

Estimating population sizes with the Rasch model

Chang Xuan Mao¹ · Cuiying Yang² ·
Yitong Yang¹ · Wei Zhuang¹

Received: 24 April 2015 / Revised: 16 January 2016 / Published online: 31 March 2016
© The Institute of Statistical Mathematics, Tokyo 2016

Abstract The Rasch model has been used to estimate the unknown size of a population from multi-list data. It can take both the list effectiveness and individual heterogeneity into account. Estimating the population size is shown to be equivalent to estimating the odds that an individual is unseen. The odds parameter is nonidentifiable. We propose a sequence of estimable lower bounds, including the greatest one, for the odds parameter. We show that a lower bound can be calculated by linear programming. Estimating a lower bound of the odds leads to an estimator for a lower bound of the population size. A simulation experiment is performed and three real examples are studied.

Keywords Capture–recapture · Nonidentifiability · Social security

1 Introduction

Estimating the number of unseen individuals, or equivalently, the population size is of great importance in studies of elusive populations (e.g., drug addicts, street children, victims of domestic violence, etc.). Various health, social, economic, and

✉ Chang Xuan Mao
mao.changxuan@mail.shufe.edu.cn

Cuiying Yang
yang.cuiying@mail.shufe.edu.cn

Yitong Yang
yitongyang@foxmail.com

Wei Zhuang
wei_zhuang@126.com

¹ School of Statistics and Management, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China

² School of Public Economics and Administration, Shanghai University of Finance and Economics, 777 Guoding Road, Shanghai 200433, China

human rights issues are associated with these populations. The problem also occurs in studies of disease populations. Particularly, many countries are facing challenges of aging society, and their social security systems and health-care systems are in need of information on chronic diseases of the elderly (e.g., dementia, diabetes, cardiovascular and cerebrovascular diseases, etc.) (Prohaska et al. 2012).

Several types of models are available (Chao 2001). For numerous applications, the Rasch model is an appropriate choice, because it can take both the list effectiveness and individual heterogeneity into account (e.g., Agresti 1994; Coull and Agresti 1999; Bartolucci and Forcina 2001). However, Coull and Agresti (1999) recognized the nonparametric nonidentifiability, and the flatness of a log-likelihood. They also found that competing models may fit the data so well that traditional goodness-of-fit criteria can not distinguish them.

We will provide a thorough investigation on the Rasch model. We reduce the problem of estimating the population size to estimating the odds that an individual is unseen. We show that the odds is nonidentifiable but admits a sequence of lower bounds, among which the greatest one usually deserves our recommendation. These lower bounds are estimated by the plug-in method and lower confidence limits are constructed by the bootstrap method.

The rest of this article is structured as follows. The methods are presented in Sect. 2. A simulation experiment is reported in Sect. 3. Real examples are studied in Sect. 4.

2 Methods

2.1 The Rasch model

A population consists of s individuals possibly included in $J \geq 2$ lists. Let $x_{ij} = 1$ if individual i is in list j and $x_{ij} = 0$ otherwise. Let $a = \text{logit}(\pi) = \log\{\pi/(1 - \pi)\}$ and $\pi = \text{expit}(a) = e^a/(1 + e^a)$. With $\Pr(x_{ij} = 1) = p_{ij}$, assume that $\text{logit}(p_{ij}) = a_i + b_j$, where a_i specifies how easily individual i is included in a list, b_j specifies how effectively list j includes an individual. To uniquely identify $\{a_i, b_j\}$ given $\{p_{ij}\}$, let $\sum_{j=1}^J b_j = \mathbf{b}^\top \mathbf{1} = 0$, where $\mathbf{b} = (b_1, b_2, \dots, b_J)^\top$. With $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iJ})^\top$, it holds that

$$\mathbf{x}_i | a_i \sim \prod_{j=1}^J p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}} = \prod_{j=1}^J \frac{\exp\{(a_i + b_j)x_{ij}\}}{1 + \exp(a_i + b_j)}.$$

If the a_i follow a distribution A (Lindsay et al. 1991), then $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_s$ arise as a random sample from $u(\mathbf{x} | \mathbf{b}, A)$, where

$$\begin{aligned} u(\mathbf{x} | \mathbf{b}, A) &= \int \prod_{j=1}^J \frac{\exp\{(a + b_j)x_j\}}{1 + \exp(a + b_j)} dA(a) \\ &= \exp(\mathbf{x}^\top \mathbf{b}) \int \frac{\exp\{a \sum_{j=1}^J x_j\}}{\prod_{j=1}^J \{1 + \exp(a + b_j)\}} dA(a). \end{aligned}$$

Let $n_x = \sum_{i=1}^s I(x_i = x)$ for $x \in \mathcal{X} = \{0, 1\}^J$. The number of observed individuals is $n_+ = \sum_{i=1}^s I(x_i \neq \mathbf{0})$, which is binomial with size s and probability $1 - u(\mathbf{0}|\mathbf{b}, A)$. Given $u(\mathbf{0}|\mathbf{b}, A)$, the maximum likelihood estimator for s is the integer part of $n_+ / \{1 - u(\mathbf{0}|\mathbf{b}, A)\}$. Estimating s becomes estimating $u(\mathbf{0}|\mathbf{b}, A)$. To estimate \mathbf{b} and A , we will consider two conditional likelihoods. Let $y_i = \mathbf{x}_i^\top \mathbf{1} = \sum_{j=1}^J x_{ij}$. By noting that

$$\Pr(\mathbf{x}_i = \mathbf{x} | y_i) = \frac{\Pr(\mathbf{x}_i = \mathbf{x})}{\sum_{\{\chi \in \mathcal{X} : \chi^\top \mathbf{1} = y_i\}} \Pr(\mathbf{x}_i = \chi)},$$

we have, from Andersen (1970),

$$\mathbf{x}_i | y_i = \mathbf{x}_i^\top \mathbf{1} \sim h(\mathbf{x}|\mathbf{b}) = \frac{u(\mathbf{x}|\mathbf{b}, A)}{\sum_{\{\chi \in \mathcal{X} : \chi^\top \mathbf{1} = \mathbf{x}^\top \mathbf{1}\}} u(\chi|\mathbf{b}, A)} = \frac{\exp(\mathbf{x}^\top \mathbf{b})}{\gamma(\mathbf{x}^\top \mathbf{1}|\mathbf{b})},$$

where $\gamma(j|\mathbf{b}) = \sum_{\{\mathbf{x} \in \mathcal{X} : \mathbf{x}^\top \mathbf{1} = j\}} \exp(\mathbf{x}^\top \mathbf{b})$. A conditional log-likelihood is

$$\ell(\mathbf{b}) = \sum_{i=1}^s \log h(\mathbf{x}_i|\mathbf{b}) = \sum_{\mathbf{x} \in \mathcal{X}} n_x \log h(\mathbf{x}|\mathbf{b}). \tag{1}$$

Because $\gamma(0|\mathbf{b}) = \sum_{\{\mathbf{x} \in \mathcal{X} : \mathbf{x}^\top \mathbf{1} = 0\}} \exp(\mathbf{x}^\top \mathbf{b}) = \exp(\mathbf{0}^\top \mathbf{b}) = 1$, and $h(\mathbf{0}|\mathbf{b}) = \exp(\mathbf{0}^\top \mathbf{b}) = 1$, one has $\log h(\mathbf{0}|\mathbf{b}) = 0$. Although \mathbf{n}_0 is unknown in (1), $\mathbf{n}_0 \log h(\mathbf{0}|\mathbf{b}) = 0$ does not contribute to $\ell(\mathbf{b})$.

Let $\delta(a)$ be degenerate at a . Given $A = \delta(a)$, y_1, y_2, \dots, y_s arise as a random sample from

$$\Pr(y = j | \delta(a)) = \sum_{\{\mathbf{x} \in \mathcal{X} : \mathbf{x}^\top \mathbf{1} = j\}} u(\mathbf{x}|\mathbf{b}, \delta(a)) \propto \gamma(j|\mathbf{b}) \exp(aj).$$

With $\pi = \text{expit}(a)$, one has

$$\Pr(y = j | \delta(\pi)) \propto \gamma(j|\mathbf{b}) \pi^j (1 - \pi)^{-j} \propto \gamma(j|\mathbf{b}) \pi^j (1 - \pi)^{J-j},$$

and consequently, $\Pr(y = j | \delta(\pi)) = \gamma(j|\mathbf{b}) \pi^j (1 - \pi)^{J-j} / \sum_{m=0}^J \gamma(m|\mathbf{b}) \pi^m (1 - \pi)^{J-m}$. Generally, write $A = \sum_{t=1}^T p_t \delta(a_t)$. Let $\pi_t = \text{expit}(a_t)$ and $G = \sum_{t=1}^T p_t \delta(\pi_t)$. Note that $\Pr(y_i = y | G) = \nu(j|\mathbf{b}, G)$, where

$$\nu(j|\mathbf{b}, G) = \int \frac{\gamma(j|\mathbf{b}) \pi^j (1 - \pi)^{J-j}}{\sum_{m=0}^J \gamma(m|\mathbf{b}) \pi^m (1 - \pi)^{J-m}} dG(\pi), \quad j = 0, 1, \dots, J.$$

Let $n_j = \sum_{i=1}^s I(y_i = j)$. Conditioning on $n_+ = \sum_{j=1}^J n_j$, $(n_1, n_2, \dots, n_J)^\top$ is multinomial with probabilities $\nu(j|\mathbf{b}, G) / \{1 - \nu(0|\mathbf{b}, G)\}$. Using the fact that

$$1 - \nu(0|\mathbf{b}, \delta(\pi)) = \frac{\sum_{m=1}^J \gamma(m|\mathbf{b}) \pi^m (1 - \pi)^{J-m}}{\sum_{m=0}^J \gamma(m|\mathbf{b}) \pi^m (1 - \pi)^{J-m}},$$

we reparameterize G by $Q = \sum_{t=1}^T q_t \delta(\pi_t)$, where

$$q_t = \frac{1 - \nu(0|\mathbf{b}, \delta(\pi_t))}{1 - \nu(0|\mathbf{b}, G)} p_t \quad \text{or} \quad dQ(\pi) = \frac{1 - \nu(0|\mathbf{b}, \delta(\pi))}{1 - \nu(0|\mathbf{b}, G)} dG(\pi).$$

We introduce a mixture $f(\cdot|\mathbf{b}, Q)$, where

$$f(j|\mathbf{b}, Q) = \int \frac{\gamma(j|\mathbf{b})\pi^j(1 - \pi)^{J-j}}{\sum_{m=1}^J \gamma(m|\mathbf{b})\pi^m(1 - \pi)^{J-m}} dQ(\pi), \quad 1 \leq j \leq J.$$

It is clear that $f(j|\mathbf{b}, Q) = \nu(j|\mathbf{b}, G) / \{1 - \nu(0|\mathbf{b}, G)\}$ as

$$\begin{aligned} \frac{\nu(j|\mathbf{b}, G)}{1 - \nu(0|\mathbf{b}, G)} &= \int \frac{\gamma(j|\mathbf{b})\pi^j(1 - \pi)^{J-j}}{\sum_{m=0}^J \gamma(m|\mathbf{b})\pi^m(1 - \pi)^{J-m}} \times \frac{1}{1 - \nu(0|\mathbf{b}, G)} dG(\pi) \\ &= \int \frac{\gamma(j|\mathbf{b})\pi^j(1 - \pi)^{J-j}}{\sum_{m=1}^J \gamma(m|\mathbf{b})\pi^m(1 - \pi)^{J-m}} \times \frac{1 - \nu(0|\mathbf{b}, \delta(\pi))}{1 - \nu(0|\mathbf{b}, G)} dG(\pi). \end{aligned}$$

This means that another conditional log-likelihood is

$$\ell(Q|\mathbf{b}) = \sum_{j=1}^J n_j \log f(j|\mathbf{b}, Q), \tag{2}$$

which arises from n_1, n_2, \dots, n_J by conditioning on n_+ .

Finally, estimating the probability $\nu(0|\mathbf{b}, G) = u(\mathbf{0}|\mathbf{b}, A)$ is equivalent to estimating its odds $\nu(0|\mathbf{b}, G) / \{1 - \nu(0|\mathbf{b}, G)\}$, which can be written as

$$\theta(\mathbf{b}, Q) = \int \frac{(1 - \pi)^J}{\sum_{m=1}^J \gamma(m|\mathbf{b})\pi^m(1 - \pi)^{J-m}} dQ(\pi). \tag{3}$$

By extending the domain of $f(\cdot|\mathbf{b}, Q)$ to $\{0, 1, \dots, J\}$, one has $\theta(\mathbf{b}, Q) = f(0|\mathbf{b}, Q)$. We will restrict the domain of $f(j|\mathbf{b}, Q)$ to be $\{1, 2, \dots, J\}$ in the following. Given \mathbf{b} , $\theta(\mathbf{b}, Q)$ is nonidentifiable in the sense that there exist mixing distributions P and Q with $f(\cdot|\mathbf{b}, Q) = f(\cdot|\mathbf{b}, P)$ but $\theta(\mathbf{b}, P) \neq \theta(\mathbf{b}, Q)$, which can be shown using the techniques in [Mao \(2007a\)](#). One may consider its upper or lower bounds. However, there is no upper bound of $\theta(\mathbf{b}, Q)$ in any neighborhood of $f(\cdot|\mathbf{b}, Q)$ ([Mao and Lindsay 2007](#)). There are lower bounds in several types of models (e.g., [Chao 1984, 1987, 1989](#); [Chao et al. 2006](#); [Mao 2006, 2007a](#); [Mao and Lindsay 2007](#); [Rivest and Baillargeon 2007](#); [Rivest 2011](#)). We will show that $\theta(\mathbf{b}, Q)$ in (3) admits some lower bounds.

2.2 Lower bounds

To introduce a lower bound, given $j = 1, 2, \dots, J$, consider

$$C = \left\{ (\theta(\mathbf{b}, \delta(\pi)), f(1|\mathbf{b}, \delta(\pi)), \dots, f(j|\mathbf{b}, \delta(\pi)))^\top : \pi \in (0, 1] \right\},$$

which is curve in the Euclidean space \mathcal{R}^{j+1} . Its convex hull is

$$\text{conv}(C) = \left\{ (\theta(\mathbf{b}, Q), f(1|\mathbf{b}, Q), \dots, f(j|\mathbf{b}, Q))^\top : \text{supp}(Q) \subset (0, 1] \right\},$$

where $\text{supp}(Q)$ contains support points of Q . Let $\varphi_j(f(\cdot|\mathbf{b}, Q))$ denote

$$\inf\{\theta(\mathbf{b}, P) : f(m|\mathbf{b}, P) = f(m|\mathbf{b}, Q), 1 \leq m \leq j, \text{supp}(P) \subset (0, 1]\}. \quad (4)$$

By doing this, we consider the set of mixing distributions P whose density values at $m = 1, 2, \dots, j$ match those of Q exactly. The corresponding set of the odds values $\theta(\mathbf{b}, P)$ will admit an infimum, which is a functional of $f(\cdot|\mathbf{b}, Q)$ because it depends on the density values $f(1|\mathbf{b}, Q), f(2|\mathbf{b}, Q), \dots, f(j|\mathbf{b}, Q)$. Clearly, $\varphi_j(f(\cdot|\mathbf{b}, Q)) \leq \theta(\mathbf{b}, Q)$, i.e., $\varphi_j(f(\cdot|\mathbf{b}, Q))$ is a lower bound because Q is also in this set of mixing distributions.

To calculate $\varphi_j(f(\cdot|\mathbf{b}, Q))$, we rewrite (4) as

$$\inf \left\{ \xi_0 : (\xi_0, \xi_1, \dots, \xi_j)^\top \in \text{conv}(C), \xi_m = f(m|\mathbf{b}, Q), 1 \leq m \leq j \right\}. \quad (5)$$

We choose a grid of support points $\mathcal{G}_K = \{\pi_k = k/K, 1 \leq k \leq K\}$ given some K , and introduce an optimization problem as follows,

$$\begin{aligned} & \min \sum_{k=1}^K \theta(\mathbf{b}, \delta(\pi_k)) \cdot w_k, \\ \text{subject to } & \begin{cases} \sum_{k=1}^K f(m|\mathbf{b}, \delta(\pi_k)) \cdot w_k = f(m|\mathbf{b}, Q), & 1 \leq m \leq j, \\ \sum_{k=1}^K w_k = 1, \\ w_k \geq 0, & 1 \leq k \leq K. \end{cases} \end{aligned} \quad (6)$$

It is clear that (6) is a linear programming problem. The problem in (6) is obtained from (5) when we replace the convex hull of C with the convex hull of \mathcal{G}_K . The lower bound $\varphi_j(f(\cdot|\mathbf{b}, Q))$ in (5) is the limit of the minimum of (6) when K goes to infinity. Numerically, the minimum of (6) for sufficiently large K will be treated as $\varphi_j(f(\cdot|\mathbf{b}, Q))$.

The lower bounds $\varphi_j(f(\cdot|\mathbf{b}, Q))$ are related because, clearly,

$$\varphi_1(f(\cdot|\mathbf{b}, Q)) \leq \dots \leq \varphi_{J-1}(f(\cdot|\mathbf{b}, Q)) = \varphi_J(f(\cdot|\mathbf{b}, Q)).$$

In particular, the greatest lower bound is $\varphi_J(f(\cdot|\mathbf{b}, Q)) = \varphi_{J-1}(f(\cdot|\mathbf{b}, Q))$, where the identity holds due to the constraint $\sum_{j=1}^J f(j|\mathbf{b}, Q) = 1$. It is possible that $\theta(\mathbf{b}, Q) = \varphi_J(f(\cdot|\mathbf{b}, Q))$, for instance, $Q = \delta(\pi)$. If Q has many support points, then $\theta(\mathbf{b}, Q) > \varphi_J(f(\cdot|\mathbf{b}, Q))$ (Mao 2008b).

Finally, note that n_+ is binomial with size s and probability $1 - \nu(0|\mathbf{b}, G)$ so that $E(n_+) = s\{1 - \nu(0|\mathbf{b}, G)\} = s/\{1 + \theta(\mathbf{b}, Q)\}$. This means that the population size s satisfies $s = E(n_+) \cdot \{1 + \theta(\mathbf{b}, Q)\}$. This motivates us to define a lower bound $s_{j, \text{LB}} \leq s$, where

$$s_{j, \text{LB}} = E(n_+) \cdot \{1 + \varphi_j(f(\cdot|\mathbf{b}, Q))\}. \tag{7}$$

The greatest lower bound $s_{J, \text{LB}}$ of s will be rewritten as s_{GLB} , which is recommended usually. When J is relatively large, small n_x can be frequently seen so that it is often difficult to estimate s_{GLB} well. If this is the case, those smaller lower bounds, in particular, $s_{2, \text{LB}}$ or $s_{3, \text{LB}}$, can be of interest.

2.3 Estimation

The conditional maximum likelihood estimator for \mathbf{b} is

$$\hat{\mathbf{b}} = \underset{\mathbf{b} \in \mathcal{R}^J, \mathbf{b}^T \mathbf{1} = 0}{\text{argmax}} \ell(\mathbf{b}), \tag{8}$$

which can be obtained easily by a commonly used optimization routine.

A nonparametric maximum likelihood estimator (NPMLE) is

$$\hat{Q} = \underset{\text{supp}(Q) \subset [0, 1]}{\text{argmax}} \ell(Q|\hat{\mathbf{b}}). \tag{9}$$

Calculating \hat{Q} can be done via the EM algorithm.

Given an estimator $\hat{\mathbf{b}}$ in (8) and an estimator \hat{Q} in (9), to obtain $\varphi_j(f(\cdot|\hat{\mathbf{b}}, \hat{Q}))$, we can replace \mathbf{b} with $\hat{\mathbf{b}}$ and $f(\cdot|\mathbf{b}, Q)$ with $f(\cdot|\hat{\mathbf{b}}, \hat{Q})$ in (6). With n_+ substituted for $E(n_+)$ in (7), $s_{j, \text{LB}}$ can be estimated by

$$\hat{s}_{j, \text{LB}} = n_+ + n_+ \cdot \varphi_j(f(\cdot|\hat{\mathbf{b}}, \hat{Q})). \tag{10}$$

It is clear that $\hat{s}_{j, \text{LB}}$ in (10) can serve as a lower bound estimator for the population size s , although $\hat{s}_{j, \text{LB}}$ estimates $s_{j, \text{LB}}$ consistently.

The nonexistence of upper bounds also presents finite nonparametric upper confidence limits from achieving their nominal confidence levels. Although two-sided confidence intervals are frequently reported in the literature, their coverage probabilities are in doubt. Consequently, we will focus on nonparametric lower confidence limits. In particular, by the bootstrap method, one can construct a $(1 - \alpha)$ lower confidence limit for a lower bound $s_{j, \text{LB}}$, which is also a $(1 - \alpha)$ lower confidence limit for s as $s_{j, \text{LB}} \leq s$.

By the way, we observe that $\hat{s}_{2, \text{LB}} = \hat{s}_{\text{GLB}}$ is numerically close to the Petersen estimator when $J = 2$. We also observe that the estimator in Rivest and Baillargeon

(2007) is numerically close to $\hat{s}_{2, LB}$, which is the greatest lower bound estimator when $J = 2$ or $J = 3$. This is an alternative insightful justification for their estimator.

3 Simulation

In our simulation experiment, $\mathbf{b} \in \{\mathbf{b}_1, \mathbf{b}_2\}$, $G \in \{G_1, G_2\}$ and $s \in \{s_1, s_2\}$, where $\mathbf{b}_1 = (0, 0, 0)^\top$, $\mathbf{b}_2 = (0.9, -0.5, -0.4)^\top$, $G_1 = 0.95\delta(0.25) + 0.05\delta(1)$, G_2 is a logit-normal distribution with $(\mu, \sigma) = (-1, 1)$, $s_1 = 500$ and $s_2 = 5000$. In each of these eight settings, 1000 samples are generated. We evaluate three estimators for s : the greatest lower bound estimator \hat{s}_{GLB} , the estimator \hat{s}_{LN} produced by the logit-normal model, and the estimator \hat{s}_{HO2} produced by the log-linear model of homogeneous two-factor association shortened as HO2 (Coull and Agresti 1999). In the HO2 model, n_x is a Poisson random variable with mean λ_x , and $\log \lambda_x = \beta_0 + \sum_{j=1}^J x_j \beta_j + \beta_{J+1} (x_2^\top \mathbf{1})$ with $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{J+1})^\top$ estimated by $\hat{\boldsymbol{\beta}}$. Note that $\hat{\boldsymbol{\beta}}$ is calculated by a Poisson regression applied to $\{n_x : \mathbf{x} \in \mathcal{X} \setminus \{\mathbf{0}\}\}$, and s is estimated by $\hat{s}_{HO2} = n_+ + \exp(\hat{\beta}_0)$. The HO2 model with $\beta_{J+1} > 0$ can also be obtained by assuming that G is a logit-reweighted-normal distribution (Lindsay 1986; Rivest 2011).

Each estimator has a target. The target of \hat{s}_{GLB} is s_{GLB} . The target of \hat{s}_{HO2} is $s_{HO2} = s\{1 - \nu(0|\mathbf{b}, G)\} + \exp(\beta_0^\dagger)$, with β_0^\dagger produced by a Poisson regression applied to $\{s \cdot u(\mathbf{x}|\mathbf{b}, G) : \mathbf{x} \in \mathcal{X} \setminus \{\mathbf{0}\}\}$. The target of \hat{s}_{LN} is $s_{LN} = s\{1 - \nu(0|\mathbf{b}, G)\} / \{1 - \nu(0|\mathbf{b}, \mu^\dagger, \sigma^\dagger)\}$, where

$$(\mu^\dagger, \sigma^\dagger) = \operatorname{argmax}_{\mu \in \mathcal{R}, \sigma \in (0, \infty)} \left\{ \sum_{j=1}^J f(j|\mathbf{b}, Q) \log \frac{\nu(j|\mathbf{b}, \mu, \sigma)}{1 - \nu(0|\mathbf{b}, \mu, \sigma)} \right\}.$$

The results are shown in Table 1, including the target, the median, the median absolute deviation as a measure of the variability, and the coverage probability of the 95 % lower confidence limit for the population size. Given an estimate, 1000 resamples are used to calculate a confidence limit.

These estimators admit some nonsignificant estimation bias; the ratio of the median to the target ranges over $[0.998, 1.006]$ (\hat{s}_{GLB}), $[0.971, 1.013]$ (\hat{s}_{LN}), and $[0.983, 1.003]$ (\hat{s}_{HO2}). They also admit some approximation bias; the ratio of the target to the population size ranges over $[0.81, 1]$ (\hat{s}_{GLB}), $[1, 4.89]$ (\hat{s}_{LN}), and $[0.99, 2.35]$ (\hat{s}_{HO2}). Because $s_{GLB} \leq s$, the coverage probabilities are close to their nominal levels. It may happen that $s_{LN} > s$ and $s_{HO2} > s$. In some settings, because the differences $s_{LN} - s$ and $s_{HO2} - s$ are substantially large, the coverage probabilities become small. Conclude that the performance of an estimator is mainly determined by the value of its target.

4 Real examples

We study three examples: street children (children), students infected by hepatitis A virus (hepatitis) and lesbians (lesbian); see Table 2.

Table 1 Simulation results with $s_1 = 500$ and $s_2 = 5000$

Setting	Greatest lower bound			Logit-normal			HO2					
	Target	Median	Mad	Cp	Target	Median	Mad	Cp	Target	Median	Mad	Cp
(G_1, b_1, s_1)	500	499	40	0.94	2444	2372	1981	0.05	1175	1155	365	0.06
(G_2, b_1, s_1)	404	405	24	1	500	496	92	0.96	493	490	83	0.96
(G_1, b_2, s_1)	500	503	35	0.95	2447	2478	2098	0.03	1175	1178	362	0.03
(G_2, b_2, s_1)	412	411	22	1	500	500	81	0.96	495	494	72	0.96
(G_1, b_1, s_2)	5000	5003	127	0.96	24440	24695	8442	0	11750	11785	1234	0
(G_2, b_1, s_2)	4043	4047	72	1	5000	5014	299	0.95	4933	4945	262	0.97
(G_1, b_2, s_2)	5000	4998	112	0.95	24472	24614	8014	0	11750	11749	1146	0
(G_2, b_2, s_2)	4115	4119	64	1	5000	5002	264	0.96	4946	4952	236	0.97

Target a parameter consistently estimated by the estimator, median the sample median of an estimator, mad the median absolute deviation, cp the coverage probability of the 95 % lower confidence limit

Table 2 The counts in three examples children, hepatitis and lesbian

x_1	x_2	x_3	Children	Hepatitis	Lesbian	
					$x_4 = 0$	$x_4 = 1$
0	0	0	?	?	?	589
0	0	1	133	63	208	143
0	1	0	64	55	281	44
0	1	1	13	18	28	19
1	0	0	83	69	534	104
1	0	1	8	17	53	48
1	1	0	12	21	64	27
1	1	1	3	28	20	23

4.1 Street children (children)

Street children live in culverts, train stations, public landfills, deprived of healthy food, sanitary facilities and education opportunities, involved in illegal activities such as robbery and drug abuse, and exposed to sexually transmitted infections (Shukla 2005). The number of street children is an important quantity in planning intervention programs. For instance, the public prosecutor office in Arapiraca, Brazil, had an official list of street children, and two street surveys were also conducted and each produced a list (Bezerra et al. 2011); see Table 2.

The estimate $\hat{\mathbf{b}} = (-0.098, -0.256, 0.354)^\top$ reflects the fact that each of the two street surveys captured fewer street children than the public prosecutor office. The likelihood ratio test statistic, approximately following χ^2_2 if $\mathbf{b} = \mathbf{0}$, is 21.8 with p value 1.8×10^{-5} (Mao 2007b). In children, $n_1 = 280$, $n_2 = 33$ and $n_3 = 3$. Three subfamilies of the Rasch mixture model are applied: the degenerate model, the logit-normal model and the nonparametric model. In the degenerate model, the estimate for $Q = \delta(\pi)$ is $\hat{Q}_0 = \delta(0.115)$; in the logit-normal model, the estimates for (μ, σ) is $(-3.543, 1.146)$; in the nonparametric model, there are infinitely many NPMLE \hat{Q} , and two of them are $\hat{Q} = 0.0032\delta(0) + 0.9913\delta(0.1061) + 0.0055\delta(1)$ and $\hat{Q} = 0.9945\delta(0.1057) + 0.0055\delta(1)$. Table 3 presents \hat{n}_j and \hat{n}_0 for each model (e.g., $\hat{n}_j = n_+ f(j|\hat{\mathbf{b}}, \hat{Q}_0)$ and $\hat{n}_0 = n_+ \theta(\hat{\mathbf{b}}, \hat{Q}_0)$ in the degenerate model). All models fit the observed counts well, but provide different value of \hat{n}_0 .

Table 3 The fitted values of frequency counts in children

	n_0	n_1	n_2	n_3
Observed	?	280	33	3
Degenerate	692	278.5	36	1.5
Logit-normal	2094.4	280	33	3
HO2	1817	280	33	3
Nonparametric	$(762.7, \infty)$	280	33	3

Table 4 The fitted values of frequency counts in hepatitis

	n_0	n_1	n_2	n_3
Observed	?	187	56	28
Degenerate	117.4	172.7	84.5	13.7
Logit-normal	4276.7	187	56	28
HO2	1042	187	56	28
Nonparametric	(207.7, ∞)	187	56	28

The greatest lower bound estimate is $\hat{s}_{GLB} = 1079$ with 95 % lower confidence limit 863. The official list only contained 157 street children; the two street surveys, together with the official list, revealed that there were at least $n_+ = 316$ street children; as $157/1079 = 14.55\%$, at least 85.45 % street children were missed by the official list.

4.2 Hepatitis A virus infection (hepatitis)

We will study another example about an outbreak of hepatitis A virus among students of a college in Taiwan (e.g., [Chao et al. 2001](#)). There are three lists, records based on a serum test, local hospital records, and records by epidemiologists; see [Table 2](#).

As $\hat{b} = (0.09, -0.07, -0.02)^\top$ with a likelihood ratio test statistic 1.09 and p value 0.58, there were no significant list effects. The counts $n_1 = 187$, $n_2 = 56$ and $n_3 = 28$ are fitted by $\delta(0.328)$ in the degenerate model, by $(\hat{\mu}, \hat{\sigma}) = (-6.76, 3.03)$ in the logit-normal model, and by infinitely many NPML \hat{Q} (e.g., $\hat{Q} = 0.917\delta(0.230) + 0.083\delta(1)$ and $\hat{Q} = 0.786\delta(0.150) + 0.214\delta(0.763)$) in the nonparametric model. From [Table 4](#), the degenerate model fits the observed counts poorly while these other three models fit the observed counts perfectly.

The true value is $s = 545$ after a screen serum test was done for all students in the college ([Chao et al. 2001](#)). The greatest lower bound estimate $\hat{s}_{GLB} = 479$ is quite close to $s = 545$ and the 95 % confidence interval $[414, \infty)$ contains 545. We also have $\hat{s}_{LN} = 4548$ and $\hat{s}_{HO2} = 1313$, and four confidence intervals, $[859, \infty)$ (95 %, s_{LN}), $[616, \infty)$ (99 %, s_{LN}), $[707, \infty)$ (95 %, s_{HO2}), and $[571, \infty)$ (99 %, s_{HO2}). The true value $s = 545$ is much less than both estimates and falls outside of all four confidence intervals.

4.3 Allegheny lesbians (lesbian)

Although the sizes of homosexual populations are of value in social security studies such as calculating the costs of civic union benefits, reliable estimates are difficult to obtain. [Table 2](#) presents an example about the lesbian population in Allegheny County, Pennsylvania ([Aaron et al. 2003](#)). There were $n_+ = 2185$ cases found in four lists: a community center, an event promoter, a foundation, and a lesbian health research project; see [Table 2](#).

The estimate for \mathbf{b} is $\hat{\mathbf{b}} = (0.457, 0.278, -0.409, -0.326)^\top$. The counts $n_1 = 1612$, $n_2 = 436$, $n_3 = 114$ and $n_4 = 23$ are fitted by $\delta(0.189)$ in the degenerate model, by $(\hat{\mu}, \hat{\sigma}) = (-3.00, 1.31)$ in the logit-normal model, and by infinitely many NPMLE (e.g., $\hat{Q} = 0.860\delta(0.117) + 0.140\delta(0.522)$ and $\hat{Q} = 0.415\delta(0) + 0.580\delta(0.287) + 0.005\delta(1)$) in the nonparametric model. Some population size estimates, with 95 % lower confidence limits in parenthesis, are $\hat{\delta}_{1, \text{LB}} = 4035$ (3848), $\hat{\delta}_{2, \text{LB}} = 4321$ (4067), $\hat{\delta}_{\text{LN}} = 8246$ (6519), and $\hat{\delta}_{\text{HO2}} = 6964$ (5967). The greatest lower bound estimate is $\hat{\delta}_{\text{GLB}} = 4910$, with 95 % lower confidence limit 4471. There were 375901 adult women in Allegheny County based on the 1990 census (Aaron et al. 2003). Consequently, we may claim that the adult lesbian population constituted at least $4910/375901=1.31$ % of the adult female population.

5 Discussion

Estimating the population size in the binomial model has the non-identifiability problem (Huggins 2001; Link 2003; Holzmann et al. 2006; Mao 2008b). Because the Rasch model includes the binomial model as a special case, the non-identifiability problem should occur in the Rasch model. Hence, we propose to estimate the greatest lower bound of the population size. This is the best possible strategy that one can take under such a circumstance (Mao and You 2009). Because there is no upper bound for the population size when a population is heterogeneous, any meaningful confidence interval must be one-sided with an infinite upper confidence limit. The use of the greatest lower bound eliminates the necessity of adopting a simple model such as the degenerate model, and the use of a lower confidence limit avoids the poor coverage problem of two-sided confidence intervals.

Acknowledgements We thank the referees, the associated editor and editor for their constructive comments. Our research is supported by the National Natural Science Foundation of China Grant No. 713-731-52.

References

- Aaron, D. J., Chang, Y. F., Markovic, N., LaPorte, R. E. (2003). Estimating the lesbian population: a capture-recapture approach. *Journal of Epidemiology and Community Health*, 57, 207–209.
- Agresti, A. (1994). Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics*, 50, 494–500.
- Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B*, 32, 283–301.
- Bartolucci, F., Forcina, A. (2001). Analysis of capture-recapture data with a Rasch-type model allowing for conditional dependence and multidimensionality. *Biometrics*, 57, 714–719.
- Bezerra, K., Gurgel, R., Ilozue, C., Castaneda, D. (2011). Estimating the number of street children and adolescents in two cities of Brazil using capture-recapture. *Journal of Paediatrics and Child Health*, 47, 524–529.
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 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. (2001). An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics*, 6, 138–155.
- Chao, A., Tsay, P. K., Lin, S.-H., Shau, W.-Y., Chao, D.-Y. (2001). The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, 20, 3123–3157.
- Chao, A., Shen, T. J., Hwang, W. H. (2006). Application of Laplace's boundary-mode approximations to estimate species and shared species richness. *Australian and New Zealand Journal of Statistics*, 48, 117–128.
- Coull, B. A., Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. *Biometrics*, 55, 294–301.
- Holzmann, H., Munk, A., Zucchini, W. (2006). On identifiability in capture-recapture models. *Biometrics*, 62, 934–936.
- Huggins, R. (2001). A note on the difficulties associated with the analysis of capture-recapture experiments with heterogeneous capture probabilities. *Statistics and Probability Letters*, 54, 147–152.
- Lindsay, B. G. (1986). Exponential family mixture models (With least-squares estimators). *The Annals of Statistics*, 14, 124–137.
- Lindsay, B. G., Clogg, C. C., Grego, J. M. (1991). Semi-parametric estimation in the Rasch model, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, 86, 96–107.
- Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, 59, 1123–1130.
- Mao, C. X. (2006). Inference of the number of species via geometric lower bounds. *Journal of American Statistical Association*, 101, 1663–1670.
- Mao, C. X. (2007a). Estimating population sizes for capture-recapture sampling with binomial mixtures. *Computational Statistics and Data Analysis*, 51, 5211–5219.
- Mao, C. X. (2007b). Testing list effects and diagnosing individual catchabilities in the Rasch model. *Statistical Methodology*, 4, 416–422.
- Mao, C. X. (2008a). Computing an NPML for a mixing distribution in two closed heterogeneous population size models. *Biometrical Journal*, 50, 983–992.
- Mao, C. X. (2008b). On the nonidentifiability of population sizes. *Biometrics*, 64, 977–979.
- Mao, C. X., Lindsay, B. G. (2007). Estimating the number of classes. *Annals of Statistics*, 35, 917–930.
- Mao, C. X., You, N. (2009). On comparison of mixture models for closed population capture-recapture studies. *Biometrics*, 65, 547–553.
- Prohaska, T. R., Anderson, L. A., Binstock, R. H. (2012). *Public Health for an Aging Society*. Baltimore: Johns Hopkins University Press.
- Rivest, L. (2011). A lower bound model for multiple record systems estimation with heterogeneous catchability. *The International Journal of Biostatistics*, 7, 1–21.
- Rivest, L., Baillargeon, S. (2007). Applications and extensions of Chao's moment estimator for the size of a closed population. *Biometrics*, 63, 999–1006.
- Shukla, P. C. (2005). *Street Children and Asphalt Life*. Delhi: Isha Books.