



Stockholms
universitet

Two ways of counting bears

Ossian Hättestrand

Kandidatuppsats 2022:13
Matematisk statistik
Juni 2022

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm



Mathematical Statistics
Stockholm University
Bachelor Thesis **2022:13**
<http://www.math.su.se>

Two ways of counting bears

Ossian Hättestrand*

June 2022

Abstract

In this thesis we investigate two different methods to estimate the Swedish brown bear population. Population estimates of this sort are often made with capture-recapture sampling in discrete-time, where each animal in the target population either is or is not captured on each sampling occasion. The data used to estimate the Swedish brown bear population consists of bear scats found continuously by hunters and volunteers, not on distinct sampling occasions. However, the data is then dichotomized into weekly binary capture histories where each week is regarded as one distinct sampling occasion on which each bear is regarded as either captured or not captured. This allows for the use of established discrete-time models but since there is no distinction made between a bear found one or ten times during a week, valuable information could be lost leading to a less accurate population estimate. An alternative approach is to use a continuous-time model based on the total number of scats found from each bear, without the dichotomization into weekly capture histories. The two models are compared with theoretical measures, simulation studies and computations on data. When applying the models to data and when investigating the theoretical measures, we confirm that the current method indeed results in a loss of information, which leads to higher variance in the population estimate. From the simulation studies however, we draw the conclusion that the current method is more robust towards overdispersion that arise when model assumptions are violated. The combination of these results indicate that the degree to which our model assumptions are violated could decide which method that is most advantageous.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: Ossian.hattestrand@gmail.com. Supervisor: Martin Sköld.

Contents

| | | |
|----------|-----------------------------------------------------------|-----------|
| 1 | Acknowledgments | 2 |
| 2 | Introduction | 3 |
| 3 | Theory | 4 |
| 3.1 | Poisson model | 5 |
| 3.2 | Binomial model | 6 |
| 3.3 | Violations of model assumptions | 7 |
| 4 | Data | 8 |
| 5 | Computations and simulations | 8 |
| 5.1 | Maximum likelihood and population estimates | 9 |
| 5.2 | Fisher information | 9 |
| 5.3 | Altering interval lengths in the Binomial model | 11 |
| 5.4 | Overdispersion | 12 |
| 6 | Discussion | 16 |
| 7 | Appendix | 19 |
| 8 | References | 21 |

1 Acknowledgments

I would like to thank my supervisor Martin Sköld for his much appreciated guidance and support during the process of writing this paper.

2 Introduction

The main methods currently used to estimate the number of brown bears in Sweden have been developed and improved continuously since the 1990s [3]. To estimate a low-density population of a large carnivore with elusive behavior is a challenging task [3]. There are plenty of possible approaches, and it can be difficult to identify the most appropriate one. When talking about different approaches, a distinction has to be made between methods used to collect data, and the statistics used to process it. Although they can be seen as two sides of the same coin (perfect data would be one observation of each animal in the population, and how actual data deviates from this is a result of data collection which in turn dictates our use of statistics), this thesis will only focus on the statistical methods taking the collected data as a given. To be more precise, the goal is to explore strengths and weaknesses of the current estimation method using binary capture-histories, compared with another method using full capture counts.

Population estimations with continuous-time sampling have become more common and easier to carry out as technology has advanced making non-invasive “capture” methods better and more accessible [6]. The Swedish brown bear survey for example, uses DNA analysis of bear scats to non invasively “catch” and identify unique bears continuously. The interest of estimating animal populations has existed longer than most technologies that enable sampling in continuous time, so a lot of the theory regarding population estimation focuses on discrete-time sampling [6]. The binary capture histories in the current Swedish brown bear estimation makes it possible to use well established capture-recapture models on the data, but it is not obvious that this is the best approach. The discretization of data comes at the cost of specificity, two bears could have the same binary capture history while being caught a radically different number of times. This loss of specificity means that less of our available data is used in the estimation, the extent and significance of this is one of the focal points in this thesis.

The binary capture-histories used in the current estimation method is generated by splitting the search period into multiple intervals, and it does not take into account how many times during each interval a given animal is found, rather the only thing it does take into account is how many intervals a given animal is found in. If we have three search intervals - an animal found ten times in the first interval, five in the second and zero in third, would have the binary capture-history $(1, 1, 0)$. In this scenario, the problem of estimating the number of animals in the population targeted by the search, becomes the problem of estimating the number of animals with binary capture-history $(0, 0, 0)$. The equivalent scenario using full counts would result in using all 15 captures, and then estimating the number of animals captured zero times. When doing the estimations with the binary capture histories, we would take the sum of $(1, 1, 0)$ being equal to two, do this for all unique individuals found in the searching period, and then try to find a probability distribution that models the distribution of these sums. If we denote the probability that a specific bear is caught during one time interval as p , and let p be equal for all bears and time intervals, the sum of one bears binary capture history will follow a $Bin(m, p)$ distribution where m is the number of intervals in the capture history. This is a simplification of the current estimation method which does not require that p is equal for all time intervals - this has to do with how we model the captures in time which is discussed in the paragraph below. When using full counts, we try to find a probability distribution that models the distribution of the total number of times each individual is caught. In Schofield (2018) [6] this is done with a Poisson distribution, and we will do the same in this thesis. Since the data never contains any individuals found zero times, independent of the model we will have to work with zero-truncated probability distributions. In both methods, when we have the zero-truncated distributions that model our data, we use the regular distributions

without truncation to estimate the number of individuals in the population found zero times.

The capture times of an animal found with continuous sampling are modeled as a non-homogeneous Poisson process in Schofield (2018) [6]. In this thesis, we will instead consider a homogeneous Poisson process, where the capture rate λ is constant across time. The method using full counts is not affected by this choice, since the only parameter we consider in the model is the total number of times each individual is caught. This is not the case when using binary capture histories, if we regard the capture times as a non-homogeneous Poisson process, we would have to estimate one capture probability p for each interval of the search period. This is done in the current estimation method for the Swedish brown bear population, but in this thesis we regard the capture times as coming from a homogeneous Poisson process for simplicity reasons. One assumption in relation to the Poisson process that affects both models, is the assumption that the capture rate λ is the same for all bears in the population. If this is violated it could manifest as overdispersion, i.e more variance in data compared to what is expected from our model [4]. Another possible source of overdispersion comes from our assumption that each scat in our data should be regarded as one catch, this may not be the case since multiple scats from one bear could be found on occasion - which is not a problem when we use binary capture histories, but will have an impact on the method using full counts.

3 Theory

The most fundamental theories in this thesis are the ones closest to the original problem of estimating a population size based on a random sample of collected data. The data used to estimate the Swedish brown bear population consists of scats collected from hunters during approximately 11 weeks in autumn, and we assume that the population size does not change during this period. The vector $\vec{b} = (b_1, b_2, \dots, b_N)$, where b_i is defined as the number of scats collected from bear i , and N the total number of bears in the population, with the assumption that each scat in our data can be presumed to belong to a bear chosen at random - will follow a multinomial distribution with probabilities $p = (\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N})$ and number of trials equal the total number of scats in our data. This distribution can be approximated with a Poisson distribution [5], and since the Poisson is easier to work with, this approximation will be utilized in this thesis. When estimating the bear population, the Poisson distribution can not be used directly since we only have data from bears we actually find. In other words we only have the zero-truncated version of \vec{b} , which means we have to use a zero-truncated Poisson distribution.

When the current population estimates are done by the Swedish brown bear survey, they use binary capture histories of bear scats instead of full counts. This means that the search period is split into m equally long time intervals, and each bears capture history is regarded as a vector $\vec{y}_i = (y_{i1}, y_{i2}, \dots, y_{im})$ where $y_{ij} \in \{0, 1\}$ and $y_{ij} = 1$ means bear i was found one or more times in interval j while $y_{ij} = 0$ means that the bear was not found in interval j . The sum $\sum_{j=1}^m y_{ij}$ will then be $Bin(m, p)$ distributed. Since all bears in our data will have $\sum_{j=1}^m y_{ij} > 0$ (found at least once), the scats in our data transformed to binary capture histories will follow a zero-truncated $Bin(m, p)$ distribution. When using full counts, we only regard the number of scats per bear in our data, which follows the zero-truncated Poisson distribution described in the paragraph above. Following this, the two approaches are referred to as the Poisson and Binomial model respectively.

3.1 Poisson model

Consider the vector $\vec{x} = (x_1, x_2, \dots, x_n)$, where x_i is defined as the number of scats collected from bear i , n the number of bears from which one or more scats has been found, and further that each x_i follows a zero-truncated Poisson distribution with rate parameter λ . This is the basic setup of our Poisson model, from which we can make an estimation of the bear population. The estimation is done by calculating the maximum likelihood estimation of λ , from that estimate the probability \hat{p} that a randomly chosen bear in our population is found zero times and then compute

$$\frac{n}{1 - \hat{p}} \quad (1.0)$$

To find the maximum likelihood estimation of λ , we first have to define the zero-truncated Poisson distribution.

If X is a stochastic variable following a Poisson distribution with rate λ , we derive the zero truncated version of X as following

$$P(X = k | X > 0) = \frac{P(X = k)}{P(X > 0)} = \frac{Po(\lambda)}{1 - P(Po(\lambda) = 0)} = \quad (1.1)$$

$$\frac{\frac{\lambda^k e^{-\lambda}}{k!}}{1 - e^{-\lambda}} = \frac{\lambda^k e^{-\lambda}}{k!(1 - e^{-\lambda})} = \frac{\lambda^k}{k!(e^\lambda - 1)}. \quad (1.1)$$

With this distribution, we can compute the likelihood function of λ based on n realizations of X

$$L(\lambda; x_{1:n}) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!(e^\lambda - 1)}. \quad (1.2)$$

To continue we compute the log-likelihood and the score function

$$l(\lambda; x_{1:n}) = \sum_{i=1}^n \ln \left(\frac{\lambda^{x_i}}{x_i!(e^\lambda - 1)} \right) \propto \sum_{i=1}^n (x_i \ln(\lambda) - \ln(e^\lambda - 1)) = \quad (1.3)$$

$$\bar{x}n \ln(\lambda) - n \ln(e^\lambda - 1), \quad (1.3)$$

$$S(\lambda) = l'(\lambda) = \frac{\bar{x}n}{\lambda} - \frac{ne^\lambda}{e^\lambda - 1} \quad (1.4)$$

from which we can derive the maximum likelihood estimation of lambda

$$\bar{x} = \frac{\hat{\lambda}_x}{1 - e^{-\hat{\lambda}_x}}. \quad (1.5)$$

When comparing the Poisson model with the Binomial, we will use their Fisher informations as a measure of variance. The Fisher information of λ_x is derived as following

$$I(\lambda) = -S'(\lambda) = \frac{n\bar{x}}{\lambda^2} + \frac{ne^\lambda}{e^\lambda - 1} - \frac{ne^{2\lambda}}{(e^\lambda - 1)^2}. \quad (1.6)$$

3.2 Binomial model

Consider the stochastic variable Y_i , which we define as the sum $\sum_{j=1}^m y_{i_j}$, where $y_{i_j} \in \{0, 1\}$ and m is the number of time intervals the search period is split in to create the binary search history. The sum $\sum_{j=1}^m y_{i_j}$ will follow a zero-truncated binomial distribution with parameters p and m , we regard the findings of scats in time as a Poisson process with λ being equal to the rate of the Poisson distribution in the Poisson model. Given the memoryless properties of a Poisson process, the number of scats found in one time interval then will be $Po(\frac{\lambda}{m})$, which give us

$$p = Pr(Po(\frac{\lambda}{m}) > 0) = 1 - Pr(Po(\frac{\lambda}{m}) = 0) = 1 - e^{-\frac{\lambda}{m}}. \quad (2.0)$$

In this model we regard p as fixed for the whole search period, which is a simplification of the model currently used in the Swedish brown bear survey, where they allow p to vary between the m intervals [3].

With this definition of the Binomial model, we can calculate the equations needed to compute an estimation of the bear population.

We derive the zero-truncated binomial distribution as follows

$$P(Y = k | Y > 0) = \frac{P(Y = k)}{P(Y > 0)} = \frac{Bin(m, 1 - e^{-\frac{\lambda}{m}})}{1 - P(Bin(m, 1 - e^{-\frac{\lambda}{m}}) = 0)} = \quad (2.1)$$

$$\frac{\binom{m}{k} (1 - e^{-\frac{\lambda}{m}})^k (e^{-\frac{\lambda}{m}})^{m-k}}{1 - e^{-\lambda}}. \quad (2.1)$$

Which gives us the likelihood

$$L(\lambda; y_{1:n}) = \prod_{i=1}^n \frac{\binom{m}{y_i} (1 - e^{-\frac{\lambda}{m}})^{y_i} (e^{-\frac{\lambda}{m}})^{m-y_i}}{1 - e^{-\lambda}}, \quad (2.2)$$

and log-likelihood

$$l(\lambda; y_{1:n}) = \sum_{i=1}^n \ln \left(\frac{\binom{m}{y_i} (1 - e^{-\frac{\lambda}{m}})^{y_i} (e^{-\frac{\lambda}{m}})^{m-y_i}}{1 - e^{-\lambda}} \right) \propto \quad (2.3)$$

$$\sum_{i=1}^n (y_i \ln(1 - e^{-\frac{\lambda}{m}}) + (m - y_i) \ln(e^{-\frac{\lambda}{m}}) - \ln(1 - e^{-\lambda})) = \quad (2.3)$$

$$\bar{y}n \ln(1 - e^{-\frac{\lambda}{m}}) + n(m - \bar{y}) \ln(e^{-\frac{\lambda}{m}}) - n \ln(1 - e^{-\lambda}). \quad (2.3)$$

Following this, we derive the score function

$$S(\lambda) = l'(\lambda) = \frac{n\bar{y}e^{-\frac{\lambda}{m}}}{m(1 - e^{-\frac{\lambda}{m}})} - \frac{ne^{-\lambda}}{1 - e^{-\lambda}} - \frac{n(m - \bar{y})}{m} \quad (2.4)$$

which gives us the maximum likelihood of λ_y

$$\bar{y} = \frac{m \left(-e^{-\frac{\lambda}{m}} + 1 \right)}{-e^{-\lambda} + 1}. \quad (2.5)$$

Since we have the score function from equation 2.4, we can compute the Fisher information

$$I(\lambda) = \frac{ne^\lambda \left(((\bar{y} - m) e^\lambda - \bar{y}) e^{\frac{\lambda}{m}} - me^\lambda \right)}{m \left(e^{\frac{\lambda}{m}} - 1 \right) (e^\lambda - 1)^2} + \quad (2.6)$$

$$\frac{ne^{\frac{\lambda}{m}} \left(((\bar{y} - m) e^\lambda - \bar{y}) e^{\frac{\lambda}{m}} + me^\lambda \right)}{m^2 \left(e^{\frac{\lambda}{m}} - 1 \right)^2 (e^\lambda - 1)} - \quad (2.6)$$

$$\frac{n \left((y - m) e^{\frac{\lambda}{m} + \lambda} + \frac{((y - m) e^\lambda - y) e^{\frac{\lambda}{m}}}{m} + me^\lambda \right)}{m \left(e^{\frac{\lambda}{m}} - 1 \right) (e^\lambda - 1)}. \quad (2.6)$$

3.3 Violations of model assumptions

In both the Poisson and the Binomial model, we view the scats found from bear i in time as a Poisson process, which brings the assumption of independence in time between events (scats found). We also assume that the rate of this process (λ) is the same for all bears in our population. There are reasons to believe that these two assumptions are not fulfilled in data. For both assumptions separately, violations of it could lead to overdispersion [4]. Hence, the amount of overdispersion alone is not enough to tell us the exact degree to which each assumption is broken, which complicates the process of adjusting for the violations.

One explanation of the overdispersion from dependent events in time, is the fact that multiple scats from one bear can be found during one visit into its territory. The two models will be affected by this in different ways. In the Binomial model, it does not matter if a bear is found one or multiple times during one occasion, since we only base our estimate on the binary capture histories. In the Poisson model however, if multiple scats from one bear are found during a short time period, it will affect the full counts and which in turn affects the population estimate. This can be simulated/modeled with a compound Poisson process, where we assume that each catch generates a random number of scats. If the random number comes from a logarithmic distribution, which has the density function

$$f(k) = \frac{-1}{\ln(1-p)} \cdot \frac{p^k}{k} \quad (3.0)$$

for $k \in \{1, 2, 3, \dots\}$, the total number of scats from bear i during the whole search period will follow a negative binomial distribution. Further, if we let $\lambda = -r \ln(1-p)$ be the rate in which catches occur in the Poisson process, and each catch result in a $Log(p)$ distributed number of scats - the total number of scats from bear i during the whole search period will be $NB(r, p)$ distributed [7].

When it comes to the overdispersion from individual captures rates (λ), one can imagine that some areas are more frequently visited by the people who are collecting the scats, and that the distance from a given bears territory to one of tease places will affect the

rate at which the bear is caught. To simulate/model data where bears have individual capture rates, we use a Gamma-Poisson mixture that also follows a negative binomial distribution. If we let each bear i have its own λ_i drawn at random from a $Gamma(\lambda k, k)$ distribution with expected value λ and variance λk . The total number of scats from bear i given λ_i will be $Poisson(\lambda_i)$ distributed, while the unconditioned distribution will be $NB(\lambda k, \frac{k}{k+1})$ [7], with expected value λ and variance $\lambda(k+1)$.

If we want to compare how the different sources of overdispersion affect our two models, we can choose the parameters that will give the same negative binomial distribution. This is done by choosing parameters that satisfies

$$\begin{cases} k\lambda = r \\ \frac{k}{k+1} = p \end{cases} \quad (3.1)$$

One can imagine a scenario where overdispersion comes partly from dependent events and partly from individual capture rates. To model/simulate this, we use the fact that $X + Y$ is $NB(r_1 + r_2, p)$ distributed if $X \sim NB(r_1, p)$ and $Y \sim NB(r_2, p)$ [1]. If we use $Gamma(qk\lambda, k)$ instead of $Gamma(k\lambda, k)$ when we model the overdispersion coming from individual capture rates, the unconditioned distribution of scats from bear i (X) will be $NB(qk\lambda, \frac{k}{k+1})$. We can then choose r and p so that our first model of overdispersion gives a number scats from bear i (Y) that is $NB((1-q)k\lambda, \frac{k}{k+1})$ distributed, by choosing parameters that satisfies

$$\begin{cases} k\lambda(1-q) = r \\ \frac{k}{k+1} = p \end{cases} \quad (3.2)$$

Adding X and Y will then give the distribution $NB(k\lambda, \frac{k}{k+1})$. When $q = 1$, the overdispersion will only come from individual capture rates, when $q = 0$ the overdispersion will only come from the logarithmic number of scats from each catch and when $0 < q < 1$ the overdispersion will come from a combination of both. Depending on q , the overdispersion will manifest itself differently in how the findings of scats are distributed across time - which will affect our estimates with the Binomial model but not the Poisson.

4 Data

The data used in this thesis consists of scats collected in conjunction with the moose hunt in the counties Västernorrland and Jämtland the year 2020 (the moose hunt takes place in autumn and the search for scats lasts approximately 11 weeks). The scats were picked up by hunters and volunteers, then sent to a laboratory where DNA extractions were made to identify unique individuals from the scats.

When dividing the search period into m intervals for binary capture histories, the interval length chosen in the current estimation method is one week. Using the same length with our data results in $m = 11$. During these 11 weeks, a total number of 3838 scats were collected from 1154 unique individuals.

5 Computations and simulations

Using the theory and data described in the previous sections, we can compare the two models (binary capture histories versus full count, which are referred to as the Binomial

and Poisson model respectively) with computations and simulations. In the calculations below regarding the Binomial model, the binary capture histories are generated with interval length one week ($m = 11$) if nothing else is stated, since this is the interval length currently used to estimate the Swedish brown bear population.

5.1 Maximum likelihood and population estimates

With equations 1.5 and 2.5 we can calculate the maximum likelihood estimate of λ for both the Poisson and the Binomial model. This is done numerically, and the results are

$$\begin{aligned}\hat{\lambda}_x &= 3.189, \\ \hat{\lambda}_y &= 2.728\end{aligned}$$

where $\hat{\lambda}_x$ is the estimation with the Poisson model and $\hat{\lambda}_y$ with the Binomial. Using these estimations of λ we can estimate the total number of bears in the population targeted by the data collection (note that this is not whole Swedish brown bear population) with the equation

$$\hat{N} = \frac{n}{1 - \hat{P}} \quad (4.1)$$

where N denotes the total number of bears in the population, n the number of unique bears in our data (1154) and P the probability that a bear in the population is found zero times. If we let $X \sim Po(\hat{\lambda}_x)$, and $Y \sim Bin(11, \hat{p})$ where $\hat{p} = 1 - e^{-\hat{\lambda}_y/11}$ (see equation 2.0), we have that

$$\begin{aligned}\hat{P}_x &= Pr(X = 0) = e^{-\hat{\lambda}_x} = 0.041 \\ \hat{P}_y &= Pr(Y = 0) = e^{-\hat{\lambda}_y} = 0.065,\end{aligned}$$

which gives the population estimates

$$\begin{aligned}\hat{N}_x &= 1204 \\ \hat{N}_y &= 1235.\end{aligned}$$

5.2 Fisher information

The maximum likelihood estimations combined with equation 1.6 and 2.6 gives us the observed Fisher informations for both models. The results are

$$\begin{aligned}I(\hat{\lambda}_x) &= 326 \\ I(\hat{\lambda}_y) &= 312.\end{aligned}$$

The interpretation of this result is that $\hat{\lambda}_x$ contains more information about λ than $\hat{\lambda}_y$, which is one of the main reasons why we might prefer the Poisson model over the Binomial. When generating the binary capture histories we ignore parts of the data that is included in the full counts, the results above confirm that this has an effect on our estimation of λ which in turn will affect the population estimate. The amount of information an estimate contains about the data is inversely proportional to the variance of the estimate [2], this fact together with the results above means that the variance of $\hat{\lambda}_y$ is greater than the variance of $\hat{\lambda}_x$. The exact (by seven decimals) values are

$$\begin{aligned}Var(\hat{\lambda}_x) &= 0.0030707 \\ Var(\hat{\lambda}_y) &= 0.0032004.\end{aligned}$$

The observed Fisher informations discussed above is dependent on our observed data. To compare the models without this dependence, we use the expected Fisher informations instead. This gives us the possibility to see how our models behave when the capture rate λ changes. Using equations 1.6 and 2.6 with the expected values instead of means for x and y , we compute and illustrate how the expected Fisher for both models change depending on the value of λ .

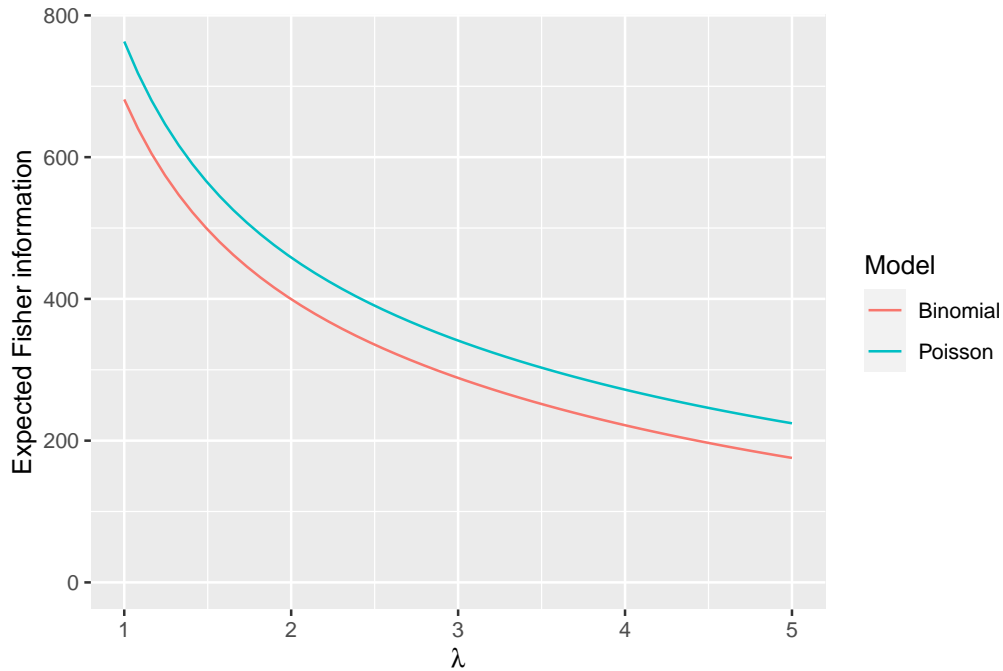


Figure 1: Expected Fisher information for both models depending on the value of lambda

The conclusions drawn from this result are that the estimates from the Poisson model contain more information (in relation to the Binomial) independent of our observed data. We can also note that the relative difference seems to depend on λ , to investigate this further we plot the quotient between the expected Fishers from the two models.

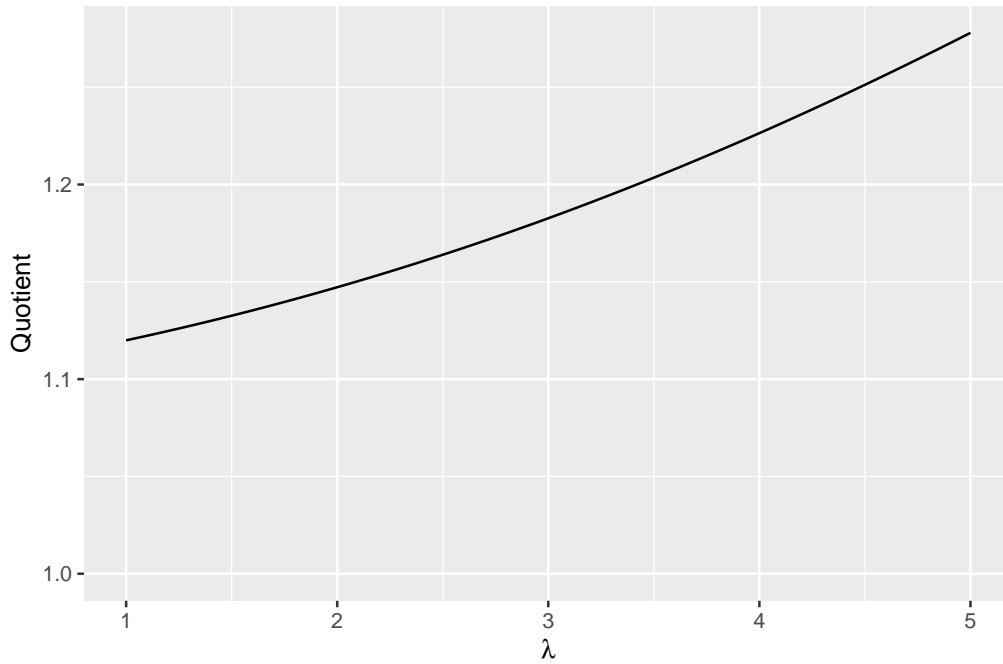


Figure 2: Quotient between expected Fisher information from the Poisson model and expected Fisher information from the Binomial model with 11 intervals

The difference in expected Fisher between the models grows from 12% to 28% when λ increases from 1 to 5. This indicates that the problem of lost information with regards to λ when generating the binary capture histories is larger for populations where the animals are captured more frequently.

5.3 Altering interval lengths in the Binomial model

To explore how the Binomial model changes depending on the number of intervals (synonymous with the interval length, higher number of intervals gives shorter intervals), we create a similar graph as in Figure 1, with lines representing the expected Fisher for the Binomial model with $m = 2$, $m = 11$ and $m = 40$.

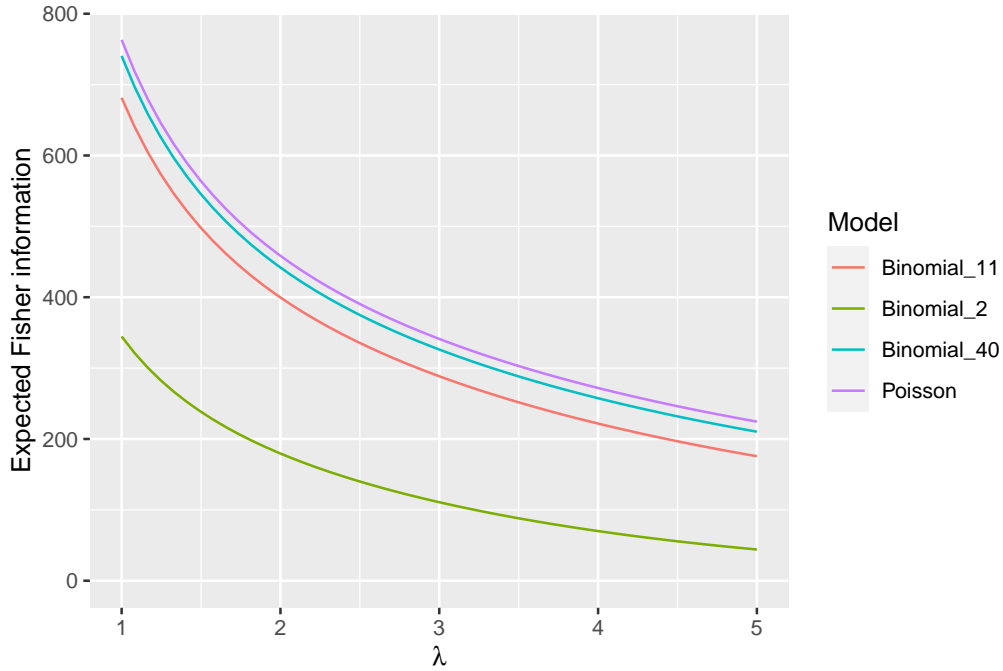


Figure 3: Expected Fisher depending on lambda with different interval lengths for the Binomial model

From Figure 2 we know that the difference in expected Fisher between the Poisson and the Binomial model changes depending on the capture rate λ . The graph above (Figure 3) shows us that this difference is also dependent on the interval length used in the Binomial model. When the interval lengths decrease, so does the difference between our two models.

5.4 Overdispersion

We want to investigate how the models perform when our assumption of λ being equal for all bears in the population is broken, which will lead to overdispersion. One way of doing this is to introduce individual heterogeneity and simulate captures from a bear population with 1000 individuals that have individual capture rates λ_i . We model this by giving each individual their own λ_i drawn at random from a $Gamma(k\lambda, \lambda)$ distribution with expected value λ and variance $k\lambda$. The total number of scats from bear i , given λ_i , will be Poisson (λ_i) distributed, while the unconditioned distribution will be $NB(k\lambda, \frac{k}{k+1})$. With this distribution, we can easily choose the amount of overdispersion by altering the value of k , where a higher value will result in more heterogeneity.

An interesting thing to examine here, is how the variance of the maximum likelihood estimate changes with k in the Poisson model compared with the Binomial with different number of intervals.

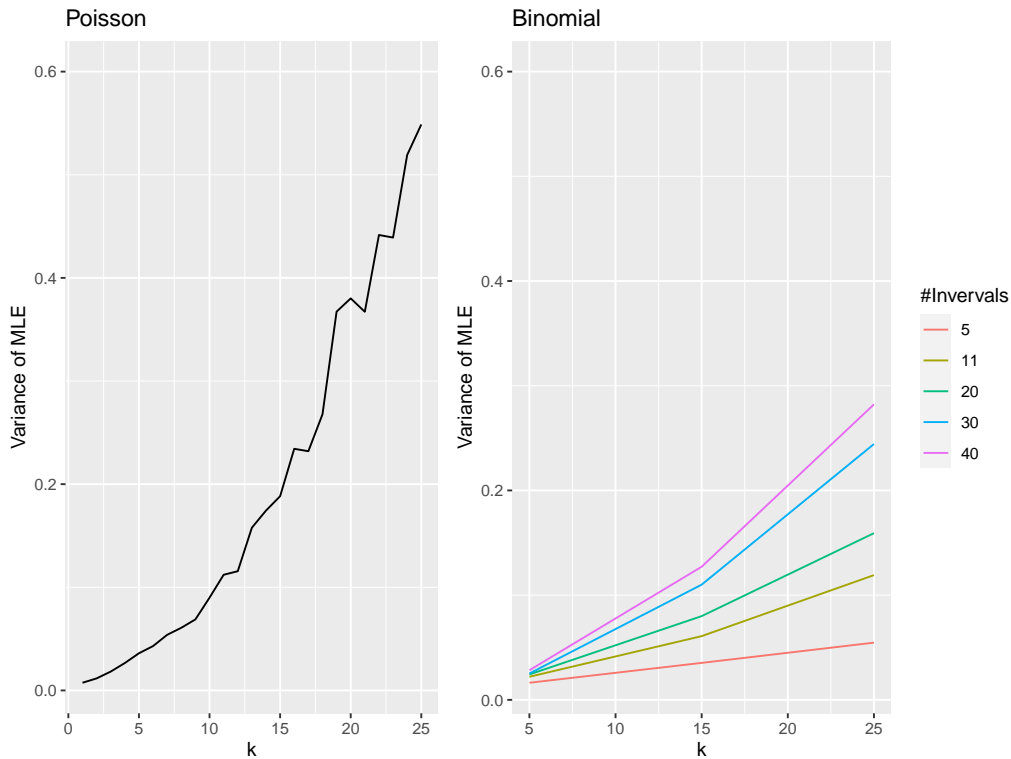


Figure 4: Variance of maximum likelihood estimation (MLE) of lambda for both models on data with individual heterogeneity

In Figure 4 we note that the variance of $\hat{\lambda}$ goes up for both models when we increase the value of k , which is expected since it leads to more individual heterogeneity. We use five different interval lengths for the Binomial model, and it looks like their differences increase exponentially with respect to k . Overall, we see that the Binomial model is more robust towards overdispersion from individual heterogeneity than the Poisson model. While it might be hard to see in the graph, when $k = 1$ the variance is approximately equal to what we get when we inverse the observed Fisher information calculated above (≈ 0.0031 for the Poisson model).

The introduction of individual heterogeneity will not only affect the variance displayed in Figure 4, it will also have an impact on the population estimates. We use the bias of the maximum likelihood estimations as a metric to measure this.

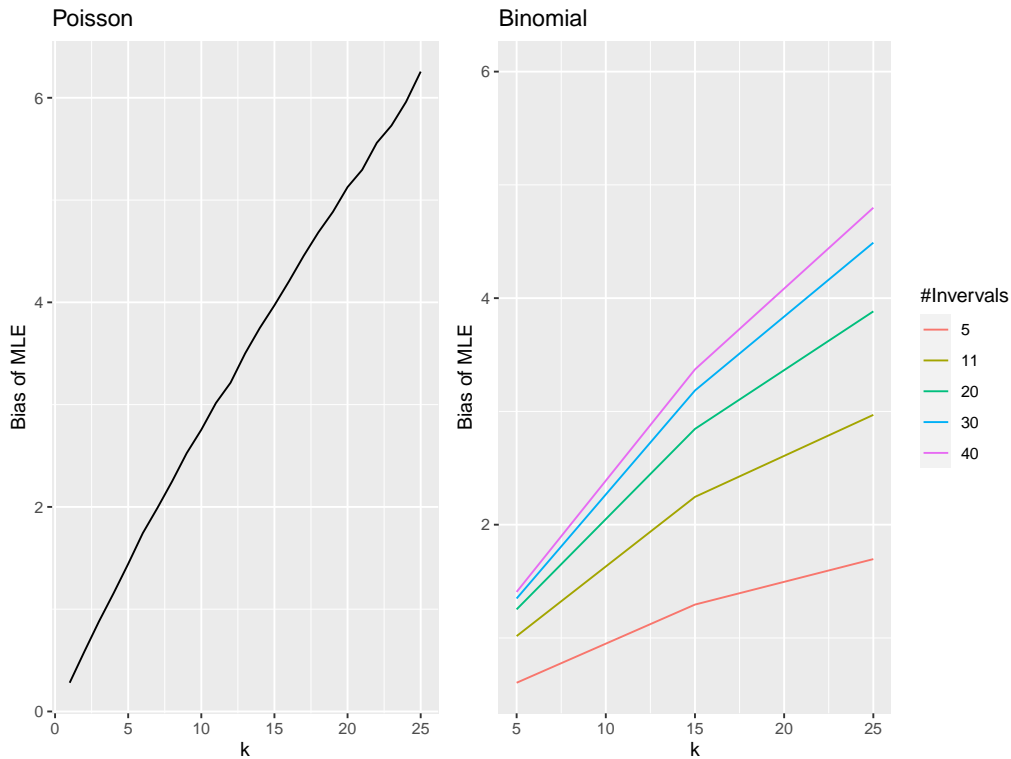


Figure 5: Bias of maximum likelihood estimation (MLE) of lambda for both models on simulated data with individual heterogeneity

In Figure 5, we see similar results as in Figure 4. The Poisson model is more sensitive to overdispersion compared to the Binomial, and that using fewer intervals makes the difference larger.

Another way of introducing individual heterogeneity, is to let all individuals have the same probability of being caught (same λ), but letting each catch result in a random number of collected scats instead of just one. Here we define a catch as finding one or more scats during one occasion. We model this by simulating a population with 1000 individuals, each one being caught with rate $\lambda = -r \ln(1 - p)$ and each catch leaving a number of scats that follow a logarithmic distribution $Log(p)$. The total number of scats from bear i in this simulation will follow a $NB(r, p)$. If we chose the parameters r and p by the Equation in 3.1, and let k go from 1 to 25 in the same manner as in the previous simulation, we should get equal results from the simulations when using the Poisson model since the distribution of total number of scats will be the same.

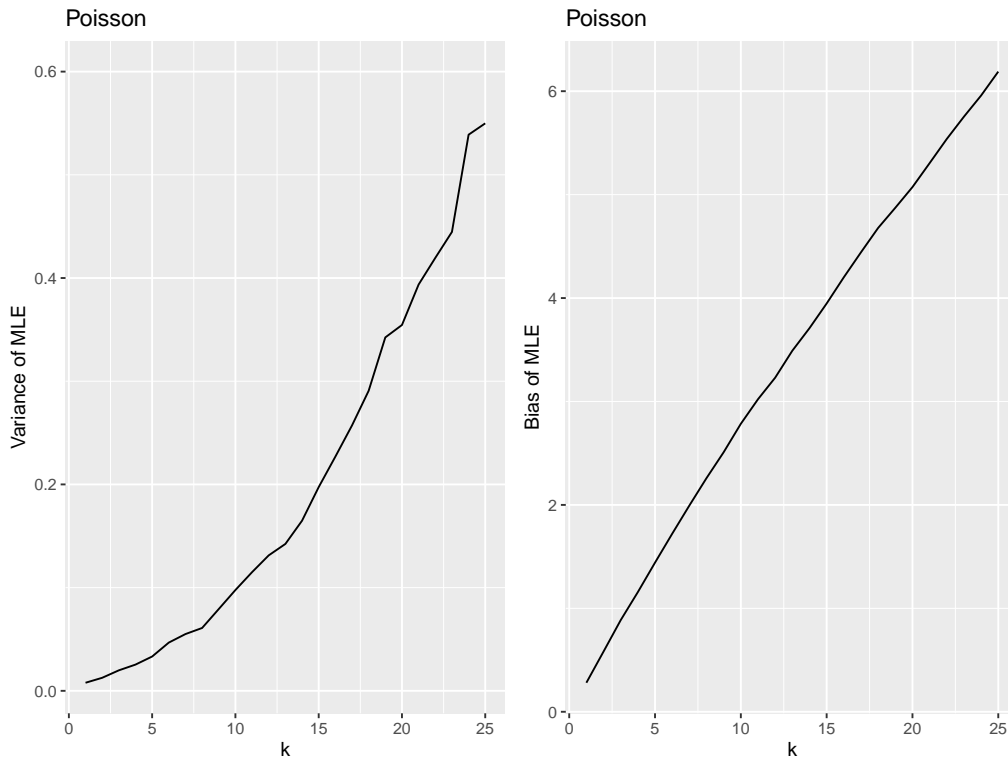


Figure 6: Variance and Bias of maximum likelihood estimation (MLE) of lambda from the Poisson model on simulated data with overdispersion from a compound Poisson process

Comparing Figure 6 with the Poisson part of Figure 4 and 5, we confirm that the source of overdispersion does not affect the population estimate when working with the Poisson model. However, the estimate will change depending on the source when using the Binomial model.

To investigate how the source of overdispersion affects the Binomial model, we combine the two methods we have used to simulate overdispersion, and use a parameter q to regulate the amount from each source. When $q = 1$ the overdispersion will only come from individual capture rates, when $q = 0$ it will only come from the compound Poisson process and when $1 > q > 0$ it will come from a combination of both. Independent of q , the total number of scats from one bear will be $NB(k\lambda, \frac{k}{k+1})$ distributed. The theory behind and details regarding this simulation are explained more thoroughly in Theory, under *Violations of model assumptions* in conjunction with Equation 3.2.

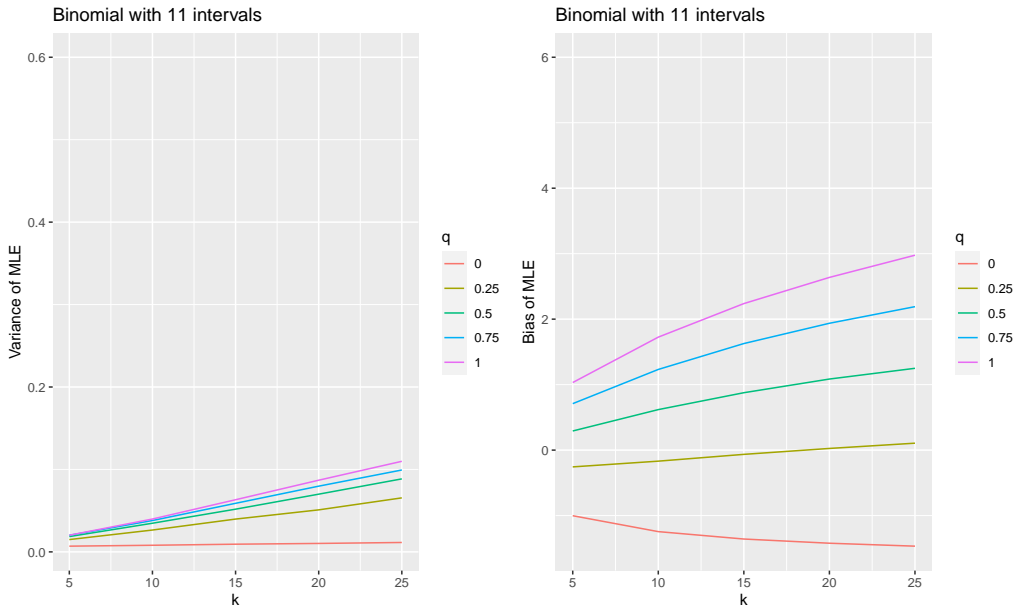


Figure 7: Variance and Bias of maximum likelihood estimation (MLE) of lambda from the Binomial model with 11 intervals on simulated data with overdispersion from both sources

To get a better understanding of Figure 7 and the simulation, note that when $q = 1$, the variance and bias are equal to the ones found in Figure 4 and 5 for the Binomial model with 11 intervals. Furthermore, we see that the source of overdispersion affects the population estimate when using the Binomial model.

For $q > 0$ the variance grows with k and when $q = 0$ the value of k does not affect the variance, which is expected since the model does not differentiate between one or multiple scats being found at the same time. When $q = 0$, an increase in k only affects the variance with regards to the number of scats found on each event (an event here is defined as one or more scats being found on one occasion).

Since the expected value of the total number of scats from each bear remains the same independent of q , when $q < 1$ a higher value of k will result in fewer events for each bear (on average), since a larger proportion of all events will result in more than one scat being found. This does not affect the bias when using the Poisson model (since the expected number of total scats remain the same), but as we see in Figure 7, it leads to lower estimates of λ in the Binomial model (since the expected number of events decrease).

Note that the individual heterogeneity increase our bias, so even though it is true that for all $q < 1$ that a higher value of k results in fewer events for each bear (on average), there seems to be a threshold when $0 < p < 0.25$ where the increase in positive bias from individual heterogeneity makes the total bias grow with k instead.

6 Discussion

There are multiple metrics one can use when comparing the two models and their utility. The importance of these metrics will depend on situational circumstances, not all of which can be considered in this thesis. One of the most basic metric to look at when

comparing the models are the observed Fisher information, we have $\hat{\lambda}_x = 312$ and $\hat{\lambda}_x = 326$ for the Binomial and Poisson respectively. The observed Fisher information measures the amount of information our variables \bar{x} and \bar{y} (from equations 1.5 and 2.5) contain regarding λ . Due to its simplicity and straightforward relevance, this measure works good as a stepping stone for more complex comparisons. From Figure 2 and 3, we see that the difference in expected Fisher is dependent on both λ and the number of intervals used in the Binomial model. Since one of the main benefits of the Poisson model is the additional information gained, and we know from Figure 2 that this gain grows exponentially with λ , using full counts seems to be more favorable in relation to binary capture histories when estimating populations with high capture rates. An alternative to replacing the current estimation method with one using full counts, could be to make the time intervals used when generating the binary capture histories shorter. In figure 3 we see that this will make the gap between the models smaller, thus making it possible to gain more information without compromising other possible benefits of using binary capture histories (besides possible statistical advantages there could be practical benefits to this, such as software accessibility, amount of established theory et cetera).

Introducing simulated data with overdispersion gives nuance to the comparison and discussion of the models. In Figure 3 and 4 when examining the simulations with individual heterogeneity, we see that the maximum likelihood estimation from the Poisson model is more sensitive to individual heterogeneity than the maximum likelihood estimation from the Binomial. We also find that the number of intervals we divide the searching period into has an impact on both variance and bias of the maximum likelihood estimation. The amount of individual heterogeneity that exists in the data is not investigated in this thesis. However, we prefer a model that is more robust towards it over a more sensitive one, everything else being equal. But as we noted in the discussion about their Fisher informations in the paragraph above, everything else is not equal between the models. We get more information when using full counts, so based on our beliefs about how much individual heterogeneity that exists in data, we might prefer one model over the other.

The overdispersion in data probably stems from multiple factors, not only individual heterogeneity. As is discussed in Theory under *Violations of model assumptions*, it is reasonable to assume that some of the catches should be clumped together and counted as one, the reason for this being that multiple scats can be found from one bear on one occasion, and this should not be counted as the bear being found multiple times. We model this with a compound Poisson process, where each event (catch) results in a random number of scats which follow a logarithmic distribution. This will affect the two models in different ways, since the binary capture histories remove the differentiation between one or multiple scats being found in one interval. When comparing Figure 6 with Figure 4 and 5, we confirm that the source of overdispersion does not affect our estimate of λ for the Poisson model. When investigating how this affects the Binomial model, we use data simulated from both scenarios that cause overdispersion, with a parameter q which dictates the proportion from each source. The total number of scats in this data will follow a negative binomial distribution independent of q , but q will affect how scats are distributed in time which in turn will have an impact on our estimates. In Figure 7 we note that the variance of the maximum likelihood estimate of λ is not affected by k when the simulated data only comes from the compound Poisson process ($q = 0$), but that the bias is. This is explained by the fact that expected number of scats found for each bear remains the same independent of q and k , and when k increase a larger proportion of the scats are found together with other scats and are thus “ignored” by the binary capture histories making the estimation of λ smaller. An effect of this is that the bias is negative when q is close to zero (how close will depend on k) and positive otherwise, which is interesting since a negative bias will lead to an overestimation of the total population

while a positive will lead to an underestimation.

In all simulations discussed above, if we consider the mean squared error (MSE) which is defined as $Variance + Bias^2$, we can note that the effect overdispersion has on bias will contribute a lot more to this error than the effect it has on variance. This indicates that the biggest problem that comes with overdispersion is increased bias. The mean errors (ME) of our estimations on the simulated data can be found in the appendix (Figure 8 and 9).

A natural expansion of this thesis could be done by examining the amount of overdispersion that arise when applying the models to the data used to estimate the Swedish brown bear population, both in terms of how the scats are distributed in time and the amount of individual heterogeneity. Together with a further investigation of different ways in which the effect that overdispersion has on our models could be mitigated.

7 Appendix

This appendix contains two figures displaying the mean error of the λ estimates from the simulation studies.

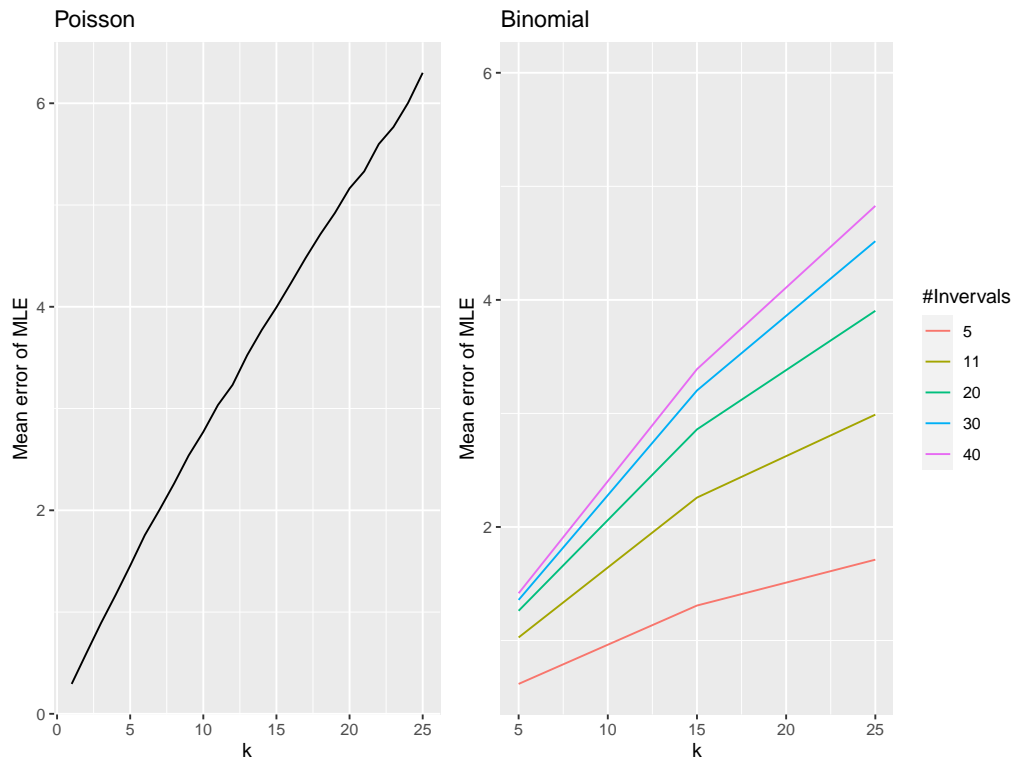


Figure 8: Mean error of maximum likelihood estimation (MLE) of lambda from both models on simulated data with individual heterogeneity

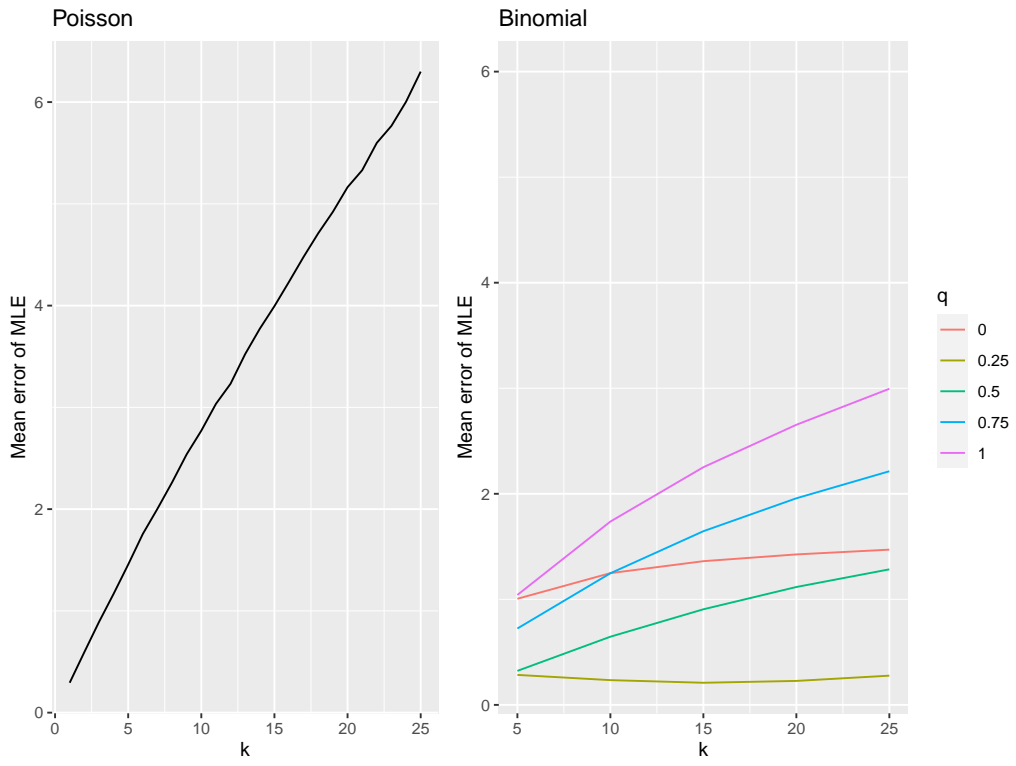


Figure 9: Mean error of maximum likelihood estimation (MLE) of lambda from both models (11 intervals in the Binomial) on simulated data with overdispersion from both sources

8 References

1. Furman, E. (2007). On the convolution of the negative binomial random variables. *Statistics & Probability Letters* *77*, 169–172.
2. Held, L., and Bove, D.S. (2013). Applied statistical inference. In (Berlin, Germany: Springer), p. 28.
3. Kindberg, J., Swenson, J.E., Ericsson, G., Bellemain, E., Miquel, C., and Taberlet, P. (2011). Estimating population size and trends of the swedish brown bear *ursus arctos* population. *Wildlife Biology* *17*, 114–123.
4. Lindsey, J.K. (1995). Modelling frequency and count data. In (New York: Clarendon Press), p. 231.
5. McDonald, D.R. (1980). On the poisson approximation to the multinomial distribution. *Canadian Journal of Statistics* *8*, 115–118.
6. Schofield, M.R., Barker, R.J., and Gelling, N. (2018). Continuous-time capture-recapture in closed populations. *Biometrics* *74*, 626–635.
7. Zhou, M., and Carin, L. (2015). Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* *37*, 307–320.