Stockholms universitet

Individual heterogeneity and identifiability in estimation of brown bear population size in Sweden

Peter Hjortskov Andersen

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2022:15
Matematisk statistik
Juni 2022

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2022:15**
http://www.math.su.se

Individual heterogeneity and identifiability in estimation of brown bear population size in Sweden

Peter Hjortskov Andersen*

June 2022

Abstract

In this thesis we investigate the estimation of the Swedish brown bear population. Especially, we look into whether problems connected to individual heterogeneity and identifiability occur for the Swedish brown bear monitoring program. In former studies it has been argued that problems connected to these things are a general problem for wildlife studies. In this thesis, based on the Swedish brown bear population we look into these kinds of problems. We fit different statistical models of individual heterogeneity to the data from the monitoring program and check whether these models provide a good fit and whether they can be distinguished.

We find that problems related to individual heterogeneity do indeed occur in the Swedish brown bear program. Several statistical models can be fit well to data but the models estimation of the population size varies a lot. This makes the estimation difficult since it is not clear which model is the optimal one to use for the estimation. This means that even though a good fit have been find you cannot know for sure that the estimate provided by that model is a good estimate. The choice of the model and method will matter a lot for the final estimate but there is no clear way to distinguish these models. Furthermore, we compare the different statistical models and find some general information of what to expect from them when using them for this kind of estimation.

---

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: peter.andersen@newmail.dk. Supervisor: Martin Sköld.

Acknowledgements

I would like to direct my thanks and gratitude to my supervisor, Martin Sköld, for his inputs, ideas and encouragement which has been useful throughout the writing process. Without him the project would not have been the same. Furthermore, I would like to thank the Swedish Museum of Natural History for providing data and making this interesting project possible.

Contents

1 Introduction

The problem of determining unknown size of wildlife population is of great importance and is relevant in many cases. Applications of such estimates of population sizes includes quality control, regulation, information of wildlife areas and so on. In general, these estimates of population sizes are commonly made using capture-recapture methods. Capture-recapture methods consist of combining information of observed individuals with plausible statistical models of the observation process. Combining these may lead to an estimate of the non-observed individuals and thereby an estimate of the whole wildlife population. The statistical theory behind the capture-recapture methods for wildlife estimation have been used for many years. Already in [Eberhardt, 1969] an approach where the observed captures is assumed to be zero-truncated was used and this problem was studied further in [Sanathanan, 1977]. This is the same kind of approach that will be used in this bachelor thesis.

Elements and choices of the method in the above described procedure may make the final estimate more or less precise. And there can be many problems involved in making these estimates in wildlife studies. This project will look into the problem of individual heterogeneity and its consequences for wildlife estimation. Individual heterogeneity occurs when the probability of being captured vary among individuals. This can be due to several things for example variation in effort from the sample collectors, some individuals may live in more easy accessible areas, some animals may spend much time outside the surveyed area and so on. There are strong evidence of individual heterogeneity in many cases. When captures of an individual is assumed to follow a Poisson distribution, individual heterogeneity manifest as a variance larger than the mean.

This project will be based on data from the Swedish brown bear monitoring program and look into whether problems related to individual heterogeneity hold for the monitoring program. Individual heterogeneity can be put into the statistical model in different ways. In [Link, 2003] he argues that different models of individual heterogeneity may lead to very different estimates and that these models can be difficult to distinguish. He shows that several models for individual heterogeneity can be fitted well to data but that the population estimates are very different in the models. Even a 95% confidence interval of the population estimate do not overlap for some of the models.

In this thesis we will mainly work with zero-truncated models. In a zero-truncated model no zero-valued observations can occur and the distribution will have support only on the strictly positive numbers. For example will a zero-truncated Poisson model, which is the basis distribution used in the analysis in this thesis, have a probability mass function on the form

$$P(Y=y) = p_\lambda(y)/(1-p_\lambda(0)), \quad y = 1, 2, \dots,$$

where $p_\lambda(y) = \exp(-\lambda)\lambda^y/y!$, $y = 0, 1, \ldots$, is the probability mass function of the Poisson(λ)-distribution. If we assume that captures from each individual follow the Poisson distribution but that observed captures follow a zero-truncated Poisson distribution then we can fit the zero-truncated model to data. Then we can estimate the zero-valued observations or the non-observed individuals as $p_{\hat{\lambda}}(0)$. So with zero-truncated models we have a clear procedure to find the probabilities to zero.

The overall problems connected to individual heterogeneity are that there are different models that can fit the captures of the observed individuals. But these fitted models put very different weight to the bears with no captures at all. In [Link, 2003] as well as in [Huggins, 2001] they argue that from a theoretical perspective, when the fitted densities of the observed individuals are close to each other then should the estimate of the non-observed individuals be close to each other as well. Essentially, they find that two candidate models that generate the same zero-truncated distribution must give the same probability to zero when not truncated. But if the captures are assumed to come from a zero-inflated distribution then it is not possible to determine the population size since it would require information about the amount of zero-inflation. In a zero-inflation model we know that zero-valued observations happen but we do not know how many zero-values to expect. This is different from the zero-truncated models where we always have a procedure to determine the probabilities to zero. However, [Link, 2003] finds that also with zero-truncated models problems of identifiability occur as explained above. Models that generate the same zero-truncated distribution provide different probabilities to zero when not truncated.

Both [Link, 2003] and [Huggins, 2001] look into circumstances and restrictions in the observation process that could make these problems of nonidentifiability between the models possible. They do indeed find some conditions that should rule out the possibility of nonidentifiability. But these conditions are not always suitable and in line with data and [Link, 2003] finds different cases and situations where problems of identifiability occur. In this thesis we investigate if these problems also occur for estimation of the size of the Swedish brown bear population.

In this project, we will do a similar approach as in [Link, 2003] to determine whether the same problems hold for the Swedish brown bear program. The Swedish brown bear monitoring program is carried out every year by volunteer sample collectors. A detailed description of the program and the current method for estimating the population size is described in [Kindberg et al., 2011].

We will use an approach where we look at total counts and assume that the capture probability follows a mixture of Poisson distributions. We the fit different kind of these mixture distributions to the data. This is different from [Link, 2003] and most earlier literature where they look at binary capture histories and look at a mixture of Binomial distributions. However, most of the mixture models in the case of using total counts corresponds to similar mixture models in the case of binary capture histories and Binomial distributions. In this thesis, some of the conclusions are the same as in [Link, 2003]. This project shows that problems related to individual heterogeneity also occur in the Swedish brown bear monitoring problem. It is very difficult to distinguish some of the models we have been working with as we will show later on. However, the final estimation of the brown bear population size is quite different for some of these models.

This means that there are some uncertainty in the final estimation and it is not clear which model and method may be the best to use for the estimation of the population size.
Everything in this bachelor thesis will be based on the Swedish brown bear monitoring program. However, the results and theories may be relevant for similar kind of wildlife population estimation and other areas where similar problems of individual heterogeneity occur.

2 Materials and Methods

2.1 Materials

As earlier mentioned, the dataset used in this bachelor project is from the Swedish brown bear monitoring program. The program usually takes place every year in different counties in Sweden. The program is based on volunteers collecting scat samples. Later, DNA-analysis will show which bears the samples comes from and how many samples there are collected for each bear [Kindberg et al., 2011].

The datasets analysed in this project are from the 2015, 2016, 2017, 2019 and 2020 monitoring program. The datasets consist of all the captures made during the program including information about place, identification number and sex. Male bears may have a tendency to behave more irregular than female bears [Kindberg et al., 2011]. To make sure that this has not caused extreme data points which could make the fitting of the models difficult and interfere with the focus of this project, samples from the male bears have been removed from the dataset. Therefore, the datasets analysed in this project only consist of samples data for the female bears.

2.2 Methods

In general, we will assume that the number of captures for each bear follows a Poisson distribution with support in the natural numbers including zero. In the dataset, we do not observe the bears with no captures. Therefore, the observed captures in the dataset will follow a zero-truncated Poisson distribution. The probability mass function is then

$$P(Y = y) = p_\lambda(y)/(1 - p_\lambda(0)), \quad y = 1, 2, \ldots,$$

where $p_\lambda(y) = \exp(-\lambda)\lambda^y/y!$, $y = 0, 1, \ldots$, is the probability mass function of the Poisson(λ)-distribution. The expected proportion of the total population that are not encountered is $p_\lambda(0) = \exp(-\lambda)$ and we may estimate the total population size as $\hat{N} = n/(1 - \exp(-\hat{\lambda}))$, where n is the number of unique individuals captured.

It is well known in the literature that ignoring individual heterogeneity as in the model above will lead to a significantly underestimation of the true population size. So we want our statistical model to include individual heterogeneity. This is, that not every bear has the same probability of being captured. The way we model this is by a mixture of Poisson distributions. So if we let n_i be the number of times individual i is captured, then we assume $n_i \sim Poisson(\lambda_i)$ where $\lambda_i \sim P_\theta$ independently for all i and where P_θ is a distribution that describes the individual variation in catchability.

2.2.1 Current approach

The current approach for estimating the size of the Swedish brown bear population is not based on total counts and mixture of Poisson distributions. Instead, it is expected that the surveys generate binary capture histories for each individual. This require that the surveys are repeated. The Swedish monitoring program is carried out during around 11-12 weeks. In the current approach each week is used as a session for capture-recapture and then there are 11-12 repeated surveys [Kindberg et al., 2011].

With binary capture histories we only look at whether the individual was captured or not. So for each survey we have a success if the individual is captured at least once and a failure if the individual is not captured at all. For example will an animal captured in the first and last survey out of five generate the capture history $y = (0, 1, 0, 0, 1)$. To estimate the population size it is then necessary to estimate the animals never encountered which is equivalent to animals with capture history $y = (0, 0, 0, 0, 0)$.

In the current approach, individual heterogeneity is taken into account as mixture of binomial distributions or similar models. Different models are fitted to data and the model with the best fit is used to estimate the population size. In the case that two or more models cannot be distinguished, a combination of the best models is used for the estimation [Kindberg et al., 2011].

There can be arguments in favour of and against using a binary capture history as in the current approach. Some issues with this approach includes that the division into weeks are arbitrary but may change the estimate. Furthermore, there may be information lost when splitting into weeks and ignoring the total count for each individual. Another problem is that weekly estimates are required which implies that many parameters have to be fitted. Some of these issued can be dealt with when using an approach of using total counts and a mixture of Poisson distributions.

2.2.2 Distributions and choices of P_θ

In the analyses in this thesis we use total counts and fit models that are a mixture of Poisson distributions as explained earlier. This is, we use different models for P_θ and fit all of these to the data. We have looked at four different choices of P_θ. These are a $Gamma(\alpha, \beta)$ distribution, a $LogNormal(\mu, \sigma^2)$ distribution, a 2-point discrete mixture and a 3-point discrete mixture. Furthermore, we have fitted the simple Poisson distribution in order to compare it with the mixture of Poisson distributions.

All of these models of individual heterogeneity have been used in a similar way in earlier capture-recapture literature. The discrete mixture models have been suggested for wildlife applications by [Norris III and Pollock, 1996]. The Poisson-Lognormal model corresponds to a logit-normal model when considering capture histories as binary which has been studied by [Coull and Agresti, 1999]. The Poisson-Gamma model corresponds to a beta-binomial model in the binary case which was suggested in wildlife applications by [Dorazio and Andrew Royle, 2003].

In the case of the simple Poisson model we have that the observed captures follows a zero-truncated Poisson model. This is, that the density is

$$P(X = x) = p_\lambda(x)/(1 - p_\lambda(0)) = \frac{\frac{\lambda^x e^{-\lambda}}{x!}}{1 - e^{-\lambda}} = \frac{\lambda^x}{(e^\lambda - 1)x!}$$

for $y = 0, 1, 2 \ldots$, where $p_\lambda(x)$ is the density of a standard Poisson model.

In the case where the captures of each individual bear follows a mixture of Poisson-Gamma distribution, we have that the probability mass function is

$$P(Y = y) = \int_0^\infty f_{Poisson(\lambda)}(y) f_{Gamma(\alpha,\beta)}(\lambda) d\lambda =$$
$$\int_0^\infty \frac{\lambda^y e^{-\lambda}}{y!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} d\lambda = \frac{\beta^\alpha}{y!\Gamma(\alpha)} \int_0^\infty \lambda^{y+\alpha-1} e^{-(1+\beta)\lambda} d\lambda =$$
$$\frac{\beta^\alpha}{y!\Gamma(\alpha)} \Gamma(y+\alpha) \frac{1}{(1+\beta)^{y+\alpha}} = \frac{\Gamma(y+\alpha)}{y!\Gamma(\alpha)} \frac{\beta^\alpha}{(1+\beta)^{y+\alpha}},$$

for $y = 0, 1, 2 \ldots,$. We recognise this as the density of the negative binomial distribution with parameter $r = \alpha$ and success probability $p = \frac{1}{\beta+1}$.

The observed captures then follow a zero-truncated negative binomial distribution. The density is

$$P(X=x)=\frac{p_{r,p}(x)}{1-p_{r,p}(0)}=\frac{\Gamma(y+r)}{y!\Gamma(r)}\frac{(1-p)^r p^x}{1-(1-p)^r},$$

for $x = 1, 2 \ldots,$. With this probability mass function we can now determine the log-likelihood function and by maximizing the log-likelihood function we get the MLE estimates for the parameters in the model r and p [Lauritzen, 2021, p. 59-60]. In a similar way we have worked with the densities of the other statistical models.

For the Poisson-Lognormal model we have that the density is

$$P(Y=y)=\int_0^\infty f_{Poisson(\lambda)}(y)f_{Lognormal(\mu,\sigma^2)}(\lambda)d\lambda=$$
$$\frac{1}{\sqrt{2\pi}\sigma y!}\int_0^\infty \lambda^{y-1}e^{-\lambda-\frac{(\log(\lambda-\mu))^2}{2\sigma^2}}d\lambda$$

for $y = 0, 1, 2 \ldots,$. The observed captures then follow a zero-truncated Poisson-Lognormal model. The density has the form $P(X = x) = \frac{p_{\mu,\sigma^2}(x)}{1-p_{\mu,\sigma^2}(0)}$ for $x = 1, 2 \ldots,$. By the method of maximum likelihood we want to fit the model and determine estimates of μ and σ.

The discrete statistical models are a discrete mixture of the Poisson model. Thus, the probability mass function of the 2-point discrete mixture model is

$$P(Y=y)=\pi_1 p_{\lambda_1}(y)+\pi_2 p_{\lambda_2}(y)$$

for $y = 0, 1, 2 \ldots$, where the π's sum to 1. The observed captures then follow a discrete mixture model of zero-truncated Poisson distributions. Thus, the density of the observed captures is

$$P(X=x)=\pi_1 p_{\lambda_1}(x)/(1-p_{\lambda_1}(0))+\pi_2 p_{\lambda_2}(x)/(1-p_{\lambda_2}(0))$$

for $x = 1, 2 \ldots$, where the π's again sum to 1. In a similar way we get that with the 3-point discrete mixture model, the density of the observed captures is

$$P(X=x)=\pi_1 p_{\lambda_1}(x)/(1-p_{\lambda_1}(0))+\pi_2 p_{\lambda_2}(x)/(1-p_{\lambda_2}(0))+\pi_3 p_{\lambda_3}(x)/(1-p_{\lambda_3}(0))$$

for $x = 1, 2 \ldots$, where the π's again sum to 1. Therefore, we have six parameters to fit in the 3-point discrete mixture model which are the three π's and the three λ's. Again, we will use the method of maximum likelihood to determine the parameters.

These are the models and distributions that have been used for the analysis in this project.

2.2.3 Confidence intervals

For every estimate of the population size we have calculated a confidence interval. Two different approaches have been used. For all five statistical models used in this project, it has been possible to make a confidence interval using a bootstrap procedure. For the simple Poisson model and the Poisson-Gamma model it has furthermore been possible to calculate theoretical confidence intervals.

The bootstrap procedure has been carried out by resampling the datasets. This has been done after the dataset has been ordered by individuals and their number of sample captures. Therefore, the total number of uniquely observed individuals will remain constant. But the number of counts per individual will be resampled. This procedure may make the confidence intervals a bit smaller than if the unordered dataset had been resampled.
We have resampled the datasets 3000 times and computed the population estimates for each of the resamples. After this, a basic bootstrap method also known as reverse percentile interval have been used to calculate the confidence interval based on the estimates of the resampled datasets [Kashin, 2013]. In some cases this leads to intervals including negative numbers. The size of the population is known to be at least as large as the number of uniquely identified individuals. So in practice the lower bound of the confidence intervals could be cut off there. But we have left the intervals as they are to illustrate the change in variance between the different models. Therefore, the results in this project contain some confidence intervals with negative values.

For the computations of the theoretical confidence intervals, asymptotic theory has been applied. In the case of the simple Poisson model, it has been possible to calculate theoretical Fischer information and thereby constructing a Wald confidence interval [Lauritzen, 2021, p. 69-73]. The probability mass function of the zero-truncated Poisson model is

$$P(X=x)=\frac{\lambda^x}{(e^\lambda-1)x!}.$$

Hence, the loglikelihood function of just one observation is

$$l_x(\lambda)=x\log(\lambda)-\log(e^\lambda-1)-\log(x!).$$

The score function and information for all n observations are then

$$S_n(x,\lambda)=\sum_{i=1}^{n}\frac{x_i}{\lambda}-\frac{e^\lambda}{e^\lambda-1}$$

$$I_n(x,\lambda)=-\frac{\partial S_n}{\partial\lambda}=-\sum_{i=1}^{n}-\frac{x_i}{\lambda^2}+\frac{e^\lambda}{(e^\lambda-1)^2}$$

Then the Fischer information can be calculated

$$E(I_n) = \frac{ne^{\lambda}}{(e^{\lambda}-1)^2} + \frac{1}{\lambda^2}\sum_{i=1}^{n} E(x_i) = n\left(-\frac{e^{\lambda}}{(e^{\lambda}-1)^2} + \frac{e^{\lambda}}{\lambda(e^{\lambda}-1)}\right)$$

The variance of the MLE parameter $\hat{\lambda}$ will then asymptotically be the inverse Fischer information [Lauritzen, 2021, p. 69-73]. So

$$var(\hat{\lambda}) = \frac{1}{n\left(-\frac{e^{\lambda}}{(e^{\lambda}-1)^2} + \frac{e^{\lambda}}{\lambda(e^{\lambda}-1)}\right)}.$$

We can then calculate the asymptotically variance of the population estimate $\hat{N}$ by using the delta-method [Lauritzen, 2021, p. 166]. For the simple Poisson model we have that $\hat{N} = \frac{n}{1-e^{-\hat{\lambda}}}$. Therefore, we have that

$$\frac{\partial \hat{N}}{\partial \hat{\lambda}} = \frac{-n}{(1-e^{-\hat{\lambda}})^2} e^{-\hat{\lambda}}.$$

By the delta-method we then have that

$$var(\hat{N}) = \left(\frac{-n}{(1-e^{-\hat{\lambda}})^2} e^{-\hat{\lambda}}\right)^2 var(\hat{\lambda})$$

Now we can construct a classic 95% confidence interval by $C_{0.95}(\hat{N}) = \hat{N} \pm z_{0.025}\sigma$ where σ is the standard error of $\hat{N}$ and $z_{0.025}$ is the 0.025 fractile of the standard normal distribution. This procedure has been used for all five surveys to construct a confidence interval of the estimate provided by the simple Poisson model.

For the Poisson-Gamma model we have not calculated the theoretical Fischer information. Instead, we have used the observed Fisher information. The observed Fisher information is provided through the Hessian matrix when we use the optimization function in [R Core Team, 2021]. The variance matrix Σ of the parameters in the model $\begin{pmatrix}\hat{p}\\ \hat{r}\end{pmatrix}$ is given as the inverse Hessian matrix. For the Poisson-Gamma model we have that $\hat{N} = \frac{n}{1-(1-\hat{p})^{\hat{r}}}$. Therefore, the Jacobian matrix of partial derivatives is given by

$$\begin{pmatrix}\frac{\partial \hat{N}}{\partial p} & \frac{\partial \hat{N}}{\partial r}\end{pmatrix} = \begin{pmatrix}\frac{-nr(1-p)^{(r-1)}}{(1-(1-p)^r)^2} & \frac{n(1-p)^r \log(1-p)}{(1-(1-p)^r)^2}\end{pmatrix}.$$

By the delta-method [Lauritzen, 2021, p. 166] the variance of $\hat{N}$ is then given by

$$var(\hat{N}) = \begin{pmatrix}\frac{\partial \hat{N}}{\partial p} & \frac{\partial \hat{N}}{\partial r}\end{pmatrix} \Sigma \begin{pmatrix}\frac{\partial \hat{N}}{\partial p}\\ \frac{\partial \hat{N}}{\partial r}\end{pmatrix}.$$

When this variance has been computed then we can calculate a standard 95% confidence interval as described before. This procedures has been used for all five surveys to construct a Wald confidence intervals for the Poisson-Gamma model.

2.2.4 Numerical methods for fitting

Since the probability mass functions are quite complex it has been necessary to use numerical methods to compute the MLE estimates. For the Poisson-Gamma distribution and Poisson-Lognormal distribution optimization functions in [R Core Team, 2021] have been used. These functions uses the Nelder-Mead method.

For the discrete mixture models we have used an EM algorithm to handle the optimization. The general setup for the EM algorithm is based on a missing data setup. We have some observed data and some missing data coming from the same distribution. So if we denote the observed data as Y and the distribution it is assumed to follow as g_θ then we have $Y \sim g_\theta$. In our cases the missing data is the non-observed individuals with no captures. We assume that the density is on the form $g_\theta(y_i) = \sum_{j=1}^{m} \pi_j \phi_j(y_i)$ where the π_j's are positive and sum to 1 and the ϕ_j's are some densities. Hence, g_θ is a discrete mixture of some densities. In our case we assume that data follows a discrete mixture of Poisson distributions. Therefore, we have that $\phi_j(y_i)$ is the density of the Poisson distribution with parameter λ_j. The density of only the observed data is different since we have some missing data. It will be on the form

$$h_\theta(y_i, z_i) = \sum_{j=1}^{m} \mathbb{1}_{z_{ij}} \pi_j \phi_j(y_i)$$

where z_{ij} indicates that individual i comes from component j since the observed data must come from one of the m components. In our case the density above reduces to a discrete mixture of zero-truncated Poisson distributions.

With an EM algorithm, instead of maximizing the log-likelihood function directly we maximize the following operator iteratively

$$Q(\theta|\theta^{(t)}) = E\left[\log h_\theta(Y, Z)|y, \theta^{(t)}\right],$$

where h_θ is the density of the complete data meaning only the observed data. The $\theta^{(t)}$ is the parameters at time t. The density has the form $h_\theta(y_i, z_i) = \sum_{j=1}^{m} \mathbb{1}_{z_{ij}} \pi_j \phi_j(y_i)$ as explained above. The π_j determines how much the j'th density is weighted. Hence, the π_j will sum to 1 [Tatiana et al., 2009, p. 3-5] .

The algorithm consist of an expectation (E) step and a maximization (M) step. In the E-step we calculate the operator mentioned above. The parameter θ include both the π's and the densities ϕ. In our case the densities will always be from a zero-truncated Poisson distribution but the parameter λ may change. In the M-step we maximize the operator with respect to λ and set the value of λ in optimum to the new $\lambda^{(t+1)}$. In practice, this is done by defining

$$\tau_j(y_i, \lambda) = \frac{\pi^{(t)} \phi_j(y_i, \lambda_j^{(t)})}{\sum_{h=1}^{m} \pi_h \phi_h(y_i, \lambda_h^{(t)})}$$

Then we have that

$$Q(\lambda|\lambda^{(t)}) = \sum_{j=1}^{m} \sum_{i=1}^{n} \tau_j(y_i, \lambda^{(t)}) \left(\log \pi_i + \log \phi_j(y_i, \lambda_j)\right),$$

where we have n data points and m different weighted densities. We can maximize the outer sum by maximizing each element in the sum with respect to λ_j. Since $\lambda^{(t)}$ is given, $\tau_j(y_i, \lambda^{(t)})$ will be a constant. We also have that π_j is a constant. Hence, the problem reduces to maximizing $\sum_{i=1}^{n} \tau_j(y_i, \lambda^{(t)}) \log \phi_j(y_i, \lambda_j)$ for each λ_j and τ_j. Since $\phi_j(y_i, \lambda_j)$ is the density of the zero-truncated Poisson distribution we get

$$\sum_{i=1}^{n} \tau_j(y_i, \lambda^{(t)}) \log \left(\frac{\lambda_j^{y_i}}{(e^{\lambda_j} - 1) y_i!} \right) =$$
$$\sum_{j=1}^{n} \tau_j(y_i, \lambda^{(t)}) \left(y_i \log(\lambda_j) - \log(e^{\lambda_j} - 1) - \log(y_i!) \right)$$

The first order condition with respect to λ_i is

$$\sum_{i=1}^{n} \tau_j(y_i, \lambda^{(t)}) \left(\frac{y_i}{\lambda_j} - \frac{e^{\lambda_j}}{e^{\lambda_j} - 1} \right) = 0$$

Solving this equation with respect to λ_j gives the λ_j that maximize $Q(\lambda|\lambda^{(t)})$. This is the new $\lambda^{(t+1)}$.
Furthermore, the algorithm consist of letting $\pi^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_j(y_i, \lambda^{(t)})}{n}$.

After this we go back to the E-step and compute the new $\tau_j(y_i, \lambda^{(t)})$. This procedure is repeated until the loglikelihood function $l_y(\lambda^{(t)}, \pi^{(t)})$ is very close to $l_y(\lambda^{(t+1)}, \pi^{(t+1)})$ [Tatiana et al., 2009, p. 3-5].

In our application of the algorithm we have set $l_y(\lambda^{(t+1)}, \pi^{(t+1)}) - l_y(\lambda^{(t)}, \pi^{(t)}) < 0,00001$ for the algorithm to stop. More details can be seen in the R-document.

3 Results

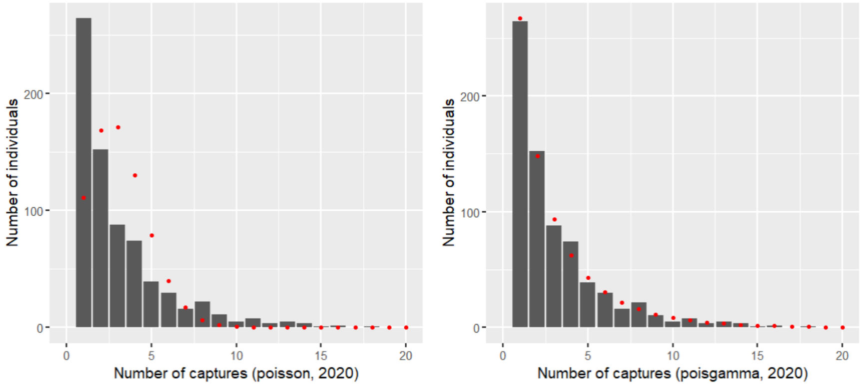
3.1 Overview

As earlier mentioned, fitting of the models have been done for each survey for the 2015, 2016, 2017, 2019 and 2020 monitoring program. An overall summary off the five surveys can be seen in Table 1.

Table 1: Summary of surveys. Samples is total number of samples collected in the survey. Unique is total number of unique bears captured in the survey.

Survey year	Regions	Samples	Unique
2015	Jämtland and Västernorrlands län	2259	581
2016	Norrbottens län	542	205
2017	Dalarnas and Gävleborgs län	1255	375
2019	Västerbotten	419	192
2020	Jämtland and Västernorrlands län	2326	728

An illustrative view off the fitting can be seen in Figure 1. Here we have plotted the densities of the fitted zero-truncated models on top of a histogram showing have many bears have been captured a specific number of times. Figure 1 shows all five statistical models fitted to the 2020 survey. The histograms have been cut off so they do not include bears that have been captured more than twenty times. For the 2020 survey this excludes one bear that have been captured 37 times. We observe that the density of the the simple Poisson model does not seem to follow the histogram very well. All the four other models seem to provide a much better fit just by looking at the illustrative fit in Figure 1. If all four models can fit the data of the observed captures well, the interesting thing is then to see how the models estimate the number of individual bears that have not been captured at all. Furthermore, it is interesting whether the four models can be distinguished at all. The overall picture of these illustrative fittings are the same for the other surveys which can be seen in Appendix. The density of the fitted simple Poisson model clearly fails to follow the histogram. However, the four other models seems to follow the histograms quite well.



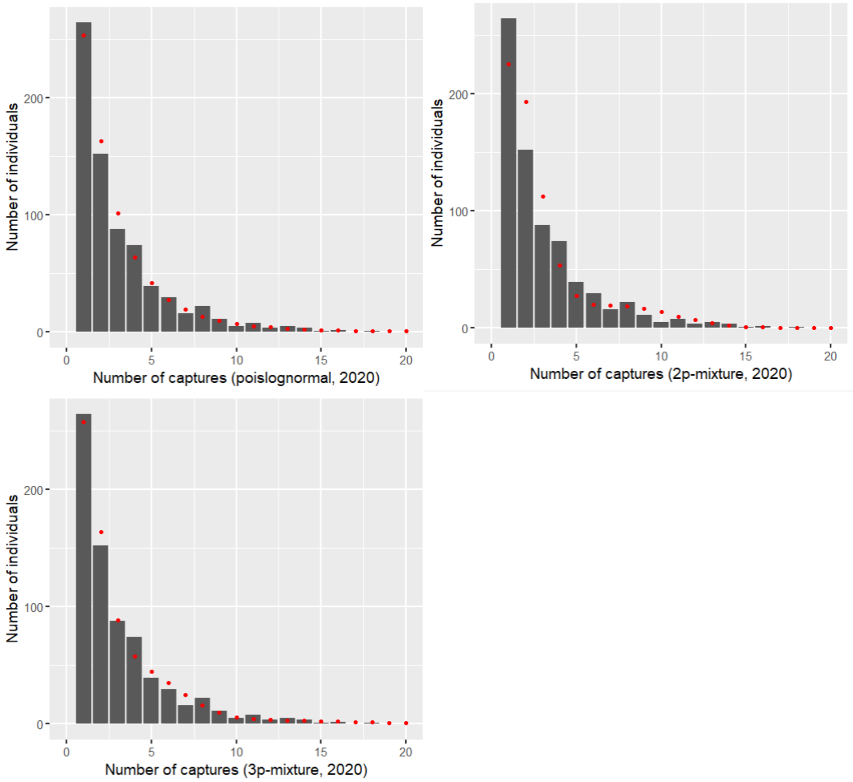


Figure 1: Illustrative plots of the fitting to the 2020 survey

3.2 Fitted models and estimates

For each model we have fitted, we have computed a range of goodness-of-fit statistics. Furthermore, we have calculated the final estimates and the confidence intervals. The results of the analyses are summarized in Table 2, Table 3, Table 4, Table 5 and Table 6. The tables contain all the values of the maximum likelihood parameters, statistics of the fits and the estimates.

Model A is a simple Poisson model and seems to be a poor fit for all of the surveys indicating that individual heterogeneity do indeed occur. That the model is a poor fit can be seen by the very low p-values and the large ΔAIC's. However, the remaining four models all provide good fits for some of the surveys. Model B and C are the Poisson-Gamma model and the Poisson-Lognormal model respectively. They seem to provide a good fit for all of the surveys and the difference in AIC for these two models is lower than 10 for all surveys. The p-values are quite high and the Δ AIC's are low for these two models indicating that they are fitting data well.

Table 2: Results for 2020 survey.

	$\hat{N}$	Bootstrap CI	Wald CI	Δ AIC	χ^2	p-value
A	764	(753, 773)	(759, 770)	691,51	$4.97 \cdot 10^23$	< 0.001
B	1549	(739, 1877)	(948, 2151)	0	202.97	0.02249
C	1040	(951, 1109)	N/A	2.63	40.81	0.5722
D	855	(810, 886)	N/A	97.15	$7.39 \cdot 10^10$	< 0.001
E	916	(833, 982)	N/A	35.42	$4.2 \cdot 10^5$	< 0.001

A: Simple Poisson model with $\lambda = 3.04$

B: Poisson-Gamma model with $r = 0.415$ and $p = 0.784$

C: Poisson-Lognormal model with $\mu = 0.323$ and $\sigma = 0.992$

D: 2-point discrete mixture placing masses $\{0.816, 0.184\}$ on $\{1.71, 8.13\}$

E: 3-point discrete mixture placing masses $\{0.642, 0.319, 0.0382\}$ on $\{1.15, 4.95, 13.6\}$

$\hat{N}$ is the maximum likelihood estimator of the population N. Bootstrap CI is a 95% confidence interval based on bootstrap procedure. Wald CI is a theoretical confidence interval which has been calculated for model A and B. ΔAIC is the difference between the AIC of the model and that of the model with the best AIC. χ^2 and p-value summarize a Pearson's chi-squared goodness-of-fit test.

Model D and E are the 2-point discrete mixture model and 3-point discrete mixture model respectively. The 2-point discrete mixture model provide a poor and unacceptable fit for the 2020, 2017 and 2015 survey as seen in Table 2, Table 4 and Table 6. However, the fit for the 2019 and 2016 survey seems to be good. Model E provide an unacceptable fit for the 2020 survey and for the 2015 survey the fit seems rather poor with a quite low p-value. But for the three remaining surveys the model provide a good fit with high p-values and low Δ AIC's.

Table 3: Results for 2019 survey.

	$\hat{N}$	Bootstrap CI	Wald CI	Δ AIC	χ^2	p-value
A	229	(212, 240)	(220, 238)	38.52	$9.16 \cdot 10^3$	< 0.001
B	387	(-154, 497)	(-200, 974)	0	7.42	0.7666
C	306	(304, 352)	N/A	0.31	5.88	0.9115
D	262	(206, 286)	N/A	6.73	20.1	0.1164
E	279	(226, 309)	N/A	7.77	5.09	0.933

A: Simple Poisson model with $\lambda = 1.83$

B: Poisson-Gamma model with $r = 0.802$ and $p = 0.574$

C: Poisson-Lognormal model with $\mu = 0.0086$ and $\sigma = 0.783$

D: 2-point discrete mixture placing masses $\{0.821, 0.179\}$ on $\{1.13, 4.46\}$

E: 3-point discrete mixture placing masses $\{0.504, 0.466, 0.0305\}$ on $\{0.609, 2.54, 7.38\}$

Column names meaning the same as for Table 2.

Taken all five surveys into account there seems to be no clear pattern in which statistical model that is the optimal one for the estimation. Only the simple Poisson model can be rejected as an acceptable model since it is not suitable for any of the surveys. In general, we have four models out of the five statistical models we have analysed that provides what we would describe as a good fit for some of the surveys. However, we have that only the two models Poisson-Gamma and Poisson-Lognormal provides a good fit for all five surveys analysed in this project.

In addition to this, it is worth noticing that some of the surveys have some extreme observations where samples from some bears have been captured a lot of times. This might make the fitting of the discrete models difficult since they convergence to zero faster than the Poisson-Gamma model and Poisson-Lognormal model. If the discrete models do not place any mass near the extreme observations then will the extreme observations make the fit a lot worse. For the 2020 survey, samples from one bear have been collected 37 times. For the 2015 survey, samples from one bear have been collected 33 times. That is a lot of samples from the same bear and these two observations are by far the most extreme observations in the datasets. We notice that these two surveys are exactly the ones where the 3-point discrete model does not provide a good fit to data. Furthermore, the 2-point discrete model do not fit data well for these two surveys. And these surveys are two out of the three surveys where the 2-point discrete model is not suitable to describe data. This indicates that if such extreme observations were removed from the dataset or if we looked at data without such extreme observations then the two discrete mixture models D and E might have an even better chance of fitting data well. Hence, these problems related to individual heterogeneity would be even more clear since more models would fit data well.

Table 4: Results for 2017 survey.

	$\hat{N}$	Bootstrap CI	Wald CI	Δ AIC	χ^2	p-value
A	391	(384, 395)	(388, 394)	322.2	$5.07 \cdot 10^8$	< 0.001
B	696	(299, 845)	(308, 1085)	0	27.39	0.4553
C	511	(452, 553)	N/A	8.35	32.83	0.6152
D	437	(382, 458)	N/A	42.6	$1.47 \cdot 10^3$	< 0.001
E	524	(472, 574)	N/A	2.19	65.33	0.03698

A: Simple Poisson model with $\lambda = 3.21$

B: Poisson-Gamma model with $r = 0.513$ and $p = 0.778$

C: Poisson-Lognormal model with $\mu = 0.454$ and $\sigma = 0.962$

D: 2-point discrete mixture placing masses $\{0.762, 0.238\}$ on $\{1.68, 7.45\}$

E: 3-point discrete mixture placing masses $\{0.331, 0.571, 0.0984\}$ on $\{0.226, 3.34, 10.2\}$

Column names meaning the same as for Table 2.

When we look at the final estimation provided by the models with a good fit then we observe that the final estimation varies a lot. For the 2020 survey, model B gives an estimate of 1549 individual female brown bears in the regions surveyed while model C gives an estimate of 1040 bears as seen in Table 2. This is a huge difference. Especially, when considering that a great deal of the estimation comes from observed individuals. In the 2020 survey there were 728 observed individual female brown bears. So the Poisson-Gamma model estimate more that double as many non-observed individuals as the Poisson-Lognormal model. The estimates of the non-observed bears can be seen in Figure 2. So for the 2020 survey there is a huge difference in the estimates of the two different models B and C but it is not possible to say which model provide the best fit since both models fit data quite well.

The same thing holds for the other surveys. For the 2019 and 2016 survey we have four suitable models to describe data. In Table 3 we have the results from the analysis of the 2019 survey. We observe that the estimate of model B is 387. This is much higher than for model C, D and E which also provide a good fit. Actually, the estimate of 387 is not even in a 95% confidence interval of the estimate for any of the three other models. So the estimates are indeed very different. But again we cannot distinguish the four different models and tell for sure which model is most likely to provide a correct estimate.

Table 5: Results for the 2016 survey.

	$\hat{N}$	Bootstrap CI	Wald CI	Δ AIC	χ^2	p-value
A	225	(217, 231)	(220, 231)	36.67	$2.56 \cdot 10^3$	0.001499
B	281	(207, 316)	(-205, 768)	0.0935	17.81	0.2569
C	262	(231, 282)	N/A	0	15.05	0.4403
D	242	(200, 255)	N/A	7.12	25.23	0.07246
E	257	(220, 278)	N/A	6.77	11.68	0.6002

A: Simple Poisson model with $\lambda = 2.41$

B: Poisson-Gamma model with $r = 1.78$ and $p = 0.52$

C: Poisson-Lognormal model with $\mu = 0.532$ and $\sigma = 0.626$

D: 2-point discrete mixture placing masses {0.821, 0.179} on {1.69, 5.24}

E: 3-point discrete mixture placing masses {0.38, 0.594, 0.0258} on {0.802, 2.96, 9.04}

Column names meaning the same as for Table 2.

Taken all five surveys into account we observe some clear patterns and we notice these kinds of problem with individual heterogeneity in all of them. The Poisson-Gamma model provide a much higher estimate that the Poisson-Lognormal model. The two discrete mixture models then have an even lower estimate. In Figure 2 this is shown clearly by using the results of the analysis. Even though, the discrete mixture models D and E are not a good fit for all surveys we still have problems related to individual heterogeneity for all five surveys. We cannot distinguish the Poisson-Gamma model and the Poisson-Lognormal model but they have a very different estimate in all surveys. And as explained earlier the poor fits of the discrete models may be due to extreme observations. So we do potentially have four models with very different estimates and we cannot distinguish the models.

Table 6: Results for the 2015 survey.

	$\hat{N}$	Bootstrap CI	Wald CI	Δ AIC	χ^2	p-value
A	594	(589, 598)	(592, 597)	613.8	$4.76 \cdot 10^16$	< 0.001
B	864	(721, 955)	(512, 1215)	0	57.5	0.1454
C	722	(681, 754)	N/A	5.56	33.64	0.7311
D	643	(623, 656)	N/A	44.7	$1.36 \cdot 10^7$	< 0.001
E	653	(585, 669)	N/A	18.2	618.9	0.003498

A: Simple Poisson model with $\lambda = 3.8$
B: Poisson-Gamma model with $r = 0.738$ and $p = 0.78$
C: Poisson-Lognormal model with $\mu = 0.746$ and $\sigma = 0.903$
D: 2-point discrete mixture placing masses $\{0.776, 0.224\}$ on $\{2.09, 9.1\}$
E: 3-point discrete mixture placing masses $\{0.719, 0.259, 0.022\}$ on $\{1.87, 7.46, 16.6\}$
Column names meaning the same as for Table 2.

For all five surveys there are models that cannot be distinguished. These models that cannot be distinguished provide a different estimate of the size of the female brown bear population in all five surveys. For a few models some of the estimates are quite close as for example in the 2016 survey as seen in Table 6. Here is the estimate obtained by using the Poisson-Lognormal model 262 while the estimate obtained by using the 3-point discrete mixture model 257. So these two estimates are close to each other. However, the Poisson-Gamma model has in the 2016 survey a quite higher estimate of 281 and the 2-point discrete mixture model has a lower estimate of 242 and we cannot distinguish any of these four models. In general, these estimates of the size of the population vary quite a lot. The 95%-confidence intervals of the estimates do in some cases contain the estimates of the other models. But this is not always the case. These things are illustrated more clearly in Figure 2.

3.3 Confidence intervals and non-observed individuals

If we take a closer look at the estimates of the non-observed bears and their confidence interval we get a clear picture of how much these estimates varies. In Figure 2 we have gathered the estimates of the number of non-observed female bears among with a bootstrap confidence intervals for all surveys and all models. We have also added the official estimate by the Swedish Museum of Natural History for every survey [Åsbrink et al., 2021], [Åsbrink et al., 2020],[Levin, 2018],[Levin, 2017b] and [Levin, 2017a]. We observe some clear patterns through all five surveys. The Poisson-Gamma model provide the highest estimate of non-observed bears in all surveys. This estimate by the Poisson-Gamma model is in several cases more than double as high as the second largest estimate. Furthermore, we observe that the Poisson-Gamma model by far has the largest 95%-confidence interval in all cases. The simple Poisson model has the lowest estimate in all cases and also the smallest confidence interval. However, the fit of the simple Poisson model is poor so we would never use these estimates anyway. But the simple Poisson model provide a estimate with quite low variance.

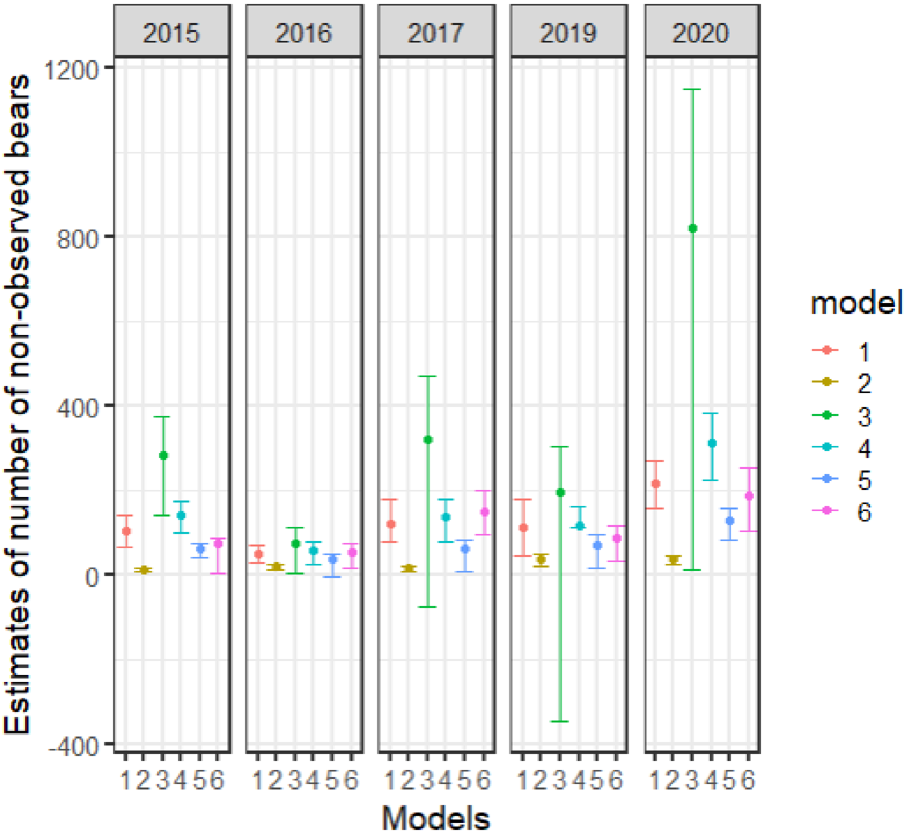


Figure 2: Bootstrap confidence intervals and estimate of non-observed individuals for all surveys and all models. 1: Official estimate. 2: The simple Poisson model. 3: The Poisson-Gamma model. 4: The Poisson-Lognormal model. 5: The 2-point discrete mixture model. 6: The 3-point discrete mixture model.

For the other models we observe that the Poisson-Lognormal model in general provide the second highest estimate. However, the official estimate by the Swedish Museum of Natural History, the Poisson-Lognormal model, the 2-point discrete mixture model and the 3-point discrete mixture model all provide estimates of the non-observed individuals that are somewhat close to each other with the 2-point mixture model providing the lowest estimate. However, the differences between these models varies from survey year to survey year and there are still differences in these estimates. In addition to this, we observe that the confidence intervals in these four situations are also of similar size in most of the surveys. The method used for the official estimate have change a bit during the surveys. However, for most of the surveys the official estimate is lower than the estimate made from the Poisson-Lognormal but greater than the estimate from the 2-point discrete mixture model and seem in general to be nearest the estimate from the 3-point discrete mixture model.

4 Discussion

The conclusions from this bachelor project is that the problems related to individual heterogeneity in wildlife estimation do occur in estimation of the Swedish brown bear population size. Several models are very difficult to distinguish but they provide very different estimates of the population sizes. This brings a lot of uncertainty to the final estimation since the estimate will depend of which statistical model is used. But it is not clear which statistical model should be used since several models all fit data well and cannot be distinguished. Overall, this means that models and methods for the capture-recapture method means a lot and should be taken into account when estimating wildlife population.

Of the four statistical models used in this thesis for describing individual heterogeneity, two of them are suitable to describe data for all five surveys. This is the Poisson-Gamma and Poisson-Lognormal model. Furthermore, the two remaining models the 2-point discrete mixture and 3-point discrete mixture are suitable to describe data in some of the surveys. That these two discrete models seem to be a worse fit in some cases may be explained due to extreme observations. The simple Poisson model do in all surveys provide a poor and unacceptable fit.

The final estimates provided by these four models of individual heterogeneity are however very different. In addition to this, the confidence intervals of the estimates are very different and of different size and length. Especially the Poisson-Gamma model provides estimates and confidence intervals that are very different from the other models. However, the Poisson-Gamma model is a good fit for all surveys analysed in this thesis.

Overall, this thesis shows and explains that it is possible to fit several models to data of the number of observed captures. However, how these models then estimate the bears with no captures seems to be very different and with no clear correlations. That is the main conclusion of this project and the contribution to future wildlife estimation studies.

Furthermore, we find that of the statistical models analysed in this project we in general get the highest estimate by using the Poisson-Gamma model. Furthermore, the Poisson-Gamma model is the model with the biggest confidence interval of the estimate. The simple Poisson model with no individual heterogeneity provides the lowest interval and the smallest confidence interval. The remaining statistical models estimates and the official estimate of the population sizes are in between these two models' estimates for all five surveys analysed.

Future research could look into whether it is in some ways possible to distinguish the different statistical models. What are the benefits and downsides for each model and how does that accord with the problems of wildlife estimation? If some models could be distinguished then the problems shown in this bachelor project would be a smaller issue and it would be more clear which methods should be used for the population estimation.

In addition to this, it could be investigated why the confidence intervals are so different and which types of datasets that makes some of models less likely to be a suitable description of the data. Would it for an example be a good idea to remove too extreme observations in order to make some models more likely to be a good fit?

There could also be done some more research about when these cases of non-identifiability occur. Would some restrictions about the data or the observation process rule out the possibility that the same problems of nonidentifiability as described in this thesis could occur? This and similar things could be relevant future research.

Appendix

A Tables and Figures

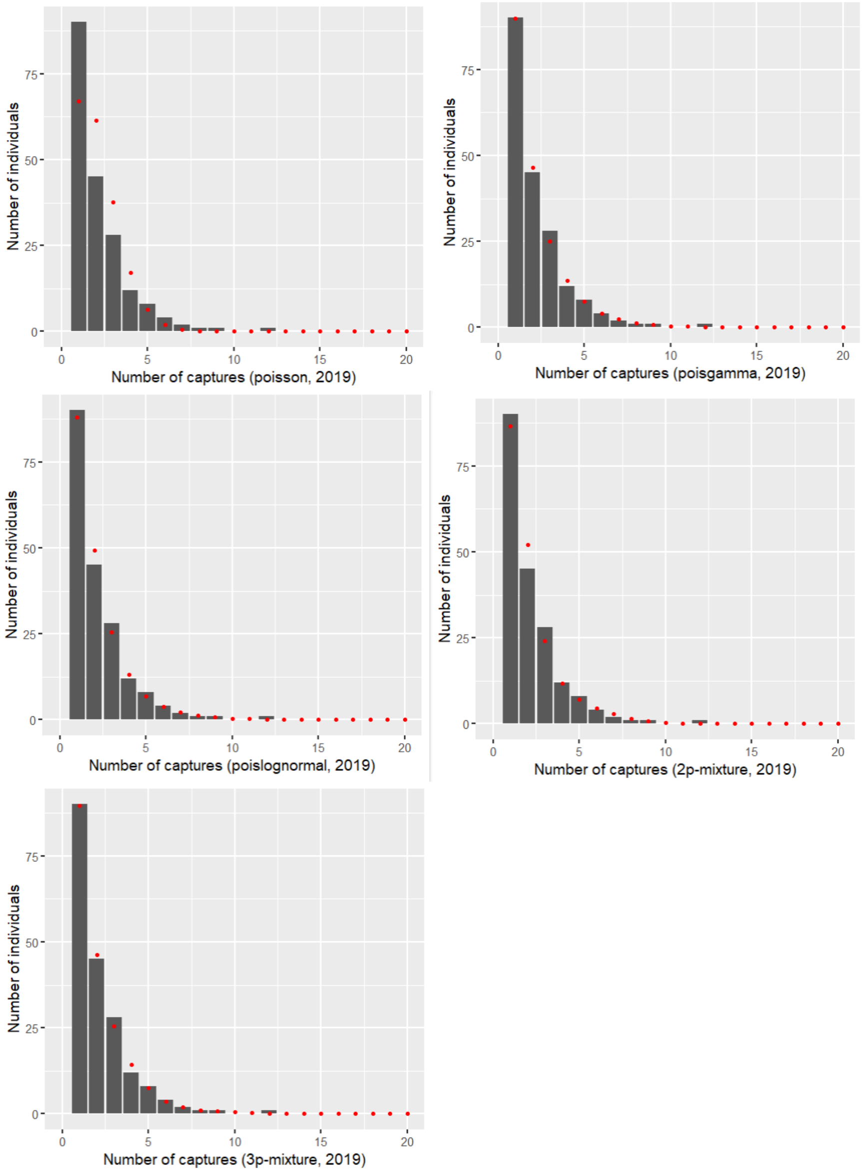


Figure 3: Illustrative plots of the fitting to the 2019 survey

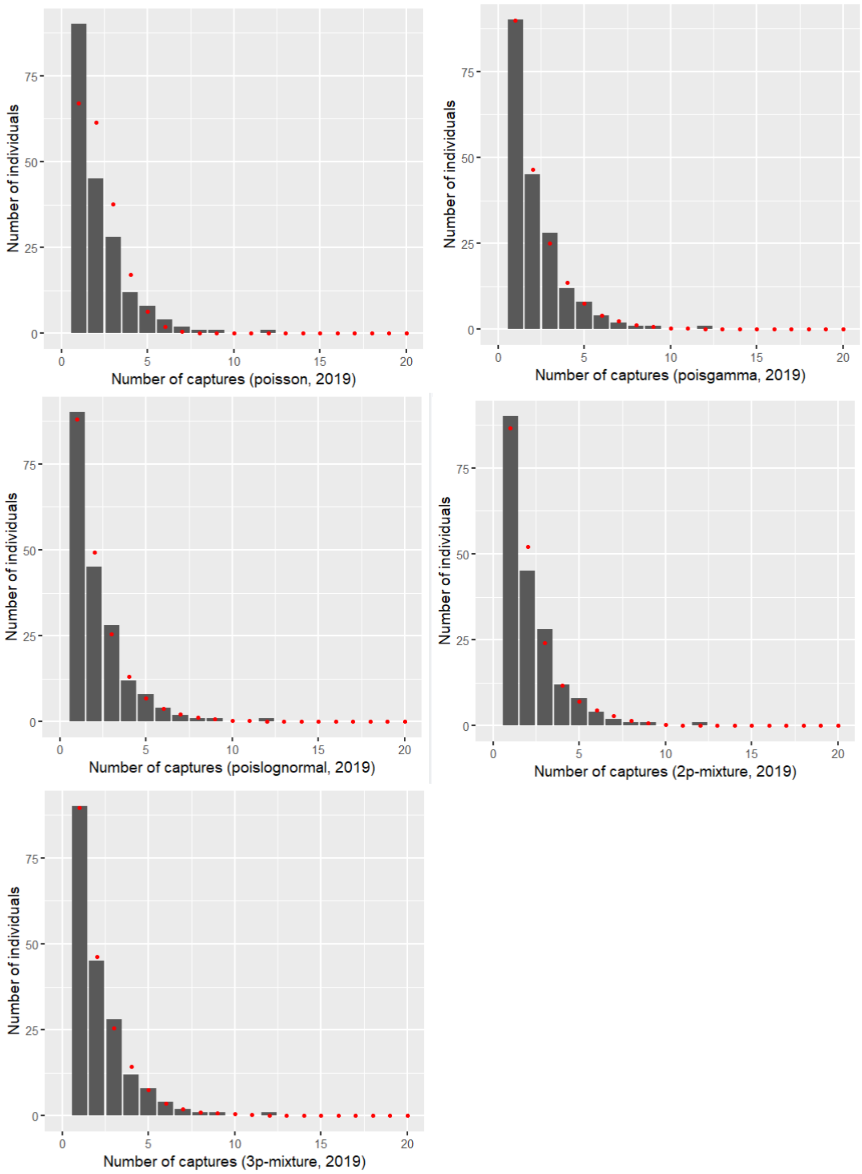


Figure 4: Illustrative plots of the fitting to the 2019 survey

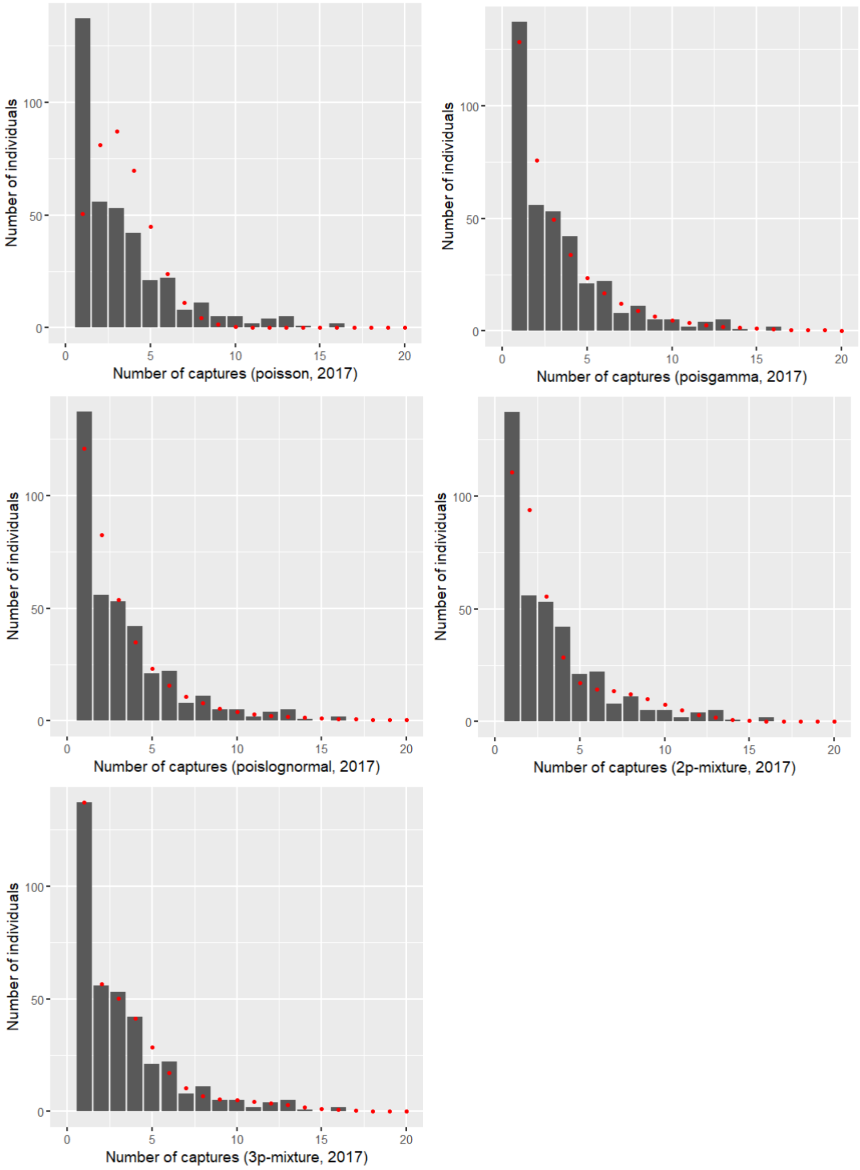


Figure 5: Illustrative plots of the fitting to the 2017 survey

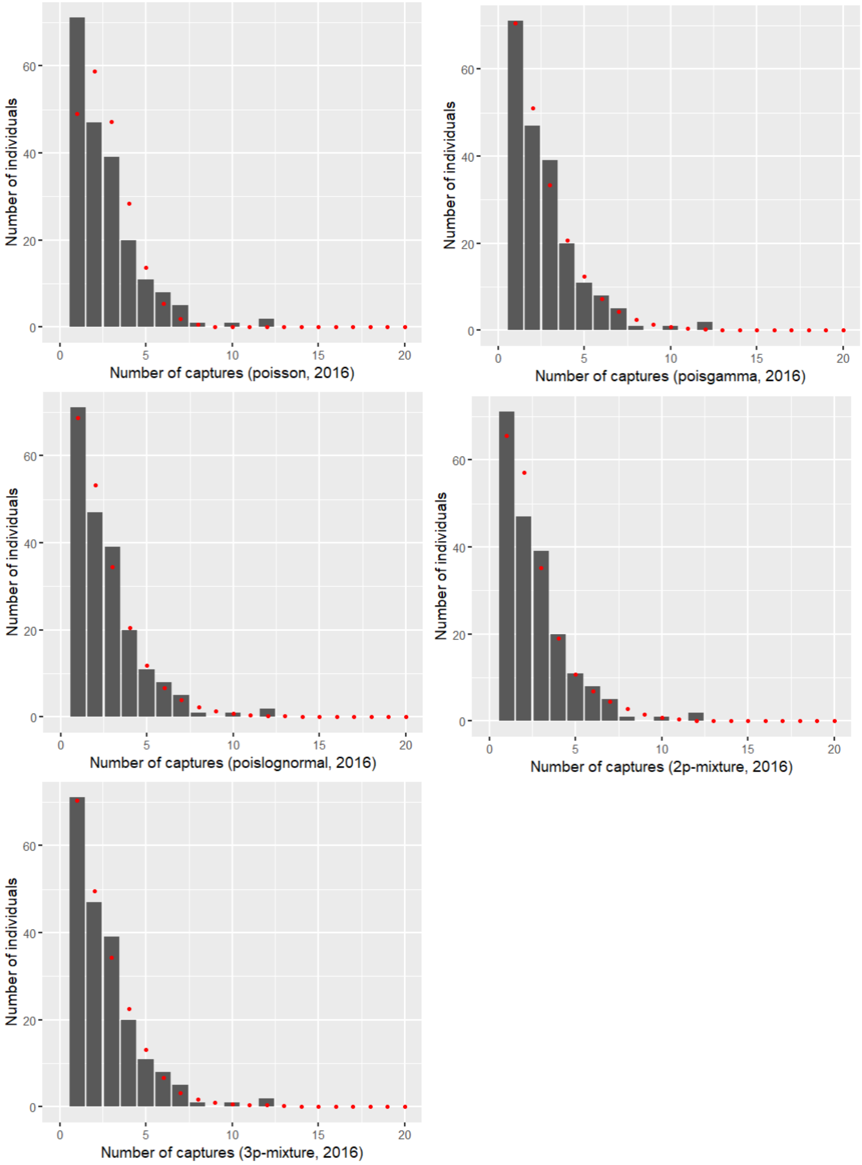


Figure 6: Illustrative plots of the fitting to the 2016 survey

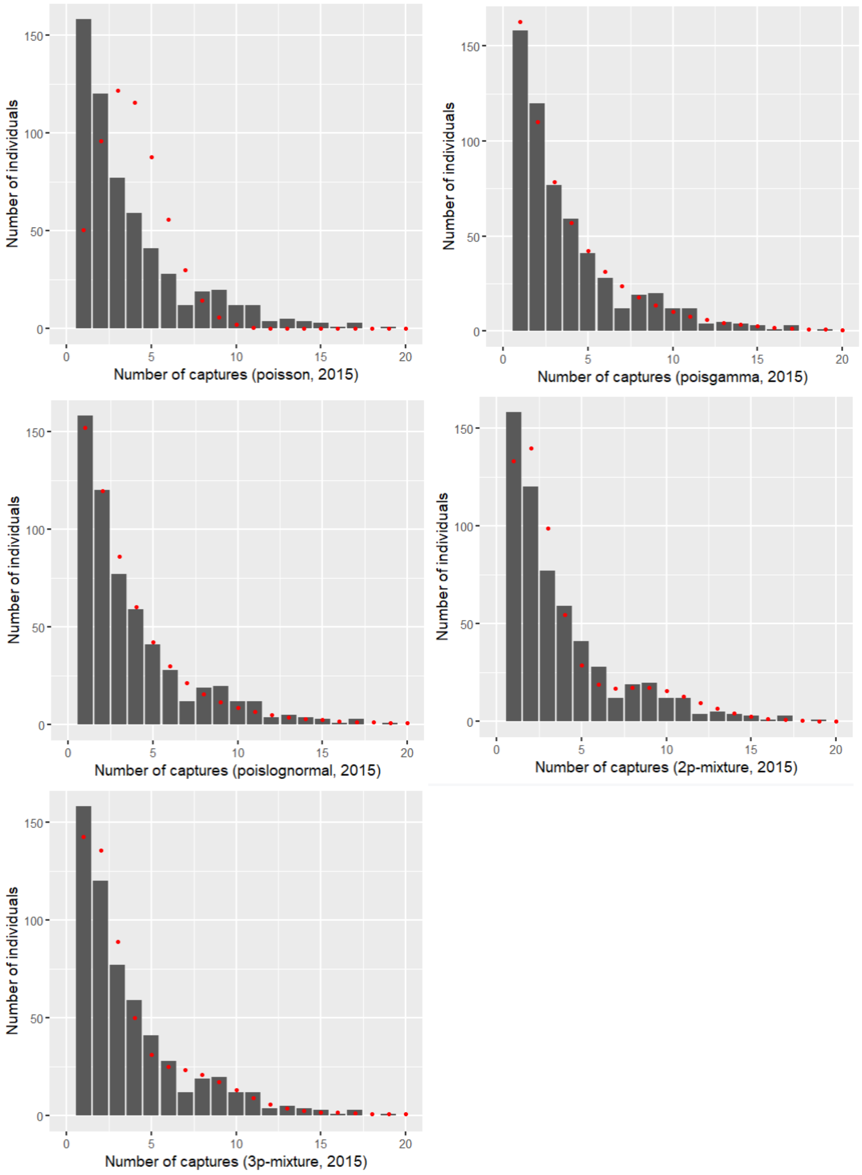


Figure 7: Illustrative plots of the fitting to the 2015 survey

B Code

All code used in this project to fit the statical models, get the estimates of the population, compute plots, etc. is available at a public GitHub repository at https://github.com/PeterHjortAnd/thesis.

References

Jessica Åsbrink, Martin Sköld, Thomas Källman, and Niclas Gyllenstrand. Resultat från inventeringen av brunbjörn i västerbottens län 2019, 2020.

Jessica Åsbrink, Martin Sköld, Källman Thomas, and Gyllenstrand Niclas. Resultat från inventering av brunbjörn i jämtlands och västernorrlands län 2020, 2021.

Brent A Coull and Alan Agresti. The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, 55(1):294–301, 1999.

Robert M Dorazio and J Andrew Royle. Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, 59(2):351–364, 2003.

LL Eberhardt. Population estimates from recapture frequencies. *The Journal of Wildlife Management*, pages 28–39, 1969.

Richard Huggins. A note on the difficulties associated with the analysis of capture–recapture experiments with heterogeneous capture probabilities. *Statistics & probability letters*, 54(2):147–152, 2001.

Konstantin Kashin. Bootstrap confidence interval methods in R, 2013. URL `https://konstantinkashin.com/2013/bootstrap-confidence-interval-methods-in-r/`.

Jonas Kindberg, Jon E Swenson, Göran Ericsson, Eva Bellemain, Christian Miquel, and Pierre Taberlet. Estimating population size and trends of the Swedish brown bear ursus arctos population. *Wildlife Biology*, 17(2):114–123, 2011.

Steffen Lauritzen. *Basic Mathematical Statistics*. Department of Mathematical Sciences, University of Copenhagen, 2021.

Maria Levin. Spillningsinventering av björn i jämtlands och västernorrlands län 2015. 2017a.

Maria Levin. Spillningsinventering av björn i norrbottens län 2016. 2017b.

Maria Levin. Spillningsinventering av björn i dalarnas, gävleborgs och värmlands län 2017. 2018.

William A Link. Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, 59(4):1123–1130, 2003.

James L Norris III and Kenneth H Pollock. Nonparametric mle under two closed capture-recapture models with heterogeneity. *Biometrics*, pages 639–649, 1996.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

Lalitha Sanathanan. Estimating the size of a truncated sample. *Journal of the American Statistical Association*, 72(359):669–672, 1977.

Benaglia Tatiana, C Didier, RH David, and Y Derek. mixtools: An r package for analyzing finite mixture models. *Journal of Statistical Software*, 32(6): 1–29, 2009.