Scandinavian Journal of Statistics, Vol. 37: 221–236, 2010
doi: 10.1111/j.1467-9469.2009.00676.x


# Some General Comparative Points on Chao's and Zelterman's Estimators of the Population Size

DANKMAR BÖHNING

*Applied Statistics, School of Biological Sciences, University of Reading*

**ABSTRACT. Two simple and frequently used capture–recapture estimates of the population size are compared: Chao's lower-bound estimate and Zelterman's estimate allowing for contaminated distributions. In the Poisson case it is shown that if there are only counts of ones and twos, the estimator of Zelterman is always bounded above by Chao's estimator. If counts larger than two exist, the estimator of Zelterman is becoming larger than that of Chao's, if only the ratio of the frequencies of counts of twos and ones is small enough. A similar analysis is provided for the binomial case. For a two-component mixture of Poisson distributions the asymptotic bias of both estimators is derived and it is shown that the Zelterman estimator can experience large overestimation bias. A modified Zelterman estimator is suggested and also the bias-corrected version of Chao's estimator is considered. All four estimators are compared in a simulation study.**

*Key words:* capture–recapture approach, Chao's lower-bound estimator, modified Zelterman estimator, Zelterman's estimator

## 1. Introduction

The size $N$ of an elusive population of interest must often be determined. Elusive populations occur, for example, in public health and medicine, agriculture and veterinary science, software engineering, illegal behaviour research, in the ecological sciences and in many other fields (Bishop *et al.*, 1975; Wilson & Collins, 1992; Bunge & Fitzpatrick, 1993; Pledger, 2000, 2005; Chao *et al.*, 2001; Hay & Smit, 2003; Roberts & Brewer, 2006). All of these situations fall under the following setting. We assume that the population is closed (no birth, death or migration), and that there is an endogenous mechanism such as a register, a diagnostic device, a set of reviewers or a trapping system, which identifies independently $n$ units from the population of size $N$ in a fixed observational period. Frequently, the identifying system produces a count $x_i > 0$ of how often the unit $i$ has been identified, where $i = 1, \ldots, n$, and $x_i = 0$ remains unobserved by the system for $i = n+1, \ldots, N$. Hence we can think of our sample $x_1, \ldots, x_n, x_{n+1}, .., x_N$ as consisting of the observed, zero-truncated part $x_1, \ldots, x_n$ and an unobserved part of unknown size $N - n$ consisting only of zero counts. Interest is in producing an estimate of $N$ on the basis of the available sample $x_1, \ldots, x_n$ and some observational model $p(x \mid \lambda)$ for count $X = x$. A typical example for $p(x \mid \lambda)$ is the Poisson $p(x \mid \lambda) = \mathrm{Po}(x \mid \lambda) = \exp(-\lambda)\lambda^x/x!$ or the binomial. However, it is more realistic to assume population heterogeneity for the distribution $p(x \mid \lambda)$ of counts of identifications $X \in \{0, 1, 2, \ldots, \}$, which allows a distribution on the parameter $\lambda$ in the population of interest:

$$p_x = \int_0^\infty p(x \mid \lambda) f(\lambda)\, \mathrm{d}\lambda,$$

with unspecified mixing density $f(\lambda)$ and a mixture kernel $p(x \mid \lambda)$ which needs to be specified, often it is the Poisson or binomial. Clearly, if $f(\lambda)$ were known then also $p_0 = \int_0^\infty p(0 \mid \lambda) f(\lambda)\, \mathrm{d}\lambda$

were known, and $N$ could be estimated by means of the Horvitz–Thompson estimator $\hat{N} = n/(1-p_0)$. However, $f(\lambda)$ is typically unknown and an estimate of $p_0$ would require an estimate of $f(\lambda)$ (see also Bunge & Fitzpatrick, 1993; Pledger, 2005).

The article considers two popular estimators of an unknown population size $N$ which allow population heterogeneity but avoid producing an estimate for the mixing distribution $f(\lambda)$: the estimators of Chao (1987) and Zelterman (1988). Both estimators seem to experience different priorities in different communities. The lower-bound estimator of Chao is well known in the fields of biology and ecology. In contrast, Zelterman's estimator is used frequently in the social sciences, in particular in illicit drug use research (see, e.g. Hay & Smit, 2003; Roberts & Brewer, 2006). Little is known on how the two approaches are connected and how these two estimators and their available modifications compare, the point precisely being addressed in this article.

The importance of the mixture $p_x = \int_0^\infty p(x \mid \lambda) f(\lambda)\,\mathrm{d}\lambda$ can be seen in the fact that it is a natural model for modelling population heterogeneity. There appears to be consensus (see, e.g. Pledger, 2005, for the discrete mixture model approach and Dorazio & Royle, 2005, for the continuous mixture model approach) that a simple model $p(x \mid \lambda)$ is not flexible enough to capture the variation in the recapture probability for the different members of most real-life populations. Every item might be different, as might be every animal or human being. However, there has also been a recent debate on the identifiability of the binomial mixture model (see Link, 2003, 2006; Holzmann *et al.*, 2006). Hence, a renewed interest has re-occurred in the lower bound approach for population size estimation suggested by Chao (1987). In the lower bound approach there is neither an identifiability problem, nor is there need to *specify or estimate* a mixing distribution. In this sense it is completely non-parametric.

To give some details of the lower bound approach consider the Poisson mixture kernel $\mathrm{Po}(x \mid \lambda) = \exp(-\lambda)\lambda^x/x!$. It follows from the Cauchy–Schwarz inequality that

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

or equivalently, $p_1^2 \le p_0(2p_2)$, from where the Chao's lower-bound estimate $f_1^2/(2f_2)$ for $f_0$ follows (see Chao, 1984, 1987, 1989). Here, $f_x$ denotes the frequency of count $x \in \{0, 1, \ldots, m\}$, where $m$ is the largest count observed in the sample. The estimate for the population size $N$ is $\hat{N} = n + f_1^2/(2f_2)$. As the Chao's estimator uses only frequencies with counts of 1 and 2, a binomial log-likelihood might be considered such as $f_1 \log(p_1) + f_2 \log(p_2)$ which is uniquely maximized for $\hat{p}_2 = 1 - \hat{p}_1 = f_2/(f_1 + f_2)$. As $p_2 = \lambda/(\lambda+2)$ and $p_1 = 2/(\lambda+2)$ in a Poisson that truncates all counts except ones and twos, the estimate $\hat{\lambda} = 2f_2/f_1$ for the Poisson parameter $\lambda$ suggested by Zelterman (1988) arises. In the approach of Zelterman the homogeneous Poisson serves only as a working model and it was shown by Zelterman that the estimate $\hat{N} = n/(1-\hat{p}_0) = n/[1-\exp(-\hat{\lambda})]$ is more robust against mis-specifications of the Poisson model than the usual maximum likelihood estimate (MLE).

When there is a fixed number $m$ of recapture occasions in the sampling period such as a number of trapping occasions a mixture with a Binomial kernel appears more appropriate than the Poisson kernel. Similarly, using the inequality of Cauchy–Schwarz again we find for the Binomial mixture kernel $\binom{m}{x}\lambda^x(1-\lambda)^{m-x} = \binom{m}{x}(\frac{\lambda}{1-\lambda})^x(1-\lambda)^m$ that

$$\left(\int_0^1 \left(\frac{\lambda}{1-\lambda}\right)(1-\lambda)^m f(\lambda)\,\mathrm{d}\lambda\right)^2 \le \int_0^1 (1-\lambda)^m f(\lambda)\,\mathrm{d}\lambda \int_0^1 \left(\frac{\lambda}{1-\lambda}\right)^2 (1-\lambda)^m f(\lambda)\,\mathrm{d}\lambda,$$

or equivalently, $p_1^2/\binom{m}{1}^2 \leq p_0 p_2/\binom{m}{2}$, or finally, $m(m-1)p_1^2/2 \leq p_0 m^2 p_2$, leading to Chao's lower-bound estimate $[(m-1)/m][f_1^2/(2f_2)]$ of $f_0$. Truncating all counts again except counts of one and two leads to a binomial likelihood which is maximized for $\hat{\lambda} = 2f_2/[2f_2 + f_1(m-1)]$, the Zelterman estimate for the binomial. This leads to the population size estimate $\hat{N} = n/[1-\hat{p}_0] = n/[1-(1-\hat{\lambda})^m]$.

The lower-bound estimator is well known in the fields of biology and ecology. In contrast, Zelterman's estimator is used frequently in the social sciences, in particular in illicit drug use research (see, e.g. Hay & Smit, 2003; Roberts & Brewer, 2006). Little is known about how the two approaches are connected, the point precisely being addressed in this article. It seems to belong to the feuilleton of the statistical literature in capture–recapture analysis (for an example, see Wilson & Collins, 1992) that Zelterman's estimate of the population size is at least as large as the Chao's lower-bound estimate. Although this is intuitively understandable and is confirmed in empirical as well as in simulation studies, a general proof of this result does not exist. We show here why such a general result cannot exist.

The article is organized as follows. In section 2, we consider the Poisson case. We show for $n = f_1 + f_2$ that the estimator of Chao is always larger than the estimator of Zelterman. In addition, we also show that this general occurrence of Zelterman's estimator being smaller than Chao's is basically restricted to when $n = f_1 + f_2$. Furthermore, if $n > f_1 + f_2$ we show that $\hat{N}^Z \geq \hat{N}^C$, if only $\hat{\lambda} = 2f_2/f_1$ is small enough. This explains why we see in empirical studies the Zelterman estimator frequently being larger than Chao's as the ratio $f_2/f_1$ is typically small. In section 3 we provide a similar analysis for the binomial case with small, but notable differences. For $m = 2$, we show that both estimators coincide. For $m > 2$, Zelterman's estimator becomes larger than Chao's if the ratio $(2f_2)/[f_1(m-1)]$ is small enough. In section 4 a two-component mixture of Poisson distribution is studied and the asymptotic biases of both estimators are derived. It is shown that for large contaminations the Zelterman estimator experiences large overestimation bias. Consequently, a modified Zelterman estimator is suggested which uses only counts of ones and twos in estimation and prediction. It is known that for the case of homogeneity Chao's estimator overestimates in small samples (Chao, 1987). A bias-corrected version of Chao's estimator has been suggested (Chao, 1989; Wilson & Collins, 1992) and is considered in section 4.2. Section 5 compares all four estimators with respect to bias and mean-squared error. The article closes with a short discussion.

## 2. The Poisson case

In the case of a Poisson kernel, we have that Chao's lower-bound estimate of the missing zero counts is $\hat{f}_0^C = f_1^2/(2f_2)$, whereas Zelterman's estimate is provided as $\hat{f}_0^Z = n/[\exp(\hat{\lambda}) - 1]$, where $\hat{\lambda} = 2f_2/f_1$. Let also $n = f_1 + f_2 + f_3 + \cdots + f_m$, where $m$ is the largest observed count. Theorem 1 describes a situation in which the Zelterman estimate is bounded above by the Chao's estimate and thus provides a counterexample to the raised hypothesis that Zelterman's estimate is always larger than the Chao's estimator.

**Theorem 1**
*Let* $p_x = \int_0^\infty \mathrm{Po}(x \mid \lambda) f(\lambda)\, \mathrm{d}\lambda$ *and* $\mathrm{Po}(x \mid \lambda)$ *be the Poisson kernel. Also, let* $n = f_1 + f_2$ *and* $f_i > 0$ *for* $i = 1, 2$. *Then,* $\hat{f}_0^C > \hat{f}_0^Z$.

*Proof.* Consider the second-order approximation $1 + x + x^2/2$ of $\mathrm{e}^x$. Note that $1 + x + x^2/2 \leq \mathrm{e}^x$, where the inequality becomes sharp except for $x = 0$. It follows that

$$\hat{f}_0^Z = \frac{n}{\exp(\hat{\lambda})-1} < \frac{f_1+f_2}{\hat{\lambda}+\hat{\lambda}^2/2} = \frac{f_1+f_2}{\frac{2f_2}{f_1^2}(f_1+f_2)} = \hat{f}_0^C,$$

which ends the proof.

As Chao's estimate provides a lower-bound estimate of the population size, Zelterman's estimator would underestimate the population size even more. As a consequence, the estimate of Zelterman should not be used if only counts of 1 or 2 have occurred.

In the following we provide evidence that the underestimation of Zelterman's estimator (relative to Chao's estimator) is basically restricted to the situation described in theorem 1. For all other situations, we will show that Zelterman's estimator provides an *upper bound*, at least if $\hat{\lambda}$ is sufficiently small. We are interested in the more general question when

$$\hat{f}_0^Z = \frac{n}{\exp(\hat{\lambda})-1} \leq \frac{f_1^2}{2f_2} = \frac{f_1}{\hat{\lambda}} = \hat{f}_0^C \tag{1}$$

holds for arbitrary integer $m>1$ and $n=f_1+f_2+f_3+\cdots+f_m$. Note that (1) can be rewritten equivalently as:

$$\frac{n}{f_1}\hat{\lambda}+1 \leq \exp(\hat{\lambda}). \tag{2}$$

Now, $\hat{\lambda}$ is a point on the line $1+bx$ where the slope $b$ is given by $n/f_1$, and for (2) to be true, it must lie in a region of the positive $x$-axis where

$$bx+1 \leq \exp(x). \tag{3}$$

For lines with slopes $b$ larger than 1, these regions always exist, as $bx+1=\exp(x)$ for $x=0$ and $(bx+1)'=b>\exp(x)'_{x=0}=1$. These lines intersect at some point $x_0$ and, for any $x \geq x_0$, (3) holds. Thus, if $\hat{\lambda}=2f_2/f_1 \geq x_0$, we have that $\hat{f}_0^C \geq \hat{f}_0^Z$. We summarize in theorem 2.

**Theorem 2**

*Let* $p_x=\int_0^\infty \mathrm{Po}(x\,|\,\lambda)f(\lambda)\,\mathrm{d}\lambda$ *and* $\mathrm{Po}(x\,|\,\lambda)$ *be the Poisson kernel. Also, let* $n=f_1+f_2+f_3+\cdots+f_m$, $f_i>0$ *for* $i=1,2$ *and* $f_i>0$ *for at least one* $i$ *in* $3,\ldots,m$. *There exists* $x_0>0$ *such that* $bx_0+1=\exp(x_0)$ *with* $b=n/f_1$. *Then, if* $\hat{\lambda}>x_0$, $\hat{f}_0^C>\hat{f}_0^Z$; *if* $\hat{\lambda}\leq x_0$, $\hat{f}_0^C \leq \hat{f}_0^Z$.

Theorem 2 guarantees the existence of a point of intersection $x_0$ of the curves $\exp(x)$ and $1+bx$ with $b=n/f_1$. Hence, if $\hat{\lambda}\leq x_0, \hat{f}_0^C \leq \hat{f}_0^Z$. Now, the point of intersection does not exist in closed form. In Fig. 1, the point of intersection $x_0$ is plotted as a function of $b=n/f_1$. If the pair $(b,\hat{\lambda})$ lies below the curve, then $\hat{f}_0^C<\hat{f}_0^Z$. If the pair $(b,\hat{\lambda})$ lies above the curve, then $f_0^C>\hat{f}_0^Z$. If the pair lies exactly on the curve we have equality.

If we use the second-order Taylor series approximation $1+x+x^2/2$ of $\exp(x)$ around 0, a point of intersection $\tilde{x}_0$ can be found in explicit form for which $1+bx=1+x+x^2/2$ is valid, or equivalently,

$$\tilde{x}_0 = 2\left(\frac{n}{f_1}-1\right).$$

Now write $\tilde{x}_0=2\{[f_1+f_2+(n-f_1-f_2)]/f_1-1\}$, so that

$$\tilde{x}_0 = \frac{2f_2}{f_1}+2\frac{n-f_1-f_2}{f_1} \geq \hat{\lambda}$$

with the inequality being strict if $n>f_1+f_2$. Now, if $\tilde{x}_0$ becomes identical to $x_0$, we have that $\hat{f}_0^Z \geq \hat{f}_0^C$. But $\tilde{x}_0$ will be close to $x_0$, if the Taylor series approximation is good which is the case if $x$ becomes small. In summary, for $\hat{\lambda}$ small enough, $\hat{f}_0^Z \geq \hat{f}_0^C$. As a rule of the thumb,

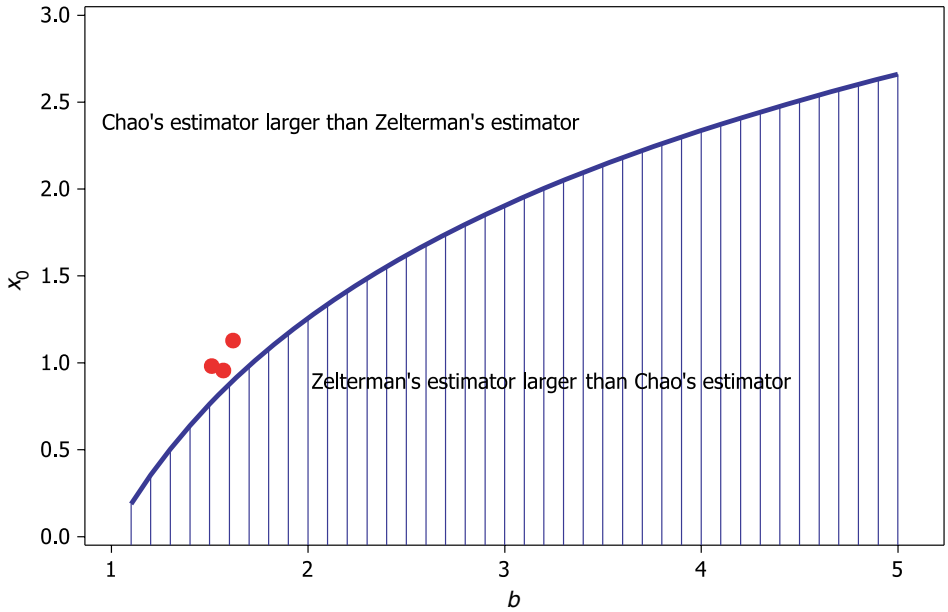


*Fig. 1.* The point of intersection as a function of $b$: in the region above the curve, Chao's estimator is larger than that of Zelterman.

we find for values of $\hat{\lambda} \leq 0.5$ the approximation good enough to force Zelterman's estimator to be larger than Chao's lower bound.

*Examples.* Roberts & Brewer (2006) present data that concern men who were arrested by the Vancouver Police Department for patronizing a prostitute. Such data can be treated as capture–recapture data when the number of re-arrests is known. For the two years including 1986 and 1987, 521 different men were arrested for patronizing a street prostitute, and 11 of these were arrested twice in the period. For the 7-year period between 1986 and 1992, 2001 different men were arrested, and 44 of these were arrested twice. No other information on the arrestees is available. The data provide an illustration for theorem 1, as $f_i = 0$ for $i > 2$. For the 2-year data, the Chao's estimate is 12,344, larger than the Zelterman estimate which is 12,340. For the 7-year period the Chao's estimator is 47,545, again larger than the Zelterman estimate of 47,531.

van der Heijden *et al.* (2003) discuss data on illegal ownership of firearms for the 1998–9 period in five administrative regions in the Netherlands. According to them 2561 people were identified exactly once for the illegal possession of a firearm, 72 people were identified exactly twice and 5 exactly thrice. We find that $\hat{\lambda} = 2f_2/f_1 = 144/2561 = 0.056$, a value close to 0. Consequently, $\hat{N}^Z = 48{,}248 > \hat{N}^C = 48{,}185$. To illustrate the importance of $\hat{\lambda}$ being small, we change $f_1$ to be 1561 and $f_2$ to be 1072. Then, $\hat{\lambda} = 2f_2/f_1 = 2144/1561 = 1.37$, a value considerably larger than 0.056. Consequently, $\hat{N}^Z = 3533 < \hat{N}^C = 3775$. Typically, $f_1$ is much larger in empirical studies than $f_2$, which leads to a small ratio $2f_2/f_1$ explaining why in most of the empirical studies the Zelterman estimator of the population provides a larger value than the Chao's estimator.

Finally, we wish to illustrate theorem 2. van Hest *et al.* (2008) study the achieved coverage of tuberculosis screening among drug users and homeless persons in Rotterdam, the Netherlands. Radiologic screening was done using a mobile digital X-ray unit which visited day and

Table 1. *Frequency $f_x$ of count $x$ of repeated entry into the screening programme by year with Chao's and Zelterman's estimates of the size of the target population*

| Year | $f_1$ | $f_2$ | $f_3+$ | $n$ | $\hat{\lambda}$ | $b$ | $\hat{N}^Z$ | $\hat{N}^C$ |
|---|---|---|---|---|---|---|---|---|
| 2003 | 1162 | 555 | 107 | 1824 | 0.96 | 1.57 | 2964 | 3040 |
| 2004 | 1058 | 597 | 57 | 1712 | 1.13 | 1.62 | 2531 | 2649 |
| 2005 | 997 | 489 | 21 | 1507 | 0.98 | 1.51 | 2411 | 2523 |

night shelters and hostels for homeless persons, methadone-dispensing centres and safe drug consumption rooms for opiate users, as well as the street prostitution zones in Rotterdam. Data on the number of persons and their repeated participation in the screening programme are provided by van Hest *et al.* (2008) and are summarized in Table 1.

As the numbers in Table 1 show, the estimate of Chao is larger than the estimate of Zelterman for all three years. This can now be easily explained by means of theorem 2. If we look at the pairs $(b, \hat{\lambda})$ for the three years – as indicated by the three bullets in Fig. 1 – we see that all three points $(b, \hat{\lambda})$ lie above the curve $(b, x_0)$. Hence, the Chao's estimator must be larger than the Zelterman's.

## 3. The binomial case

In the binomial case, we have that Chao's lower-bound estimate of the missing zero counts is $\hat{f}_0^C = [(m-1)/m][f_1^2/(2f_2)]$, whereas Zelterman's estimate of $f_0$ is provided as $\hat{f}_0^Z = n/[(1+2f_2/\{(m-1)f_1\})^m - 1]$ (see section 1).

First we consider the case where $m=2$, so that $n=f_1+f_2$. Chao's estimate of the frequency of the zero counts in this case is given by $\hat{f}_0^C = f_1^2/(4f_2)$. Comparing this with the Zelterman's estimate using $\hat{\lambda}' = 2f_2/f_1$, we have:

$$\hat{f}_0^Z = \frac{n}{(1+\hat{\lambda}')^2 - 1} = \frac{n}{\frac{4f_2}{f_1} + \frac{4f_2^2}{f_1^2}} = \frac{n}{\frac{4f_2}{f_1^2}(f_1+f_2)} = \frac{f_1^2}{4f_2} = \hat{f}_0^C$$

showing that the Zelterman estimator is identical to the Chao's estimator in the two-source binomial situation. We summarize in theorem 3.

**Theorem 3**
*Let* $p_x = \binom{m}{x} \int_0^1 \lambda^x (1-\lambda)^{m-x} f(\lambda)\,\mathrm{d}\lambda$. *For* $m=2$, *we have that* $\hat{N}^Z = \hat{N}^C$.

More generally we wish to know if there exists a condition in the binomial case when

$$\hat{f}_0^Z = \frac{n}{(1+\hat{\lambda}')^m - 1} \leq \frac{m-1}{m}\frac{f_1^2}{2f_2} = \hat{f}_0^C \tag{4}$$

holds for $m>2$ and $n=f_1+f_2+f_3+\cdots+f_m$. Now, (4) can be written equivalently as:

$$\frac{n}{(1+\hat{\lambda}')^m - 1} \leq \frac{f_1}{\hat{\lambda}' m}$$

and re-arranging this expression gives:

$$(1+\hat{\lambda}')^m \geq 1 + \frac{\hat{\lambda}' m n}{f_1}.$$

Hence, we need to compare the functions $(1+x)^m$ and $1+bmx$ with $b=n/f_1$. There exists again a point of intersection $x_0>0$ such that $(1+x)^m<1+bmx$ for all $x\in(0,x_0)$ and $(1+x)^m>1+bmx$ for all $x>x_0$. We summarize this in theorem 4.

**Theorem 4**
*Let* $p_x=\binom{m}{x}\int_0^1\lambda^x(1-\lambda)^{m-x}f(\lambda)\,\mathrm{d}\lambda$. *Also, let* $n=f_1+f_2+f_3+\cdots+f_m, f_i>0$ *for* $i=1,2$ *and* $f_i>0$ *for at least one* $i$ *in* $3,\ldots,m$. *There exists* $x_0>0$ *such that* $bmx_0+1=(1+x_0)^m$ *with* $b=n/f_1$. *Then, if* $\hat{\lambda}'>x_0$, $\hat{f}_0^C>\hat{f}_0^Z$; *if* $\hat{\lambda}'\le x_0$, $\hat{f}_0^C\le\hat{f}_0^Z$.

Consider the Taylor approximation $1+mx+m(m-1)x^2/2$ of $(1+x)^m$ around $x=0$. Equating $1+mx+m(m-1)x^2/2$ to $1+bmx$ leads to

$$\tilde{x}_0=\frac{2}{m-1}\left(\frac{f_2}{f_1}+\frac{n-f_1-f_2}{f_1}\right),$$

which is an upper bound for $\hat{\lambda}'=\frac{2}{m-1}\frac{f_2}{f_1}$. Now, if $\tilde{x}_0$ becomes identical to $x_0$, we have that $\hat{f}_0^Z\ge\hat{f}_0^C$. But $\tilde{x}_0$ will be close to $x_0$, if the Taylor series approximation is good which is the case if $x$ becomes small. In summary, for $\hat{\lambda}'$ small enough, $\hat{f}_0^Z\ge\hat{f}_0^C$.

## 4. A two-component heterogeneity model in the Poisson case

We are considering here a special case of heterogeneity distribution in which the heterogeneity distribution is represented by a discrete two-component mixture: $f(\lambda)=(1-q)\delta_{\lambda_1}+q\delta_{\lambda_2}$, where $\delta_x$ corresponds to the one-point measure at $x$. In consequence, the marginal distribution is given as a two-component mixture $p_x=(1-q)\mathrm{Po}(x\,|\,\lambda_1)+q\mathrm{Po}(x\,|\,\lambda_2)$. In capture–recapture analysis frequently two-component mixtures are sufficient to represent the observed heterogeneity (see also Pledger, 2000, 2005). In addition, a two-component model of this kind is common in modelling a contaminated distribution. Hence we think of $\lambda_1$ as the distributional part which becomes contaminated by the distributional part represented by $\lambda_2$. In the following we simply write $\lambda_1=\lambda$ and $\lambda_2=\mu$.

**Theorem 5**
*Let* $p_x=qp(x\,|\,\lambda)+(1-q)p(x\,|\,\mu)$ *be a discrete, two-component mixture with* $p(x\,|\,\theta)=\mathrm{Po}(x\,|\,\theta)$ *being the Poisson kernel with parameter* $\theta$ *and* $0\le q\le 1$. *Then,*

$$\lim_{N\to\infty}\frac{E(\hat{N}^Z)}{N}=\frac{1-[q\exp(-\lambda)+(1-q)\exp(-\mu)]}{1-\exp(-\tilde{\lambda})}\to_{\mu\to\infty}\frac{1-q\exp(-\lambda)}{1-\exp(-\lambda)}\ge 1,$$

*with*

$$\tilde{\lambda}=\frac{q\exp(-\lambda)\lambda^2+(1-q)\exp(-\mu)\mu^2}{q\exp(-\lambda)\lambda+(1-q)\exp(-\mu)\mu}.$$

*The inequality is strict if* $\lambda>0$ *and* $q\in(0,1)$.

*Proof.* We have that $\hat{N}^Z=n/[1-\exp(-2f_2/f_1)]$. As

$$E(f_j)=N[q\exp(-\lambda)\lambda^j/j!+(1-q)\exp(-\mu)\mu^j/j!]$$

for $j=1,2$ and

$$E(n) = N[1 - q\exp(-\lambda) - (1-q)\exp(-\mu)]$$

we have that $E(N^Z)$ becomes for large $N$

$$N\frac{1-[q\exp(-\lambda)+(1-q)\exp(-\mu)}{1-\exp(-2\frac{1}{2}\tilde{\lambda})},$$

with

$$\frac{1}{2}\tilde{\lambda} = \frac{q\exp(-\lambda)\lambda^2/2 + (1-q)\exp(-\mu)\mu^2/2}{q\exp(-\lambda)\lambda + (1-q)\exp(-\mu)\mu}.$$

The second part of the theorem follows from the fact that $(1-q)\exp(-\mu)\mu^j/j!$ converges to zero with $\mu \to \infty$.

The theorem implies that large contaminations have a persistent, potentially strongly overestimating effect. To give an example, consider $q = 0.5$ and any $\lambda \leq 0.4$, then the factor $[1 - q\exp(-\lambda)]/[1 - \exp(-\lambda)]$ is larger than 2.

On the contrary, it is a remarkable property of Chao's estimator that it is not affected by large contaminations as the following result in theorem 6 shows.

**Theorem 6**
*Let* $p_x = qp(x \mid \lambda) + (1-q)p(x \mid \mu)$ *be a discrete, two-component mixture with* $p(x \mid \theta) = \mathrm{Po}(x \mid \theta)$ *being the Poisson kernel with parameter* $\theta$ *and* $0 \leq q \leq 1$. *Then,*

$$\lim_{N\to\infty}\frac{E(\hat{N}^C)}{N} = [1 - q\exp(-\lambda) - (1-q)\exp(-\mu)] + \frac{\{q\exp(-\lambda)\lambda + (1-q)\exp(-\mu)\mu\}^2}{q\exp(-\lambda)\lambda^2 + (1-q)\exp(-\mu)\mu^2}$$
$$\to_{\mu\to\infty} 1.$$

*Proof.* We have that $\hat{N}^C = n + f_1^2/(2f_2)$. The proof then follows along similar arguments as used in the proof of the previous theorem.

### *4.1. The modified Zelterman estimator*

The overestimation bias involved in the Zelterman estimator suggests considering the modification

$$\hat{N}^M = (f_1 + f_2)/[1 - \exp(-\hat{\lambda})] + (n - f_1 - f_2).$$

The motivation behind this modification lies in the idea to use only those frequency counts $f_1, f_2$ in the prediction $(f_1 + f_2)/[1 - \exp(-\hat{\lambda})]$ which have been used also in the estimation of $\lambda$. As a consequence of theorem 1 we have the following *lower-bound property* of the modified Zelterman estimator.

**Corollary 1**
*Let* $p_x = \int_0^\infty \mathrm{Po}(x \mid \lambda) f(\lambda)\,\mathrm{d}\lambda$ *and* $\mathrm{Po}(x \mid \lambda)$ *be the Poisson kernel. Then,* $\hat{N}^M \leq \hat{N}^C$.

*Proof.* Consider that

$$\hat{N}^M = \frac{f_1+f_2}{1-\exp(-\hat{\lambda})} + n - (f_1+f_2) = n + \frac{f_1+f_2}{\exp(\hat{\lambda})-1}$$
$$\leq n + \frac{f_1+f_2}{\hat{\lambda}+\hat{\lambda}^2/2} = n + \frac{f_1+f_2}{2f_2/f_1^2(f_1+f_2)} = \hat{N}^C,$$

using that $\hat{\lambda} = 2f_2/f_1$, very similar to the argument used in theorem 1.

As a consequence we see in theorem 7 that with this modification the overestimation bias of the Zelterman estimator disappears.

**Theorem 7**

*Let* $p_x = qp(x \mid \lambda) + (1-q)p(x \mid \mu)$ *be a discrete, two-component mixture with* $p(x \mid \theta) = \mathrm{Po}(x \mid \theta)$ *being the Poisson kernel with parameter* $\theta$ *and* $0 \leq q \leq 1$. *Then,*

$$\lim_{N\to\infty} \frac{E(\hat{N}^M)}{N} = 1 - [q\exp(-\lambda) + (1-q)\exp(-\mu)]$$
$$+ \frac{q\exp(-\lambda)\lambda + (1-q)\exp(-\mu)\mu + q\exp(-\lambda)\lambda^2/2 + (1-q)\exp(-\mu)\mu^2/2}{\exp(\tilde{\lambda})-1}$$
$$\to_{\mu\to\infty} 1 + q\,\mathrm{e}^{-\lambda}\left(\frac{\lambda+\lambda^2/2}{\exp(\lambda)-1} - 1\right) \leq 1,$$

*where*

$$\tilde{\lambda} = \frac{q\exp(-\lambda)\lambda^2 + (1-q)\exp(-\mu)\mu^2}{q\exp(-\lambda)\lambda + (1-q)\exp(-\mu)\mu}.$$

*Proof.* We have that

$$\hat{N}^M = [f_1+f_2]/[1-\exp(-2f_2/f_1)] + n - f_1 - f_2 = n + (f_1+f_2)/(\exp(\hat{\lambda})-1).$$

Replacing observed frequencies by their expected values $E(\hat{N}^M)$ becomes for large $N$

$$N(1 - [q\exp(-\lambda) + (1-q)\exp(-\mu)])$$
$$+ N\frac{q\exp(-\lambda)\lambda + (1-q)\exp(-\mu)\mu + q\exp(-\lambda)\lambda^2/2 + (1-q)\exp(-\mu)\mu^2/2)}{\exp(\tilde{\lambda})-1}. \quad (5)$$

The second part of the theorem follows from the fact that $(1-q)\exp(-\mu)\mu^j/j!$ converges to zero with $\mu \to \infty$ and, hence, (5) becomes

$$N - Nq\mathrm{e}^{-\lambda} + Nq\,\mathrm{e}^{-\lambda}\frac{\lambda+\lambda^2/2}{\mathrm{e}^{\lambda}-1} \leq N,$$

which ends the proof (by noting that $\mathrm{e}^{\lambda} = \sum_{j=0}^{\infty}\lambda^j/j! \geq 1+\lambda+\lambda^2/2$).

Note that $\lambda+\lambda^2/2$ is the second-order McLaurin series approximation of $\exp(\lambda)-1$. Hence, the bias will be small. This will also be seen in the simulation study given in the next section.

To illustrate the modified Zelterman estimator (and its closeness to the Chao's estimator) we consider again the data of section 2 on illegal ownership of firearms for the 1998–9 period in five administrative regions in the Netherlands (van der Heijden *et al.*, 2003). We had $f_1 = 2561$, $f_2 = 72$ and $f_3 = 5$, so that $\hat{N}^Z = 48{,}248$, $\hat{N}^C = 48{,}185$ and $\hat{N}^M = 48{,}161$, showing that the latter two are rather close.

### 4.2. The bias-corrected Chao's estimator

A bias-corrected version of Chao's estimator was suggested by Chao (1987)

$$\hat{N}^B = n + \frac{f_1(f_1-1)}{2(f_2+1)} \tag{6}$$

and also discussed in Wilson & Collins (1992) and Chao (2005). Note that $\hat{N}^B \leq N^C$ by definition of the estimator. As $N^C$ is giving already a lower bound under the Poisson mixture model, the question arises why this corrected estimator is necessary. The motivation for this correction stems mainly from the case of homogeneity of the Poisson model where $N^C$ is asymptotically unbiased for $N$, but experiences overestimation bias for small sample sizes as we shall work out in detail next. The conventional estimator of Chao is $n+f_1^2/(2f_2)$ with expected value

$$E[n+f_1^2/(2f_2)] = E(n) + E[f_1^2/(2f_2)] = N(1-\exp(-\lambda)) + E[f_1^2/(2f_2)].$$

However, $E[f_1^2/(2f_2)]$ becomes close to

$$N[\exp(-\lambda)\lambda]^2/[2\exp(-\lambda)\lambda^2/2] = N\exp(-\lambda)$$

only for large $N$, so that the approximation

$$E[f_1^2/(2f_2)] \approx N\exp(-\lambda)$$

is only valid for large $N$. In fact, $E[f_1^2/f_2]$ overestimates $E(f_1^2)/E(f_2)$, potentially considerably. This becomes clear when investigating numerator and denominator in $f_1^2/f_2$. $E(f_1^2) = \mathrm{var}(f_1^2) + [E(f_1)]^2$, so that $E(f_1^2)$ overestimates $[E(f_1)]^2$ by $\mathrm{var}(f_1^2)$ which can be approximated by $E(f_1)$ leading to the bias correction $f_1^2 - f_1$ in the numerator of $f_1^2/(2f_2)$. Likewise the denominator expected value $E(1/f_2)$ overestimates $1/E(f_2)$ by Jensen's inequality (assuming that the expected values exist). In contrast, $1/(f_2+1)$ always exists and provides a bias reduction as

$$\frac{1}{1+E(f_2)} \leq E\left(\frac{1}{1+f_2}\right) \leq E\left(\frac{1}{f_2}\right),$$

where we have used Jensen's inequality once more to achieve the first inequality. This bias reduction can have maximal bias

$$1/[1+E(f_2)] - 1/E(f_2) = -1/[(1+E(f_2))E(f_2)].$$

Simulations show that the denominator correction is more important than the numerator correction. In the case of heterogeneity, the correction (6) is less important. However, Chao (2005) points out that $N^B$ is also defined for $f_2 = 0$ so that the denominator correction might always be advisable.

To illustrate the modified Chao's estimator we consider again the data of section 2 on illegal ownership of firearms for the 1998–9 period in five administrative regions in the Netherlands. We had $f_1 = 2561, f_2 = 72$ and $f_3 = 5$, so that $\hat{N}^B = 47{,}543$ which is considerably smaller than the other three: $\hat{N}^Z = 48{,}248$, $\hat{N}^C = 48{,}185$ and $\hat{N}^M = 48{,}161$.

### 4.3. The source for the overestimation bias in Zelterman's estimator

The question arises which is the source for the overestimation bias in Zelterman's estimate and is approached in theorem 8 which notes that the Zelterman uses the *wrong* expected value in predicting $f_0$. It was pointed out in the introduction that both estimators use only frequencies of counts of one and two. Hence, we might consider a log-likelihood truncated for all counts except ones and twos, namely $\log L(\lambda) = f_1\log(p_1) + f_2\log(p_2)$ with $p_2 = \lambda/(\lambda+2) =$

$1-p_1$. This binomial-type likelihood is uniquely maximized for $\hat{p}_2 = 1-\hat{p}_1 = f_2/(f_1+f_2)$. Note that $p_2 = \lambda/(\lambda+2)$ and $p_1 = 2/(\lambda+2)$ are the truncated Poisson probabilities $p_2 = [\exp(-\lambda)\lambda^2/2]/[\exp(-\lambda)\lambda^2/2 + \exp(-\lambda)\lambda]$ for count two and $p_1 = [\exp(-\lambda)\lambda]/[\exp(-\lambda)\lambda^2/2 + \exp(-\lambda)\lambda]$ for count 1. Hence, $\log L(\lambda)$ is a truncated Poisson likelihood that truncates all counts except ones and twos. As the MLE for $p_2$ is $f_2/(f_1+f_2)$, the estimate $\hat{\lambda} = 2f_2/f_1$ for the Poisson parameter $\lambda$ suggested by Zelterman (1988) arises. In the approach of Zelterman the homogeneous Poisson serves only as a working model and it was shown by Zelterman that the estimate $\hat{N} = n/(1-\hat{p}_0) = n/[1-\exp(-\hat{\lambda})]$ is less sensitive against mis-specifications of the Poisson model than the usual MLE. The major problem of Zelterman's estimate can be identified as follows. Whereas in a homogeneous Poisson model with counts larger than two *not* truncated, the conditional expectation of $f_0$ is $E(f_0 \mid f_1, \ldots, f_m; \hat{\lambda}) = (f_1 + \cdots + f_m)/(\exp(\hat{\lambda}) - 1)$ with $\hat{\lambda}$ being the MLE with respect to the zero-truncated likelihood, in the case of a homogeneous Poisson model with counts larger than two truncated, the conditional expectation of $f_0$ is $E(f_0 \mid f_1, f_2; \hat{\lambda}) = (f_1+f_2)/(\hat{\lambda} + \hat{\lambda}^2/2)$ with $\hat{\lambda} = 2f_2/f_1$. This latter conditional expectation turns out to be the lower bound of Chao: $E(f_0 \mid f_1, f_2; \hat{\lambda}) = f_1^2/(2f_2)$, as shown by theorem 8.

**Theorem 8**

(i) *Let* $\log L(\lambda) = f_1 \log(p_1) + f_2 \log(p_2)$ *with*

$$p_1 = \frac{e^{-\lambda}\lambda}{e^{-\lambda}\lambda + e^{-\lambda}\lambda^2/2} = \frac{2}{\lambda+2} \text{ and } p_2 = \frac{e^{-\lambda}\lambda^2/2}{e^{-\lambda}\lambda + e^{-\lambda}\lambda^2/2} = \frac{\lambda}{\lambda+2}$$

*being the Poisson probabilities truncated to counts of ones and twos. Then,* $\log L(\lambda)$ *is maximized for* $\hat{\lambda} = 2f_2/f_1$.

(ii) $E(f_0 \mid f_1, f_2; \hat{\lambda}) = f_1^2/(2f_2)$, *for* $\hat{\lambda} = 2f_2/f_1$.

*Proof.* For the first part, it is clear that $f_1 \log(p_1) + f_2 \log(p_2)$ is maximal for $\hat{p}_1 = f_1/(f_1+f_2)$, which is attained for $\hat{\lambda} = 2f_2/f_1$. For the second part, we see that with $e_x = E(f_x \mid f_1, f_2; \lambda) = \mathrm{Po}(x \mid \lambda)N$:

$$e_x = \mathrm{Po}(x \mid \lambda)N = \mathrm{Po}(x \mid \lambda)N = \mathrm{Po}(x \mid \lambda)\left(e_0 + f_1 + f_2 + \sum_{j=3}^{\infty} e_j\right)$$

so that

$$e_0 + e_3^+ = [1 - \mathrm{Po}(1 \mid \lambda) - \mathrm{Po}(2 \mid \lambda)](e_0 + e_3^+) + [1 - \mathrm{Po}(1 \mid \lambda) - \mathrm{Po}(2 \mid \lambda)](f_1 + f_2)$$

with $e_3^+ = \sum_{j=3}^{\infty} e_x$. Hence,

$$e_0 + e_3^+ = \frac{1 - \mathrm{Po}(1 \mid \lambda) - \mathrm{Po}(2 \mid \lambda)}{\mathrm{Po}(1 \mid \lambda) + \mathrm{Po}(2 \mid \lambda)}(f_1 + f_2)$$

and

$$\begin{aligned} e_0 &= \mathrm{Po}(0 \mid \lambda)(f_1 + f_2 + e_0 + e_3^+) \\ &= \mathrm{Po}(0 \mid \lambda)(f_1 + f_2) + \mathrm{Po}(0 \mid \lambda)\frac{1 - \mathrm{Po}(1 \mid \lambda) - \mathrm{Po}(2 \mid \lambda)}{\mathrm{Po}(1 \mid \lambda) + \mathrm{Po}(2 \mid \lambda)}(f_1 + f_2) \\ &= \frac{\mathrm{Po}(0 \mid \lambda)}{\mathrm{Po}(1 \mid \lambda) + \mathrm{Po}(2 \mid \lambda)}(f_1 + f_2) = \frac{f_1 + f_2}{\lambda + \lambda^2/2}. \end{aligned}$$

Plugging in the MLE $\hat{\lambda} = 2f_2/f_1$ for $\lambda$ yields the desired result.

Theorem 8 establishes a close connection between the approach by Zelterman and Chao's estimator. It shows that Zelterman's estimator of the Poisson parameter $\lambda$ arises

when all counts are truncated except counts of ones and twos and when the resulting likelihood is maximized. If the correct prediction for $f_0$ is used, namely the conditional expectation for the truncated Poisson model, the Chao's estimator arises. Hence, the strong overestimation of the original Zelterman estimator stems from using a *wrong* conditional expectation.

## 5. Simulation study

To investigate the behaviour of estimators further we have executed a comparative simulation study. Two cases were distinguished: the homogeneity case in which samples of size $N = 100$ were generated from a Poisson distribution with parameter $\lambda \in \{0.5, 1, 2, 3, 4, 5\}$. Any occurring zeros were truncated and population sizes were estimated with the four estimators according to Zelterman, Chao and the two proposed modifications. In the second case, heterogeneous samples were generated arising from 50 per cent:50 per cent mixture of Poisson distributions (high amount of contamination) where the first component was chosen with parameter 0.5 and the second component parameter $\mu$ varied in $\{1, 2, 3, 4, 5, 6, 7\}$. In addition, a 90 per cent:10 per cent mixture of Poisson distributions (small amount of contamination) was studied where the first component was chosen with parameter 0.5 and the second component parameter $\mu$ varied in $\{1, 2, 3, 4, 5\}$. Expected values and root mean-squared error were determined in the conventional way:

$$\widehat{E(\hat{N})} = \Big(\sum_i \hat{N}_i\Big)/10{,}000,$$

$$\widehat{\mathrm{var}(\hat{N})} = \sum_i [\hat{N}_i - \widehat{E(\hat{N})}]^2/10{,}000$$

and

$$\mathrm{RMSE} = \sqrt{(\widehat{E(\hat{N})} - N)^2 + \widehat{\mathrm{var}(\hat{N})}}.$$

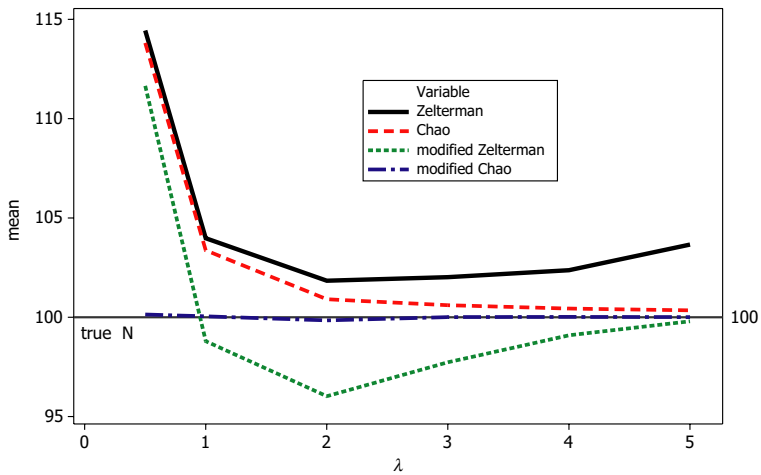


*Fig. 2.* Expected values for the four population size estimators for $N = 100$ and homogeneous Poisson population with parameter $\lambda$.

To eliminate any random error owing to the simulation a replication size of 10,000 was used.

We summarize here the salient findings of the simulation study.

- In the homogeneous case, the Zelterman estimator tends to overestimate more than the Chao's estimator (see Fig. 2). The modified Zelterman estimator is less overestimating than the Chao's estimator. Clearly, the modified Chao's estimator is the least biased (almost unbiased). Chao's estimator and the modified Zelterman's estimator have similar standard deviations, but again the modified (bias-corrected) Chao's estimator has smaller standard deviation (see Fig. 3).
- In the heterogeneous case, the Zelterman estimator overestimates strongly for large contaminations (see Figs 4 and 6). The Chao's estimator and the modified Zelterman's estimator have less overestimation bias and both have smaller mean-squared error than

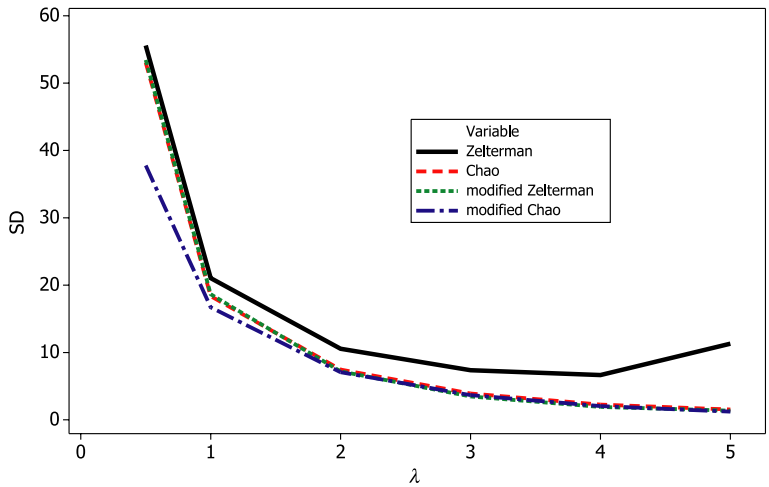


*Fig. 3.* Standard deviations for the four population size estimators for $N = 100$ and homogeneous Poisson population with parameter $\lambda$.

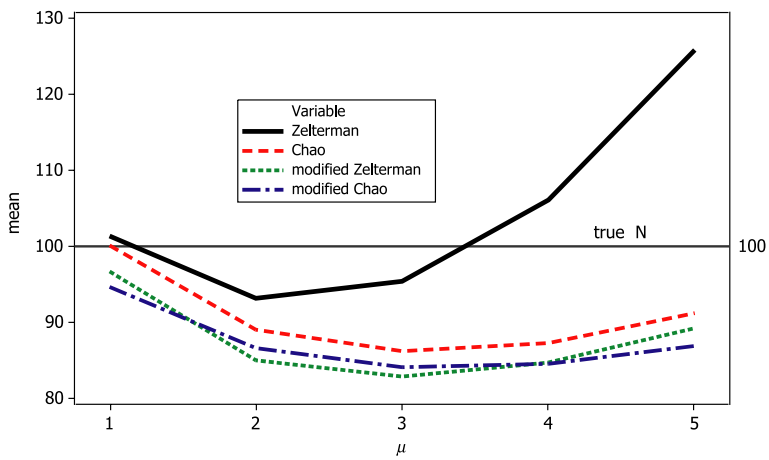


*Fig. 4.* Expected values for the four population size estimators for $N = 100$ and heterogeneous Poisson population with mixing distribution $0.5\delta_{0.5} + 0.5\delta_{\mu}$.

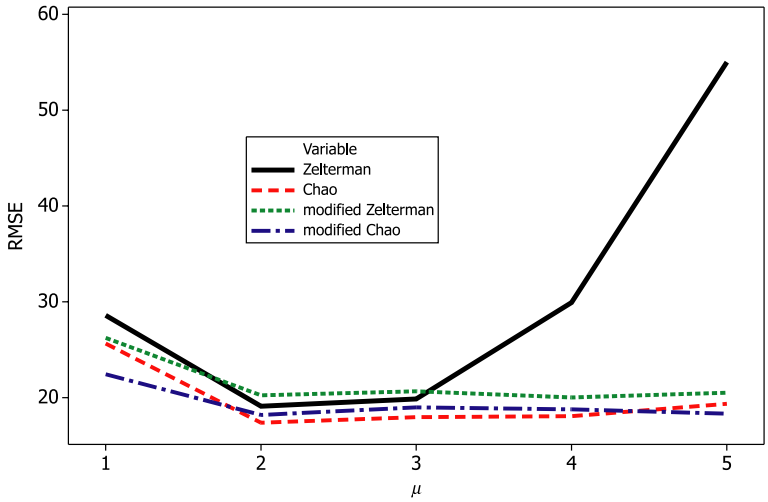


*Fig. 5.* Root mean-squared error for the four population size estimators for $N = 100$ and heterogeneous Poisson population with mixing distribution $0.5\delta_{0.5} + 0.5\delta_{\mu}$.

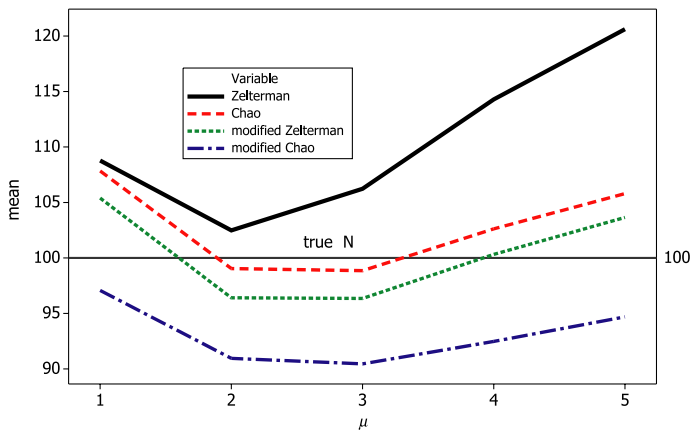


*Fig. 6.* Expected values for the four population size estimators for $N = 100$ and heterogeneous Poisson population with mixing distribution $0.9\delta_{0.5} + 0.1\delta_{\mu}$.

the Zelterman's estimator (see Figs 4 and 5). In all cases, the modified Chao's estimator is below the true population size. This also means that it has a larger bias than Chao's conventional and the modified Zelterman estimator (see Figs 4 and 6). If the root mean-squared error is considered it seems to perform best among all the four estimators (see Figs 5 and 7).

## 6. Discussion

For the Poisson case, van der Heijden *et al.* (2006) is one of the few papers to discuss the relationship between Chao's and Zelterman's estimator. Their analysis is based on asymptotic considerations concluding that Chao's estimator is smaller than Zelterman's estimator if only the ratio $f_2/f_1$ is small enough. This conclusion is correct. However, it overlooks the existing

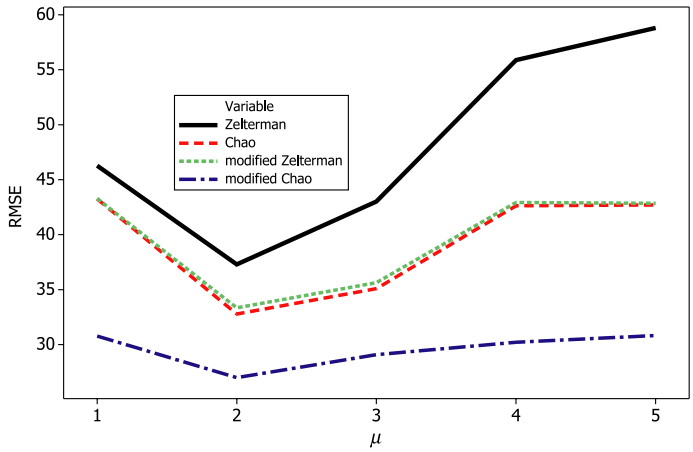


*Fig. 7.* Root mean-squared error for the four population size estimators for $N = 100$ and heterogeneous Poisson population with mixing distribution $0.9\delta_{0.5} + 0.1\delta_{\mu}$.

exact relationships for $m = 2$ as well as for the more general case $m > 2$, which is presented here. It was also shown that Zelterman's and Chao's estimators are close if the ratio $f_2/f_1$ is small. However, it was also demonstrated that Zelterman's estimator can overestimate considerably. A modification of the Zelterman's estimator was suggested which behaves similar to Chao's estimator but still shares the simple features of Zelterman's estimator; in particular, it can be generalized to allow the incorporation of covariates as suggested for the conventional Zelterman estimator in Böhning & van der Heijden (2009). The bias-corrected estimator of Chao appears to be performing well for small samples and small amounts of heterogeneity as the simulation study has shown. It also has smallest variance among all considered estimators, an important aspect when constructing confidence intervals.

**Acknowledgements**

The author is most grateful to the editor, the associate editor and a referee for their helpful comments and suggestions.

**References**

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1975). *Discrete multivariate analysis*. MIT Press, Cambridge, MA.

Böhning, D. & van der Heijden, P. G. M. (2009). A covariate adjustment for zero-truncated approaches to estimating the size of hidden and elusive populations. *Ann. Appl. Stat.* **3**, 595–610.

Bunge, J. & Fitzpatrick, M. (1993). Estimating the number of species: a review. *J. Amer. Statist. Assoc.* **88**, 364–373.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Statist.* **11**, 265–270.

Chao, A. (1987). Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* **43**, 783–791.

Chao, A. (1989). Estimating population size for sparse data in capture–recapture experiments. *Biometrics* **45**, 427–438.

Chao, A. (2005). Species estimation and applications. In *Encyclopedia of statistical sciences*, 2nd edn., Vol. 12 (eds N. Balakrishnan, C. B. Read & B. Vidakovic), 7907–7916. Wiley, New York.

Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. (2001). Tutorial in biostatistics: the applications of capture–recapture models to epidemiological data. *Stat. Med.* **20**, 3123–3157.

Dorazio, R. M. & Royle, J. A. (2005). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**, 351–364.

Hay, G. & Smit, F. (2003). Estimating the number of drug injectors from needle exchange data. *Addict. Res. Theory* **11**, 235–243.

van der Heijden, P. G. M., Cruyff, M. & van Houwelingen, H. C. (2003). Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statist. Neerlandica* **57**, 1–16.

van der Heijden, P. G. M., Van Putten, W. & Van Rongen, R. (2006). A comparison of Zelterman's and Chao's estimators for the size of an unknown population by capture–recapture frequency data. Personal communication with van der Heijden.

van Hest, N. A. H., De Vries, G., Smit, F., Grant, A. D. & Richardus, J. H. (2008). Estimating the coverage of tuberculosis screening among drug users and homeless persons with truncated models. *Epidemiol. Infect.* **136**, 628–635.

Holzmann, H., Munk, A. & Zucchini, W. (2006). On identifiability in capture–recapture models. *Biometrics* **62**, 934–939.

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

Link, W. A. (2006). Response to a paper by Holzmann, Munk and Zucchini. *Biometrics* **62**, 936–939.

Pledger, S. A. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* **56**, 434–442.

Pledger, S. A. (2005). The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics* **61**, 868–876.

Roberts, J. M. & Brewer, D. D. (2006). Estimating the prevalence of male clients of prostitute women in Vancouver with a simple capture–recapture method. *J. Roy. Statist. Soc. Ser. A* **169**, 745–756.

Wilson, R. M. & Collins, M. F. (1992). Capture–recapture estimation with samples of size one using frequency data. *Biometrika* **79**, 543–553.

Zelterman, D. (1988). Robust estimation in truncated discrete distributions with applications to capture–recapture experiments. *J. Statist. Plann. Inference* **18**, 225–237.

*Received January 2008, in final form July 2009*

Dankmar Böhning, Applied Statistics, School of Biological Sciences, University of Reading, Whiteknights, 305 Lyle Building, RG6 6BX, Reading, UK.
E-mail: d.a.w.bohning@reading.ac.uk