
Param4	$\kappa = \alpha$, $\mu = \kappa\beta$	$a = \frac{1}{\alpha}$, $u = \frac{\beta}{\alpha}$	$k = \alpha$, $p = \frac{1}{\beta+1}$		$\gamma = \frac{\alpha\beta}{\beta+1}$, $\delta = 1 + \beta$
Param5	$\kappa = \frac{\gamma\delta}{\delta-1}$, $\mu = \gamma\delta$	$a = \frac{\delta-1}{\gamma\delta}$, $u = \delta\gamma$	$k = \frac{\gamma\delta}{\delta-1}$, $p = \frac{1}{\delta}$	$\alpha = \frac{\gamma\delta}{\delta-1}$, $\beta = \delta - 1$	

Table 3.3: Translation table

Table 3.4 links all articles of Table 3.2 to the parametrization that arose from the Poisson gamma mixture. All probability mass functions are exactly in the form they are stated in the articles of Table 3.2. Since the link is made to the Poisson-gamma mixture model (parametrization 4), one can derive the parameters for the other parametrziations. Some articles have a probability density function that is hard to read. So in Table 3.4 the probability mass function are rewritten to a simpler form. From these forms it is easier to see which parametrization it is.

Literature	**PDF**	**Parameters**	**Expected value, variance**
Piegorsch (1)	$\frac{\Gamma(y+\kappa)}{\Gamma(\kappa)y!}\left(\frac{\mu}{\mu+\kappa}\right)^y\left(1+\frac{\mu}{\kappa}\right)^{-\kappa}$	$\alpha=\kappa,\ \beta=\frac{\mu}{\kappa}$	$\mathbb{E}[Y]=\mu$, $\text{Var}(Y)=\mu(1+\frac{\mu}{\kappa})$
Piegorsch (2)	$\frac{\Gamma(y+\frac{1}{a})}{\Gamma(\frac{1}{a})y!}\left(\frac{a\mu}{a\mu+1}\right)^y(1+a\mu)^{-\frac{1}{a}}$	$\alpha=\frac{1}{a},\ \beta=a\mu$	$\mathbb{E}[Y]=\mu$, $\text{Var}(Y)=\mu(1+a\mu)$
Anscombe	$\frac{\Gamma(r+k)}{\Gamma(k)r!}\left(\frac{m}{m+k}\right)^r\left(1+\frac{m}{k}\right)^{-k}$	$\alpha=k,\ \beta=\frac{m}{k}$	$\mathbb{E}[R]=m$, $\text{Var}(R)=m(1+\frac{m}{k})$
Guenther (1)	$\frac{\Gamma(x+k)}{\Gamma(k)x!}(p)^k(1-p)^x$	$\alpha=k,\ \beta=\frac{1-p}{p}$	$\mathbb{E}[X]=\frac{k(1-p)}{p}$, $\text{Var}(X)=\frac{k(1-p)}{p^2}$
Guenther (2)	$\frac{\Gamma(x+\frac{ab}{b-1})}{\Gamma(\frac{ab}{b-1})x!}\left(\frac{1}{b}\right)^{\frac{ab}{b-1}}\left(1-\frac{1}{b}\right)^x$	$\alpha=\frac{ab}{b-1},\ \beta=b-1$	$\mathbb{E}[X]=ab$, $\text{Var}(X)=ab^2$
Scheaffer	$\frac{\Gamma(x+k)}{\Gamma(k)x!}\left(\frac{p}{1+p}\right)^x\left(\frac{1}{1+p}\right)^k$	$\alpha=k,\beta=p$	$\mathbb{E}[X]=kp$, $\text{Var}(X)=kp+kp^2$
Pieters	$\frac{\Gamma(x+k)}{\Gamma(k)x!}\left(\frac{p}{q}\right)^x\left(\frac{1}{q}\right)^k$	$\alpha=k,\beta=p=q-1$	$\mathbb{E}[X]=kp$, $\text{Var}(X)=kp+kp^2$
Clark (1)	$\frac{\Gamma(y+k)}{\Gamma(k)y!}\left(\frac{\mu}{\mu+k}\right)^y\left(1+\frac{\mu}{k}\right)^{-k}$	$\alpha=k,\beta=\frac{\mu}{k}$	$\mathbb{E}[Y]=\mu$, $\text{Var}(Y)=\mu(1+\frac{\mu}{k})$
Clark (2)	$\frac{\Gamma(y+\frac{1}{a})}{\Gamma(\frac{1}{a})y!}\left(\frac{\mu}{\mu+\frac{1}{a}}\right)^y(1+a\mu)^{-\frac{1}{a}}$	$\alpha=\frac{1}{a},\beta=a\mu$	$\mathbb{E}[Y]=\mu$, $\text{Var}(Y)=\mu(1+a\mu)$
Hoffman	$\frac{\Gamma(x+\frac{1}{k})}{\Gamma(\frac{1}{k})x!}\left(\frac{k\mu}{1+k\mu}\right)^x\left(\frac{1}{1+k\mu}\right)^{\frac{1}{k}}$	$\alpha=\frac{1}{k},\beta=k\mu$	$\mathbb{E}[X]=\mu$, $\text{Var}(X)=\mu(1+k\mu)$
Anraku	$\frac{\Gamma(x+\frac{1}{\theta})}{\Gamma(\frac{1}{\theta})x!}\left(\frac{\theta\mu}{1+\theta\mu}\right)^x\left(\frac{1}{1+\theta\mu}\right)^{\frac{1}{\theta}}$	$\alpha=\frac{1}{\theta},\beta=\theta\mu$	$\mathbb{E}[X]=\mu$, $\text{Var}(X)=\mu(1+\frac{\mu}{\theta})$
Blom	$\frac{\Gamma(N+A)}{\Gamma(\frac{1}{N})A!}\left(\frac{P}{Q}\right)^A\left(\frac{1}{Q}\right)^N$	$\alpha=N,\beta=P=Q-1$	$\mathbb{E}[A]=NP$, $\text{Var}(A)=NP(1+P)$
Best and Gipps	$\frac{\Gamma(x+k)}{x!\Gamma(k)}p^k(1-p)^x$	$\alpha=k,\ \beta=\frac{1-p}{p}$	$\mathbb{E}[X]=\frac{k(1-p)}{p}$, $\text{Var}(X)=\frac{k(1-p)}{p^2}$
Willson	$\frac{\Gamma(x+k)}{\Gamma(k)x!}\left(\frac{\mu}{\mu+k}\right)^x\left(1+\frac{\mu}{k}\right)^{-k}$	$\alpha=k,\ \beta=\frac{\mu}{k}$	$\mathbb{E}[X]=\mu$, $\text{Var}(X)=\mu(1+\frac{\mu}{k})$

Table 3.4: Parameters in literature

Although there are different ways of writing the negative binomial distribution, there are many articles that have written about the same parametrization, but it was written in a different form. So when using an article of the negative binomial distribution it is always helpful to check whether the parametrization is already known and whether it is the same as in the article of another author.

Chapter 4

Estimators

There are several ways to estimate the parameters of the negative binomial distribution. The most common methods are the method of moments and the method of maximum likelihood. Because there are many possibilities to parametrize the negative binomial distribution, there may be differences in the parameters. It may also give different results if there is a lack of invariance. For some parametrizations it may be harder to use certain methods than for others. Estimating the parameters in a good way is very important because bad estimation can lead to bad results for the quantiles if we use the same parameters to estimate the quantiles later. Little difference in parameters might lead to large differences in the quantiles. So a good estimation method is necessary. In this chapter we describe serveral methods to find the parameters of the negative binomial distribution. Also a simple example is given for the need of all this knowledge of estimators. This example will give an explicit expression for the estimators of the exponential distribution and it shows very well the problem we will also have for the negative binomial distribution.

4.1 Method of Moments (MME)

The easiest method for estimating parameters is the method of moments. We denote the j-th sample moment by w_j. Then the sample moments can be calculated by $w_j = \frac{1}{n}\sum_{i=1}^{n}(x_i^j)$, where $(x_1, x_2, ..., x_n)$ is an i.i.d. sample of length n. We define the sample mean m as $m = w_1$, and the sample variance s^2 as $s^2 = w_2 - w_1^2$. We will express all parameters in terms of m and s^2 because these quantities can easily be derived from the data. E.g., Clark and Perry (1989) calculated a moment estimator for Parametrization 2 of (3.2). The moment estimators can be solved by solving two equations that make use of the mean and the variance. The moment estimator for a is given by $\hat{a} = \frac{s^2-m}{m^2}$, and the moment estimator for u is given by $\hat{u} = m$. The reason that they use their second parametrization instead of their first is that there are no problems for the estimator for α when the variance is equal to the mean. So division by zero is prevented. For the other parametrizations the moment estimators can also be calculated. This results in the estimators stated in Table 4.1.

Parametrization 1	$\hat{\kappa} = \frac{m^2}{s^2-m}$	$\hat{\mu} = m$
Parametrization 2	$\hat{a} = \frac{s^2-m}{m^2}$	$\hat{u} = m$
Parametrization 3	$\hat{k} = \frac{m^2}{s^2-m}$	$\hat{p} = \frac{m}{s^2}$
Parametrization 4	$\hat{\alpha} = \frac{m^2}{s^2-m}$	$\hat{\beta} = \frac{s^2-m}{m}$
Parametrization 5	$\hat{\gamma} = \frac{m^2}{s^2}$	$\hat{\delta} = \frac{s^2}{m}$

Table 4.1: Moment estimators

The advantage of the method of moments is that these parameters can be calculated very easily. There are no complicated functions and therefore there will be no numerical errors except in the case that $s^2 = m$ for some parameterizations. However the estimators are definitely not optimal. The estimators may be biased and give therefore results that can quite differ from the original parameters. Also for each parametrization we have to calculate the moment estimators separately. There is no invariance principle known for these estimators, so we can not derive the estimator for parametrization i from parametrization j by simple substitution.

4.2 Maximum Likelihood Estimator (MLE)

Maximum likelihood is a estimating method that determines which parameters of the distribution are most likely. Suppose that $x_1, ..., x_n$ are i.i.d. samples of a negative binomial distribution with probability mass function $f(x; a, \mu)$. The likelihood function is defined as:

$$L(a, \mu) = \prod_{i=1}^{n} f(x_i; a, \mu) \tag{4.1}$$

Then the log-likelihood function is given by:

$$l(a, \mu) = \log(L(a, \mu)) = \log \left(\prod_{i=1}^{n} f(x_i; a, \mu) \right) = \sum_{i=1}^{n} \log \left(f(x_i; a, \mu) \right) \tag{4.2}$$

We find the parameters a and μ by maximizing the log-likelihood function for both parameters. In other words we have to calculate the following two equations and check whether it is indeed a maximum:

$$\frac{\partial l(a, \mu)}{\partial a} = 0 \qquad \frac{\partial l(a, \mu)}{\partial \mu} = 0 \tag{4.3}$$

Piegorsch (1990) derived the maximum likelihood estimator for Parametrization 2 of (3.2). Using the equation (4.2), we know that the log-likelihood function (with irrelevant parts for the parameters omitted) is proportional to:

$$l(a, \mu) = \frac{1}{n} \sum_{i=1}^{n} \frac{\Gamma(y_i + a^{-1})}{\Gamma(a^{-1})} + \overline{y} \log(\mu) - (\overline{y} + a^{-1}) \log(1 + a\mu) \tag{4.4}$$

For calculations we use the following:

$$\frac{\Gamma(y + a^{-1})}{\Gamma(a^{-1})} = a^{-1}(1 + a^{-1}) \cdot ... \cdot (y_i - a^{-1}) \tag{4.5}$$

So if we take the logarithm we get that:

$$\log(\frac{\Gamma(y + a^{-1})}{\Gamma(a^{-1})}) = \sum_{\nu=0}^{y_i - 1} \log \left(\frac{1 + a\nu}{a} \right) \tag{4.6}$$

The estimators can be found by solving the following two equations:

$$\begin{aligned} \frac{\overline{y}}{\mu} - \frac{1 + a\overline{y}}{1 + a\mu} &= 0 \\ \frac{1}{n} \sum_{i=1}^{n} \sum_{\nu=0}^{y_i - 1} \frac{\nu}{1 + a\nu} + \frac{1}{a^2} \log(1 + a\mu) - \frac{\mu(\overline{y} + a^{-1})}{1 + a\mu} &= 0 \end{aligned} \tag{4.7}$$

This results in $\hat{\mu} = \overline{y}$. The estimator of a should be solved by substituting $\hat{\mu}$ into the second formula of (4.7). Willson et al. (1984) stated that Anscombe (1948) had the conjecture that there was only one solution in the case that $s^2 > m$. It seems that this property has not been proven yet. But if this is true then it is numerically easy to obtain the right solution.
A special property of the maximum likelihood estimator is the invariance principle. This property is stated by Zehna (1966)

Theorem 1 *If $\hat{\theta} \in \mathbb{R}^d$ is the maximum likelihood estimator for the parameter $\theta \in \mathbb{R}^d$, then $\hat{g} = \hat{g}(\hat{\theta})$ is the maximum likelihood Estimator of $g(\theta)$, where g is an arbitrary function.*

So for each parametrization we can get the parameters using the estimated parameters and the invariance principle. This is a property that does not hold for the Method of Moments estimator. Therefore the maximum likelihood estimator is easier to use when one can derive the parameters for one parametrization. For this reason we do not state the maximum likelihood method for the other parametrizations. Those likelihoods can be derived just by subsitute all parameters according to Table 3.3 in the equations that have to be solved.

4.3 Maximum Quasi-Likelihood Estimator (QMLE)

The Maximum Quasi-Likelihood Estimator is an estimator that is comparable with a maximum likelihood estimator. The difference is that this estimator can be used when the total distribution is not known but only the variance function is known. Nelder and Pregibon (1987) give the quasi-likelihood function. This function is stated in (4.8)

$$Q^+(y;\mu,v,c) = -\frac{1}{2}\log(2\pi\varphi V(y,v;c)) - \frac{1}{2}\cdot\frac{1}{\varphi}D(y;\mu) \tag{4.8}$$

For the negative binomial distribution they defined V as follows:

$$V(\mu,v) = \frac{\mu\,(\mu+v)}{v} \qquad V(y,v;c) = \frac{(y+v)^2\,(y+c)\,(v+c)}{v^2\,(y+v+c)} \tag{4.9}$$

In this formula $V(\mu,v)$ is the modified variance function. The deviance $D(y;\mu)$ is given by:

$$D_\theta(y;\mu) = -2\int_y^\mu \frac{y-\mu}{V(\mu,v)}\,\mathrm{d}\mu \tag{4.10}$$

and c is chosen in the way that it optimizes Stirling's formula:

$$k! \simeq \sqrt{(2\pi(k+c))}\,k^k\,e^{-k} \tag{4.11}$$

The optimal value for c is apparently $\frac{1}{6}$. Clark and Perry (1989) gave the quasi likelihood function for Parametrization 2 of Table 3.3 of the negative binomial distribution. The likelihood function is:

$$\begin{aligned} Q^+\left(y;\mu,a,\frac{1}{6}\right) =& \sum_{i=1}^{n}\left(y_i\log\left(\frac{\mu}{y_i}\right) - \frac{1+ay_i}{a}\log\left(\frac{1+a\mu}{1+ay_i}\right) - \frac{1}{2}\log(2\pi)\right. \\ & - \log(1+ay_i) - \frac{1}{2}\log\left(y_i+\frac{1}{6}\right) - \frac{1}{2}\log\left(1+\frac{a}{6}\right) \\ & \left. + \frac{1}{2}\log\left(ay_i+1+\frac{a}{6}\right)\right) \end{aligned} \tag{4.12}$$

The estimator for $\hat{a}$ can be found by solving the following equation with $m=\hat{\mu}$.

$$\sum_{i=1}^{n}\left(\frac{1}{a^2}\log\left(\frac{1+am}{1+ay_i}\right) - \frac{y_i}{1+ay_i} + \frac{1+6y_i}{2(a+6+6ay_i)}\right) = \frac{n}{2(a+6)} \tag{4.13}$$

This equation can be solved numerically.

But for comparing the different parametrizations we also need to compute the equations for the other parametrizations in order to compare the parametrizations. For Parametrization 1 we have:

$$\begin{aligned} Q^+\left(y;\mu,k,\frac{1}{6}\right) =& \sum_{i=1}^{n}\left(y_i\log\left(\frac{\mu}{y_i}\right) - (k+y_i)\log\left(\frac{k+\mu}{k+y_i}\right) - \frac{1}{2}\log(2\pi)\right. \\ & - \log(k+y_i) - \frac{1}{2}\log\left(y_i+\frac{1}{6}\right) - \frac{1}{2}\log\left(k+\frac{1}{6}\right) + \log(k) \\ & \left. + \frac{1}{2}\log\left(y_i+k+\frac{1}{6}\right)\right) \end{aligned} \tag{4.14}$$

For Parametrization 3 we have:

$$\begin{aligned} Q^+\left(y;\mu,p,\frac{1}{6}\right) =& \sum_{i=1}^{n}\left(p\left(y_i\log\left(\frac{\mu}{y_i}\right) - (\mu+y_i)\right) - \frac{1}{2}\log(2\pi)\right. \\ & - \log\left(y_i\left(1+\frac{p}{1-p}\right)\right) - \frac{1}{2}\log\left(y_i+\frac{1}{6}\right) - \frac{1}{2}\log\left(\frac{p\mu}{1-p}+\frac{1}{6}\right) \\ & \left. + \log\left(\frac{p\mu}{1-p}\right) + \frac{1}{2}\log\left(y_i+\frac{p\mu}{1-p}+\frac{1}{6}\right)\right) \end{aligned} \tag{4.15}$$

For Parametrization 4 we have:

$$Q^{+}\left(y;\mu,\beta,\frac{1}{6}\right)=\sum_{i=1}^{n}\left(\frac{y_i}{\beta+1}\log\left(\frac{\mu}{y_i}\right)-\frac{1}{\beta+1}\left(\mu+y_i\right)-\frac{1}{2}\log\left(2\pi\right)\right.$$
$$-\log\left(\frac{\mu}{\beta}+y_i\right)-\frac{1}{2}\log\left(y_i+\frac{1}{6}\right)-\frac{1}{2}\log\left(\frac{\mu}{\beta}+\frac{1}{6}\right)+\log\left(\frac{\mu}{\beta}\right)$$
$$\left.+\frac{1}{2}\log\left(y_i+\frac{\mu}{\beta}+\frac{1}{6}\right)\right) \tag{4.16}$$

For Parametrization 5 we have:

$$Q^{+}\left(y;\mu,\delta,\frac{1}{6}\right)=\sum_{i=1}^{n}\left(\frac{y_i}{\delta}\log(\frac{\mu}{y_i})-\frac{1}{2}\log(2\pi)\right.$$
$$-\log\left(y_i+\frac{\mu}{\delta-1}\right)-\frac{1}{2}\log\left(y_i+\frac{1}{6}\right)-\frac{1}{2}\log\left(\frac{\mu}{\delta-1}+\frac{1}{6}\right)$$
$$\left.+\log\left(\frac{\mu}{\delta-1}\right)+\frac{1}{2}\log\left(y_i+\frac{\mu}{\delta-1}+\frac{1}{6}\right)\right) \tag{4.17}$$

After studying those likelihood functions we noticed the following: The difference between those likelihood functions is that we have to solve two equations in a numerical way for formula (4.15), (4.16) and (4.17). In formula (4.14) and (4.12) we can substitute the estimator for μ into the second equation. In these cases we only have to solve one equation numerically. Calculations for formula (4.15), (4.16) and (4.17) become harder. For this reason, we will not take this method into account for estimating the quantiles. However, in the next section we will do a short simulation study to check how the quasi likelihood estimator behaves comparing to the method of moments estimator. If this estimator behaves worse than the method of moments, then there is no sense in using this method. If not, than there should be done much more research on this estimator so that it can be used for all parametrizations. This will not be done in this report.

4.4 Simulation study MME and QMLE

To compare the estimators we reproduced the simulation that Clark and Perry (1989) did in their article. For several combinations of μ and κ they calculated the estimator of α. To estimate this α they used the estimators described above using parametrization 2. The simulation is done for 10000 runs and for each run 50 observations. The results of the moment estimator can be found in (4.2). The results of the QMLE can be found in (4.3).

μ	κ	α	$\hat{\alpha}$	Std. Deviation	No. Neg	Median	Q1	Q3
1	1	1	0.957070	0.539646	75	0.870348	0.591837	1.218577
1	3	0.333	0.330831	0.316891	1253	0.293245	0.109940	0.506661
1	5	0.2	0.196799	0.264365	2238	0.162250	0.016940	0.347516
3	1	1	0.976647	0.360471	0	0.919180	0.728388	1.170254
3	3	0.333	0.329393	0.154840	22	0.311217	0.224683	0.412390
3	5	0.2	0.199693	0.116532	221	0.189655	0.117966	0.269004
5	1	1	0.976437	0.321264	0	0.930832	0.753323	1.146298
5	3	0.333	0.327343	0.120782	1	0.314137	0.243109	0.398671
5	5	0.2	0.197921	0.087876	21	0.190001	0.135540	0.250293
10	1	1	0.976783	0.295417	0	0.930152	0.775218	1.128039
10	3	0.333	0.330545	0.098145	0	0.319557	0.261063	0.38846
10	5	0.2	0.198790	0.065257	1	0.191296	0.153202	0.236413
15	1	1	0.978381	0.285488	0	0.934562	0.781445	1.126233
15	3	0.333	0.331341	0.091438	0	0.320231	0.266110	0.384964
15	5	0.2	0.198740	0.058447	0	0.193587	0.158377	0.232834
20	1	1	0.975792	0.276547	0	0.934161	0.780376	1.120854
20	3	0.333	0.332061	0.088535	0	0.320416	0.270048	0.383174
20	5	0.2	0.199260	0.054401	0	0.194394	0.160805	0.232043

Table 4.2: Simulation results MME

μ	κ	α	$\hat{\alpha}$	Std. Deviation	No. Neg	Median	Q1	Q3
1	1	1	1.030396	0.539971	45	0.959690	0.652737	1.326152
1	3	0.333	0.344948	0.317004	945	0.306345	0.130110	0.528174
1	5	0.2	0.210335	0.282590	1939	0.186766	0.037074	0.368556
3	1	1	1.012163	0.312815	0	0.981828	0.790066	1.197690
3	3	0.333	0.324985	0.145481	20	0.310561	0.221536	0.414827
3	5	0.2	0.195085	0.113752	174	0.184501	0.114168	0.265556
5	1	1	1.001383	0.265575	0	0.975524	0.810497	1.161726
5	3	0.333	0.327623	0.114431	0	0.318647	0.247386	0.396156
5	5	0.2	0.193124	0.083561	19	0.186891	0.133688	0.244847
10	1	1	1.000677	0.223623	0	0.984639	0.843970	1.140245
10	3	0.333	0.326948	0.088210	0	0.319032	0.265582	0.381209
10	5	0.2	0.195785	0.060597	0	0.191487	0.153397	0.234134
15	1	1	0.997759	0.213563	0	0.981506	0.847287	1.130618
15	3	0.333	0.327526	0.079828	0	0.321260	0.269887	0.378509
15	5	0.2	0.195562	0.053194	0	0.191903	0.157760	0.229445
20	1	1	0.999709	0.203423	0	0.985879	0.857468	1.127290
20	3	0.333	0.328474	0.077010	0	0.323189	0.274927	0.377598
20	5	0.2	0.195096	0.049561	0	0.191431	0.160731	0.226800

Table 4.3: Simulation results QMLE

We see that in most cases the method of moments estimator gives a result closer to the true value than the quasi maximum likelihood estimator. But this is not a conclusion we can take for all values. But even when there are no negative values of α the method of moments estimator seems to perform better. In this case there were no numerical errors in the quasi maximum likelihood estimator. Method of moments estimators can have a bias. That is also true in this case since the estimators are significantly different to the parameters. But the quasi maximum likelihood estimator is therefore also biased. With the choice that we do not test the quasi maximum likelihood estimator for each parametrization we will also not test it for the two we easily can test. Because the error is often big, we have the risk that the difference in the quantiles is even bigger. So from now on we only consider the method of moments and the maximum likelihood estimator.

4.5 Example: Exponential distribution

In this section we give an example why we need a good analysis for the estimators of parameters. This example shows that the transformation to another parametrization is not trivial. We take the exponential distribution as an example. Also this distribution has more than one parametrization. We distinguish two parametrizations. The first is used the most. The random variable X has the following probability density function

$$f(x;\lambda) = \lambda e^{-\lambda x}\ x > 0 \tag{4.18}$$

and the random variable Y has probability density function:

$$f(y;\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}}\ y > 0 \tag{4.19}$$

Note that the only difference is the parameter. We have that $\lambda = \frac{1}{\theta}$. We know that $\mathbb{E}[X] = \frac{1}{\lambda}$ and $\mathbb{E}[Y] = \theta$. We now define $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. From the maximum likelihood method we know that $\mathbb{E}[\hat{\theta}] = \overline{X}$, and $\mathbb{E}[\hat{\lambda}] = \frac{1}{\overline{X}}$ .We also know that if $X_i \sim \text{Exp}(\lambda)$, then $\sum_{i=1}^{n}(X_i) \sim \Gamma(n,\lambda)$. With these properties we can calculate $\mathbb{E}[\overline{X}]$:

$$\begin{aligned}
\mathbb{E}[\bar{X}] =& \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i)\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n}(X_i)\right] = \frac{1}{n}\int_0^\infty x\,\frac{1}{\Gamma(n)}e^{-\lambda x}x^{n-1}\lambda^n dx \\
=& \frac{1}{n}\int_0^\infty \frac{1}{\Gamma(n)}e^{-\lambda x}x^n\lambda^n dx = \frac{1}{\lambda}\int_0^\infty \frac{1}{\Gamma(n+1)}e^{-\lambda x}x^n\lambda^{n+1}dx \\
=& \frac{1}{\lambda} = \theta
\end{aligned} \tag{4.20}$$

So from here we conclude that $\mathbb{E}[\hat{\theta}] = \theta$. But now we show that $\mathbb{E}[\hat{\lambda}] \neq \lambda$. Clearly a transformation in the parameters does not need to hold in their expectations. We calculate now $\mathbb{E}[\frac{1}{\bar{X}}]$

$$\begin{aligned}\mathbb{E}[\frac{1}{\bar{X}}] =&\mathbb{E}\left[\frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i)}\right] = n\mathbb{E}\left[\frac{1}{\sum_{i=1}^{n}(X_i)}\right] = n\int_0^\infty \frac{1}{x}\frac{1}{\Gamma(n)}e^{-\lambda x}x^{n-1}\lambda^n dx \\ =&n\int_0^\infty \frac{1}{\Gamma(n)}e^{-\lambda x}x^{n-2}\lambda^n dx = \frac{\lambda n}{n-1}\int_0^\infty \frac{1}{\Gamma(n-1)}e^{-\lambda x}x^{n-2}\lambda^{n+1}dx \\ =&\frac{\lambda n}{n-1} \neq \lambda\end{aligned} \tag{4.21}$$

So we conclude that if we have $\lambda = \frac{1}{\theta}$, that $\mathbb{E}[\hat{\lambda}] \neq \frac{1}{\mathbb{E}[\hat{\theta}]}$. This is the main problem that occurs when using different parametrizations. For this reason some parametrizations may be better to estimate parameters and quantiles, and therefore research does make sense. In the example of the exponential distribution we saw that the estimator for θ is unbiased, and the estimator for λ biased. Therefore we should expect better estimations when using the parametrization with θ. On the other hand, if the sample size is very large, then the estimator for λ is asymptotically unbiased, because $\lim_{n\to\infty}\frac{n}{n-1} = 1$. So if the sample size is large, there is almost no difference between the two estimators. Only small sample sizes give significant differences. In the case of counting bacteria, there are not that much observations because it costs too much time to do for example 1000 observations. So n will not be that large in this case, and therefore we should expect that $\hat{\theta}$ should perform better than $\hat{\lambda}$.

Chapter 5

Quantiles

In Chapter 4 we derived estimators for all different parametrizations of the negative distribution. The next goal is to find quantiles of the negative binomial distribution. There are two methods for estimating quantiles that we will discuss. We first discuss transformation of negative binomial data, then we will try to inverse the cumulative distribution function, and finally we use an estimation method based on a gamma distribution.

5.1 Transformation of negative binomial data

A simple way of dealing with negative binomial data is to use transformations. Also for quantiles we have an invariance principle when we transform the data. Estimating quantiles can easily be done when we have a normal distribution with a constant mean and variance. This is therefore the goal of this section. In the article of Anscombe (1948), a transformation for negative binomial data is considered for Parametrization 1. If r is an outcome for the negative binomial distribution with parameters k and m, then the transformation for Parametrization 1 works as follows.

$$y = \sqrt{\left(k - \frac{1}{2}\right)} \sinh^{-1} \sqrt{\left(\frac{r + \frac{3}{8}}{k - \frac{3}{4}}\right)} \tag{5.1}$$

To check how this transformation behaves, we do a small simulation in R. We simulated 5000 outcomes from a negative binomial distribution. In R the following input is needed for the function to create random negative binomial numbers: the amount of random numbers, the number of successes, and the succes probability. The output gives us a vector which contains the number of failures for each simulation. In the article of Anscombe (1948), the number of successes is equal to k, and the number of failures is equal to r. The standard parametrization in R is Parametrization 3 of (3.2). So we have to define the succes probability by $p = \frac{k}{m+k}$. An example of transformed data in a histogram can be found in Figure 5.1.

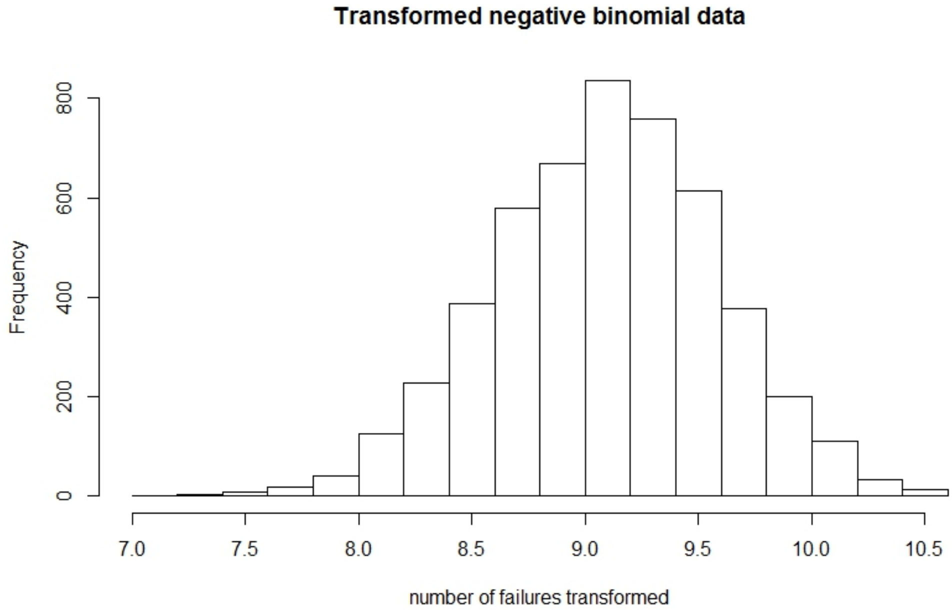

Figure 5.1: Transformed Negative Binomial Data, m=160, k=40 ,n=5000

The next thing to performed is to make this histogram multiple times and estimating the variance of each sample of 5000 random negative binomial numbers. We calculate the mean of all these variances and also the variance of this number. Another thing to do is to test for normality. First we test the variance. The result is that the mean variance is 0.2499985, which is almost $\frac{1}{4}$. The variance of this number is $2.53 \cdot 10^{-5}$. So we could say that the variance is constant. So the first goal has been accomplished. However the normality test is quite bad. The Shapiro-Wilk test gives a p-value of $7.881 \cdot 10^{-5}$. So this transformation is good for a constant variance but bad for normality. But when we look at the quantile of normal probability plot of the data in Figure 5.2, it does not look bad. Only in the extreme values it differs from normality. These extreme values are the highest of lowest values that can occur. So we can question whether this transformation can be used. It depends of what question you are interested in. If the research is focused on the extreme values, it is bad idea to choose this transformation. Since we are interested in quantiles for different levels, it is not a good idea to estimate the quantiles with this transformation. Stabilizing the variance by transformation of the data always leads to bad normality.

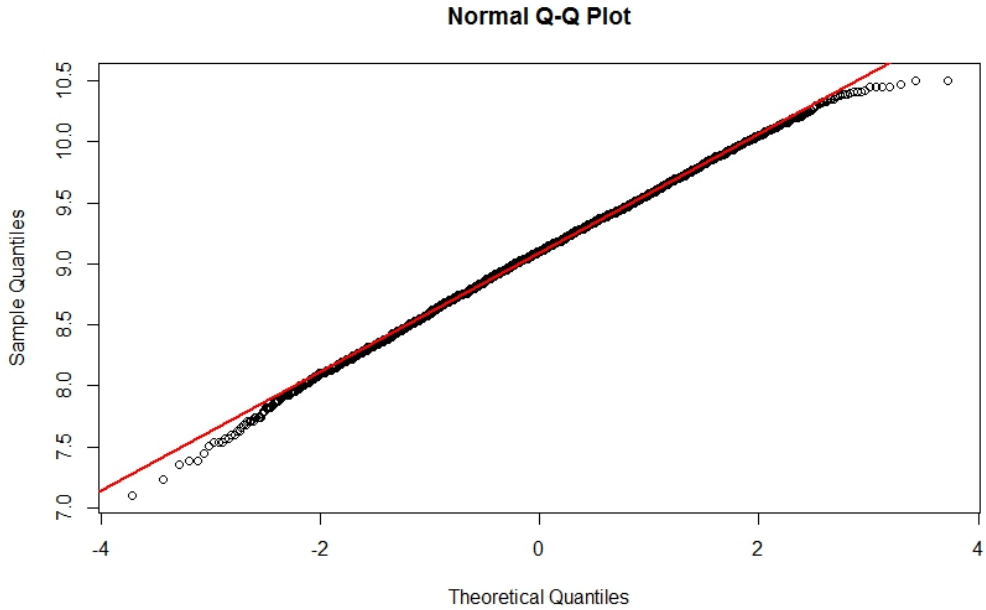

Figure 5.2: Transformed Negative Binomial Data QQPlot, m=160, k=40 ,n=5000

5.2 Inverting cumulative distribution function

The second method for estimating the quantiles is to invert the cumulative distribution function. When the estimators have been derived, we can derive exact quantiles for the distribution function with these estimated parameters. Hoffman (2003) stated the exact way of deriving the quantiles. The exact quantiles can now be found by using (5.2)

$$q_\alpha = \inf \left\{ n | \sum_{x=0}^{n} p(x|\hat{\mu}, \hat{k}) \geq 1 - a \right\} \tag{5.2}$$

with $\hat{p}(x|\hat{\mu}, \hat{k})$ the probability mass function of the negative binomial distribution with parameters $\hat{\mu}$ and $\hat{k}$. A specification is needed which parameterization is used. The formula (5.2) gives the first integer value that exceeds the cumulative probability mass function. We have to take in mind that the negative binomial distribution is discrete. So there might be no value x for the random variable X which satisfies the following property:

$$\mathbb{P}(X \leq x) = 1 - a \quad \wedge \quad \mathbb{P}(X > x) = a \tag{5.3}$$

For this reason we take the infimum of all the values x that exceeds the first probability of (**??**). So the x^{th} percent quantile is the first value which has at least x procent of mass below this value. The software R uses the same definition of these quantiles. We will use R to calculate q_α, because these can not be calculated by hand.

5.3 Control charts

Control charts can be used to analyze data that is ordered in time. By plotting the data points in the chart we get a clear vision of how the data is distributed over time. The goal of this control chart is to find whether the system is under control with the same determined parameters. If it happens too often that the data is described outside the control chart, then it is necessary to modify the model by choosing other parameters. Outside the control chart means that data exceeds a certain level. This level is called a quantile. The x^{th} percentage quantile gives the value of the distribution of which there is x procent of probability that the random variable takes a value below this quantile. An example for this is the count of bacteria. When we count bacteria we do not want to exceed a certain number of bacteria because this is dangerous. If this number exceeds the maximum we want to have some alert to warn that there are too many bacteria in the room. With the use of quantiles we can decide a certain level (a) that tells us how often we want to have an alarm. This a is the type-1 error, that is the probability of a false positive. For example if $a = 0.001$, then in expectation we have in 1 of 1000 samples an alarm that tells us that something extreme is happening. We would like to know what this number

of bacteria will be when there is an alarm. With this number we can determine an upper bound for the number of bacteria, and make sure that in almost every case we have less bacteria then this upper bound. However if it happens more than expected that the data exceeds the upper bound, then there may be a shift in the data and the model is not sufficient anymore. In our case we deal with data that is assumed to be independent and identically distributed. So for now we do not look at the data as function over time, but we only estimate the quantiles in the i.i.d. case. In other words: we are interested in the upper bound of the negative binomial distribution which has only a probability of a that the upper bound will be exceeded. The upper quantile we will call the upper control limit (UCL) and the lower quantile the lower control limit (LCL).
Now we will approximate these quantiles. Hoffman (2003) states that the following approximation can be done for Parametrization 2:

$$\mathbb{P}(X \leq r) \approx \mathbb{P}\left(\chi^2_\nu \leq \frac{2r+1}{1+uk}\right) \tag{5.4}$$

So we try to approximate the negative binomial distribution with a chi-squared distribution. The idea comes from Guenther (1972), who used a parametrization such that the mean and the variance can be written in such a form that is the same as the mean and variance of the gamma distribution. Guenther (1972) uses Parametrization 5 which also shows that the mean and variance similar with the gamma distribution. With a continuity correction we get that:

$$\mathbb{P}(X \leq r) \approx \mathbb{P}\left(Y < r + \frac{1}{2}\right) \tag{5.5}$$

with Y a gamma distribution with parameters α and β such that $\mathbb{E}[Y] = \alpha\beta$, and $\text{Var}(Y) = \alpha\beta^2$. Now we calculate $\mathbb{P}\left(Y < r + \frac{1}{2}\right)$ and we will see that we have a chi-squared distribution. We make use of the following formulas for the probability density functions for the gamma distribution and the chi-squared distribution. For the gamma distribution we have:

$$f(y) = \frac{1}{\beta^\alpha}\frac{1}{\Gamma(\alpha)}y^{\alpha-1}e^{-\frac{y}{\beta}}\; y > 0 \tag{5.6}$$

and for the chi-squared distribution χ^2_k we have:

$$f(y) = \frac{1}{2^{\frac{k}{2}}}\frac{1}{\Gamma(\frac{k}{2})}y^{\frac{k}{2}-1}e^{-\frac{y}{2}}\; y > 0 \tag{5.7}$$

We see that the chi-squared distribution is a special case of the gamma distribution. Apparently, if we take $\beta = 2$ and $k = 2\alpha$ for the gamma distribution, we get a chi-squared distribution. For the calculations we also make use of the transformation $\frac{y}{\beta} = \frac{t}{2}$.

$$\begin{aligned}
\mathbb{P}(X \leq r) &\approx \mathbb{P}(Y < r + \frac{1}{2}) \\
&= \int_0^{r+\frac{1}{2}} \frac{1}{\beta^\alpha}\frac{1}{\Gamma(\alpha)}y^{\alpha-1}e^{-\frac{y}{\beta}}\,\mathrm{d}y \\
&= \int_0^{\frac{2(r+\frac{1}{2})}{\beta}} \frac{\frac{\beta}{2}}{\beta^\alpha}\frac{1}{\Gamma(\alpha)}(\frac{t\beta}{2})^{\alpha-1}e^{-\frac{t}{2}}\,\mathrm{d}t \\
&= \int_0^{\frac{2(r+\frac{1}{2})}{\beta}} \frac{1}{2^\alpha}\frac{1}{\Gamma(\alpha)}t^{\alpha-1}e^{-\frac{t}{2}}\,\mathrm{d}t \\
&= \Gamma(\alpha, 2) \\
&= \chi^2_{2\alpha}
\end{aligned} \tag{5.8}$$

So in 5.8 we indeed get a $\chi^2_{2\alpha}$ distribution. Since we have $\beta = 2$, we can also write:

$$\mathbb{P}(X \leq r) \approx \mathbb{P}\left(Y < r + \frac{1}{2}\right) = \mathbb{P}\left(Y < \frac{2(r+1)}{2}\right) = \mathbb{P}\left(Y < \frac{2(r+1)}{\beta}\right) \tag{5.9}$$

Now we can apply this rule for every parametrization by translating the parameters to $\alpha(=\gamma)$ and $\beta = (\delta)$ of parametrization 5. We obtain the following results:

For Parametrization 1 we have:

$$\mathbb{P}(X \leq r) \approx \mathbb{P}\left(Y < r + \frac{1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{\frac{\mu}{\kappa}+1}\right) \tag{5.10}$$

with $Y \sim \chi^2$ with $v = 2 \cdot \frac{\kappa\mu}{\mu+\kappa}$.

For Parametrization 2 we have:

$$\mathbb{P}(X \leq r) \approx \mathbb{P}\left(Y < r + \frac{1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{ua+1}\right) \tag{5.11}$$

with $Y \sim \chi^2$ with $v = 2 = \cdot \frac{u}{ua+1}$.

For Parametrization 3 we have:

$$\mathbb{P}(X \leq r) \approx \mathbb{P}\left(Y < r + \frac{1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{\frac{1}{p}}\right) \tag{5.12}$$

with $Y \sim \chi^2$ with $v = 2(1-p)k$.

For Parametrization 4 we have:

$$\mathbb{P}(X \leq r) \approx \mathbb{P}\left(Y < r + \frac{1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{2}\right) = \mathbb{P}\left(Y < \frac{2r+1}{1+\beta}\right) \tag{5.13}$$

with $Y \sim \chi^2$ with $v = 2 \cdot \frac{\alpha\beta}{\beta+1}$.

With these approximations we will perform a simulation study to determine which parametrization performs best.

Chapter 6

Simulation and Results

6.1 Simulation

In this section we will describe a simulation so that we can analyze the different parametrizations with respect to parameters and quantiles. Since calculations are really hard to do by hand, a simulation is needed to be able to draw conclusions about the parametrizations. The most important steps of the simulation are described below:

First we have to determine how many observations we have. The amount of observations can have much influence on the estimators or the quantiles. Good estimators should also perform quite well when the number of observations is small. So in this simulation we take 3 cases: 20 observation, 50 observations and 100 observations. We also have to determine the number of runs of the simulation and the type-1 error for the quantiles. The more runs, the more accurate the estimates are. We choose 5000 runs. We are interested in quantiles for extreme cases, so $\alpha = 0.001$ should be a good choice. After this step maybe the most important is to choose the true parameters. For this we use the outcome of the simulation of Hoffman (2003). In this simulation parametrization 2 is used. The outcome was that $\hat{k} = 0.7902$ and $\hat{\mu} = 10.611$. Although these parameters were the simulation results of Hoffman (2003), we will use these as the true parameters. In the simulation we have to be very consequent in which parameter is first and which is the second. These parameters are sometimes called a and b respectively, or $u1$ and $u2$ respectively. The order of which parameter is the first one is the same as in Table 3.3. When we start the simulation we give the input parameters and decide of which parametrization we want to use. Then a sample is generated with the number of observations we want. Then for all different parametrizations we can calculate both the moment estimators and maximum likelihood estimators. To do this it is very helpful to create a function that automatically translate all parameters. From the estimators we also calculate the variance and 95% confidence intervals. These confidence intervals will also be used for calculating the quantiles. When we have the estimates for the parameters, these estimates are used to calculate the quantiles with a chi-squared distribution. This method is described in Chapter 5. There is also a function created for this. These estimates for the quantiles are also done for the boundaries of the confidence intervals of the parameters. Note that these intervals are not exact 95% confidence intervals since we use two lower bounds and two upper bounds from both parameters. But the interval we created gives a good estimate for the interval of the parameter. For exact intervals more research should be done. To compare the different parametrizations we need some measure such that the comparison is fair. A measure based on percental deviation seems fair. Absolute difference can not be applied because all estimators have another scale. So one way to compare the estimators is to calculate the following:

$$d = \frac{\mathbb{E}[\hat{\theta}] - \theta}{\theta} \tag{6.1}$$

where d is the percental difference between the true parameter θ, and the estimator $\hat{\theta}$. The other measure is based on the standard deviation of the estimator $\hat{\theta}$. Again we have to take into account that is the estimator has a low value that the variance could also be a lot smaller. So here we again correct this by dividing by θ. So we have:

$$v = \frac{\sqrt{(\mathrm{Var}\hat{\theta})}}{\theta} \tag{6.2}$$

The best situation is when d and v are both very close to zero. For the parameters we have to look for both measures if it is close to zero. For the quantiles we look at the intervals and check whether the true value is in that interval. Since the quantile function of R takes the smallest value that has at least x percent of the mass below this value, it could also be that the true value is almost one integer less then the value R gives. With these measures we are now able to do the simulation. The results of this simulation are described in the next section.

6.2 Results

In this section we give the solutions for the simulation. The simulation is done for for 20, 50 and 100 observations. For each number of observations there are 4 tables produced:

- Estimation for the first parameter
- Estimation for the second parameter
- Estimation for the upper quantile (0.999)
- Estimation for the lower quantile (0.001)

For the first two tables the structure of the colums is as follows:

- 1. The parametrization that is used for the estimate
- 2. The method that is used for estimating parameters (MME or MLE)
- 3. The true parameter
- 4. The estimate for the true parameter
- 5. The variance for the estimator
- 6. The lower bound of the 95% confidence interval for the estimator
- 7. The upper bound of the 95% confidence interval for the estimator
- 8. The translation of the estimate of the parameter to parametrization 4
- 9. The scaled bias of the estimate with respect to the true parameter (of own parametrization)
- 10. The measure based on variance of the estimate for the parameter

The third and fourth table have the following structure for the columns:

- 1. The parametrization that is used for the estimate
- 2. The method that is used for estimating parameters (MME or MLE)
- 3. The degree of freedom for the used chi-squared distribution
- 4. The estimator of the quantile based on the chi-squared distribution (estimated parameters are used)
- 5. The estimator of the quantile based on the quantile function in R (estimated parameters are used)
- 6. The true quantile (the true parameters are used)
- 7. The estimated quantile with the chi-squared distribution when both parameters are minimal
- 8. The estimated quantile with the chi-squared distribution when both parameters are maximal

6.2.1 20 observations

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	1.2655	1.5811	0.5957	1.5646	1.5976	1.5811	0.1996	0.6099
2	1	MLE	1.2655	1.5159	0.5117	1.5018	1.5301	1.5159	0.1652	0.5652
3	2	MME	0.7902	0.7597	0.1101	0.7567	0.7628	1.3163	-0.0401	0.4199
4	2	MLE	0.7902	0.7656	0.0786	0.7634	0.7678	1.3061	-0.0321	0.3549
5	3	MME	1.2655	1.5811	0.5957	1.5646	1.5976	1.5811	0.1996	0.6099
6	3	MLE	1.2655	1.5159	0.5117	1.5018	1.5301	1.5159	0.1652	0.5652
7	4	MME	1.2655	1.5811	0.5957	1.5646	1.5976	1.5811	0.1996	0.6099
8	4	MLE	1.2655	1.5159	0.5117	1.5018	1.5301	1.5159	0.1652	0.5652
9	5	MME	1.1307	1.3349	0.2575	1.3278	1.3421	1.5012	0.1530	0.4488
10	5	MLE	1.1307	1.2906	0.2140	1.2846	1.2965	1.4499	0.1239	0.4091

Table 6.1: 20 observations, 5000 runs, parameter 1

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	10.6110	10.5884	4.8586	10.4537	10.7231	6.6969	-0.0021	0.2077
2	1	MLE	10.6110	10.5884	4.8586	10.4537	10.7231	6.9847	-0.0021	0.2077
3	2	MME	10.6110	10.5884	4.8586	10.4537	10.7231	8.0440	-0.0021	0.2077
4	2	MLE	10.6110	10.5884	4.8586	10.4537	10.7231	8.1067	-0.0021	0.2077
5	3	MME	0.1066	0.1311	0.0032	0.1310	0.1312	6.6290	0.1871	0.5344
6	3	MLE	0.1066	0.1268	0.0028	0.1268	0.1269	6.8841	0.1599	0.4983
7	4	MME	8.3848	8.0294	14.8888	7.6167	8.4420	8.0294	-0.0443	0.4602
8	4	MLE	8.3848	8.0989	11.5561	7.7785	8.4192	8.0989	-0.0353	0.4054
9	5	MME	9.3848	9.0294	14.8888	8.6167	9.4420	8.0294	-0.0394	0.4112
10	5	MLE	9.3848	9.0989	11.5561	8.7785	9.4192	8.0989	-0.0314	0.3622

Table 6.2: 20 observations, 5000 runs, parameter 2

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.7513	59.8672	59.0000	67.0000	59.4774	60.2545
2	1	MLE	2.6522	61.1819	60.0000	67.0000	60.7512	61.6098
3	2	MME	2.3415	65.9331	65.0000	67.0000	65.0033	66.8666
4	2	MLE	2.3254	66.2102	65.0000	67.0000	65.3173	67.1058
5	3	MME	2.7477	59.3010	58.0000	67.0000	59.0849	59.5159
6	3	MLE	2.6473	60.3588	59.0000	67.0000	60.1641	60.5524
7	4	MME	2.8119	71.4626	70.0000	67.0000	67.7299	75.2249
8	4	MLE	2.6987	70.7939	69.0000	67.0000	67.9111	73.6965
9	5	MME	2.6699	69.9431	69.0000	67.0000	66.5993	73.3007
10	5	MLE	2.5811	69.5117	68.0000	67.0000	66.9387	72.0936

Table 6.3: 20 observations, 5000 runs, upper quantiles

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.7513	-0.4411	0.0000	0.0000	-0.4446	-0.4374
2	1	MLE	2.6522	-0.4502	0.0000	0.0000	-0.4531	-0.4473
3	2	MME	2.3415	-0.4734	0.0000	0.0000	-0.4734	-0.4734
4	2	MLE	2.3254	-0.4744	0.0000	0.0000	-0.4746	-0.4742
5	3	MME	2.7477	-0.4420	0.0000	0.0000	-0.4452	-0.4387
6	3	MLE	2.6473	-0.4514	0.0000	0.0000	-0.4539	-0.4488
7	4	MME	2.8119	0.0779	0.0000	0.0000	0.0680	0.0887
8	4	MLE	2.6987	0.0626	0.0000	0.0000	0.0558	0.0698
9	5	MME	2.6699	0.0584	0.0000	0.0000	0.0541	0.0630
10	5	MLE	2.5811	0.0485	0.0000	0.0000	0.0455	0.0515

Table 6.4: 20 observations, 5000 runs, lower quantiles

We see that for the first parameter in Table 6.1 we have the lowest scaled bias for parametrization 2. Also the measure based on variance is the lowest for parametrization 2. But this only holds when we analyze the methods separately. For example: The value for the MME of parametrization 2 is bigger than for the MLE of parametrization 5. If we look at the translated parameter than we also see that the MLE of parametrization 2 is the closest to the true parameter. But the true parameter is never in the confidence interval. For the second parameter in Table 6.2 we see that parametrization 1 and 2 have the same estimate, and that it has the lowest scaled bias and scaled variance measure. Now we have that except for parametrization 3 the true value is in the confidence interval. So these estimators may be better. Note that if the estimators are the same, the translated are not the same. The reason for this is that the first parameter differs. If we take this into account, we see again that the MLE of parametrization 2 gives the best result. When we look at the upper quantiles in Table 6.3 we see that the R function estimate is the best for the MLE parametrization 5. On the other hand, the chi-squared distribution is the best for MLE parametrization 2. The estimations of parametrization 1 and 3 are really bad. We see also that parametrization 4 and 5 estimate to high, while parametrization 1, 2 and 3 are below the true value. The same result we see in Table 6.4 for the lower quantile. The best estimation is MLE parametrization 5, based on both scaled bias and scaled variance.

6.2.2 50 observations

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	1.2655	1.3919	0.1591	1.3875	1.3963	1.3919	0.0908	0.3151
2	1	MLE	1.2655	1.3536	0.1043	1.3507	1.3565	1.3536	0.0651	0.2552
3	2	MME	0.7902	0.7793	0.0537	0.7778	0.7808	1.2833	-0.0140	0.2933
4	2	MLE	0.7902	0.7789	0.0317	0.7780	0.7798	1.2838	-0.0145	0.2253
5	3	MME	1.2655	1.3919	0.1591	1.3875	1.3963	1.3919	0.0908	0.3151
6	3	MLE	1.2655	1.3536	0.1043	1.3507	1.3565	1.3536	0.0651	0.2552
7	4	MME	1.2655	1.3919	0.1591	1.3875	1.3963	1.3919	0.0908	0.3151
8	4	MLE	1.2655	1.3536	0.1043	1.3507	1.3565	1.3536	0.0651	0.2552
9	5	MME	1.1307	1.2179	0.0926	1.2154	1.2205	1.3653	0.0717	0.2691
10	5	MLE	1.1307	1.1915	0.0614	1.1898	1.1932	1.3357	0.0510	0.2192

Table 6.5: 50 observations, 5000 runs, parameter 1

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	10.6110	10.6099	1.9761	10.5551	10.6647	7.6225	-0.0001	0.1325
2	1	MLE	10.6110	10.6099	1.9761	10.5551	10.6647	7.8382	-0.0001	0.1325
3	2	MME	10.6110	10.6099	1.9761	10.5551	10.6647	8.2680	-0.0001	0.1325
4	2	MLE	10.6110	10.6099	1.9761	10.5551	10.6647	8.2643	-0.0001	0.1325
5	3	MME	0.1066	0.1166	0.0011	0.1166	0.1167	7.5747	0.0863	0.3051
6	3	MLE	0.1066	0.1141	0.0008	0.1141	0.1141	7.7650	0.0660	0.2587
7	4	MME	8.3848	8.2667	7.2658	8.0653	8.4681	8.2667	-0.0143	0.3215
8	4	MLE	8.3848	8.2621	4.7164	8.1314	8.3929	8.2621	-0.0148	0.2590
9	5	MME	9.3848	9.2667	7.2658	9.0653	9.4681	8.2667	-0.0127	0.2872
10	5	MLE	9.3848	9.2621	4.7164	9.1314	9.3929	8.2621	-0.0132	0.2314

Table 6.6: 50 observations, 5000 runs, parameter 2

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.4610	64.1113	63.0000	67.0000	63.9086	64.3135
2	1	MLE	2.4009	65.0736	64.0000	67.0000	64.8323	65.3145
3	2	MME	2.2896	66.9751	66.0000	67.0000	66.5803	67.3706
4	2	MLE	2.2905	66.9589	66.0000	67.0000	66.5925	67.3256
5	3	MME	2.4592	63.7342	63.0000	67.0000	63.6692	63.7991
6	3	MLE	2.3984	64.5025	64.0000	67.0000	64.4595	64.5454
7	4	MME	2.4834	69.6935	68.0000	67.0000	68.0177	71.3734
8	4	MLE	2.4149	68.8794	68.0000	67.0000	67.8019	69.9587
9	5	MME	2.4359	69.1526	68.0000	67.0000	67.5924	70.7153
10	5	MLE	2.3830	68.5128	67.0000	67.0000	67.5072	69.5195

Table 6.7: 50 observations, 5000 runs, upper quantiles

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.4610	-0.4654	0.0000	0.0000	-0.4662	-0.4647
2	1	MLE	2.4009	-0.4696	0.0000	0.0000	-0.4701	-0.4690
3	2	MME	2.2896	-0.4764	0.0000	0.0000	-0.4764	-0.4764
4	2	MLE	2.2905	-0.4764	0.0000	0.0000	-0.4764	-0.4763
5	3	MME	2.4592	-0.4658	0.0000	0.0000	-0.4664	-0.4651
6	3	MLE	2.3984	-0.4700	0.0000	0.0000	-0.4704	-0.4696
7	4	MME	2.4834	0.0392	0.0000	0.0000	0.0370	0.0415
8	4	MLE	2.4149	0.0331	0.0000	0.0000	0.0318	0.0343
9	5	MME	2.4359	0.0349	0.0000	0.0000	0.0337	0.0361
10	5	MLE	2.3830	0.0304	0.0000	0.0000	0.0297	0.0311

Table 6.8: 50 observations, 5000 runs, lower quantiles

In Table 6.7 we see for the first parameter that the translated estimate for the second parametrization is the closest to the true parameter. The measure for the scaled bias confirms this. However the measure for the scaled variance is better for parametrization 5. But this difference is very small. In Table 6.6 we see that the MLE for parametrization 2 gives the best results. The MME for parametrization 2 gives almost the same. Parametrization has the worst results, because it has the biggest bias and variance measure. In Table 6.7 we see that the R function for MLE parametrization 5 gives the right result. But again parametrization 2, has the best results for the chi-squared distribution estimate. We see also again that parametrization 4 and 5 estimate to high, and parametrization 1, 2 and 3 estimate to low. For the lower quantile in Table 6.7 we see that parametrization 5 has the best estimates.

6.2.3 100 observations

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	1.2655	1.3362	0.0770	1.3340	1.3383	1.3362	0.0529	0.2193
2	1	MLE	1.2655	1.3098	0.0467	1.3085	1.3111	1.3098	0.0338	0.1707
3	2	MME	0.7902	0.7809	0.0269	0.7802	0.7817	1.2805	-0.0119	0.2074
4	2	MLE	0.7902	0.7836	0.0156	0.7831	0.7840	1.2762	-0.0085	0.1583
5	3	MME	1.2655	1.3362	0.0770	1.3340	1.3383	1.3362	0.0529	0.2193
6	3	MLE	1.2655	1.3098	0.0467	1.3085	1.3111	1.3098	0.0338	0.1707
7	4	MME	1.2655	1.3362	0.0770	1.3340	1.3383	1.3362	0.0529	0.2193
8	4	MLE	1.2655	1.3098	0.0467	1.3085	1.3111	1.3098	0.0338	0.1707
9	5	MME	1.1307	1.1803	0.0466	1.1790	1.1816	1.3230	0.0420	0.1909
10	5	MLE	1.1307	1.1615	0.0287	1.1607	1.1623	1.3015	0.0266	0.1497

Table 6.9: 100 observations, 5000 runs, parameter 1

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	10.6110	10.5865	0.9574	10.5600	10.6130	7.9230	-0.0023	0.0922
2	1	MLE	10.6110	10.5865	0.9574	10.5600	10.6130	8.0826	-0.0023	0.0922
3	2	MME	10.6110	10.5865	0.9574	10.5600	10.6130	8.2674	-0.0023	0.0922
4	2	MLE	10.6110	10.5865	0.9574	10.5600	10.6130	8.2951	-0.0023	0.0922
5	3	MME	0.1066	0.1124	0.0005	0.1124	0.1124	7.8987	0.0518	0.2137
6	3	MLE	0.1066	0.1106	0.0003	0.1105	0.1106	8.0453	0.0362	0.1737
7	4	MME	8.3848	8.2704	3.6645	8.1688	8.3719	8.2704	-0.0138	0.2283
8	4	MLE	8.3848	8.2950	2.3533	8.2298	8.3603	8.2950	-0.0108	0.1830
9	5	MME	9.3848	9.2704	3.6645	9.1688	9.3719	8.2704	-0.0123	0.2040
10	5	MLE	9.3848	9.2950	2.3533	9.2298	9.3603	8.2950	-0.0097	0.1635

Table 6.10: 100 observations, 5000 runs, parameter 2

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.3729	65.3924	65.0000	67.0000	65.2944	65.4902
2	1	MLE	2.3312	66.0991	65.0000	67.0000	65.9785	66.2196
3	2	MME	2.2847	66.9136	66.0000	67.0000	66.7211	67.1064
4	2	MLE	2.2779	67.0356	66.0000	67.0000	66.8572	67.2142
5	3	MME	2.3720	65.2036	64.0000	67.0000	65.1705	65.2367
6	3	MLE	2.3300	65.8121	65.0000	67.0000	65.7920	65.8321
7	4	MME	2.3841	68.5864	67.0000	67.0000	67.7553	69.4185
8	4	MLE	2.3377	68.2332	67.0000	67.0000	67.7048	68.7620
9	5	MME	2.3605	68.3152	67.0000	67.0000	67.5372	69.0939
10	5	MLE	2.3230	68.0623	67.0000	67.0000	67.5663	68.5586

Table 6.11: 100 observations, 5000 runs, upper quantiles

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.3729	-0.4715	0.0000	0.0000	-0.4718	-0.4711
2	1	MLE	2.3312	-0.4740	0.0000	0.0000	-0.4743	-0.4738
3	2	MME	2.2847	-0.4767	0.0000	0.0000	-0.4767	-0.4768
4	2	MLE	2.2779	-0.4771	0.0000	0.0000	-0.4772	-0.4771
5	3	MME	2.3720	-0.4716	0.0000	0.0000	-0.4719	-0.4713
6	3	MLE	2.3300	-0.4742	0.0000	0.0000	-0.4744	-0.4741
7	4	MME	2.3841	0.0305	0.0000	0.0000	0.0296	0.0314
8	4	MLE	2.3377	0.0270	0.0000	0.0000	0.0265	0.0276
9	5	MME	2.3605	0.0287	0.0000	0.0000	0.0282	0.0292
10	5	MLE	2.3230	0.0260	0.0000	0.0000	0.0257	0.0263

Table 6.12: 100 observations, 5000 runs, lower quantiles

For the parameters we see the same results as with 20 runs or 50 runs. The MLE of parametrization 2 performs best. But the difference become smaller. Parametrization 4 and 5 give results that are very close of the results from parametrization 2. In general parametrization 1 and 3 have the worst results, and parametrization 2 the best. We see for each parametrization that the maximum likelihood method gives better results than the method of moments.

6.2.4 Second simulation for control

The previous simulation was only one example. Of course there are many possibilities to choose parameters. Now we will another simulation to see whether parametrization 2 gives also in this case the best results. Now we only to 20 observations and 5000 runs. We again use parametrization 2, but now with $k = 0.95$ and $\mu = 60$. Another change that we made is that we translate all parameters to parametrization 1 instead of parametrization 4. The results are in the 4 tables below:

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	1.0526	1.2862	0.2692	1.2787	1.2936	1.2862	0.1816	0.4929
2	1	MLE	1.0526	1.2117	0.1691	1.2070	1.2164	1.2117	0.1313	0.3906
3	2	MME	0.9500	0.9038	0.1375	0.9000	0.9076	1.1064	-0.0511	0.3903
4	2	MLE	0.9500	0.9112	0.0785	0.9091	0.9134	1.0974	-0.0425	0.2950
5	3	MME	1.0526	1.2862	0.2692	1.2787	1.2936	1.2862	0.1816	0.4929
6	3	MLE	1.0526	1.2117	0.1691	1.2070	1.2164	1.2117	0.1313	0.3906
7	4	MME	1.0526	1.2862	0.2692	1.2787	1.2936	1.2862	0.1816	0.4929
8	4	MLE	1.0526	1.2117	0.1691	1.2070	1.2164	1.2117	0.1313	0.3906
9	5	MME	1.0345	1.2533	0.2403	1.2466	1.2600	1.2765	0.1746	0.4739
10	5	MLE	1.0345	1.1837	0.1525	1.1795	1.1879	1.2054	0.1261	0.3776

Table 6.13: 20 observations, 5000 runs, parameter 1

	paramet	method	Param1	Estimate	Variance	L conf	U conf	Translate	Bias	Variance
1	1	MME	60.0000	59.8315	175.4016	54.9697	64.6933	59.8315	-0.0028	0.2207
2	1	MLE	60.0000	59.8315	175.4016	54.9697	64.6933	59.8315	-0.0028	0.2207
3	2	MME	60.0000	59.8315	175.4016	54.9697	64.6933	59.8315	-0.0028	0.2207
4	2	MLE	60.0000	59.8315	175.4016	54.9697	64.6933	59.8315	-0.0028	0.2207
5	3	MME	0.0172	0.0220	0.0001	0.0220	0.0220	57.1232	0.2170	0.5933
6	3	MLE	0.0172	0.0208	0.0001	0.0208	0.0208	57.0952	0.1704	0.4903
7	4	MME	57.0000	54.0972	651.3721	36.0424	72.1519	69.5776	-0.0537	0.4478
8	4	MLE	57.0000	54.5492	440.9939	42.3257	66.7727	66.0987	-0.0449	0.3684
9	5	MME	58.0000	55.0972	651.3721	37.0424	73.1519	69.0529	-0.0527	0.4400
10	5	MLE	58.0000	55.5492	440.9939	43.3257	67.7727	65.7541	-0.0441	0.3621

Table 6.14: 20 observations, 5000 runs, parameter 2

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.5182	358.9030	358.0000	405.0000	331.1844	386.4357
2	1	MLE	2.3753	371.6678	371.0000	405.0000	342.5677	400.6353
3	2	MME	2.1727	392.3137	391.0000	405.0000	359.8367	424.9516
4	2	MLE	2.1553	394.2434	393.0000	405.0000	362.0005	426.5782
5	3	MME	2.5157	342.8380	342.0000	405.0000	342.0752	343.5990
6	3	MLE	2.3731	354.8502	354.0000	405.0000	354.3353	355.3641
7	4	MME	2.5256	417.2138	416.0000	405.0000	278.8130	556.2620
8	4	MLE	2.3798	410.6860	409.0000	405.0000	319.1594	502.4934
9	5	MME	2.5066	415.9356	415.0000	405.0000	279.0353	553.4211
10	5	MLE	2.3674	409.8300	408.0000	405.0000	319.1919	500.7245

Table 6.15: 20 observations, 5000 runs, upper quantiles

	paramet	method	Degr freedom	Estimate	R function	True value	Lower value	Upper value
1	1	MME	2.5182	-0.2812	0.0000	0.0000	-0.3064	-0.2548
2	1	MLE	2.3753	-0.3378	0.0000	0.0000	-0.3553	-0.3197
3	2	MME	2.1727	-0.4011	0.0000	0.0000	-0.4078	-0.3948
4	2	MLE	2.1553	-0.4056	0.0000	0.0000	-0.4129	-0.3985
5	3	MME	2.5157	-0.2922	0.0000	0.0000	-0.2993	-0.2849
6	3	MLE	2.3731	-0.3460	0.0000	0.0000	-0.3497	-0.3423
7	4	MME	2.5256	0.2581	0.0000	0.0000	0.1588	0.3644
8	4	MLE	2.3798	0.1809	0.0000	0.0000	0.1333	0.2309
9	5	MME	2.5066	0.2467	0.0000	0.0000	0.1607	0.3381
10	5	MLE	2.3674	0.1751	0.0000	0.0000	0.1335	0.2185

Table 6.16: 20 observations, 5000 runs, lower quantiles

For the parameters we can conclude the same as in the first simulation parametrization 2 gives the best results, and in general MME gives better results than MLE. This is also true when we look at the scaled bias and scaled variance. Also for the quantiles we see again that the MLE of parametrization 5 gives the best estimate when the R function is used. But for the quantiles based on the chi-squared distribution there is something remarkable. We see that the estimate of parametrization 4 and 5 are very close to the true value. This result is different from the previous simulation, because parametrization 2 was performing better then. But the remarkable is that the interval constructed for this estimated is much wider for parametrization 4 and 5 than for parametrization 2. So it seems that the estimate for parametrization is more accurate then for the others. So one can doubt whether parametrization 4 and 5 are better than 2. But since we should construct exact 95% confidence intervals we can not draw conclusions from this.

Chapter 7

Conclusions

In this report we saw many different parametrizations for the negative binomial distribution. We know now that there are at least 5 different parametrizations. One of the parametrizations was a negative binomial based on successes and failures, and another parametrization was the Poisson-gamma mixture. All the parameters of the 5 parametrizations can be estimated with the method of moments and the method of maximum likelihood. Also the quasi-maximum likelihood estimator is possible but this estimator we did not test. For the quantiles three estimating methods are possible. The first method is to transform the data to a normal distribution and determine the quantiles of the normal distribution. The second method is to invert the cumulative distribution function, and the third method is to use a chi-squared distribution. After doing two simulations we saw that Parametrization 2 is the most suitable for estimating quantiles. Parametrization 1 and 3 give the worst results, and should therefore not be used. Also Parametrization 4 and 5 give results that are too high, and the other parametrizations give results that are too low. In general it is a good idea to use Parametrization 2. For all simulations the result was that the maximum likelihood estimator performs better than the method of moments. So Parametrization 2 combined with the maximum likelihood estimator is the best option to estimate quantiles of the negative binomial distribution.

Chapter 8

Future Research

There are several topics that could merit further research:

- We now left the Anscombe transform out of the simulation because of the lack of normality. But in a plot we saw that this gives only problems in extreme cases. But for other applications this transform may work when the goal is not to work with the extreme cases. Since we did need the extreme cases, we left this transform out of the simulation.
- We tried to create some confidence interval for the estimate of the upper quantile. But we had to do this using the confidence intervals of both two parameters. But now we got intervals that are not exact 95% confidence intervals, because using two lower bounds of two upper bounds do not give the perfect upper bound for the quantiles. Optimzing this may be hard, because we do not only need to optimize functions of two variables, but we only have the degree of freedom of the chi-squared distribution.
- We left the maximum-quasi likelihood out of perspective because it was not easy for Parametrization 3, 4 and 5 to calculate the parameters. We had to solve two equations numerically what we did not do in this report. For further research we can also calculate the estimators for Parametrization 3, 4 and 5.

Bibliography

K. Anraku and T. Yanagimoto. Estimation for the negative binomial distribution based on the conditional likelihood. *Communication in Statistics. Part B, Simulation and Computation*, 19(3):771–786, 1990.

F. T. Anscombe. The transformation of Poisson, Binomial and Negative-Binomial Data. *Biometrika*, 35(3/4): 246–254, 1948.

D. J. Best and P. G. Gipps. An Improved Gamma Approximation to the Negative Binomial. *Technometrics*, 16(4):621–624, 1974.

G. Blom. Transformations of the Binomial, Negative Binomial, Poisson and χ^2 Distributions. *Biometrika*, 41 (3/4):302–316, 1954.

S. J. Clark and J. N. Perry. Estimation of the Negative Binomial Parameter-Kappa by Maximum Quasi-Likelihood. *Biometrics*, 45(1):309–316, 1989.

WC Guenther. A simple approximation to the negative binomial (and regular binomial). *Technometrics*, 14(2): 385–389, 1972.

David Hoffman. Negative binomial control limits for count data with extra Poisson variation. *Pharmaceutical Statistics*, 132(2):127–132, 2003.

J. A Nelder and D. Pregibon. An Extended Quasi-Likelihood Function. *Biometrika*, 74(2):221–232, 1987.

W. W. Piegorsch. Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, 46(3):863–867, 1990.

E.P. Pieters, C.E. Gates, J.H. Matis, and W.L. Sterling. Small Sample Comparison of Different Estimators of Negative Binomial Parameters. *Biometrics*, 33(4):718–723, 1977.

R. L. Scheaffer. A note on approximate confidence limits for the negative binomial model. 5(2):149–158, 1976.

L. J. Willson, J. Leroy Folks, and J. H. Young. Multistage Estimation Compared with Fixed-Sample-Size Estimation of the Negative Binomial Parameter k. *Biometrics*, 40(1):109–117, 1984.

P.W. (1966) Zehna. Invariance of Maximum Likelihood Estimators. *The Annals of Mathematical Statistics*, 3 (37):744, 1966.