Comparison of Confidence Intervals for the TG Estimator in Capture-recapture Data

Orasa Anan\(^1\), Preedaporn Kanjanasamranwong\(^1\), W. Chantarangsri\(^2\), and A’yunin Sofro \(^3\),
\(^1\)Department of Mathematics and Statistics, Faculty of Science, Thaksin University, Phatthalung, Thailand
\(^2\)Department of Mathematics, Faculty of Sciences and Technology, Nakhon Pathom Rajabhat University, Nakhon Pathom, Thailand
\(^3\)Mathematics Department, Universitas Negeri Surabaya, Indonesia

E-mail: aorasa@tsu.ac.th

Abstract. Capture-recapture techniques are very powerful tool and widely used for estimating an elusive target population size. Capture-recapture count data is presented in form of frequencies of frequencies data. They consist of the frequency of unites detected exactly once, twice, and so on, and the frequency of undetected unites is unknown. As consequence, the resulting distribution is a zero-truncated count distribution. The binomial distribution is selected as a simple model if the maximum number of counting occasions is known. If counting occasions are not known in advance, the series of frequencies assumed to be the Poisson distribution. In fact, the target population might be heterogeneous because it has different characteristics, resulting in over or under dispersion based on the basic models. The mixed Poisson, which is the exponential-Poisson mixture model, have been widely used to construct population size estimator for capture-recapture data. The original Turing estimator provides a good performance under the Poisson distribution. Additionally, an extension of Turing estimator, called the Turing-based geometric distribution with non-parametric approach was proposed (TG) for the heterogeneous population. It gives an easy way to estimate the target population size. In this work, we derived uncertainty measures for the TG estimator by considering two sources of variance (M1), and the second way is using only one source of variance (M2). It is emphasised that although the analytic approaches to compute uncertainty measures can be easily used in practice, there are valid asymptotically and requires a large sample size. Therefore, re-sampling approaches, true bootstraps (M3), imputed bootstrap (M4) and reduced bootstrap (M5), are proposed as alternative methods to get uncertainty measures. The study compares performance of variance and confidence interval of paralytics and re-sampling methods by using a simulation study. Overall, the imputed bootstrap is the best choice for estimating variance and constructing confidence interval for the TG estimator. The analytic approach with two sources of variance remains successful to estimate variance and calculate confidence interval in the case of large. It is very clear that the reduced bootstrap and the analytic approach with one source of variance are not appropriate in all situations. For the true bootstrap, the true value of population size is often unknown in nature; therefore, it quite useless for capture-recapture study.

Keywords: capture-recapture; uncertainty; population size estimator
1. Introduction

Capture-recapture techniques have been originally used in ecological science with the object to estimate the elusive wildlife population size.\(^1\). However, this technique has been extended in many areas aim to deal with under report problem such as epidemiology and public health \(^2,3\), social science \(^4\), computer system \(^5\), marketing \(^6\). In this paper aims to investigate the uncertainty in the TG estimator that was proposed in \(^7\). This non-parametric estimator is suitable for unobserved heterogeneity based geometric distribution in capture-recapture data. In this work we compare five uncertainty measurements from analytic and bootstrap approaches. A large-scale of simulation study is provided to lean about behaviour of variance estimation methods. Their comparisons also show how often the confidence interval covers the true value of population sizes. The article is organized as follows: we introduce the Poisson mixture model and the TG estimator in the section 2. In section 3, we show detail of variance estimation methods. Following by the simulation results in the section 4. In section 5, we present some real data examples. Finally, we summarize and point out some remarks in section 6.

2. The Poisson mixture model

Let \(x_i, i = 1, 2, 3, 4, ..., N\) denote the number of time individual \(i\) is observed covering the \(m\) occasions, and let \(g_x = Pr(X = x)\), \(f_x\) is the frequency of individuals observed exactly \(x\) times so that \(x = 0, 1, 2, 3, ..., m\). Note that \(f_0\) is missing and it can be estimated by \(\hat{N}_p\). The Poisson distribution often assumed to be a basic model in capture-recapture study; however, the unobserved heterogeneity usually occurs as the variance is greater/\(\text{lower}\) than mean. The Poisson mixture model is more realistic than the original model. Refer to \(^8\) suggested a mixed Poisson model with arbitrary mixing \(h(x)\) to model the form of heterogeneity as

\[
g_x = \int_0^\infty \frac{\exp(-\lambda)\lambda^x}{x!}h(\lambda; \theta)d\lambda. \quad (2.1)
\]

The exponential distribution was selected as the arbitrary mixing density; \(h(x) = \frac{1}{\theta} exp\left(-\frac{x}{\theta}\right)\) in the mixed Poisson model. The marginal of \(x\) achieved that

\[
g(x) = \int_0^\infty \frac{\exp(-\lambda)\lambda^x}{x!} \exp\left(-\frac{\lambda}{\theta}\right) d\lambda = \frac{1}{\theta^x} \int_0^\infty \lambda^x exp\left(-\lambda(1 + \theta^{-1})\right)d\lambda = \frac{1}{1 + \theta} \left(\frac{\theta}{1 + \theta}\right)^x = p(1-p)^x,
\]

where \(p = \frac{\theta}{1 + \theta}\) is the probability of success based on the geometric distribution. The geometric distribution is potentially flexible candidate for unobserved heterogeneity in capture-recapture study, it can be applied to many real life situations. The population size estimator based-geometric distribution was proposed by \(^7\) to deal with an violation assumption of Poisson model, and \(^9\) reconstructed in 2.1

2.1. A Turing based-geometric distribution

The Turing based-geometric distribution estimator was proposed by \(^7\), let \(X \sim Geo(p)\), \(g_0 = p\) and \(E(X) = \frac{1-p}{p}\),

\[
g_0 = p = \sqrt{p^2} = \sqrt{\frac{(1-p)p^2}{(1-p) \frac{1-p}{p}}} = \sqrt{\frac{1-p}{(1-p)/p}} = \sqrt{\frac{g_1}{E(X)}}. \quad (2.3)
\]

In practice, the \(g_x\) can be estimated by the \(\frac{L_x}{n}\), leading to

\[
g = \sqrt{\frac{f_1/n}{S/n}} = \sqrt{\frac{f_1}{S}}, \quad (2.4)
\]
where $S = \sum_{x=0}^{m} xf_x = \sum_{x=1}^{m} xf_x$. As a consequence, the Turing based geometric distribution estimator (TG) was given as

$$\hat{N}_{TG} = \frac{n}{1 - \sqrt{\frac{f_1}{S}}}.$$  \hspace{1cm} (2.5)

The interesting points of the TG estimator are that the formula is not complicated for users in practice, and it includes all of information of sample, by using $S$.

**Theorem 2.1.** The TG estimator is the asymptotically unbiased estimator under the geometric distribution

$$\lim_{N \to \infty} \frac{E(\hat{N}_{TG})}{N} \to 1.$$  

See [7] for more detail.

3. Variance estimation techniques

3.1. Analytical approaches

$M1$: Analytic approaches with two sources of variances. According to [10], the variance of TG estimator was derived under a conditional technique mixed with the delta method. As the $\hat{N}_{TG} = \frac{n}{1 - \sqrt{\frac{f_1}{S}}}$, the variance of $\hat{N}_{TG}$ arise from two sources; the first one is influenced by the random variable $n$, and the second one come from the predictive value $\hat{g}_0$ based on the observed $n$ individuals. Thus

$$Var(\hat{N}) = Var_n \left\{ E(\hat{N}|n) \right\} + E_n \left\{ Var(\hat{N}|n) \right\}.$$  \hspace{1cm} (3.1)

From the first term in (3.1), the approximation $E(\hat{N}|n)$ can be justified by the delta-method, so that $E(\hat{N}|n) \approx \frac{n}{1-g_0}$,

$$\tilde{Var}_n \left\{ E(\hat{N}|n) \right\} = \frac{1}{(1-g_0)^2} Var(n) = \frac{N(1-g_0)g_0}{(1-g_0)^2} = \frac{n\sqrt{\frac{f_1}{S}}}{(1 - \sqrt{\frac{f_1}{S}})^2}. \hspace{1cm} (3.2)$$

Since $E(n) = N(1-g_0)$ and $\hat{g}_0(TG) = \sqrt{\frac{f_1}{S}}$. Additionally, for the second term of (3.1) is estimated as

$$E_n \left\{ Var(\hat{N}|n) \right\} = E_n \left\{ Var \left( \frac{n}{1 - \sqrt{\frac{f_1}{S}}} |n \right) \right\} = E_n \left\{ n^2 Var \left( \frac{1}{1 - \sqrt{\frac{f_1}{S}}} \right) \right\} = n^2 \left\{ \frac{S + f_1}{4S^2 \left( 1 - \sqrt{\frac{f_1}{S}} \right)^4} \right\}. \hspace{1cm} (3.3)$$

Finally, the variance of TG estimator is given as

$$\tilde{Var}(\hat{N}_{TG}) = \frac{n\sqrt{\frac{f_1}{S}}}{(1 - \sqrt{\frac{f_1}{S}})^2} + n^2 \left\{ \frac{S + f_1}{4S^2 \left( 1 - \sqrt{\frac{f_1}{S}} \right)^4} \right\}.$$  \hspace{1cm} (3.4)
M2: Analytic approach with one source of variance. The second way of variance estimation, we treat \( n \) as fixed therefore the variance of \( \hat{N}_{TG} \) come from only one source as following:

\[
\text{Var}(\hat{N}_{TG}) = n^2 \text{Var} \left( \frac{1}{1 - \frac{1}{S}} \right) = n^2 \left\{ \frac{S + f_1}{4S^2 \left( 1 - \sqrt{\frac{f_1}{S}} \right)^2} \right\}. \quad (3.5)
\]

This approach is expected more realistic for capture-recapture data in practice since it avoids a computational software system error [9], [11].

3.2. Resampling approaches

The resampling techniques was proposed as alternative choices in the literature [12] to estimate variance and construct confidence interval in capture-recapture framework. Since the variance of the TG estimation by the normalized approximation method is valid asymptotically, large sample sizes are required as a basic assumption. In reality, a target population size can be small and this might lead to underestimation or overestimation of the variance. Additionally, model uncertainty in the analytic approaches usually occur due to the model structure depends on data from only one sample. These limitations of analytic variance approach might lead to invalid confidence intervals for the true population size and achieve an unsatisfactory coverage probability. This study, the resampling approaches; True Bootstrap (M3); Imputed Bootstrap (M4); Reduced Bootstrap (M5) are suggested as an alternative choices to measure variance and construct confidence intervals. The advantages of using the bootstrap approaches are not reverent to the consideration model and straightforward to implement by using computer programming. The basic assumption of those approaches is that the capture-recapture history can be defined by the multinomial likelihood as

\[
\left( N, f_0, f_1, f_2, f_3, \ldots, f_m \right) \sim p_0^{f_0} p_1^{f_1} p_2^{f_2} \ldots p_m^{f_m}
\]

The algorithms for variance estimation are given as follows:

Step 1: Estimating population size: \( \hat{N}_{TG} = \frac{n}{1 - \sqrt{\frac{1}{S}}} \).

Step 2: For \( T \) samples consist of individuals from the ordinary data as:

(i) True Bootstrap (M3)

Let \( \hat{p}_{M3} = \{ \frac{f_0}{N}, \frac{f_1}{N}, \frac{f_2}{N}, \ldots, \frac{f_m}{N} \} \). This method can be use to estimate variance of population size estimator only if the population size is known in advance. Each individual is drawn from the multinomial distribution with parameter \( N \) and \( \hat{p}_{M3} \).

(ii) Imputed Bootstrap (M4)

Let \( \hat{p}_{M4} = \{ \frac{\hat{f}_0}{\hat{N}_{TG}}, \frac{\hat{f}_1}{\hat{N}_{TG}}, \frac{\hat{f}_2}{\hat{N}_{TG}}, \ldots, \frac{\hat{f}_m}{\hat{N}_{TG}} \} \). Under the M4, each individual is drawn from the multinomial distribution with parameter \( \hat{N}_{TG} \) and \( \hat{p}_{M4} \).

(iii) Reduced Bootstrap (M5)

Let \( \hat{p}_{M5} = \{ \frac{\hat{f}_1}{n}, \frac{\hat{f}_2}{n}, \ldots, \frac{\hat{f}_m}{n} \} \). Under the M5, each individual is drawn from the multinomial distribution with parameter \( n \) and \( \hat{p}_{M5} \).

Step 3: For each sample, \( \hat{N}_{TG} \) is estimated based on the geometric distribution model.

Step 4: Estimate the variance of \( \hat{N}_{TG} \) under the \( T \) bootstrapped samples as

\[
\text{Var}(\hat{N}) = \frac{1}{T - 1} \sum_{t=1}^{T} \left( \hat{N}_{TG,t} - E(\hat{N}_{TG}) \right) \quad (3.6)
\]
4. Simulation

The simulation study is taken following the geometric distribution with different level of heterogeneity. The objective of this is to investigate the performance of proposed methodologies. Then $X \sim \text{Geo}(p)$ where $p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7$ and $0.8$. The population size $N$ is set at $N = 50, 80$ and $100$ for small sizes, $N = 250, 500$ and $750$ for medium sizes, and $N = 1,000, 5,000, 10,000$ and $50,000$ for large sizes. It is $1,000$ samples from the true model.

4.1. Bias and Variance of TG estimator

Figure 1 shows the behaviour of TG estimator under different setting. The performance of estimator is evaluated in term of relative bias; $\text{RBias} = \frac{1}{N}E(\widehat{N} - N)$ and relative variance; $\frac{1}{N}\text{Var}(\widehat{N})$ As can be seen that the TG estimator performs very well under the $\text{Geo}(p)$ generated data. Increasing population sizes lead to a better performance of estimation in term of both variance and bias. Hence, it can be confirmed that the TG estimator has an asymptotic unbiased property with respect to population size. Another interesting point is that the TG estimator are likely to be identical with the parametric estimator under the geometric distribution, namely MLEGeo. The simulation results were shown that they are the best choices for estimating population size in CR data if and only if data follow the geometric distribution.

4.2. Comparing variance estimation

The objective of this section is to compare the performance of five variance approximation methods. For the convenience, the ratio of standard error of estimation, which is defined as the estimated standard error from each approach divided by the true standard error. The simulation result is represented in Figure 2. The findings suggest that imputed bootstrap might be the best choice for estimating variance of TG estimator, especially when the population size is greater than $250$. The analytic approach with two sources of variances remains realistic and valid method. The analytic method with one source of variance and reduced bootstrap underestimate in all situations. We do not suggest both of them to estimate the variance of the TG estimator. The true bootstrap tends to identical estimation with the imputed bootstrap but it can be use when the population size is known.

4.3. Comparing confidence intervals

The performance of a confidence interval (CI) method is assessed by how often the confidence interval covers the true value of $N$. The coverage probability is used for comparing the
Figure 2. Ratio of standard errors from five methods to the true standard error when data are generated under a geometric distribution; M1: Analytic approach with two source of variance, M2: Analytic approach with one source of variance M3: True bootstrap, M4: Imputed bootstrap and M5: Reduced bootstrap.

performance of five variance estimation methods with the nominal coverage probability, setting at 0.95 or 95%. A common procedure of 95% CI is construed by \( \hat{N} \pm Z_{0.975} \sqrt{\text{Var}(\hat{N})} \). For the bootstrap methods, considering all of \( \hat{N}_t, t = 1, 2, 3, ..., 1000 \), the standard error can be computed by taking the sample standard deviation from the bootstrap samples results. The approximate of 95% CI of \( \hat{N} \) can be achieved by using the percentile methods. Since the simulation has \( R \) replication runs, the coverage probability (Cov) can be calculated as \( \text{Cov} = \left( \frac{1}{R} \sum_{r=1}^{R} A_t \right) \times 100 \), where \( A_t \) equal to 1 if the true value \( N \) is in the target confidence interval, and 0 otherwise.

Figure 3. Coverage probabilities from five methods to the true standard error when data are generated under a geometric distribution; M1: Analytic approach with two source of variance, M2: Analytic approach with one source of variance M3: True bootstrap, M4: Imputed bootstrap and M5: Reduced bootstrap.

We compare five methods to get confidence interval of the TG estimator. The simulation results as Figure 3 suggest that the imputed bootstrap performs the best to construct the confidence intervals of the TG estimator. Overall, this method provides coverage probabilities.
close to the nominal 95%. For the analytic approach with one source of variance and reduce bootstrap methods suffer in providing a good performance of coverages for all situations. The confidence intervals of analytic approach with two sources of variances might affect the coverage probability for a small population size in some cases; however, it remains useful in practice. Since it is not only has a good performance but also easy to compute and do not require computational programming method. We recommend to use the imputed bootstrap and analytic approach with two sources of variances for investigating uncertainty of the TG estimator.

5. Real data examples

5.1. Golf tees data

The first example is the golf tees data, \( N = 250 \) groups of golf tees were placed in survey regain at the University of St Andrews, see [13]. This example is useful for comparing the performance of several variance estimation methods in the case of a true value of population size is known. A total of \( n = 162 \) groups of golf tees were observed, so the number of hidden in the experiment need to be estimated. The frequency distribution is given as \((f_0, f_1, f_2, \ldots, f_m) = (88, 46, 28, 21, 13, 23, 14, 6, 11)\). We use the ratio plots [14] for investigating the suitable models as Figure 4, we expect that the TG estimator would perform well. The TG estimator provides \( \hat{N} = 228 \). Interestingly, variance estimation from M4 is slightly greater than M1 and M4. As we expect M2 and M5 are not appropriated for constructing confidence intervals, they do not cover the true value \( N \) because of underestimate of variance of estimation. We suggest to use M1 or M4 to estimate variance and construct the confidence interval for golf tees data.

![Frequency Distribution of Observed Data](image)

![Ratio of frequency](image)

**Figure 4.** Frequency distribution (left) and the ratio plot of Poisson and Geometric models

| Method | \((\hat{N})_{M_{true}}\) | \((\hat{N})_{M_{est}}\) | S.E. \((\hat{N})\) | 95% CI | Length |
|--------|-----------------|-----------------|----------------|-------|-------|
| M 1    | 228             | -               | 12.03          | (204 – 252) | 48    |
| M 2    | 228             | -               | 7.165          | (214 – 242) | 28    |
| M 3    | 228             | 228             | 13.20          | (203 – 255) | 52    |
| M 4    | 228             | 228             | 12.35          | (204 – 253) | 49    |
| M 5    | 229             | 228             | 7.77           | (214 – 245) | 31    |

**Table 1.** Comparison of five methods of variance estimation for golf tees data
5.2. Heroin users in Bangkok data

A further example based upon the geometric distribution is the heroin drug user in Bangkok data which was analysed in [15] using the Poisson mixture model to estimate the number of heroin drug users in Bangkok in 2003. This data set is collected from 61 health treatment centres between 1 October and 31 December 2001. From Figure 5, the log ratios suggest to select the population size estimator based on the geometric model. The resulting estimated population size by using the TG estimator is $\hat{N} = 12,175$. Since the true value of the number of drug users is unknown in advance, so that M3 can not use for estimating variance and produce confidence interval in this case. M1 and M3 are more realistic than M2 and M5, which often underestimate of variance, supporting by the simulation study.

![Frequency Distribution of Observed Data](image_url)

**Figure 5.** Frequency distribution (left) and the ratio plot of Poisson and geometric models

| Table 2. Comparison of five methods of variance estimation for heroin users in Bangkok data |
|---------------------------------------------------------------|
| **Method** | $(\hat{N})_{Mean}$ | $(\hat{N})_{Med}$ | $S.E.(\hat{N})$ | 95%CI Length |
| M 1 | 12,175 | - | 73.99 | (12,030 – 12,320) | 290 |
| M 2 | 12,175 | - | 41.41 | (12,094 – 12,256) | 162 |
| M 3 | - | - | - | - | - |
| M 4 | 12,176 | 12,176 | 76.48 | (12,026 – 12,325) | 299 |
| M 5 | 12,175 | 12,175 | 45.57 | (12,087 – 12,266) | 179 |

5.3. Dengue village in Chaing Mai in 2013

A case study on dengue fever in Chaing Mai was analysed in [16]. The data were collected from the laboratory to confirm dengue patients in 2013. To estimate total of patients by using the TG estimator that is confirmed by the ratio plots as Figure 6. Finding that an estimate of dengue patients is 1,617, which is larger than using the MLEGeo estimator as original analysis. We suggest M3 and M4 to investigate the uncertainty for this data that was supporting by the simulation study.

6. Conclusion

Capture-recapture approaches are widely us to estimate the hidden population size in many areas, and the TG estimator was proposed as an alternative choice to solve an unobserved heterogeneity in CR data. The TG estimator is mentioned as an asymptotically unbiased estimator with respect to the population size under the geometric distribution. This work we compare an analytic approaches and bootstrap techniques to estimate uncertainty of population
Figure 6. Frequency distribution (left) and the ratio plot of Poisson and geometric models

Table 3. Comparison of five methods of variance estimation for Dengue per village in Chaing Mai in 2013

| Method | $(\hat{N})_{\text{mean}}$ | $(\hat{N})_{\text{Med}}$ | $S.E. (\hat{N})$ | 95% CI        | Length |
|--------|--------------------------|--------------------------|----------------|----------------|--------|
| M 1    | 1,617                    | -                        | 19.88          | (1,578 – 1,656) | 78     |
| M 2    | 1,617                    | -                        | 9.27           | (1,599 – 1,635) | 36     |
| M 3    |                          |                           |                |                |        |
| M 4    | 1,663                    | 1,663                    | 25.82          | (1,615 – 1,715) | 100    |
| M 5    | 1,663                    | 1,661                    | 17.28          | (1,632 – 1,700) | 68     |

size in the capture-recapture data. The further result is that the analytic approach with one source of variance, which was often suggested for investigating uncertainty in capture-recapture study [9], [17], presents a poor coverage probability. The analytic approach and the reduce bootstrap should discontinue in capture-recapture study. The imputed bootstrap is the most accurate but it require a computer intensive programming. Overall, the analytic approach with two sources of variance behaves close to the imputed bootstrap. Therefore, we suggest to use the analytic approach with two sources of variance to estimate variance and construct confidence interval in practice.

Acknowledgement(s) : I would like to thank the referees for his comments and suggestions on the manuscript. This work was supported by Thaksin University.

[1] Darroch, J. N., and D. Ratcliff, A note on capture-recapture estimation, Biometrics (1980) 149-153.
[2] Wesson, Paul, Richard Lechtenberg, Arthur Reingold, Willi McFarland, and Neena Murgai, Evaluating the Completeness of HIV Surveillance Using Capture–Recapture Models, Alameda County, California, AIDS and Behavior. 22 (7) (2018) 2248-2257.
[3] Baujat, Geneviève, Rémy Choquet, Stéphane Bouée, Viviane Jeambat, Laurène Courouve, Amélie Ruel, Caroline Michot et al, Prevalence of fibrodysplasia ossificans progressiva (FOP) in France: an estimate based on a record linkage of two national databases, Orphanet journal of rare diseases. 12 (1) (2017) 123.
[4] Van Der Heijden, Peter GM, Maarten Cruyff, and Hans C. Van Houwelingen, Estimating the size of a criminal population from police records using the truncated Poisson regression model, Statistica Neerlandica. 57(3) (2003) 289-304.
[5] Accettura, Nicola, Giovanni Neglia, and Luigi Alfredo Grieco, The Capture-Recapture approach for population estimation in computer networks, Computer Networks. 89 (2015) 107-122.
[6] Yauck, Mamadou, Louis-Paul Rivest, and Greg Rothman, Capture-Recapture Methods for Data on the Activation of Applications on Mobile Phones, Journal of the American Statistical Association. (2018) 1-33.
[7] Anan, Orasa, Dankmar Böhning, and Antonello Maruotti, On the Turing estimator in capture-recapture count data under the geometric distribution, Metrika. (2018) 1-24.
[8] Chao, Anne, Estimating the population size for capture-recapture data with unequal catchability, Biometrics. (1987) 783-791.
[9] Niwitpong, Sa-aat, Dankmar Böhning, Peter GM van der Heijden, and Heinz Holling, Capture-recapture
estimation based upon the geometric distribution allowing for heterogeneity, Metrika. 76 (4) (2013) 495-519.

[10] Böhning, Dankmar, A simple variance formula for population size estimators by conditioning, Statistical Methodology. 5 (5) (2008) 410-423.

[11] Böhning, Dankmar, Ratio Plot and Ratio Regression with Applications to Social and Medical Sciences, Statistical Science. 31 (2016) 205-218.

[12] Anan, Orasa, Böhning, Dankmar, and Antonello Maruotti, Uncertainty estimation in heterogeneous capture-recapture count data, Journal of Statistical Computation and Simulation. 87 (10) (2017) 2094-2114.

[13] Anan, Orasa, Böhning, Dankmar, and Antonello Maruotti, Population size estimation and heterogeneity in capture-recapture data: a linear regression estimator based on the Conway–Maxwell–Poisson distribution, Statistical Methods and Applications. 26 (1) (2017) 49-79.

[14] Böhning, Dankmar, Ratio plot and ratio regression with applications to social and medical sciences, Statistical Science. 31 (2) (2016) 205-218.

[15] Böhning, Dankmar, Busaba Suppawattanabodee, Wilai Kusolvisitkul, and Chukiat Viwatwongkasem, Estimating the number of drug users in Bangkok 2001: A capture-recapture approach using repeated entries in one list, European Journal of Epidemiology. 19(12) (2004) 1075.

[16] Böhning, Dankmar, Peter GM van der Heijden, and John Bunge, eds. Capture-recapture Methods for the Social and Medical Sciences. CRC Press, 2017.

[17] Böhning, Dankmar, Power series mixtures and the ratio plot with applications to zero-truncated count distribution modelling, Metron. 73 (2) (2015) 201-216.