Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

On two occasions capture and recapture designs: Comparing and discussing some estimators

Sueli Aparecida Mingoti

Relatório Técnico
RTP-02/2002
Série Pesquisa

ON TWO OCCASIONS CAPTURE AND RECAPTURE DESIGNS: COMPARING AND DISCUSSING SOME ESTIMATORS

Sueli Aparecida Mingoti
sueli@est.ufmg.br
Professora Adjunto - Departamento de Estatística da UFMG

2002

ON TWO OCCASIONS CAPTURE AND RECAPTURE DESIGNS: COMPARING AND DISCUSSING SOME ESTIMATORS

SUMMARY

Capture and recapture designs with only two occasions are still being used in a variety of situations especially in Epidemiological studies. Usually, the well-known Lincoln-Petersen (1896,1930) estimator is used to estimate the population size. Because only two occasions are being used some concern is devoted to the reliability of the final estimate. In this paper we will address the problem by comparing the Lincoln-Petersen estimator with three others proposed in the literature: Chapman (1951), Bailey (1951) and the stepwise Bayesian estimators (Mingoti,2000). By using a mathematical relationship it will be shown that the Bayes estimator has an expected value smaller than Lincoln-Petersen's and higher variability unless the sample sizes are large in which case the variability is almost the same. The simulation study shows that for small sample fractions, in both capture and recapture stages, the Bayes estimator has a better performance, but for larger sample fractions the Bayes estimator had a similar or worse performance than the other three. Chapman and Bailey are good alternatives in several cases.

Key-words: Capture-Recapture; Stepwise Bayesian, Lincoln-Petersen, Chapman and Bailey Estimators, Small Sample Sizes.

ON TWO OCCASIONS CAPTURE AND RECAPTURE DESIGNS: COMPARING AND DISCUSSING SOME ESTIMATORS

Index

1 Introduction

The estimation of population size using capture and recapture methods has been an issue of many papers along the years. Good reviews in this subject are Seber and Schwartz (2002) and Tsay and Chao (2001). Although these models are frequently used to estimate the size of animals population, more recently they have become popular in other areas such as Epidemiology, Social Sciences and Industry for example. Some interesting papers in this field are McKeganey *et. al.* (1992), Abeni *et. al.* (1994) and Mastro *et. al* (1994). In these papers capture and recapture methods were used to estimate the size of the HIV infected drug users in Bangkok and Lazio in Italy, and the size the female streetworking prostitution population and HIV infection in Glasgow. Capture and recapture methods have also been used to estimate the prevalence and the underreporting of certain diseases such as Aids and Diabetes for example (Ismail, *et. al*, 2000; Bernillon *et. al*, 2000; Hartnoll *et. al.*, 1985), to perform an adjustment for the undercount of census (Bell, 1993) or to estimate the total number of different defects types in a software inspection (Briand, *et. al.*1997; Miller, 1998; Eick *et. al.* 1993). Although nowdays the modern practical interest is to focus in models that allow for heterogeneity and trap response which require more than just one recapture, it is important to point out that the Lincoln-Petersen, Chapman and Bailey estimators have still being very used in situations where the high cost, time consumed and the difficulty to collect the data, make more than one recapture prohibited. A good example of this situation is related with surveys of difficult to access populations. In Caiaffa, Mingoti *et. al.* (2000, 2001) a study was conducted with the injection drug users in many cities of Brazil. The main objective was to obtain a profile of the drug users and to estimate the size of the population attended by a Drug Reduction Program supported by the Brazilian government. The city of Porto Alegre located at the south of Brazil, was used as a test for the capture-recapture methodology. The population involved in this study was very difficult to interview and the interviewers needed a special training. The cost of the whole survey was high. The time involved in collecting the data would be very considerable if we would make an option for many recaptures stages. The interesting part is that the results obtained with just one recapture was very reasonable indicating that probably there was no need for more recaptures to fulfill the main objective of the survey. Some other applications in Epidemiology are good examples of similar situations.

Probabilistic (Otis *et. al.* 1978; Seber, 1992) and log-linear methods (Bishop,1988) have been used to construct sensible estimators for the population size. In both methodologies the estimators are derived considering the presence of some variation factors in the capture occasions such as: time, environment, particular behavior of the element, as well as the three sources together.

A very well known estimator for a closed population and a capture and recapture design with only two capture occasions is the Lincoln-Petersen model (1896,1930) which was first used by Laplace in 1786 to estimate the population size of France. Lincoln-Petersen estimator has an infinite expectation. To correct this problem Chapman (1951) and Bailey (1951) proposed two other estimators which are modifications of the Lincoln-Petersen and which have finite expectation and variance. Recently, Mingoti (2000) suggested an estimator for population size which is a particular case of the stepwise Bayesian estimator derived in her paper to estimate the total number of distinct species in the population when sampling by elements is used to collect the data. An example was presented in Mingoti's paper in which the stepwise Bayesian estimator resulted in a better estimate than the Lincoln-Petersen. However, no study was performed comparing these two estimators. The purpose of this paper is to compare Mingoti, Lincoln-Petersen, Chapman and Bailey estimators presenting the main relationship among them. More than that we will discuss the reliability of these estimators by means of simulation.

2 The Estimators to be Compared

In this section we present the four estimators that are the subject of this paper. More attention will be given to the stepwise Bayesian estimator since the other three estimators are very well known in the field (Pollock, 1991; Seber 1986,1992).

We will suppose that the studied population is closed and that capture and recapture design with two stages of capture is used to collect the sample data. The first sample (capture) has m distinct elements of the population which were captured, tagged and returned to the population. The second sample has n elements from which s were already observed in the first sample (recaptured elements). Both samples were collected according to a simple random sampling. We will also assume that the marks of the elements captured in the first sample cannot be lost in the period of time between the collection of the two samples, so that the elements captured in the first sample can be recognized without error in the second sample. The true population size is denoted by N.

2.1 Lincoln-Petersen, Chapman and Bailey Estimators

The Lincoln-Petersen estimator of N is very simple. It is based on the fact that the proportion of marked elements of the second sample is an estimator of the marked elements of the population, before the second sample is collected. By making this two proportions equal, the Lincoln-Petersen estimator ($\hat{N}_P$) is obtained as follows:

$$\frac{s}{n} = \frac{m}{N} \Rightarrow \hat{N}_P = \frac{n\, m}{s} \qquad (2.1)$$

The smaller the number of recaptured elements the larger is the value of $\hat{N}_P$ and if $s = 0$ then $\hat{N}_P$ is infinite. Because s can take a zero value with probability different than zero, the expectation of $\hat{N}_P$ is infinite. To overcome this problem Chapman (1951) and Bailey (1951) proposed some estimators that have finite expectation and that are basically the Lincoln-Petersen estimator modified. They are respectively defined as:

$$\hat{N}_C = \frac{(m+1)(n+1)}{(s+1)} - 1 \qquad (2.2)$$

$$\hat{N}_B = \frac{(m)(n+1)}{s+1} \qquad (2.3)$$

Chapman estimator ($\hat{N}_C$) is based on the assumption that the second sample is taken without replacement and that the number of recaptured elements (s) has a Hypergeometric distribution. Bailey estimator $(\hat{N}_B)$ assumes that the second sample is taken with replacement and that s has a Binomial distribution. In practical problems there is no much difference between the Chapman and Bailey estimators (Seber, 1982).

2.2 Mingoti's Estimator

The stepwise Bayes estimator for N proposed by Mingoti (2000) is defined as:

$$\hat{N}_M = n + \frac{(m-s)(n-s+1)}{s} \qquad (2.4)$$

It is basically derived as a special case of a stepwise Bayesian estimator constructed for the true number of distinct species in a population when sampling by elements is used.

Originally, the construction of the estimator (2.4) assumes that the "species" of the population are independent and are divided in two groups: *group 1* which contains a list of the "species" the researcher believes are present in the population, and *group 2* which contains the "species" that are present in the population but does not belong to the researcher previous list. Every "species" has its own abundance value and a probability θ to belong to *group 1* and ($1-\theta$) to belong to *group 2*, $0 \leq \theta \leq 1$. Considering a model that gives a prior distribution for θ , for the abundance value of each species, and for the unknown true number of "species" S in the population, a stepwise Bayesian estimator is constructed for *S*.

To understand how this model can be applied in the estimation of the population size, each element of the population has to be considered as a distinct "species". When the capture and recapture procedure is used to collect the data we have the following situation: after the first sample is observed the researcher has the information about some elements ("species") that are present in the population of study. Before the second sample is taken the researcher does not know exactly if all the elements of the first sample are still present in the population of study or not. Therefore, a list containing the *m* elements of the first sample is available before the second sample is chosen. These *m* elements would then constitute the members of *group 1*. The group of new species (*group 2*) would be constituted by those elements that are present in the population but did not appear in the first sample. The estimator given in (2.4) is derived under a negative binomial distribution (Taylor *et. al*., 1979) for the true number of elements in the population N.

Generally speaking, the stepwise Bayesian is an estimation procedure which uses a partition of the main parameter set Θ in *c* mutually exclusive sets Θ_i, *i=1,2,...,c*. For every set Θ_i a Bayesian estimator is constructed according to some prior distribution π_i, considering the samples that have positive probability under π_i. If a sequence of mutually orthogonal prior densities π_i is chosen than the final estimator, which combines the results obtained for every Θ_i, is admissible (Hsuan, 1979). In Mingoti (2000) all the steps and prior distributions involved in the construction of the stepwise Bayesian estimator discussed in this technical report are presented.

3 An Example of Application

To ilustrate the use of the four estimators presented in section 2.0 we will report to the data presented in Caiaffa, Mingoti *et.al.* (2000,2001) mentioned in section 1 of this paper. In that study conducted in 1999, 55 injected drug users were interviewed in the first occasion and 99 in the second. A total of 17 individuals were capture in both occasions. The period of time for each occasion was about two months. Considering these results we would obtain the following estimates for N:

Lincoln-Petersen

$$\hat{N}_P = \frac{(55)(99)}{17} \approx 320$$

Chapman

$$\hat{N}_C = \frac{(56)(100)}{18} - 1 \approx 310$$

Bailey

$$\hat{N}_B = \frac{(55)(100)}{18} \approx 305$$

Mingoti

$$\hat{N}_M = 99 + \frac{(55-17)(99-17+1)}{17} \approx 285$$

As we can see there is no much difference among Lincoln-Petersen, Chapman and Bailey estimates. However, the Bayesian estimator resulted in a smaller value than the other three.

4 The Relationship between Mingoti and Lincoln-Petersen Estimators

Using some simple mathematics it is very easy to show that the following equations are true:

$$\hat{N}_M = \frac{nm}{s} - [(m-s)\left(\frac{s-1}{s}\right)] = \hat{N}_P - [(m-s)\left(\frac{s-1}{s}\right)]$$
$$= \hat{N}_P - c(m,s) \tag{4.1}$$

$$\hat{N}_M = \left(\frac{m(s+1)}{(m+1)s}\right)\hat{N}_C - (m-s+1) + \left(\frac{m(s+1)}{(m+1)s}\right) \tag{4.2}$$

$$\hat{N}_M = \left(\frac{s+1}{s}\right)\hat{N}_B - m + s + 1 \tag{4.3}$$

Although Chapman and Bailey estimators are very important in the literature, in this section we will focus the attention in the mathematical relationship between the Bayesian and the Lincoln-Petersen. The equation (4.1) shows that the Bayesian estimator is equal to the Lincoln-Petersen when $s=1$ or $s=m$. In any other situation the Bayesian estimator will take a smaller value than the Lincoln-Petersen. Let $s = \alpha_1 m$, and $s = \alpha N$, $0 < \alpha_1 < 1$, $0 < \alpha < 1$. Then the following result holds for the correction term $c(m,s)$ in equation (4.1):

$$c(m,s) = \frac{(1-\alpha_1)}{\alpha_1} [\, \alpha N - 1 \,] \qquad (4.\ 4)$$

and therefore, the numerical difference between the Bayesian and the Lincoln-Petersen estimator depends upon the true value *N*, and the proportions of recaptured elements, according to the size of the first sample *m* and to the true population size *N*. For some values of (α_1, α) the Bayesian estimator can take a value much smaller than the Lincoln-Petersen estimator. If the number of recaptured elements *s* is small then the Lincoln-Petersen estimator tends to result in a very large number. Therefore, in that case the correction term $c(m,s)$ might do some improvement in the quality of the estimation of N because it will decrease the bias of the Lincoln-Petersen. In this case the Bayesian estimator is expected to give better results than the Lincoln-Petersen. On the other hand, if the number of recaptured elements *s* is large, then the correction term $c(m,s)$ will affect negatively the quality of the Bayes estimator because its value might be much smaller than the true N in some cases. The simulation study in section 5.0 will show this fact.

Because the probability of s being zero is not zero then the Mingoti's estimator has an infinite expectation. If however, we suppose that the true population size *N* is large enough so that the probability of *s* being zero is approximately zero, and we use the Hypergeometric probability model for the variable *s*, it can be easily proved that the expectation and variance of $\hat{N}_P$ and $\hat{N}_M$ are approximated by:

$$E[\,\hat{N}_P\,] \approx N + \frac{N(N-m)}{n\,m} \qquad (4.5)$$

$$Var\,[\,\hat{N}_P\,] \approx N^2 \left(\frac{N-m}{nm} \right) \tag{4.6}$$

$$\begin{aligned} E[\,\hat{N}_M\,] &\approx E[\,\hat{N}_P\,] - m - 1 + \frac{nm}{N} + \frac{N}{nm} - \frac{N(N-m)}{(nm)^2} = \\ &= N - m - 1 + \frac{nm}{N} + \frac{N}{nm} + \frac{N(N-m)}{nm}[\,1 - \frac{1}{nm}\,] \end{aligned} \tag{4.7}$$

$$\begin{aligned} Var[\,\hat{N}_M\,] &\approx Var[\,\hat{N}_P\,] + N^2\,(N-m)\,[\,\frac{1}{m\,n^3} + \frac{2}{m\,n^2}\,] + \\ &+ (N-m)\,[\frac{n\,m}{N^2} - 2\,] \end{aligned} \tag{4.8}$$

As we can see from equations (4.5) to (4.8) **t**he $\hat{N}_M$ expectation is always smaller than the Lincoln-Petersen estimator and the variance is always larger. For large values of (*m,n,)* the variance of both estimators are about the same.

By replacing the true value of N by its respective estimate ($\hat{N}_M$, $\hat{N}_P$) in equations (4.5) to (4.8) one can obtain an estimate for the respective expectation and variance of these estimators.

5 Simulation Results

In this section we present the results obtained for a simulation study conducted with the purpose of comparing the four estimators presented in section 2.0. Three different population sizes were considered, *N*=250, *N*=500 and *N*=1000. The capture and recapture method was used for each population considering different sample sizes *m* and *n*. The four estimators were evaluated for samples that presented at least one element recaptured (*s* > 0). For each *N* a total of 800 random samples were selected with

$s > 0$. The mean error (*ME*) and the square root of the mean square error (*SRSME*) were also evaluated. To perform the simulation the *"Resampling Stats"* software (Simon *et. al*,1995) was used. The obtained results are presented in Tables 1, 2 and 3 (see Appendix).

For *N=250* Table 1 shows that for samples of same size and small sample fraction (6 up to 10 %) the Bayes estimator performs better than the Lincoln-Petersen, Chapman and Bailey. For a sample fraction of 20% Bayes is better than Lincoln-Petersen but it is worse than Chapman and Bailey. For higher sample fractions (over 20%) the best estimators were Chapman and Bailey, and Bayes was the worst. For samples of different sizes Table 2 shows that the Bayes estimator had a good performance compared to Lincoln-Petersen, being worse than Chapman and Bailey in some cases. The same general conclusions are obtained when *N=500* and *N=1000* are considered (Tables 3 to 6). However, it is interesting to notice that for *N=1000* the best estimators in all the sample fractions considered were Chapman and Bailey. In all cases the values of *SRMES* were very large for small sample fractions which is expected. The analysis of samples of different sizes is important because in many situations the capture and recapture design is applied considering the period of time to collect information as a marking point to stop the experiment. This fact makes the samples sizes to be random and, in the majority of cases, they would be different. That is the case of the example presented in section 3.0 of this paper.

It is natural to understand the results of the Bayes estimator in this simulation study. When the sample fraction is high in both samples, it is expected to see a higher number of recaptured elements which affects negatively the performance of the Bayes estimator. On the other hand, for small sample sizes the expected number of recaptured elements is small and therefore the Bayes estimator tends to give better results. If we observe the original Mingoti's paper (2000) we will see that the stepwise Bayes estimator is derived aiming to create an estimator that is sensitive to the "rare" species present in the population. In the capture-recapture methodology the information about a "rare" species, i.e the elements that are present in the population but were not caught in the samples, would be represented by the value of s. Small values of s would indicate that the population size N is large, and high values of s would indicate that the N is small.

From Tables 1 to 6 it can be seen that for a similar situation as the Andreawartha's example (1961) mentioned by Mingoti (2000), the Bayes estimator had a better performance for all the studied populations. In that example the first sample and the second samples corresponded respectively to 4 % and 1% of the true population size. The sample fractions were 24.9% (first sample) and 5.6% (second sample). Therefore, the samples were unbalanced and the fractions of recaptured elements were very small.

6 Final Remarks

The purpose of this paper was to compare the four estimators used to estimate the population size when capture-recapture experiments with two occasions is used. As we could see the Bayes estimator can be a better alternative for situations where the sample fractions are small and the proportion of recaptured elements is small compared to the first sample and the true population size *N*. For larger sample fractions Chapman, Bailey and Lincoln-Petersen estimators were better. For samples of same size the sample fraction cutting point is 20% being the Bayesian estimator recommended for situations with sample fraction smaller than that. For samples with different sizes the Bayes estimator could be considered as an alternative in the cases were both of the sample fractions are not higher than 20%. In order to decide which estimator is more appropriate in a practical situation, the estimated sample fractions can be used as a helpful tool to take the final decision. In all cases, the square root of the mean error square (*SRMES*) are very large unless the samples sizes are closer to the true population size. Therefore, the price of using only two occasions is paid by using estimators with large variability in the majority of practical situations. However, in many cases is almost impossible to collect data of human populations in more than two occasions due mainly to the limitations of time and budget. Even for studies of certain diseases such as Aids, Diabetics, Meningococis, where the population size is estimated by matching lists of records obtained by some governmental source, the use of more than two lists could be complex due to the time consuming in collecting and preparing the data to be analysed by capture and recapture procedures. It is important to consider that in many countries the mecanism of keeping records about the people who has certain diseases or who died from them is not as well organized and reliable as it should be. Therefore, the first step in any study that would use these sources as a data base for a capture and recapture experiment, would be to analyse very carefully the quality of the data, correcting typing errors, wrong or missing information, and so on (*Seber*, et. al., 2000).

Considering the time, the difficulty and the cost involved in collecting data in some difficult to access populations, in particular human populations, the existence of a good estimator that would require small samples sizes to estimate the population size is very important.

The Andreawartha's example presented in Mingoti's paper (2000) and which had motivated the use of the stepwise Bayesian estimator in capture and recapture designs had some special characteristics that affected the Bayes estimator positively in the sense that it resulted in a value closer to the true population size. In that example the second sample size was much smaller than the first and consequently the proportion of recaptured elements was very small. In all the cases considered in the simulation study which were similar to the situation of Andreawartha's example, the Bayes estimator had a better performance than Lincoln-Petersen's, Chapman's and Bailey's.

The stepwise Bayesian estimator treated in this paper is a particular case of the more general estimator derived in Mingoti's paper (2000). The negative binomial distribution was used as a prior for the true population size *N*. However, other prior distributions can be chosen and other alternative estimators for capture-recapture designs can be constructed by using the equation (2.4, pp. 655) of Mingoti's original paper. This fact would give more flexibility to the researcher who could use some previous knowledge that he or she had about the population and which could result in a different prior from the negative binomial distribution.

7 References

ANDREAWARTHA, H.G.(1961) *Introduction of the Study of Animal Populations* (Chicago,University of Chicago Press).

ABENI, D. D., BRANCATO, G., PERUCCI, C. A. (1994) Capture-recapture to estimate the size of the population with human immunodeficiency virus type I infection, *Epidemiology*, 5, 4. 410-414.

BRIAND, L. C., EMAN, K. E., FREIMUT, B., LAITENBERGER, O. (1997) Quantitative evaluation od capture-recapture models to control software inspection, *International Software Engineering Research Network Technical Report.*

CAIAFFA, W., MINGOTI, S. A., PROIETTI, F., CARNEIRO-PROIETTI, A.B.,MARQUES, L., DONEDA, D., SILVA, R.C., PROJECT AjUDE BRASIL (2000) Estimating HIV-HCV single and co-infected injecting drug users (IDUs) who attended an outreach syringe exchange program (SEP) located in south of Brazil, *Abstracts of the XII International Aids Conference,* Durban, South Africa, vol. II, p. 349.

CAIAFFA, W., MINGOTI, S. A., PROIETTI, F., CARNEIRO-PROIETTI, A.B.,MARQUES,L.F, DONEDA, D., LOPES, D., SILVA, R.C., PROJECT AjUDE BRASIL (2001) Estimating of injecting drug users (IDUs) attendeding an outreach syringe exchange program (SEP) in the city of Porto Alegre, south of Brazil, and number of infected with HIV and hepatitis C (HVC) - The AjUDE BRASIL PROJECT, *Abstracts of the XII International Conference on the Reduction of Drug related Harm,* New Delli,India, p. 64.

BAILEY, N.T.J. (1951) On estimating the size of mobile populations from capture-recapture data. *Biometrika*, 38, 293-306.

BELL, W.R. (1993) Using information from demographic analysis in post-enumeration survey estimation. *Journal of the American Statistical Association*, 88, 423, 1106-1118.

BERGER , J. O. (1985) *Statistical Decision Theory and Bayesian Analysis*. 2nd. edn. (New York, Springer Verlag).

BERNILLON, P., LIEVRE, L., PILLONEL, J., LAPORTE, A., COSTAGLIOLA, D. & THE CLINICAL EPIDEMIOLOGY GROUP FROM CISIH. (2000) Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of Aids cases: France 1990-1993, *International Journal of Epidemiology*, 29, 168-174.

BISHOP, Y. M. M., FIENBERG, S. E. , HOLLAND, P. W. (1988) *Discrete multivariate analysis* (London: The MIT Press).

CHAPMAN, D. G.(1951) Some properties of the hypergeometric distribution with applications to zoological censuses. *University of California Publications Statistics*, 1, 131-160.

EICK, S. G., LOADER, C. R., LONG, M. D., VOTTA, L. G., WIEL, S. V. (1993) Estimating software fault content before coding, *Proceedings of the 15th International Conference of Software Engineering, Melbourne*.

HSUAN, F.C. (1979) A stepwise Bayesian procedude. *Annals of Statistics*, 7:860-868.

ISMAIL, A. A., BEECHING, N. J., GILL, G. V., BELLIS, M. A. (2000) How many data sources are needed to determine diabetes prevalence by capture-recapture? *International Journal of Epidemiology*, 29, 536-541.

LINCOLN, F. C. (1930) Calculating waterfowl abundance on the basis of banding returns, *U.S. Department of Agricultural Circular,* 118, 1-4.

MASTRO, T. D., KITAYAPORN, D., WENIGER, B. G., VANICHSENI, S., LAOSUNTHORN, V., UNEKLABH, T., UNEKLABH, C., CHOOPANYA, K., LIMPAKARNJANARAT, K. (1994) Estimating the number of HIV-infected drug users in Bangkok: a capture-recapture method, *American Journal of Public Health*, 7, 1094-1099.

McKEGANEY, N.,BARNARD M.,LEYLAND, A., COOTE,I., FOLLET, E. (1992) Female streetworking prostitution and HIV infection in Glasgow, *British Medical Journal*, 305, 801-804.

MILLER, J. (1998) Estimating the number of remaining defects after inspection, Technical Report, Departament of Computer Science, University of Stratheclyde, 38 p.

MINGOTI, S. A. (2000) A stepwise Bayesian estimator for the total number of distinct species in finite populations: sampling by elements, *Journal of Applied Statistics* 27, 5, 651-670.

OTIS, D., BURNHAN, K.P., WHITE, G.C. & ANDERSON, D.R. (1978) Statistical inference from capture data on closed populations, *Wildlife Monographs*, 62, 1-135.

PETERSEN, G.G.J. (1896) The yearly immigration of young plaice into the limfjord from the german sea, *Report of Danish Biology Statistics*, 6, 1-48.

POLLOCK, K. H. (1991) Modeling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future, *Journal of the American Statistical Association*, 86, 413, 225-238.

SEBER, G.A.F. (1982) *The estimation of Animal Abundance and Related Parameters,* 2nd. edn. (London, Griffin).

SEBER, G. A. F. (1986) A review of estimating animal abundance, *Biometrics*, 42, 267-292.

SEBER, G. A. F. (1992) A review of estimating animal abundance II, *International Statistical Review*, 60, 2, 129-166.

SEBER, G. A. F., HUAKAU, J. T., SIMMONS, D. (2000) Capture-Recapture. Epidemiology, and list mismatches: two lists, *Biometrics*, 56, 1227-1232.

SEBER, G. A, F. & SCHWARTZ, C. J. (2002) Capture-recapture: before and after Euring 2000. *Journal of Applied Statistics*, 29, 1-4, 5-18.

SIMON, J., BRUCE P. & OSWALD, T.(1995) *Resampling Stats User's Guide* (Virginia, Resampling Stats, Inc).

TAYLOR, L.R., WOIWOD, I.P. & PERRY, J.N. (1979) The negative binomial as a dynamic ecological model for aggregation, and the density dependence of *k, Journal of Animal Ecology,* 48, 289-304.

TSAY, P. K. & CHAO, A. (2001) Population size estimation for capture-recapture models with applications to epidemiological data. *Journal of Applied Statistics*, 28, 1, 25-36.

8 Appendix

The Tables 1 to 6 mentioned in section 5 are presented in this Appendix.

Table 1: Simulation results for population size *N*=250 and equal sample sizes

$(m,n;\ f_1,f_2)$	$(s';\ \hat{\alpha}_1,\hat{\alpha})$	*Estimator*	$\hat{N}$	*ME*	SRMES
(15, 15; 6; 6)	(1.31; 8.7,0.5)	Bayes	181.37	- 68.632	92.604
		Petersen	183.69	- 66.312	88.692
		Chapman	110.66	- 139.34	141.350
		Bailey	104.68	- 145.32	147.010
(20, 20; 8,8)	(1.96; 9.8,0.78)	Bayes	253.22	3.2198	123.650
		Petersen	259.30	9.2967	118.950
		Chapman	163.71	- 86.286	99.152
		Bailey	156.87	- 93.130	104.870
(25, 25; 10,10)	(2.75; 11, 0.11)	Bayes	281.09	37.094	178.870
		Petersen	298.41	48.408	176.140
		Chapman	203.96	- 46.037	87.524
		Bailey	197.08	- 52.920	89.015
(50, 50; 20,20)	(10; 20, 4)	Bayes	235.07	- 14.932	92.582
		Petersen	270.65	20.653	94.076
		Chapman	250.73	0.730	71.259
		Bailey	246.80	- 3.200	69.932
(100, 100; 40,40)	(39.94; 39, 16)	Bayes	194.17	- 55.829	59.697
		Petersen	252.70	2.702	24.820
		Chapman	250.37	0.371	23.919
		Bailey	248.88	- 1.118	23.705
(150, 150; 60,60)	(90.03; 60, 36)	Bayes	191.06	- 58.939	59.345
		Petersen	250.36	0.359	10.658
		Chapman	249.91	- 0.088	10.557
		Bailey	249.25	- 0.750	10.514
(200, 200; 80,80)	(150.96; 75, 60)	Bayes	210.34	- 39.663	39.690
		Petersen	250.12	0.124	4.013
		Chapman	250.06	0.061	4.002
		Bailey	249.81	- 0.187	3.986

Note: $f_1=m/N$ *and* $f_2=n/N$; $\hat{\alpha}_1=s'/m$ *and* $\hat{\alpha}=s'/N$ *(all in percentages)*
ME: mean error; SRMES: square root of the mean error square

Table 2: Simulation results for population size N=250 and different sample sizes

$(m,n;\ f_1,f_2)$	$(s';\ \hat{\alpha}_1,\hat{\alpha})$	*Estimator*	$\hat{N}$	*ME*	SRMES
(20, 25; 8,10)	(2.32; 11.6,0.9)	Bayes	274.41	24.406	152.960
		Petersen	281.82	31.818	149.620
		Chapman	184.32	- 65.679	90.330
		Bailey	176.50	- 73.504	94.290
(25, 20; 10, 8)	(2.25; 9, 0.9)	Bayes	279.33	29.327	155.990
		Petersen	288.64	38.638	151.890
		Chapman	187.39	- 62.610	88.100
		Bailey	181.14	- 68.856	91.060
(25, 50; 10,20)	(5.48; 22, 2.2)	Bayes	284.82	34.816	186.680
		Petersen	299.77	49.773	187.930
		Chapman	244.46	- 5.538	96.560
		Bailey	236.02	- 13.979	93.730
(50, 100; 20,40)	(19.92; 39, 7.9)	Bayes	228.91	- 21.090	45.410
		Petersen	257.42	7.417	43.390
		Chapman	250.89	0.886	39.530
		Bailey	246.48	- 3.053	38.870
(50, 25; 20, 10)	(4.94; 9.8, 1.97)	Bayes	272.42	22.421	184.110
		Petersen	306.23	56.230	185.850
		Chapman	249.16	- 0.844	96.630
		Bailey	245.25	- 4.749	94.850
(50, 10; 20, 4)	(2.23; 4.5, 0.89)	Bayes	271.69	21.688	164.800
		Petersen	291.32	41.317	155.050
		Chapman	193.45	- 56.553	85.740
		Bailey	190.63	- 59.365	86.690
(60, 14; 24, 5.6)	(3.40; 5.6, 1.36)	Bayes	284.69	34.689	210.580
		Petersen	319.47	69.966	207.040
		Chapman	235.84	- 14.163	93.260
		Bailey	232.95	- 17.046	92.250

Note: $f_1=m/N$ *and* $f_2=n/N$; $\hat{\alpha}_1 = s'/m$ *and* $\hat{\alpha} = s'/N$ *(all in percentages)*
ME: mean error; SRMES: square root of the mean error square

Table 3: Simulation results for population size N=500 and equal sample sizes

(m,n; f_1, f_2)	*(s'; $\hat{\alpha}_1, \hat{\alpha}$)*	*Estimator*	$\hat{N}$	*ME*	*SRMES*
(30, 30; 6, 6)	(2.14; 7.1, 0.4)	Bayes	539.06	39.063	281.980
		Petersen	549.60	49.603	275.730
		Chapman	342.78	- 157.22	192.350
		Bailey	332.69	- 167.31	198.730
(40, 40; 8, 8)	(3.31; 8.3, 0.7)	Bayes	609.60	109.60	406.760
		Petersen	631.50	131.50	405.330
		Chapman	445.33	- 54.67	182.040
		Bailey	435.44	- 64.56	181.290
(50, 50; 10, 10)	(4.84; 9.7, 0.9)	Bayes	633.83	133.83	498.550
		Petersen	666.65	166.65	500.880
		Chapman	512.51	12.52	232.790
		Bailey	503.45	3.45	227.920
(100, 100; 20, 20)	(20.27; 20.27,4)	Bayes	432.57	- 67.427	110.740
		Petersen	508.22	8.221	90.649
		Chapman	491.65	- 8.348	83.278
		Bailey	487.95	- 12.226	82.944
(200, 200; 40, 40)	(80.00; 40,16)	Bayes	381.72	- 116.280	119.810
		Petersen	502.20	2.006	34.073
		Chapman	499.92	- 0.082	33.484
		Bailey	498.43	- 1.574	33.355
(300, 300; 60, 60)	(180.26; 60, 36)	Bayes	380.64	- 119.360	119.730
		Petersen	499.72	- 0.278	14.690
		Chapman	499.28	- 0.721	14.641
		Bailey	498.62	- 1.384	14.640
(400, 400; 80, 80)	(320.16; 80, 64)	Bayes	420.22	- 79.782	79.806
		Petersen	499.81	- 0.191	5.419
		Chapman	499.75	- 0.253	5.415
		Bailey	499.50	- 0.501	5.419

Note: $f_1 = m/N$ *and* $f_2 = n/N$; $\hat{\alpha}_1 = s'/m$ *and* $\hat{\alpha} = s'/N$ *(all in percentages)*
ME: mean error; SRMES: square root of the mean error square

Table 4: Simulations results for population size *N*=500 and different sample sizes

(m,n; f_1, f_2)	*(s'; $\hat{\alpha}_1$,$\hat{\alpha}$)*	*Estimator*	$\hat{N}$	*ME*	*SRMES*
(40, 50; 8, 10)	(3.96; 9.9, 0.8)	Bayes	628.74	128.740	435.960
		Petersen	652.73	152.730	437.290
		Chapman	483.54	- 16.456	198.020
		Bailey	472.73	- 27.274	194.640
(50, 40; 10, 8)	(3.99; 7.9, 0.8)	Bayes	629.10	129.100	464.640
		Petersen	659.62	159.620	465.120
		Chapman	484.15	- 15.847	207.620
		Bailey	475.64	- 24.360	204.410
(50, 100; 10, 20)	(10.07; 20.14, 2)	Bayes	504.89	4.886	180.190
		Petersen	540.42	40.416	185.210
		Chapman	496.77	- 3.229	143.220
		Bailey	488.01	- 11.989	140.890
(100, 200; 20,40)	(39.83; 39.83, 7,9)	Bayes	410.72	- 89.280	89.537
		Petersen	500.30	0.305	11.630
		Chapman	500.02	0.018	11.587
		Bailey	499.35	- 0.646	11.567
(100, 50; 20, 10)	(10.02; 10, 2)	Bayes	463.69	- 36.313	184.150
		Petersen	543.80	43.795	184.700
		Chapman	499.65	- 0.352	142.920
		Bailey	495.69	- 4.308	141.590
(100, 20; 20, 4)	(4.03; 4, 0.8)	Bayes	574.39	74.387	428.370
		Petersen	639.39	139.390	426.550
		Chapman	483.08	- 16.924	198.930
		Bailey	479.28	- 20.716	197.340
(120, 28; 24, 5.6)	(6.73; 5.6, 1.3)	Bayes	481.79	- 18.206	291.220
		Petersen	575.51	75.509	292.300
		Chapman	499.42	- 0.582	180.410
		Bailey	496.28	- 3.718	178.960

Note: $f_1 = m/N$ *and* $f_2 = n/N$; $\hat{\alpha}_1 = s'/m$ *and* $\hat{\alpha} = s'/N$ *(all in percentages)*
ME: mean error; SRMES: square root of the mean error square

Table 5: Simulation results for population of size *N=1000* and equal sample sizes

$(m,n;\ f_1,f_2)$	$(s';\ \hat{\alpha}_1,\hat{\alpha})$	*Estimator*	$\hat{N}$	*ME*	*SRMES*
		Bayes	1270.30	270.290	893.090
(60, 60; 6, 6)	(3.66; 6.1, 0.4)	Petersen	1305.90	305.860	892.670
		Chapman	924.92	- 75.082	391.000
		Bailey	910.74	- 89.261	387.840
		Bayes	1160.40	160.390	712.440
(80, 80; 8, 8)	(6.27; 7.8, 0.62)	Petersen	1219.90	219.870	721.680
		Chapman	1013.60	13.602	404.340
		Bailey	1002.10	2.076	399.130
		Bayes	1071.10	71.083	572.050
(100, 100; 10, 10)	(9.69; 9.7, 0.96)	Petersen	1150.90	150.880	583.870
		Chapman	1034.00	33.986	365.380
		Bailey	1024.70	24.739	361.040
		Bayes	852.34	- 147.660	196.120
		Petersen	1008.00	7.980	133.460
(200, 200; 20, 20)	(40.32; 20.16, 4)	Chapman	991.65	- 8.353	127.810
		Bailey	987.71	- 12.292	127.490
		Bayes	763.73	236.270	239.920
		Petersen	1002.20	2.183	49.397
(400, 400; 40, 40)	(160.04; 40, 16)	Chapman	999.92	- 0.084	48.976
		Bailey	998.42	- 1.500	48.880
		Bayes	761.13	- 238.870	239.270
		Petersen	1000.50	0.455	21.570
(600, 600; 60, 60)	(360.00; 60, 36)	Chapman	1000.00	0.001	21.517
		Bailey	999.34	- 0.656	21.491
		Bayes	840.21	- 159.790	159.810
		Petersen	999.78	- 0.219	8.027
(800, 800; 80, 80)	(640.18; 80, 64)	Chapman	999.72	- 0.282	8.024
		Bailey	999.47	- 0.531	8.027

Note: $f_1 = m/N$ *and* $f_2 = n/N$; $\hat{\alpha}_1 = s'/m$ *and* $\hat{\alpha} = s'/N$ *(all in percentages)*
ME: mean error; SRMES: square root of the mean error square

Table 6: Simulation results for population of size *N=1000* and different sample sizes

(m,n; f_1,f_2)	*(s'; $\hat{\alpha}_1$,$\hat{\alpha}$)*	*Estimator*	$\hat{N}$	*ME*	*SRMES*
(80, 100; 8, 10)	(7.99; 9.9, 0.79)	Bayes	1066.80	66.766	462.320
		Petersen	1128.50	128.480	472.960
		Chapman	994.03	- 5.974	329.000
		Bailey	982.74	-17.258	325.350
(80, 100; 8, 10)	(7.99; 9.9, 0.79)	Bayes	1035.40	35.406	548.660
		Petersen	1114.20	114.230	554.800
		Chapman	977.76	- 22.241	365.080
		Bailey	969.07	- 30.931	362.120
(100, 200; 10, 20)	(19.77; 19.77; 1.97)	Bayes	972.51	- 27.486	203.200
		Petersen	1048.50	48.504	209.560
		Chapman	1008.50	8.471	185.760
		Bailey	999.48	- 0.524	183.730
(200, 400; 20, 40)	(80.17; 40, 8)	Bayes	885.62	- 114.380	135.750
		Petersen	1003.90	3.935	79.158
		Chapman	997.87	- 2.349	77.685
		Bailey	993.90	- 6.104	77.510
(200, 100; 20, 10)	(20.12; 10.06, 2)	Bayes	859.72	- 140.280	252.390
		Petersen	1030.30	30.293	213.340
		Chapman	991.65	- 8.351	190.970
		Bailey	987.71	- 12.289	190.230
(200, 40; 20, 4)	(4.03; 2, 0.4)	Bayes	954.00	- 45.996	571.480
		Petersen	1119.00	118.960	570.070
		Chapman	989.27	- 10.734	351.830
		Bailey	985.34	- 14.661	350.220
(240, 56; 24, 5.6)	(6.73; 2.8, 0.67)	Bayes	856.31	- 143.690	317.220
		Petersen	1064.90	64.892	288.150
		Chapman	1002.90	2.861	240.690
		Bailey	999.70	- 0.305	239.670

Note: $f_1 = m/N$ and $f_2 = n/N$; $\hat{\alpha}_1 = s'/m$ and $\hat{\alpha} = s'/N$ (all in percentages)

ME: mean error; SRMES: square root of the mean error square