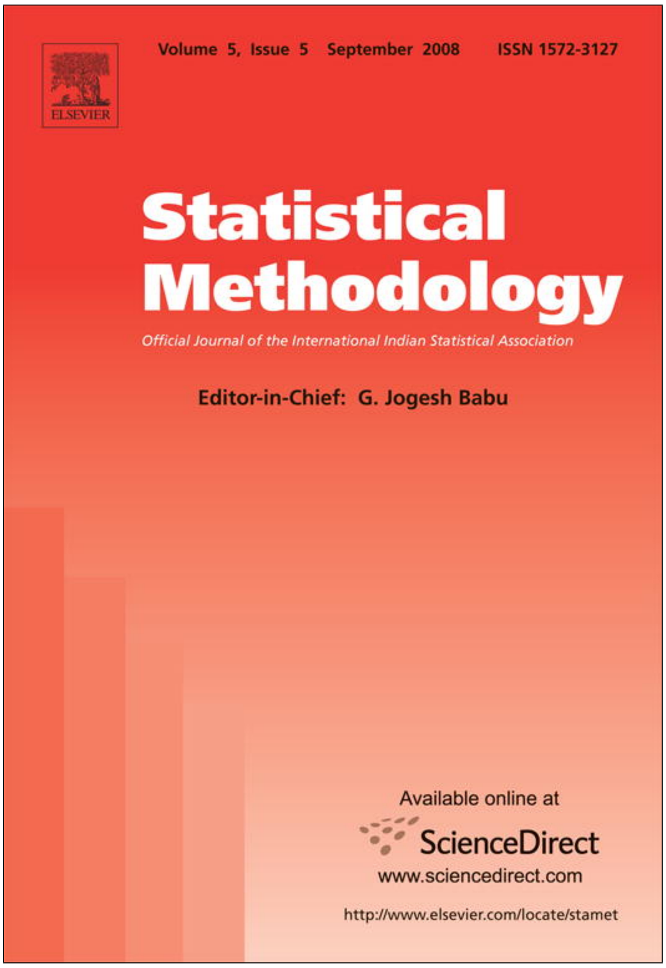
Volume 5, Issue 5 September 2008 ISSN 1572-3127
ELSEVIER
Statistical
Methodology
Official Journal of the International Indian Statistical Association
Editor-in-Chief: G. Jogesh Babu
Available online at
ScienceDirect
www.sciencedirect.com
http://www.elsevier.com/locate/stamet

Available online at www.sciencedirect.com
ScienceDirect
Statistical Methodology 5 (2008) 410–423
Statistical Methodology
Official Journal of the International Indian Statistical Association
www.elsevier.com/locate/stamet

A simple variance formula for population size estimators by conditioning

Dankmar Böhning*

Applied Statistics, School of Biological Sciences, University of Reading, Reading, RG6 6FN, UK

Received 15 December 2006; received in revised form 30 August 2007; accepted 15 October 2007

Abstract

This note considers the variance estimation for population size estimators based on capture–recapture experiments. Whereas a diversity of estimators of the population size has been suggested, the question of estimating the associated variances is less frequently addressed. This note points out that the technique of conditioning can be applied here successfully which also allows us to identify sources of variation: the variance due to estimation of the model parameters and the binomial variance due to sampling n units from a population of size N. It is applied to estimators typically used in capture–recapture experiments in continuous time including the estimators of Zelterman and Chao and improves upon previously used variance estimators. In addition, knowledge of the variances associated with the estimators by Zelterman and Chao allows the suggestion of a new estimator as the weighted sum of the two. The decomposition of the variance into the two sources allows also a new understanding of how resampling techniques like the Bootstrap could be used appropriately. Finally, the sample size question for capture–recapture experiments is addressed. Since the variance of population size estimators increases with the sample size, it is suggested to use relative measures such as the observed-to-hidden ratio or the completeness of identification proportion for approaching the question of sample size choice.

Keywords: Capture-recapture; Population size; Variance estimation; Horvitz–Thompson estimator

1. Introduction

The size N of some population is frequently wished to be determined. In the biological sciences this is often a wildlife population whereas in the life or social sciences this population

* Tel.: +44 1183786211.
E-mail address: d.a.w.bohning@reading.ac.uk.

1572-3127/$ - see front matter
doi:10.1016/j.stamet.2007.10.001

might be a group of people difficult to sample such as illegal drug users or car drivers without a license. Suppose that a specific mechanism identifies some, say n, but not all units of a population of size N. Furthermore, assume that identification occurs independently for each population unit with probability $1 - p_0$. This stochastic situation can be described by tuples of size N

$$(\delta_1, \delta_2, \ldots, \delta_N)$$

where $\delta_i = 1$ indicates that the ith unit is identified (and observed) and $\delta_i = 0$ otherwise (and the unit remains unobserved). Each of these tuples occurs with probability $(1-p_0)^{\sum_{i=1}^{N}\delta_i} p_0^{N-\sum_{i=1}^{N}\delta_i}$. We are interested in the probability that exactly n units are identified. Since there are $\binom{N}{n}$ tuples $(\delta_1, \delta_2, \ldots, \delta_N)$ with $\sum_{i=1}^{N}\delta_i = n$ the probability of observing exactly n units is a simple *binomial distribution*:

$$\binom{N}{n}(1 - p_0)^n p_0^{N-n}. \tag{1}$$

Then, the well-known Horvitz–Thompson estimator of the population size is given as

$$\hat{N} = \frac{n}{(1 - p_0)}. \tag{2}$$

Note that (1) can be viewed as a likelihood function in N which is maximized for N being the integer part of (2). Note that all sources of random variation occurring in (2) are due to the random structure of n, in other words, depend on its probability distribution which in fact is binomial with success parameter $(1 - p_0)$ and sample size parameter N. Hence, its variance is readily available as

$$\mathrm{Var}(\hat{N}) = \frac{N(1 - p_0)p_0}{(1 - p_0)^2}. \tag{3}$$

As a consequence, relative measures such as the *observed/hidden ratio* $n/(N - n)$ or the *completeness of identification* measure n/N are *free* of any random error when estimated: $n/(\hat{N} - n) = 1/[(1 - p_0)^{-1} - 1] = p_0/(1 - p_0)$ and $n/\hat{N} = (1 - p_0)$. This implies that these measures will have reduced variation (if the unknown parameter is estimated) if compared to estimates of population size.

Furthermore, since $E(n) = N(1 - p_0)$, the variance given in (3) can readily be estimated as

$$\widehat{\mathrm{Var}}(\hat{N}) = n\frac{p_0}{(1 - p_0)^2}. \tag{4}$$

However, p_0 will be known only in exceptional cases and usually an estimate of p_0 will be required for practical use. This will add additional variation and the variance estimate (4) (with $\hat{p}_0$ replacing p_0) will no longer be valid. In addition, some modeling for p_0 will be required. We will address this question in the following sections.

1.1. Estimators based upon counts arising from capture–recapture experiments in continuous time (CRECT)

The mechanism that identifies units with probability $1 - p_0$ can be quite general. It might be that several sources identify the units leading to a log-linear modeling approach [1]. Another common approach for deriving an estimator of p_0 is based upon counting repeated identifications of the same unit by the mechanism over a given time span. This is usually referred to as a capture–recapture experiment in continuous time (CRECT). For example, in a CRECT repeated occurrences of dolphins are counted by some mechanism or it may be counted how often a patient comes for the treatment of a certain disease to a treatment institution. We will denote by f_0, f_1, $f_2, \ldots f_m$ the frequency of those units identified exactly 0, 1, 2, ..., m times where m is the largest count occurred. Also, we will denote with $p_0, p_1, p_2, \ldots p_m$ the probability of exactly 0, 1, 2, ... m identifications. Clearly, f_0 is unobserved and is the target of the inference. We have that $n = f_1 + f_2 + \cdots + f_m$ and $N = n + f_0$.

Example 1. To illustrate, we look at the following CRECT: Oremus [8] tried to estimate the size of a small community of spinner dolphins which are residing 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 mis-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 (see also Table 1).

It is interesting to recall the likelihoods involved in CRECT. The CRECT can be described with a multinomial likelihood

$$\begin{pmatrix} & & N & \\ f_0 & f_1 & \ldots & f_m \end{pmatrix} p_0^{f_0} p_1^{f_1} \ldots p_m^{f_m} \tag{5}$$

which can be written as the product of the two likelihoods

$$\binom{N}{f_0} p_0^{f_0}(1-p_0)^{N-f_0} \times \begin{pmatrix} & n & \\ f_1 & \ldots & f_m \end{pmatrix} \left(\frac{p_1}{1-p_0}\right)^{f_1} \cdots \left(\frac{p_m}{1-p_0}\right)^{f_m}. \tag{6}$$

Note that the first binomial likelihood in (6) is identical to (1) with $f_0 = N - n$, also using that $\binom{N}{f_0} = \binom{N}{n}$. Also note that $1 - p_0 = p_1 + \cdots + p_m$. The second, truncated multinomial likelihood is independent of N and this fact is exploited in the *conditional maximum likelihood* approach where the second likelihood is maximized in p_j separately from the first, the binomial likelihood.

Now some parametric or at least semi-parametric structure needs to be imposed on $p_j = p_j(\lambda)$. Since, if the p_js are left nonparametric, the re-parameterized $p_j^* = p_j/(1 - p_0)$ would simply be estimated by f_j/n and carry no information on p_0. Having found an estimate for $p_0(\lambda) = 1 - g(\lambda)$, the estimate of N is readily available with the Horvitz–Thompson estimator. To illustrate the conditional approach in a simple model, one can assume that the counts arise from a Poisson with parameter λ, so that $p_j = \exp(-\lambda)\lambda^j/j!$ and $g(\lambda) = 1 - p_0 = 1 - \exp(-\lambda)$. The estimation could be done by maximizing the likelihood function based upon the zero-truncated Poisson density

$$L(\lambda) = \prod_{i=1}^{m} \left\{ \frac{\mathrm{Po}(i, \lambda)}{1 - \exp(-\lambda)} \right\}^{f_i} \tag{7}$$

with respect to λ, where m is the largest observed count and $\mathrm{Po}(i, \lambda) = \exp(-\lambda)\lambda^i/i!$ The simple zero-truncated Poisson requires homogeneity of the Poisson parameter which appears to be unlikely in CRECT. It might be more appropriate to incorporate unobserved heterogeneity leading to $p_j(\lambda) = \int_0^\infty t^j \exp(-t)/j!\lambda(t)\mathrm{d}t$ for $j = 0, 1, 2, \ldots$. with $\lambda(t)$ being the density of the heterogeneity distribution and now the parameter of interest. Chao [4] uses the Cauchy–Schwartz inequality to deduce

$$\left(\int_0^\infty t\exp(-t)\lambda(t)\mathrm{d}t\right)^2 \le \int_0^\infty \exp(-t)\lambda(t)\mathrm{d}t \int_0^\infty t^2\exp(-t)\lambda(t)\mathrm{d}t$$

or simply $p_1^2 \le p_0(2p_2)$, from which the Chao lower bound estimator $\hat{f}_0 = f_1^2/(2f_2)$ or $N_C = n + f_1^2/(2f_2)$ follows.

Zelterman [16] was inspired by developing a more robust approach being valid even if contaminations in the Poisson model occurred. His suggestion was based upon

$$\lambda = (i+1)\frac{\mathrm{Po}(i+1,\lambda)}{\mathrm{Po}(i,\lambda)} \tag{8}$$

so that the simple estimator $\hat{\lambda}_i = (i+1)f_{i+1}/f_i$ can be constructed, with the typical choice $i = 1$ leading to $\hat{\lambda}_1 = 2f_2/f_1$ and associated population size estimator $\hat{N}_Z = \frac{n}{1-\exp(-2f_2/f_1)}$. Evidently, $\hat{\lambda}_1 = 2f_2/f_1$ will not be affected by counts larger than 2, so that it will retain its unbiased property if only the f_2/f_1 ratio remains invariant which can be expected for a wider class of Poisson mixture models. The Zelterman's estimator turns out to be less dependent on the Poisson assumption, so that it might still be considered even if the Poisson model is not valid [16] and it is likely that for this reason it has found popularity in many applied areas [6,9]. For the data of Example 1 we find the Zelterman estimate and the Chao lower bound estimate to be $\hat{N}_Z = 180$ and $\hat{N}_C = 177$, respectively, both estimates being quite similar.

2. A general approach by conditioning

We are interested in developing an expression for the variance of $\hat{N} = \frac{n}{1-p_0(\hat{\lambda})} = \frac{n}{1-p_0(\hat{\lambda})}$. It is important to take two sources of random variation into account: the binomial random variation induced by sampling n units out of N with N unknown, and the random variation due to the estimation of λ. It is assumed that λ is estimated by some well-motivated estimator $\hat{\lambda}$ such as the estimator of Chao [4] or Zelterman [16], though not necessarily by maximum likelihood. For conditional and unconditional maximum likelihood estimations, Sanathan [11,12] provided variance estimates of the population size estimator based upon the full information matrix with respect to the truncated likelihood. In the following we apply a technique for computing moments usually referred to as conditioning [10 (p. 92),7 (p. 191),13 (p. 9)] to population size estimation. The technique provides a simple formula for variance computation of population size which can be applied to a general estimator of λ. If maximum likelihood estimation is used the variance estimating formula based upon conditioning appears to be simple in its application. Let $f(\hat{\lambda}, n)$ be the joint distribution of λ and n, which can be written as $f(\hat{\lambda}, n) = f(\hat{\lambda}|n)b(n)$, the product of the conditional distribution of $\hat{\lambda}$ given n, and the marginal distribution (binomial) $b(n)$ of n. Let

$$E_{\hat{\lambda},n}\left(\frac{n}{g(\hat{\lambda})}\right) = \sum_n \int_{\hat{\lambda}} \frac{n}{g(\hat{\lambda})} f(\hat{\lambda}|n)\mathrm{d}\hat{\lambda}\; b(n), \tag{9}$$

$$E_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right) = \int_{\hat{\lambda}} \frac{n}{g(\hat{\lambda})} f(\hat{\lambda}|n)\mathrm{d}\hat{\lambda} \tag{10}$$

denote the *unconditional mean* (9), the mean with respect to the joint distribution, and the *conditional mean* (10). We are interested in computing the variance with respect to the joint distribution $f(\hat{\lambda}|n)b(n)$ and will use the following lemma (see [10, (p. 129)]).

Lemma. *Let X and Y be two random variables. Then,*

$$\mathrm{Var}(X) = E[\mathrm{Var}(X|Y)] + \mathrm{Var}(E[X|Y]). \tag{11}$$

Now, choose $X = \frac{n}{g(\hat{\lambda})}$ and $Y = n$. Then,

$$\mathrm{Var}_{\hat{\lambda},n}\left(\frac{n}{g(\hat{\lambda})}\right) = \sum_n \left\{\int_{\hat{\lambda}} \left(\frac{n}{g(\hat{\lambda})} - E_{\hat{\lambda},n}\left(\frac{n}{g(\hat{\lambda})}\right)\right)^2 f(\hat{\lambda}|n)\mathrm{d}\hat{\lambda}\right\} b(n) \tag{12}$$

$$= \sum_n \mathrm{Var}_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right) b(n) \quad (= E[\mathrm{Var}(X|Y)])$$

$$+ \sum_n \left(E_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right) - E_{\hat{\lambda},n}\left(\frac{n}{g(\hat{\lambda})}\right)\right)^2 b(n)$$

$$(= \mathrm{Var}(E[X|Y])). \tag{13}$$

In conclusion, we may summarize the finding from (13) as the following

Theorem.

$$\mathrm{Var}_{\hat{\lambda},n}\left(\frac{n}{g(\hat{\lambda})}\right) = \mathrm{Var}_n\left\{E_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right)\right\} + E_n\left\{\mathrm{Var}_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right)\right\}, \tag{14}$$

where E_n and Var_n refer to the marginal distribution $b(n)$ of n.

Clearly, the two terms in (14) reflect the *two sources* of random variation. The first term arises from the (binomial) random variation involved in sampling the n units from the population of size N with probability $g(\lambda)$ each, the second term stands for the random variation arising from estimating λ on the basis of the observed n units.

2.1. $\mathrm{Var}_n\{E_{\hat{\lambda}|n}(\frac{n}{g(\hat{\lambda})})\}$

If we assume that $E_{\hat{\lambda}|n}(\frac{n}{g(\hat{\lambda})})$ can be estimated by $\frac{n}{g(\lambda)}$, then the variance becomes

$$\mathrm{Var}_n\left\{E_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right)\right\} \approx \mathrm{Var}_n\left(\frac{n}{g(\lambda)}\right) \tag{15}$$

$$= N\frac{g(\lambda)(1-g(\lambda))}{g(\lambda)^2}, \tag{16}$$

so that a variance estimate of this term can be reached by replacing λ by $\hat{\lambda}$ and $Ng(\lambda)$ by n leading to

$$\widehat{\mathrm{Var}}_n\left\{E_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right)\right\} \approx n\frac{(1-g(\hat{\lambda}))}{g(\hat{\lambda})^2}. \tag{17}$$

Note that the approximation $E_{\hat{\lambda}|n}(\frac{n}{g(\hat{\lambda})}) \approx \frac{n}{g(\lambda)}$ can be justified by the δ-method which is applied here to the first moment and states that the expected value of the transformed random variable can be approximated by the transformation of the expected value (see [1 (p. 493),3 (p. 240)]).

2.2. $E_n\{\mathrm{Var}_{\hat{\lambda}|n}(\frac{n}{g(\hat{\lambda})})\}$

We let λ be a vector of parameters and consider the term $n^2\mathrm{Var}_{\hat{\lambda}|n}(\frac{1}{g(\hat{\lambda})})$ further and will use the multivariate δ-method ([1 (p. 493),3 (p. 240)]) to achieve that

$$\mathrm{Var}_{\hat{\lambda}|n}\left(\frac{1}{g(\hat{\lambda})}\right) \approx \left(\frac{1}{g(\lambda)^2}\right)^2 \nabla g(\lambda)^T \mathbf{Cov}_{\hat{\lambda}|n}(\hat{\lambda}) \nabla g(\lambda). \tag{18}$$

Replacing $\mathbf{Cov}_{\hat{\lambda}|n}$ with an estimate $\widehat{\mathbf{Cov}}_{\hat{\lambda}|n}$, then a final estimate of $E_n\{\mathrm{Var}_{\hat{\lambda}|n}(\frac{n}{g(\hat{\lambda})})\}$ can again be achieved as

$$\left(\frac{n}{g(\hat{\lambda})^2}\right)^2 \nabla g(\hat{\lambda})^T \mathbf{Cov}_{\hat{\lambda}|n}(\hat{\lambda}) \nabla g(\hat{\lambda}). \tag{19}$$

If λ is a scalar, (19) simplifies to

$$n^2 \left(\frac{g'(\hat{\lambda})}{g(\hat{\lambda})^2}\right)^2 \widehat{\mathrm{Var}}_{\hat{\lambda}|n}(\hat{\lambda}). \tag{20}$$

3. Variance of $\hat{N}$ based upon a CRECT

Wilson and Collins [15] compare a number of estimators including Zelterman's estimator and the lower bound estimator of Chao [4,5]. We consider them here in more detail, since both do not require the assumption that the number of recapture occasions is fixed and known. Therefore, they are both suitable for a CRECT.

3.1. Zelterman's estimator

Let us come back to the estimators suggested in Section 1.1 for a CRECT. We consider $\hat{\lambda}_i = (i+1)f_{i+1}/f_i$, in particular $\hat{\lambda}_1 = 2f_2/f_1$ for $i = 1$, as suggested by Zelterman [16]. It is now straightforward to provide a variance estimate for $\hat{\lambda} = \hat{\lambda}_i$. According to (20) and with $g(x) = 1 - \exp(x)$

$$E_n\left\{\mathrm{Var}_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right)\right\} \approx n^2\left(\frac{g'(\hat{\lambda})}{g(\hat{\lambda})^2}\right)^2 \widehat{\mathrm{Var}}_{\hat{\lambda}|n}(\hat{\lambda}) = n^2\left(\frac{\exp(-\hat{\lambda})}{\left(1-\exp(-\hat{\lambda})\right)^2}\right)^2 \widehat{\mathrm{Var}}_{\hat{\lambda}|n}(\hat{\lambda})$$

so that only $\widehat{\mathrm{Var}}_{\hat{\lambda}|n}(\hat{\lambda})$ is left to be evaluated. Two routes are possible. Assuming a truncated Poisson model in which all counts are truncated except counts i and $i+1$, we achieve the truncated probability for count i as

$$\frac{\exp(-\lambda)\lambda^i/i!}{\exp(-\lambda)\lambda^i/i! + \exp(-\lambda)\lambda^{i+1}/(i+1)!} = \frac{1}{1+\lambda/(i+1)}$$

and similarly for count $(i+1)$ as $\frac{\lambda/(i+1)}{1+\lambda/(i+1)}$, so that the (truncated) log-likelihood is provided as

$$f_{i+1}\log(\lambda) - (f_i + f_{i+1})\log(1+\lambda/(i+1))$$

with maximum likelihood estimate $\hat{\lambda} = (i+1)f_{i+1}/f_i$ and second derivative $-f_{i+1}/\lambda^2 + (f_i + f_{i+1})/(i+1+\lambda)^2$, from where the variance estimate as inverse of the observed Fisher information $f_{i+1}/\lambda^2 - (f_i + f_{i+1})/(i+1+\lambda)^2$ evaluated at $\lambda = \hat{\lambda}$

$$\widehat{\mathrm{Var}}_{\hat{\lambda}|n}(\hat{\lambda}) = \frac{(i+1)^2(f_i + f_{i+1})f_{i+1}}{f_i^3} = \hat{\lambda}_i^2\left(\frac{1}{f_i} + \frac{1}{f_{i+1}}\right) \tag{21}$$

is obtained. This variance estimate implicitly assumes the validity of the Poisson distribution which the Zelterman estimate allows to be violated. Therefore it seems wise to consider a distribution-free estimate of the variance. Alternatively, we might use the bivariate δ-method to achieve that

$$\widehat{\mathrm{Var}}_{\hat{\lambda}|n}((i+1)f_{i+1}/f_i) = \nabla g(f_i, f_{i+1})^T \mathbf{Cov}\begin{pmatrix} f_i \\ f_{i+1}\end{pmatrix} \nabla g(f_i, f_{i+1})$$

with $g(x, y) = y/x$. The conventional estimator of the covariance matrix of the multinomial provides the covariance matrix for $\begin{pmatrix} f_i \\ f_{i+1}\end{pmatrix}$ as

$$\widehat{\mathbf{Cov}}\begin{pmatrix} f_i \\ f_{i+1}\end{pmatrix} = \begin{pmatrix} f_i(1-f_i/n) & -f_i f_{i+1}/n \\ -f_i f_{i+1}/n & f_{i+1}(1-f_{i+1}/n)\end{pmatrix}.$$

This gives

$$\widehat{\mathrm{Var}}_{\hat{\lambda}|n}((i+1)f_{i+1}/f_i) = (i+1)^2\frac{f_{i+1}^2}{f_i^2}\left[\frac{1-f_i/n}{f_i} + 2/n + \frac{1-f_{i+1}/n}{f_{i+1}}\right] \tag{22}$$

which simplifies further to

$$\widehat{\mathrm{Var}}_{\hat{\lambda}|n}((i+1)f_{i+1}/f_i) = (i+1)^2\frac{f_{i+1}^2}{f_i^2}\left[\frac{1}{f_i} + \frac{1}{f_{i+1}}\right]. \tag{23}$$

Note that (23) is identical to (21) which is a remarkable result: Both routes (assuming a Poisson likelihood versus the nonparametric multinomial) lead to the same variance estimator. We put all the terms together:

Corollary 1. *Consider the Zelterman's estimator $\hat{\lambda} = (i+1)(f_{i+1}/f_i)$. Then, an (unconditional) variance estimator is provided as:*

$$\widehat{\mathrm{Var}}_{\hat{\lambda},n}\left(\frac{n}{g(\hat{\lambda})}\right) = nG(\hat{\lambda})\left[1 + nG(\hat{\lambda})\hat{\lambda}^2\left(\frac{1}{f_i} + \frac{1}{f_{i+1}}\right)\right], \tag{24}$$

where $G(\hat{\lambda}) = \frac{\exp(-\hat{\lambda})}{(1-\exp(-\hat{\lambda}))^2}$.

Zelterman [16] provides also a variance estimator which ignores the first term in (14). In addition, Wilson and Collins [15, (p. 549)] point out an error in the variance computation:

> This is what happened with Zelterman's [16] analysis: his Tables 1 and 2 have some anomalous values caused by the mean squared errors he calculates being close to zero, when in fact it is only the highest order term which vanishes.

We will come back to this point and compare the variance estimator (24) with the one given by Zelterman [16].

3.2. Chao's estimator under heterogeneity

Chao [4] suggested the estimator $\hat{N} = n + f_1^2/(2f_2)$ based on a mixed Poisson model with $p_i = \int_t \mathrm{e}^{-t} t^i / i! \lambda(t) \mathrm{d}t$ for $i = 0, 1, 2, \ldots$ and arbitrary density $\lambda(t)$. Then, using the inequality of Cauchy–Schwartz, Chao [4] arrived at $2p_0 p_2 \le p_1^2$ from where the estimator follows. The important result here is that the Chao's estimator provides a lower bound for the population size independent of the form of the heterogeneity distribution $\lambda(t)$. Using the theorem of Section 2 we have that

$$\mathrm{Var}_{\hat{\lambda}_0, n}(n + \hat{\lambda}_0) = E_n \left\{ \mathrm{Var}_{\hat{\lambda}_0 | n}(n + \hat{\lambda}_0) \right\} + \mathrm{Var}_n \left\{ E_{\hat{\lambda}_0 | n}(n + \hat{\lambda}_0) \right\}, \tag{25}$$

where E_n and Var_n refer to the marginal distribution $b(n)$ of n and $\hat{\lambda}_0 = \hat{f}_0 = f_1^2/(2f_2)$. Assuming that $E_{\hat{\lambda}_0|n}(n + \hat{\lambda}_0)$ in the second term in (25) can be estimated by $n + \hat{\lambda}_0$, we have that

$$\mathrm{Var}_n(n + \hat{\lambda}_0) = \mathrm{Var}_n(n) = N p_0 (1 - p_0).$$

Since $E(f_0) = N p_0$ and $1 - p_0 = 1 - E(f_0)/N$, we can estimate this variance as

$$\widehat{\mathrm{Var}}_n(n) = \frac{f_1^2}{2f_2} \left(1 - \frac{f_1^2}{2f_2 \hat{N}} \right) = \frac{f_1^2}{2f_2} \left(1 - \frac{f_1^2}{2f_2 n + f_1^2} \right).$$

Assume again that $E_n\{\mathrm{Var}_{\hat{\lambda}_0|n}(n + \hat{\lambda}_0)\}$ can be estimated by $\mathrm{Var}_{\hat{\lambda}_0|n}(n + \hat{\lambda}_0) = \mathrm{Var}_{\hat{\lambda}_0|n}(\frac{f_1^2}{2f_2})$. Using the δ-method similar to the way it was used in the previous section we arrive at

$$\widehat{\mathrm{Var}}_{\hat{\lambda}_0|n} \left(\frac{f_1^2}{2f_2} \right) = \frac{f_1^3}{f_2^2} \left(1 + \frac{1}{4} \frac{f_1}{f_2} (1 - f_2/n) \right).$$

Adding both variances we obtain

Corollary 2. *Consider the estimator $\hat{N} = n + f_1^2/(2f_2)$. Then:*

$$\widehat{\mathrm{Var}}_{\hat{\lambda}_0, n} \left(n + \frac{f_1^2}{2f_2} \right) = \frac{1}{2} \frac{f_1^2}{f_2} \left(1 - \frac{f_1^2}{2f_2 n + f_1^2} \right) + \frac{f_1^3}{f_2^2} \left(1 + \frac{1}{4} \frac{f_1}{f_2} (1 - f_2/n) \right) \tag{26}$$

$$= \frac{1}{4} \frac{f_1^4}{f_2^3} + \frac{f_1^3}{f_2^2} + \frac{1}{2} \frac{f_1^2}{f_2} - \frac{1}{4} \frac{f_1^4}{(f_2^2 n)} - \frac{1}{2} \frac{f_1^4}{f_2(2f_2 n + f_1^2)}. \tag{27}$$

Note that the first three terms in (27) correspond to the variance estimate given in Chao [4].

Example 1 (*Continued*). Let us illustrate the variance formula at the spinner dolphin data set used in Section 2.1. Zelterman's estimator provides 180 dolphins with a wide confidence interval (Table 2). Chao's estimator gives a value of 177 dolphins with a smaller (but still wide) confidence interval.

Table 1
Observed frequencies of repeated counts of identifications for $n = 51$ spinner dolphin's in the Moorea island community

f_0	f_1	f_2	f_3
–	42	7	2

Table 2
Zelterman's and Chao's estimates of the population size of spinner dolphin's in the Moorea island community

Estimator	$\hat{N}$	$\widehat{SE}$	95% CI
Zelterman	180	65.45	52–308
Chao	177	59.20	61–293

3.3. A simulation study

To investigate the variance estimates in more detail the following simulation was done. $N = 100$ counts y were sampled from a two-component mixture $q\text{Po}(y \mid \lambda_1) + (1-q)\text{Po}(y \mid \lambda_2)$ with $0 < q < 1$. Using the zero-truncated counts only, the population size N was estimated with the Zelterman's estimator as well as Chao's estimator. True variances were estimated and compared with the estimated variances according to (23) and (27). Four scenarios were considered (for all four $\lambda_1 = 0.5$): experiment 1 used $q = 0.5$ and $\lambda_2 = 5$, experiment 2 used $q = 0.9$ and $\lambda_2 = 5$, experiment 3 used $q = 0.5$ and $\lambda_2 = 1$, and experiment 4 used $q = 0.9$ and $\lambda_2 = 1$.

Zelterman [16] provided the variance approximation $\text{Var}(e^{\hat{\lambda}_1}) \approx \frac{1}{n}\text{e}^{-\lambda}(1 - \text{e}^{-\lambda})(2 + \lambda)$ with $\hat{\lambda}_1 = 2f_2/f_1$ which we use to provide a corrected version of the variance of the population size estimate based upon the Zelterman estimate:

$$\widehat{\text{Var}}(\hat{N}) = \widehat{\text{Var}}\left(\frac{n}{g(\hat{\lambda}_1)}\right) = nG(\hat{\lambda}_1)\left(1 + \frac{\hat{\lambda}_1 + 2}{1 - \text{e}^{-\hat{\lambda}_1}}\right). \tag{28}$$

The simulation results of Table 3 indicate that Zelterman's estimator and the estimator of Chao are quite close with that of Zelterman always being larger than the estimator of Chao. The simulation study also supports the *bona-fide* knowledge that the variance of Zelterman's estimator is rather large, at least larger than Chao's estimator. Wilson and Collins [15] write:

> For although it often does have a smaller bias than the other estimators (*used by Wilson and Collins in their comparison*), it does so at the cost of having a larger standard deviation which overwhelms the reduced bias.

Most importantly, it can be seen that both variance formulas provide reasonable approximations of the true variances with the approximating formula providing conservative approximations of the true variances. In particular, it appears that the variance formula for Chao's estimator (27) gives a slightly better approximation than Chao's variance formula. The variance approximation (23) is conservative, whereas the one in (28) appears to be very good, but underestimates occasionally drastically. A more detailed analysis explains when this can happen. Let us consider the two variance approximations for $n/\hat{N} = 1 - \exp(-\hat{\lambda})$ with $\hat{\lambda} = \hat{\lambda}_1 = 2f_2/f_1$:

$$n \times \text{Var}(n/\hat{N}) \approx \exp(-\lambda)^2\lambda^2\,(2 + \lambda/2 + 2/\lambda) \tag{29}$$

Table 3
Results of a simulation study comparing Zelterman's estimator with Chao's estimator under Poisson model violations; true N is 100

Experiment	Zelterman			
	Mean	True SE	SE (23)	Zelterman's SE
1	128	58	67	40
2	124	68	78	69
3	101	27	29	28
4	109	46	51	51
	Chao			
Experiment	Mean	True SE	SE (27)	Chao's SE
1	91	20	22	22
2	109	52	58	60
3	100	25	25	26
4	108	42	46	48

Table 4
Comparing the two variance estimators: $\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))(2+\hat{\lambda})/n$ and $\exp(-\hat{\lambda})^2\hat{\lambda}^2\,(1/f_1+1/f_2)$

Experiment	True	$\exp(-\hat{\lambda})(1-\exp(-\hat{\lambda}))(2+\hat{\lambda})$	$\exp(-\hat{\lambda})^2\hat{\lambda}^2\,(1/f_1+1/f_2)$
1	0.02282	0.00923	0.02321
2	0.01634	0.01336	0.01628
3	0.01347	0.01293	0.01295
4	0.01466	0.01444	0.01437

$$n \times \mathrm{Var}(n/\hat{N}) \approx \exp(-\lambda)(1-\exp(-\lambda))\,(2+\lambda) \tag{30}$$

where the second approximation (30) goes back to Zelterman [16]. Consider Fig. 1: clearly, both approximations are close if λ is small whereas they differ considerably for larger values of λ. A simulation study again helps us to illustrate that using $\widehat{\mathrm{Var}}(n/\hat{N}) = \exp(-\hat{\lambda})^2\hat{\lambda}^2\,(1/f_1+1/f_2)$ should be preferred. The results of the simulation study for which we have used the same parameter constellations as used previously are shown in Table 4.

Most of the differences occur when the Poisson assumption is strongly violated (Experiments 1 and 2), whereas both variance approximations are close (and close to the true variance) when the Poisson assumption is mildly violated (Experiments 3 and 4). The suggested variance estimate $\exp(-\hat{\lambda})^2\hat{\lambda}^2\,(1/f_1+1/f_2)$ appears to provide a quite reasonable approximation to the true variance.

4. Some consequences

4.1. The implication of using resampling procedures

Frequently, resampling procedures such as the Bootstrap are used to estimate the variance of $\hat{N}$ [14,2]. This is done in the following way: from the original sample the estimator $\hat{N} = n/g(\hat{\lambda})$ is constructed. Then, binomial resamples are generated using the size parameter $\hat{N}$ and event parameter $g(\hat{\lambda})$. This will lead to B samples of sizes $n^{(1)}, n^{(2)}, \ldots, n^{(B)}$ from which estimates $\hat{\lambda}^{(b)}$ and $\hat{N}^{(b)} = n^{(b)}/g(\hat{\lambda}^{(b)})$ for $b = 1, \ldots, B$ can be constructed and their sample variance $\frac{1}{B}\sum_b(\hat{N}^{(b)} - \overline{\hat{N}})^2$ provides the variance estimate of $\hat{N}$. This procedure can be improved upon in terms of a technical simplification.

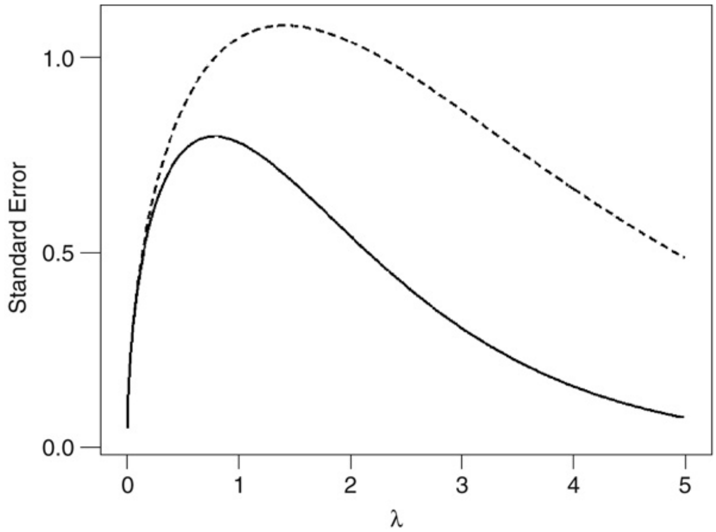

Fig. 1. The two standard errors according to (29) and (30) as functions of λ.

Table 5
Bootstrap mean and variance of Zelterman's and Chao's population size estimates of spinner dolphin's in the Moorea Island community

Estimator	Mean		Variance			SE
	$\hat{N}$	$E^*(\hat{N})$	$\hat{I}$	$\widehat{\text{II}}^*$	$\hat{I}+\widehat{\text{II}}^*$	$\sqrt{\hat{I}+\widehat{\text{II}}^*}$
Zelterman	180	183.52	455	1599	2054	45
Chao	177	177.75	36	1433	1469	38

Let us consider again the variance decomposition

$$\text{Var}_{\hat{\lambda},n}\left(\frac{n}{g(\hat{\lambda})}\right) = \text{Var}_n\left\{E_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right)\right\} + E_n\left\{\text{Var}_{\hat{\lambda}|n}\left(\frac{n}{g(\hat{\lambda})}\right)\right\} = \text{I} + \text{II},$$

where the first part I (binomial variance) can always be estimated by $n\frac{(1-g(\hat{\lambda}))}{g(\hat{\lambda})^2}$, for which only an estimate $\hat{\lambda}$ of λ is required. To illustrate we find for Zelterman's estimate $\hat{\lambda} = 2f_2/f_1$ that $\hat{I} = n\frac{(1-g(\hat{\lambda}))}{g(\hat{\lambda})^2} = n\frac{\exp(-\hat{\lambda})}{[1-\exp(\hat{\lambda})]^2}$ (see Section 3.1), and for Chao's lower bound estimate $f_1^2/(2f_2)$ that the corresponding part I can be estimated by $\hat{I} = \frac{f_1^2}{2f_2}(1 - \frac{f_1^2}{2f_2n+f_1^2})$ (see Section 3.2).

Now, the estimate $n^2\text{Var}_{\hat{\lambda}|n}(\frac{1}{g(\hat{\lambda})})$ of the second part is *conditional* upon n, so that a conventional, nonparametric Bootstrap might be used by sampling n units with replacement and calculating the variance of B such samples. This procedure will provide a more efficient estimation of the variance. An application of this modified Bootstrap procedure to Spinner Dolphin's data is provided in Table 5.

4.2. *A weighted estimator*

As was mentioned before, the Zelterman's estimator is known to have a small bias but large variance. Chao's estimator is a lower bound estimator (thus negatively biased) with small variance. In addition, reasonable variance approximations were provided for both estimators.

This suggests the need to combine the positive aspects of both estimators by construction of a weighted estimator

$$\hat{N}_W = (w_1\hat{N}_Z + w_2\hat{N}_C)/(w_1 + w_2), \tag{31}$$

where $\hat{N}_C$ and $\hat{N}_Z$ are the estimators of Chao and Zelterman, respectively. The weights could be chosen according to the inverse variances. However, the true variances are unknown and replacing them by estimates will introduce additional variation and loss in efficiency can be expected. Instead we will use equal weights and motivate this choice as follows. Consider the Zelterman's estimator $\hat{N}_Z = n + \frac{n}{\exp(2f_2/f_1)-1}$ and the Taylor-series approximation of the exponential around zero $e^x \approx 1 + x + x^2/2$, so that

$$\hat{N}_Z = n + \frac{n}{\exp(2f_2/f_1) - 1} \approx n + \frac{n}{2f_2/f_1 + 2f_2^2/f_1^2} \tag{32}$$

$$= n + \frac{f_1^2}{2f_2}\frac{n}{f_1 + f_2} \tag{33}$$

which provides a representation of the Zelterman's estimator in terms of the Chao's estimator since $\hat{N}_C = n + \frac{f_1^2}{2f_2}$. Note that (33) also provides an explanation for the fact that the Zelterman's estimator is often larger than Chao's estimator since $\frac{n}{f_1+f_2} \geq 1$. Let us define

$$\hat{N}_W = \frac{1}{2}(\hat{N}_Z + \hat{N}_C) = n + \frac{f_1^2}{2f_2}\left(\frac{1}{2} + \frac{n}{2(f_1 + f_2)}\right) \tag{34}$$

which has two advantages. For one, we can incorporate the low bias of the Zelterman's estimator in a weighted estimator which turns out to be an inflated Chao's estimator. Hence, we can retain to a certain degree the lower variance of the Chao's estimator. Secondly, the inflation representation allows a closed form expression for the variance of $\hat{N}_W$. Assuming again that the second term in (25) can be estimated by $Np_0(1 - p_0)$, where $1 - p_0$ can be further estimated by $n/\hat{N}_W$, we arrive at

$$\widehat{\text{II}} = n(\hat{N}_W - n)/\hat{N}_W.$$

To evaluate the first term in (25) we assume once more that $E_n\{\text{Var}_{\hat{\lambda}_0|n}(n + \hat{\lambda}_0)\}$ for $\hat{\lambda}_0 = \frac{f_1^2}{2f_2}(\frac{1}{2} + \frac{n}{2(f_1+f_2)})$ can be estimated by

$$\text{Var}_{\hat{\lambda}_0|n}(n + \hat{\lambda}_0) = \left(\frac{1}{2} + \frac{n}{2(f_1 + f_2)}\right)^2 \text{Var}_{\hat{\lambda}_0|n}\left(\frac{f_1^2}{2f_2}\right).$$

Using the δ-method for $\text{Var}_{\hat{\lambda}_0|n}(\frac{f_1^2}{2f_2})$ similar to the way it was used in the previous sections we arrive at

$$\widehat{\text{Var}}_{\hat{\lambda}_0,n}(\hat{N}_W) = \frac{n}{\hat{N}_W}(\hat{N}_W - n) + \left(\frac{1}{2} + \frac{n}{2(f_1 + f_2)}\right)^2 \frac{f_1^3}{f_2^2}\left(1 + \frac{f_1}{4f_2}\left(1 - \frac{f_2}{n}\right)\right). \tag{35}$$

To illustrate the beneficial behavior of the new, weighted estimator it was included in the previously described simulation experiment. The results are provided in Table 6. Though Chao's estimator is well-behaved in most situations, it underestimates strongly occasionally such as in

Table 6
Results of a simulation study comparing Zelterman's, Chao's estimator and the weighted estimator under Poisson model violations; true N is 100

Exp.	Zelterman			Chao			Weighted		
	Mean	True SE	SE (23)	Mean	True SE	SE (27)	Mean	True SE	SE (35)
1	124	39	46	91	15	16	107	27	32
2	117	48	54	103	38	40	110	43	48
3	101	28	30	99	25	25	100	27	28
4	108	41	45	107	39	41	107	40	44

Experiment 1, whereas the new weighted estimator provides a better balance between bias and variance.

4.3. Size of a CRECT

Frequently, in a CRECT the lower confidence limit of the population size estimator is close or even below the observed number of units (see Table 5). This raises questions of designing the CRECT with an appropriate sample size. It is quite clear by considering (4), however, that the variance of the population size estimator increases with the sample size. Therefore, a relative measure such as the proportion $n/\hat{N}$ of completeness of identification should be considered. Let us consider this measure in connection with Zelterman's estimator. We have that $n/\hat{N} = 1 - \exp(-\hat{\lambda})$ and the variance can be simply estimated as

$$\widehat{\mathrm{Var}}(n/\hat{N})^{1/2} = \exp(-\hat{\lambda})\hat{\lambda}\sqrt{\frac{1}{f_i} + \frac{1}{f_{i+1}}}$$

and $n\widehat{\mathrm{Var}}(n/\hat{N})$ converges to $\exp(-\lambda)^2\lambda^2\left(\frac{1}{p_i} + \frac{1}{p_{i+1}}\right)$. Thus, confidence intervals for this measure can be made arbitrarily small by increasing n. For the spinner dolphin data we find $n/\hat{N} = 51/180 = 0.28$ with $\widehat{\mathrm{Var}}(n/\hat{N}) = 0.0095$ and an approximate 95% CI of (0.0883–0.4783) for the capture probability of this particular CRECT. Clearly, this confidence interval is rather large. How many dolphins need to be captured to achieve a desired margin of error? This is answered as follows. For achieving a $(1-\alpha)100\%$ confidence interval of length ϵ one needs to solve the equation

$$2 \times z\widehat{\mathrm{Var}}(n/\hat{N})^{1/2} = (2z\exp(-\lambda)\lambda)\sqrt{2 + \lambda/2 + 2/\lambda} = \epsilon \tag{36}$$

for n leading to

$$n = (2z\exp(-\lambda)\lambda)^2(2 + \lambda/2 + 2/\lambda)/\epsilon^2$$

with $z = \Phi^{-1}(1-\alpha/2)$ and Φ being the standard normal distribution function. For the spinner dolphin data, using $\hat{\lambda} = 0.28$, we have that $\epsilon = 0.1, 0.2, 0.3, 0.4$ lead to the associated sizes $n = 715, 180, 80, 45$. If no prior knowledge on the value for λ is available, then an upper bound for the standard error (see also Fig. 2)

$$\exp(-\lambda)\lambda\sqrt{2 + \lambda/2 + 2/\lambda} \leq 0.8 \tag{37}$$

might be used, leading to required sample sizes of 983, 246, 109, 61 associated with $\epsilon = 0.1, 0.2, 0.3, 0.4$, respectively.

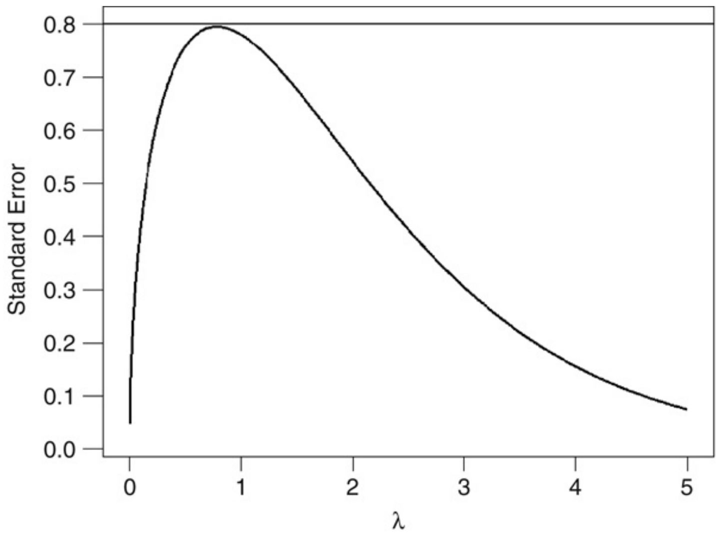

Fig. 2. Standard error $\exp(-\lambda)\lambda\sqrt{2+\lambda/2+2/\lambda}$ as a function of λ.

Acknowledgments

The author is indebted to Professor John Whitehead (Medical & Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Fylde College Lancaster, University Lancaster) for a critical reading and discussion of this paper. The author also thanks an Associate Editor for helpful comments leading to the substantial improvement of this paper.

References

[1] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge, 1975.
[2] D. Böhning, E. Dietz, R. Kuhnert, D. Schön, Mixture models for capture–recapture count data, Statistical Methods & Applications, Journal of the Italian Statistical Society 14 (2005) 29–43.
[3] G. Casella, R.L. Berger, Statistical Inference, 2nd ed., Duxbury Press, Belmont, CA, 2002.
[4] A. Chao, Estimating the population size for capture–recapture data with unequal capture probabilities, Biometrics 43 (1987) 783–791.
[5] A. Chao, Estimating population size for sparse data in capture–recapture experiments, Biometrics 45 (1989) 427–438.
[6] G. Hay, F. Smit, Estimating the number of drug injectors from needle exchange data, Addiction Research and Theory 11 (2003) 235–243.
[7] M.G. Kendall, A. Stuart, The Advanced Theory of Statistics, 3rd ed., vol. 3, Griffin, London, 1976.
[8] M. Oremus, Personal communication, 2005.
[9] J.M. Roberts, D.D. Brewer, Estimating the prevalence of male clients of prostitute women in Vancouver with a simple capture–recapture method, Journal of the Royal Statistical Society A 169 (2006) 745–756.
[10] S.M. Ross, Introduction to Probability Models, 3rd ed., Academic Press, Orlando, 1985.
[11] L. Sanathanan, Estimating the size of a multinomial population, Journal of the American Statistical Association 72 (1972) 669–672.
[12] L. Sanathanan, Estimating the size of a truncated sample, The Annals of Mathematical Statistics 43 (1977) 142–152.
[13] G.A.F. Seber, The Estimation of Animal Abundance and Related Parameters, 2nd ed., The Blackburn Press, Caldwell, 2002.
[14] P.G.M. van der Heijden, R. Bustami, M. Cruy, G. Engbersen, H.C. van Houwelingen, Point and interval estimation of the population size using the truncated Poisson regression model, Statistical Modelling — An International Journal 3 (2003) 305–322.
[15] R.M. Wilson, M.F. Collins, Capture-recapture estimation with samples of size one using frequency data, Biometrika 79 (1992) 543–553.
[16] D. Zelterman, Robust estimation in truncated discrete distributions with applications to capture–recapture experiments, Journal of Statistical Planning and Inference 18 (1988) 225–237.