

ROBERT KOCH INSTITUT



Originally published as:

**Kuhnert, R., Böhning, D.**

**CAMCR: Computer-assisted mixture model analysis for capture-recapture count data (2009) AStA Advances in Statistical Analysis, 93 (1), pp. 61-71.**

**DOI: 10.1007/s10182-008-0092-z**

The definitive version is available at: <http://www.springerlink.com>

## **CAMCR: Computer-Assisted Mixture model analysis for Capture-Recapture count data**

**Ronny Kuhnert · Dankmar Böhning**

Received: 29 July 2008 / Accepted: 16 September 2008

**Abstract** Population size estimation with discrete or nonparametric mixture models is considered and reliable ways of construction of the nonparametric mixture model estimator are reviewed and set into perspective. Construction of the maximum likelihood estimator of the mixing distribution is done for any number of components up to the global nonparametric maximum likelihood bound using the EM algorithm. In addition, the estimators of Chao and Zelterman are considered with some generalizations of Zelterman's estimator. All computations are done with **CAMCR**, a special software developed for population size estimation with mixture models. Several examples and data sets are discussed and the estimators illustrated. Problems using the mixture model based estimators are highlighted.

**Keywords** CAMCR · capture-recapture · Chao's and Zelterman's estimator of population size · mixture of truncated Poisson distributions

### **1 Introduction**

The estimation of the size of a specific population has become an important role in the last decade. The population of interest could be a wildlife population in the biological sciences or an illicit drug user population in the social sciences. In each population an identifying mechanism is required. A police data base could be used as identifying mechanism to estimate the number of car drivers without a licence. We restrict ourselves in this contribution to a setting in which the identifying mechanism is based upon counting repeated identifications of the same unit within a given time span. This is usually referred as capture-recapture data in the form of frequencies of frequencies.

---

Ronny Kuhnert  
Robert Koch-Institute, Division for Health of Children and Adolescents, Prevention Concepts,  
Seestr. 10, 13353 Berlin, Germany  
E-mail: kuhnert@rki.de

Dankmar Böhning  
Quantitative Biology and Applied Statistics, School of Biological Sciences, Harry Pitt Building,  
Whiteknights, Reading, RG6 6FN, UK  
E-mail: d.a.w.bohning@reading.ac.uk

**Table 1** Data from the Bangkok heroin users capture–recapture study Böhning et al. (2004)

$i$	1	2	3	4	5	6	7	8	9	10
$f_i$	2955	1186	803	611	416	338	278	180	125	74
$i$	11	12	13	14	15	16	17	18	19	
$f_i$	38	20	14	11	4	1	3	4	1	

The objective of this paper is to introduce the theory and background of the software CAMCR for Windows which we have developed for population size estimation based on capture-recapture data and is freely available (<http://www.personal.reading.ac.uk/~sns05dab/Software.html>). Before we come in more detail to CAMCR, we provide a brief introduction of unobserved population size estimation using mixture models as it is designed in CAMCR. For a more general introduction into the capture-recapture methodology see Bunge and Fitzpatrick (1993). We will begin with an example to illustrate the capture–recapture approach in form of frequencies of frequencies.

### 1.1 A capture–recapture study on illicit drug users in Bangkok

In this study the identifying mechanism are hospital registers. Included are all public and private hospitals in the Bangkok metropolitan area which are delivering treatment to drug users. More details of the study are provided in Böhning et al. (2004). Table 1 shows the frequencies  $f_1, \dots, f_{19}$  of the treatment episodes. For example, the table means that  $f_1 = 2955$  of the  $n = 7048$  heroin users had one treatment episode up to one case with 19 treatment episodes ( $f_{19} = 1$ ). The objective is to estimate the number of drug users without observed treatment episode.

## 2 Estimation of population size

We will denote by  $f_0, f_1, \dots, f_j, \dots$  the frequency of those units identified exactly 0, 1, ...,  $j$  times and let  $m$  denote the largest occurring count. Furthermore, we will denote with  $p_0, p_1, \dots, p_j, \dots$  the probability of exactly 0, 1, ...,  $j, \dots$  identifications. We have that  $n = f_1 + f_2 + \dots + f_m$  and  $N = n + f_0$ .

A general population size estimator for  $N$  is available by means of the well-known Horvitz-Thompson estimator (Horvitz and Thompson, 1952)

$$\hat{N} = \frac{n}{(1 - p_0)}. \quad (1)$$

However,  $p_0$  is usually unknown and an estimate will be required for practical use. Several modelling strategies have been developed. In the next section we will introduce in the modelling of the count distribution by means of mixture of Poissons as well as the Zelterman estimator and the Chao estimator.

### 2.1 Modelling the count distribution as mixture of Poisson distributions

A simple count distribution is the Poisson distribution given as  $Po(j, \lambda) = e^{-\lambda} \frac{\lambda^j}{j!}$ . The Poisson density does often not provide enough flexibility to give an adequate fit. More

flexible are discrete mixture models (Norris and Pollock, 1996, 1998; Mao and Lindsay, 2002, 2003) of the form

$$f(j, Q_k) = \sum_{\ell=1}^k Po(j, \lambda_\ell) q_\ell, \quad (2)$$

where the mixing distribution  $Q_k$  is giving weight  $q_\ell \geq 0$  to parameters  $\lambda_\ell$  for  $\ell = 1, \dots, k$ , and  $k$  is the number of components in the mixture. Note that  $q_1 + \dots + q_k = 1$ . The likelihood analysis focuses on the zero-truncated mixture log-likelihood

$$\log L(Q_k) = \sum_{j=1}^m f_j \log[f(j, Q_k)] - n \log[1 - f(0, Q_k)]. \quad (3)$$

Equivalently, a log-likelihood based upon mixtures of zero-truncated Poissons could be considered as in Böhning and Kuhnert (2006). In this situation the log-likelihood can be maximised in the set of all discrete probability distributions, leading to the nonparametric maximum likelihood estimate (NPMLE). A variety of numerical algorithms exist to find the global maximum likelihood estimator, if it exists. The EM algorithm with gradient function update is included in **CAMCR**. The EM algorithm (Dempster et al., 1977) has become very popular in connection with mixture models, see McLachlan and Krishnan (1997); McLachlan and Peel (2000). This algorithm has the additional advantage of providing a maximum likelihood solution conditional upon the number of mixture components  $k$  although there is no guarantee for a global solution. To proceed in the EM context we need the *complete data log-likelihood* which is given in this case as

$$\sum_{j=1}^m f_j \sum_{\ell=1}^k z_{j\ell} \log f_+(j, \lambda_\ell) + \sum_{j=1}^m f_j \sum_{\ell=1}^k z_{j\ell} \log q_\ell \quad (4)$$

with

$$\begin{aligned} f_+(j, \lambda_\ell) &= \frac{Po(j, \lambda_\ell)}{1 - \exp(0; \lambda_\ell)} \\ f_+(j, Q) &= \sum_{\ell} q_\ell f_+(j, \lambda_\ell) \end{aligned} \quad (5)$$

and where the unobserved covariate  $z_{j\ell}$  is 1 if  $j$  belongs to component  $\ell$  and 0 otherwise. The EM algorithm replaces in the *E-step* the unobserved indicator variates  $z_{j\ell}$  by their expected values conditional upon the observed data and current values of  $\lambda_\ell, q_\ell, \ell = 1, \dots, k$  leading to

$$e_{j\ell} = E(z_{j\ell} | f_j; q_\ell, \lambda_\ell, \ell = 1, \dots, k) = \frac{f_+(j, \lambda_\ell) q_\ell}{\sum_{i=1}^k f_+(j, \lambda_i) q_i}. \quad (6)$$

In the *M-step* new values  $\hat{\lambda}_1, \dots, \hat{\lambda}_k, \hat{q}_1, \dots, \hat{q}_k$  are found by maximising the expected version of (4) leading to

$$\hat{q}_\ell = \frac{1}{n} \sum_{j=1}^m f_j e_{j\ell}, \quad \text{for } \ell = 1, \dots, k \quad (7)$$

as new estimates for the weights. The new estimates  $\hat{\lambda}_1, \dots, \hat{\lambda}_k$  need to be found as solutions of

$$\hat{\lambda}_\ell = \frac{\sum_{j=1}^{\nu} j n_j e_{j\ell}}{\sum_{j=1}^{\nu} n_j e_{j\ell}} (1 - e^{-\hat{\lambda}_\ell}), \quad \text{for } \ell = 1, \dots, k. \quad (8)$$

Note that (8) does not provide a closed form solution for  $\hat{\lambda}_\ell$ , but rather suggests an iterative solution of the form  $\hat{\lambda}_\ell^{new} = \frac{\sum_{j=1}^m j n_j e^{j\lambda_\ell}}{\sum_{j=1}^m n_j e^{j\lambda_\ell}} (1 - e^{-\hat{\lambda}_\ell^{old}})$  which needs to be iterated until convergence. The benefit of working with a mixture model of zero-truncated Poisson densities (4) (instead of using a zero-truncated mixture of Poisson densities) can be seen in the fact that an existing global maximisation theory can be used. This was developed by various authors including Simar (1976); Laird (1978); Böhning (1982); Lindsay (1983); Leroux (1992); Böhning (2000), among others. The log-likelihood with respect to  $f_+(j; Q)$

$$\log L_+(Q) = \sum_j f_j \log(f_+(j; Q))$$

is a *concave functional* on the set of *all* discrete probability distributions (though it is not concave on the set of all discrete probability measures with exactly  $k$  support points). This is the main reason for achieving the following global results. An important, analytical tool is the *gradient function* defined for any discrete distribution

$$Q = \begin{pmatrix} \lambda_1 & \dots & \lambda_k \\ q_1 & \dots & q_k \end{pmatrix} \text{ as}$$

$$d(\lambda, Q) = \frac{1}{n} \sum_{j=1}^m f_j \frac{f_+(j, \lambda)}{f_+(j, Q)}$$

where  $f_+(j, Q) = q_1 f_+(j, \lambda_1) + q_2 f_+(j, \lambda_2) + \dots + q_k f_+(j, \lambda_k)$ . With the help of the gradient function, the NPMLE can be characterized. The general mixture maximum likelihood theorem (Lindsay, 1983; Böhning, 1982) states that the NPMLE  $\hat{Q} = \begin{pmatrix} \lambda_1 & \dots & \lambda_k \\ q_1 & \dots & q_k \end{pmatrix}$  can be uniquely characterised by an upper bound condition for the gradient function:

$$\hat{Q} \text{ is NPMLE} \Leftrightarrow d(\lambda, \hat{Q}) \leq 1 \text{ for all } \lambda > 0. \quad (9)$$

In addition,  $d(\lambda, \hat{Q}) = 1$  for  $\lambda \in \{\hat{\lambda}_1, \dots, \hat{\lambda}_k\}$ , the set of all support points of  $\hat{Q}$ . The mixture maximum likelihood theorem for count densities like the truncated Poisson can be used to determine if a given mixture is or not is the NPMLE. This supplements simple diagnostic techniques like overdispersion tests (Böhning, 1994) which can be used to identify violations of homogeneity but are unable to give further characterizations of heterogeneity that the mixture maximum likelihood theorem provides.

If a maximum likelihood estimator  $\hat{Q}$  has been identified (this might be the non-parametric maximum likelihood estimator or the one for a mixture model with a specific number of components), then the population size is estimated as (Böhning and Kuhnert (2006))

$$\hat{N} = n \sum_{\ell=1}^k \frac{\hat{q}_\ell}{1 - f(0, \hat{\lambda}_\ell)}.$$

## 2.2 Zelterman's estimator

Zelterman (Zelterman, 1988) argued that the Poisson assumption might not be valid over the range of possible values for the count variable  $Y$  taking values in the set of integers  $\{0, 1, 2, 3, \dots\}$ . Nevertheless the Poisson assumption might be valid for small ranges of  $Y$  such as from  $j$  to  $j + 1$ , so that it would be meaningful to use only the

frequencies  $f_j$  and  $f_{j+1}$  in estimating  $\lambda$ . Zelterman derived the estimator using the fact that for any  $j$  both, the truncated and the untruncated Poisson distribution, have the property that  $\frac{Po(j+1, \lambda)}{Po(j, \lambda)} = \frac{\lambda}{j+1}$  and  $\frac{f_+(j+1, \lambda)}{f_+(j, \lambda)} = \frac{\lambda}{j+1}$ . Hence,  $\lambda$  can be written as

$$\lambda = \frac{(j+1)Po(j+1, \lambda)}{Po(j, \lambda)} = \frac{(j+1)f_+(j+1, \lambda)}{f_+(j, \lambda)}. \quad (10)$$

An estimator for  $\lambda$  is obtained by replacing  $f_+(j, \lambda)$  by the empirical frequency  $f_j$ :

$$\hat{\lambda}_j = \frac{(j+1)f_{j+1}}{f_j}. \quad (11)$$

There are two reasons to take  $\hat{\lambda}_1$ . For one,  $\hat{\lambda}_1$  is using frequencies in the vicinity of  $f_0$  which is the target of prediction. And for two, in many application studies for estimating  $f_0$ , the majority of counts fall into  $f_1$  and  $f_2$ . The counts larger than 2 do not affect the estimator, a fact largely contributing to its robustness.

### 2.2.1 Generalizing the idea of Zelterman

The idea of Zelterman can be extended to the inclusion of higher counts by rewriting  $\lambda$  as

$$\lambda = \lambda \frac{\overbrace{\sum_{i=1}^j \lambda^i}^1}{\sum_{i=1}^j \lambda^i} = \frac{\sum_{i=1}^j (i+1) \frac{\lambda^{i+1}}{(i+1)!}}{\sum_{i=1}^j \frac{\lambda^i}{i!}} = \frac{\sum_{i=1}^j (i+1) Po(i+1, \lambda)}{\sum_{i=1}^j Po(i, \lambda)}.$$

Replacing theoretical probabilities by sample frequencies we achieve the estimator

$$\hat{\lambda}_j = \frac{\sum_{i=1}^j (i+1) f_{i+1}}{\sum_{i=1}^j f_i}$$

and in particular for  $j = 1, \dots, 4$

$$\hat{\lambda}_1 = \frac{2f_2}{f_1}, \hat{\lambda}_2 = \frac{2f_2 + 3f_3}{f_1 + f_2}, \hat{\lambda}_3 = \frac{2f_2 + 3f_3 + 4f_4}{f_1 + f_2 + f_3}, \hat{\lambda}_4 = \frac{2f_2 + 3f_3 + 4f_4 + 5f_5}{f_1 + f_2 + f_3 + f_4}$$

Notice that the form of these estimators, achieved by taking sums before ratios, will provide some stability. Clearly there is a trade-off between bias and variance. If we compare  $\hat{\lambda}_j$  with  $\hat{\lambda}_{j+1}$  the latter will have the smaller variance, whereas the former the smaller bias. Limited simulation studies show that a better mean squared error is achieved by using  $\hat{\lambda}_2$  or  $\hat{\lambda}_3$ . For larger values of  $j$  in  $\hat{\lambda}_j$ , the bias becomes large. In CAMCR these four Zelterman estimators are provided.

### 2.3 Chao's estimator

Another popular population size estimator that also only uses the counts  $f_1$  and  $f_2$  has been proposed by Chao (1987, 1989). Chao suggested the estimator  $\hat{N}_C = n + \frac{f_1^2}{2f_2}$ . The estimator is based upon the Cauchy-Schwartz inequality (Wilson and Collins, 1992) for the nonparametric mixture of a Poisson, namely

$$\left( \int_0^\infty \lambda e^{-\lambda} d\lambda \right)^2 \leq \int_0^\infty e^{-\lambda} d\lambda \int_0^\infty \lambda^2 e^{-\lambda} d\lambda$$

where the inequality of Cauchy-Schwartz  $(\int uv)^2 \leq (\int u^2)(\int v^2)$  is used with  $u(\lambda) = \sqrt{e^{-\lambda}}$  and  $v(\lambda) = \lambda\sqrt{e^{-\lambda}}$  and leading to  $p_1^2 \leq p_0 \times 2p_2$ , so that  $\frac{f_1^2}{2f_2}$  estimates a lower bound for  $f_0$ . Chao suggested to use this bound as an estimator if higher frequency counts are small.

### 3 Working with CAMCR

Existing statistical packages such as STATA, S-plus, MINITAB, or others do not include mixture modelling of zero-truncated Poisson densities as an option so that it seemed appropriate to develop a separate piece of software. It was decided to use the Microsoft Fortran Power Station to develop the software tool CAMCR (Computer Assisted Mixture Model Analysis for Capture-Recapture Count Data). Very fast and efficient computation of estimation problems requiring complex algorithmic solutions is an advantage of FORTRAN (in comparison of developing a macro in one of the standard packages). Nevertheless, limited possibilities are also available to create a user friendly interface. A detailed description of CAMCR can be found on the web site:

<http://www.personal.reading.ac.uk/~sns05dab/Software.html>

CAMCR provides three model selection criteria for modeling the count distribution; a) NPMLE (Nonparametric Maximum Likelihood Estimator) b) BIC (Bayesian Information Criterion) c) specific maximum number of mixed components. The NPMLE can be verified by the gradient function criterion, see section 2.1 or Böhning (2000). If the gradient function is bounded above by 1 over the full parameter space, then the unique maximum of the log-likelihood function is obtained. The BIC criterion goes back to Schwartz (1978) and penalizes the log-likelihood with a term representing model complexity.

#### 3.1 Analysis and results for the introductory drug user population in Bangkok

Table 2 shows the result of estimating the population size for the Bangkok heroin user's data (see also Böhning et al., 2004). The top of the table shows the results for the analysis according to the nonparametric maximum likelihood estimation. The first column gives the number of mixture components  $k$ . The first row contains the results in the homogeneity case up to the fourth row with the results for the four-component mixture model. The second and third column gives the  $\lambda$  and weight values of these models. The fourth and fifth column show the values for the log-likelihood function and the BIC, respectively. For the objective of the analysis the sixth and seventh column are most interesting. They give the estimates for the hidden and for the total population size. The last column gives the value of the maximum of the gradient function. For the Bangkok heroin users data the NPMLE is given by four components with a population size of 18367. In this case the BIC criterion chooses as well four components. In the lower part of the table the Zelterman estimate and the Chao estimate are provided. Chao's estimate and the first two Zelterman-estimate as well as the two-component mixture model are close together. Limited simulation studies have provided some evidence that the higher Zelterman estimates are providing estimates that experience considerable bias and might be better avoided. In addition, often the mixture model likelihood (and the associated BIC value) are close for close component models (like in the case

**Table 2** Results from the Bangkok heroin users capture-recapture study Böhning et al. (2004)

$k$	$\hat{\lambda}_j$	$\hat{q}_j$	Log-L	BIC	$\hat{f}_0$	$\hat{N}$	$\max_{\lambda} d(\lambda, \hat{Q})$
1	2.7507	1.0	-15462.4	30927.8	482	7544	>10000
2	0.8840	0.6341	-13214.0	26436.9	3164	10226	29.1
	5.3955	0.3659					
3	0.4001	0.4345	-13134.3	26283.3	6391	13453	3.9217
	2.9459	0.3882					
	6.7876	0.1772					
4	0.1894	0.3214	-13120.4	26261.3	11305	18367	1.00
	2.0818	0.4114					
	5.8254	0.2598					
	12.157	0.0073					
Zelterman	$\hat{\lambda}_1 = \frac{2f_2}{f_1}$	$\hat{\lambda}_2 = \frac{2f_2+3f_3}{f_1+f_2}$	$\hat{\lambda}_3 = \frac{2f_2+3f_3+4f_4}{f_1+f_2+f_3}$	$\hat{\lambda}_4 = \frac{2f_2+3f_3+4f_4+5f_5}{f_1+f_2+f_3+f_4}$			
$\hat{f}_0$	5734	3250	2132	1628			
$\hat{N}$	12796	10312	9194	8690			
Chao	$f_0 = \frac{f_1 \times f_1}{2f_2}$						
$\hat{f}_0$	3681						
$\hat{N}$	10743						

of the 3 and 4 component model here). However, the estimated population sizes differ considerably and experience instability. Hence, also caution must accompany the choice of the mixture model here.

We now look at some other data sets.

### 3.2 Spinner dolphin data

Oremus (2005) tried to estimate the size of a small community of spinner dolphins which are resident around the island of Moorea (near Tahiti). In 2002, using an interval of 8 months, skin samples were randomly taken and 12 microsatellite loci were genotyped which makes miss-matching of dolphins very unlikely.  $f_1 = 42$  dolphins were sampled only once,  $f_2 = 7$  dolphins were sampled exactly twice and  $f_3 = 2$  dolphins were sampled exactly three times. This leads to  $n = 51$  different dolphins that were observed in the experiment.

Table 3 shows the results of all population estimators with CAMCR. The nonparametric likelihood estimator is provided for two components leading to a clearly *spurious* estimate of 111,729. This example shows that the results based on the nonparametric mixture maximum likelihood estimator should be considered with great caution. Evidently, the one-component model is the right choice leading to an estimate of 153 for the population size. The alternative estimators of Chao with 177 and Zelterman 180 are close and provide a consistent picture.



**Table 3** Analysis of the spinner dolphin data Oremus (2005)

$k$	$\hat{\lambda}_j$	$\hat{q}_j$	Log-L	BIC	$\hat{f}_0$	$\hat{N}$	$\max_{\lambda} d(\lambda, \hat{Q})$
1	0.4042	1.0	-29.1	59.2	102	153	1.0149
2	0.6037 0.0002	0.6495 0.3505	-28.9	61.1	111678	111729	1.00
Zelterman	$\hat{\lambda}_1 = \frac{2f_2}{f_1}$	$\hat{\lambda}_2 = \frac{2f_2+3f_3}{f_1+f_2}$	$\hat{\lambda}_3 = \frac{2f_2+3f_3+4f_4}{f_1+f_2+f_3}$	$\hat{\lambda}_4 = \frac{2f_2+3f_3+4f_4+5f_5}{f_1+f_2+f_3+f_4}$			
$\hat{f}_0$	129	101	106	106			
$\hat{N}$	180	152	157	157			
Chao	$f_0 = \frac{f_1 \times f_1}{2f_2}$						
$\hat{f}_0$	126						
$\hat{N}$	177						

**Table 4** Analysis of illegal immigrants in the Netherlands

$k$	$\hat{\lambda}_j$	$\hat{q}_j$	Log-L	BIC	$\hat{f}_0$	$\hat{N}$	$\max_{\lambda} d(\lambda, \hat{Q})$
1	0.3086	1.0	-902.0	1805.7	5200	7080	28.25
2	0.1308 1.2244	0.8570 0.1430	-872.2	1749.8	11638	13518	1.00
Zelterman	$\hat{\lambda}_1 = \frac{2f_2}{f_1}$	$\hat{\lambda}_2 = \frac{2f_2+3f_3}{f_1+f_2}$	$\hat{\lambda}_3 = \frac{2f_2+3f_3+4f_4}{f_1+f_2+f_3}$	$\hat{\lambda}_4 = \frac{2f_2+3f_3+4f_4+5f_5}{f_1+f_2+f_3+f_4}$			
$\hat{f}_0$	7545	6306	5732	5716			
$\hat{N}$	9425	8186	7612	7596			
Chao	$f_0 = \frac{f_1 \times f_1}{2f_2}$						
$\hat{f}_0$	7394						
$\hat{N}$	9274						

### 3.3 The illegal immigrant's study

In the next example the number of illegal immigrants in four cities in the Netherlands is estimated from police records. The data have been analyzed previously by means of the truncated Poisson regression model by van der Heijden et al. (2003). The analysis focus on those illegal immigrants that, once apprehended, cannot be effectively expelled by the police, for example because their home country does not cooperate in receiving them back. In this case the police requested them to leave the country, but it is unlikely that will abide by this request. Hence they can be apprehended multiple times. The observed frequencies are  $f_1 = 1645$ ,  $f_2 = 183$ ,  $f_3 = 37$ ,  $f_4 = 13$ ,  $f_5 = 1$ ,  $f_6 = 1$ .

Table 4 shows the results with **CAMCR**. By the illegal immigrants data the NPMLE and the BIC is given by two components with a population size of 13518. This estimate is considerably larger than those by Zelterman with 9425 and by Chao with 9274 indicating again that nonparametric mixture models need to be used with great caution in population size estimation.

## 4 Discussion and Conclusion

Discrete mixture models offer a wide and flexible modelling framework to cope with heterogeneity in the parameters representing capture-recapture probabilities. They are potentially the most suitable models for fitting recapture counts – as has been demonstrated by many authors (Norris and Pollock, 1996, 1998; Mao and Lindsay, 2002, 2003; Bunge and Fitzpatrick, 1993; Chao, 1987, 1989). However, the computation of the NPMLE for a discrete mixture model is not straightforward. Hence it seems helpful to have software that can reliably (or more reliably than other software products) compute the maximum likelihood estimator. In fact, **CAMCR** computes all mixture models from one component to the largest number of possible components delivered by the NPMLE. To ease choice of model the associated BIC-value is provided as well. The BIC-criterion turned out to be the better choice in comparison with the AIC-criterion when dealing with mixtures (see McLachlan and Peel, 2000). It is crucial to do model selection since – as has been demonstrated in the examples – relatively small changes in the likelihood can be accompanied by large changes in the population size estimates. If there is doubt which one of two competing models to choose, it seems wise to choose the one with fewer parameters. In any case, mixture model estimates should be seen in the context to other estimators. For this reason, **CAMCR** provides also the estimators of Chao and Zelterman which are simple to compute. Mixture model based estimators should always be contemplated together with these simple estimators since Chao’s estimator give a lower bound for the population size whereas Zelterman’s estimator frequently provides an upper bound (both statements are correct up to random error). More trust can be attached to situations where all estimators provide similar results whereas doubt might remain for situations with vary different resulting estimators.

**Acknowledgements** The authors would like to thank the Editor of the journal *AStA Advances in Statistical Analysis*, Prof. Dr. Wilfried Seidel, for the possibility to publish this work. The work has been done with support from the German Research Foundation.

## References

- Böhning, D.: Convergence of simar’s algorithm for finding the maximum likelihood estimate of a compound poisson process. *Annals of Statistics* **10**, 1006–1008 (1982)
- Böhning, D.: A note on a test for poisson overdispersion. *Biometrika* **81**, 418–419 (1994)
- Böhning, D.: *Computer-Assisted Analysis of Mixtures and Applications. Meta-Analysis, Disease Mapping and Others.* Chapman & Hall/CRC, Boca Raton (2000)
- Böhning, D., Kuhnert, R.: The equivalence of truncated count mixture distributions and mixtures of truncated count distributions. *Biometrics* **62**, 1207–1215 (2006)
- Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W., Viwatwongkasem, C.: Estimating the number of drug users in bangkok 2001: a capture-recapture approach using repeated entries in one list. *European Journal of Epidemiology* **19**, 1075–1083 (2004)
- Bunge, J., Fitzpatrick, M.: Estimating the number of species: A review. *Journal of the American Statistical Association* **88**, 364–373 (1993)
- Chao, A.: Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791 (1987)

- 
- Chao, A.: Estimating population size for sparse data in capture-recapture experiments. *Biometrics* **45**, 427–438 (1989)
- Dempster, A., Laird, N., Rubin, D.B.: Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B* **39**, 1–38 (1977)
- Horvitz, D.G., Thompson, D.J.: A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* **47**, 663–685 (1952)
- Laird, N.: Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**, 805–811 (1978)
- Leroux, B.G.: Consistent estimation of a mixing distribution. *Annals of Statistics* **20**, 1350–1360 (1992)
- Lindsay, B.G.: The geometry of mixture likelihoods, Part I: A general theory. *Annals of Statistics* **11**, 783–792 (1983)
- Mao, C.X., Lindsay, B.G.: Diagnostics for the homogeneity of inclusion probabilities in a bernoulli census. *Sankhyā: The Indian Journal of Statistics, Series A* **64**, 626–639 (2002)
- Mao, C.X., Lindsay, B.G.: Tests and diagnostics for heterogeneity in the species problem. *Computational Statistics and Data Analysis* **41**, 389–398 (2003)
- McLachlan, G., Krishnan, T.: *The EM Algorithm and Extensions*. Wiley, New York (1997)
- McLachlan, G., Peel, D.: *Finite Mixture Models*. Wiley, New York (2000)
- Norris, J.L.I., Pollock, K.H.: Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* **52**, 639–649 (1996)
- Norris, J.L.I., Pollock, K.H.: Non-parametric MLE for poisson species abundance models allowing for heterogeneity between species. *Environmental and Ecological Statistics* **5**, 391–402 (1998)
- Oremus, M.: Personal communication. (2005)
- Schwartz, G.: Estimating the dimension of a model. *Annals of Statistics* **6**, 461–164 (1978)
- Simar, L.: Maximum likelihood estimation of a compound poisson process. *Annals of Statistics* **4**, 1200–1209 (1976)
- van der Heijden, P.G.M., Bustami, R., Cruy, M., Engbersen, G., van Houwelingen, H.C.: Point and interval estimation of the population size using the truncated poisson regression model. *Statistical Modelling - An International Journal* **3**, 305–322 (2003)
- Wilson, R.M., Collins, M.F.: Capture-recapture estimation with samples of size one using frequency data. *Biometrika* **79**, 543–553 (1992)
- Zelterman, D.: Robust estimation in truncated discrete distributions with application to capture recapture experiments. *Journal of Statistical Planning and Inference* **18**, 225–237 (1988)