Computer Engineering and Intelligent System Journal
ISSN: 2997-6812 |
Volume 10 Issue 3, July-September, 2022
Journal Homepage: https://ethanpub.online/Journals/index.php/E16
Official Journal of Ethan Publication

TAILORED POPULATION ESTIMATION: LEVERAGING HETEROGENEITY PATTERNS FOR IMPROVED ACCURACY

[1] Laura Elizabeth Green, [2] Michael Thomas Reid and [3] Emily Rose Carter
[1]Department of Mathematics and Statistics, James Madison University, Harrisonburg, VA
[2,3]Student, Department of Mathematics and Statistics, James Madison University, Harrisonburg, VA

Abstract:
Population size estimation is a critical endeavor, often requiring innovative methods when counting every individual isn't feasible. Capture-recapture, initially developed for wildlife population estimation, offers a valuable approach. In the context of human populations, this method involves creating lists of individuals sampled from the population, providing the foundation for estimating the total population size. In scenarios where the sampling period is relatively short, it is reasonable to assume a closed population—devoid of births, deaths, immigration, or emigration—resulting in a constant population size throughout the study period.
Parametric and nonparametric capture-recapture methods have been proposed, with the choice depending on how capture probabilities are specified for different individuals across various sample occasions. Models of the "time-only" type consider variations in capture probabilities over time, while "individual-only" models account for heterogeneity among individuals. In practice, it's often observed that capture probabilities fluctuate both across individuals and sampling occasions, necessitating models that embrace such complexity.
Both parametric and nonparametric models exist in the latter category, addressing these variations. Non-parametric approaches, such as sample coverage, have been developed by Chao, Lee, and Jeng, while martingales approaches have been explored by Lloyd and Yip. Parametric models, like Sanathanan's mixed logit model, attempt to express capture probabilities as additive functions of subject and sampling effect parameters. However, caution is needed to ensure the model's consistency, as the introduction of subject parameters can challenge the estimation of population size.

Keywords: Population size estimation, capture-recapture, closed population, parametric models, nonparametric models.

1 Introduction

The size of a population can be estimated by repeatedly taking samples from the same population when it is infeasible or unnecessary to count every individual in the target population. This method is called capture-recapture due to its origin in the estimation of wildlife populations. In estimating human population sizes, the samples are often called lists and the individuals are called subjects. When the sampling period is relatively short, it is often reasonable to assume the population is closed with no birth, immigration, death or emigration and thus the population size is constant in the study period.
Both parametric and nonparametric capture-recapture methods have been proposed for estimating the size of a closed population. See Seber (2002) for a general reference. Most estimators in the literature can be classified into M_t or M_h classes depending on how the capture probabilities for different individuals at

different sample occasions are specified. Models of M_t type assume capture probabilities vary with time or sample occasions only, while M_h models assume capture probabilities vary across individuals only. In capture-recapture studies, it is commonly observed that capture probabilities vary across both individuals and sampling occasions and consequently models of M_{th} type are needed to reflect varying time effects as well as heterogeneity among individuals.

Both parametric and nonparametric models have been developed in the M_{th} class. Chao, Lee and Jeng (1992), Chao and Tsay (1998) developed a non-parametric sample coverage approach. Lloyd and Yip (1991) used martingales approaches. For parametric models, Sanathanan (1972a,b) used a mixed logit model which expressed the logit of the capture probability as an additive function of a subject effect parameter and a sampling effect parameter. This setup could introduce as many subject parameters as the population size into the model and the maximum likelihood estimation of model parameters and population size will not be consistent.

There are two common approaches to dealing with this issue. One approach uses a continuous distribution which uses only one or a few parameters to model these subject parameters (e.g., Coull and Agresti 1999). Another approach uses a latent class or finite mixture models, i.e., the subjects are partitioned into several groups with homogeneous subjects within each group so only a limited number of subject parameters need to be estimated (e.g., Agresti 1994; Pledger 2000).

With a limited number of subject parameters, the population size N can be estimated by fitting a proper model through maximum likelihood. The loglinear model is a common choice if the complete capture history of each individual is known. Assume the complete capture history of an individual is represented by a vector $(i_1, \cdots, i_T)$ where, for $t = 1, \cdots, T, i_t = 1$ denotes "captured" and $i_t = 0$ denotes "not captured". There are 2^T capture patterns and the corresponding frequencies can be represented as a contingency table with one missing cell corresponding to the $(0,0, \cdots, 0)$ capture pattern. The model parameters can be estimated through maximizing the multinomial likelihood of the cell counts with the cell probabilities expressed as a function of the model parameters. The missing cell count can be estimated based on its estimated cell probability. This is the basis of the commonly used loglinear modeling method (Bishop et. al 1977; Cormack 1989).

A more convenient approach can adopt a quasi-symmetric loglinear model without explicitly modeling the subject parameters. It is shown that (e.g., Duncan 1985) under the additive logit specification, the expected cell counts of the contingency table satisfy the property of quasi-symmetry with some constraints, and thus N can be estimated through fitting this particular loglinear model. The quasi-symmetric loglinear model is a valuable tool for accounting for subject heterogeneity. It is easy to fit using statistical software such as R or SAS and it generally performs better than the mixed model in some simulation studies (Coull and Agresti 1999).

The additive logit formulation assumes the heterogeneity pattern is constant across the samples. That is, given a subject, the parameter reflecting the subject effect is constant across the sampling occasions. However, the heterogeneity pattern may change across samples with variation in sampling efforts or methods and therefore the catchability of a given subject may vary across the samples (Darroch et. al 1993). In this paper, we will explore using loglinear models of particular structures to estimate the population size under this relaxed condition. We will characterize the structure of the loglinear models that incorporate varying heterogeneity pattern. Darroch et. al (1993) termed the new model partial quasi-symmetry loglinear model and derived its design matrix for $T = 3$ using a two stage construction technique.

Computer Engineering and Intelligent System Journal
ISSN: 2997-6812 |
Volume 10 Issue 3, July-September, 2022
Journal Homepage: https://ethanpub.online/Journals/index.php/E16
Official Journal of Ethan Publication

We will generalize the results to $T \geq 3$ cases and present the R command that fits the model without the need to derive the design matrix.

The rest of the paper is as follows. Section 2 will explore the structure of the loglinear model under the assumption of varying heterogeneity patterns. Section 3 will demonstrate the value and scope of the model through a simulation study and real data applications. Discussions and conclusions are in section.

2 Loglinear Modeling

2.1 Quasi-symmetric loglinear model

Assume we have N subjects and T samples (captures). The capture history of subject s is $y_s = (y_{s1}, \cdots, y_{sT})$, where $y_{st}, t = 1, \cdots T$ takes value 0 or 1 with $y_{st} = 1$ indicating capture in sample t. Let $p_{st} = P(y_{st} = 1)$.

We incorporate heterogeneity using the model

$$\text{o}\ \text{t}(p_{st}) = \alpha_s + \beta_t, \alpha_s \sim F, s = 1, \cdots, N; t = 1, \cdots, T. \quad (1)$$

The parameter α_s reflects the subject catchability and a greater α_s value indicates a higher capture probability for subject s given a sample. We treat $\{\alpha_s\}$ as random effects having a certain distribution F. Greater variability in $\{\alpha_s\}$ indicates more heterogeneous individuals. The parameter β_t indicates the sampling effect and a greater β_t value implies a higher capture probability at sample t for a given subject. This model developed by Rasch (1961) was originally used in educational testing where N subjects were tested on T items with binary response to each question. In that setting, α_s reflects the ability or attitude of subject s and β_t reflects the difficulty of item t. This model assumes no interaction between the subject and sampling effect. In fitting the model, one typically assumes independent responses by the same subject (termed local independence) and independence among subjects.

For a possible capture history $i = (i_1, \cdots, i_T)$, let $n_i = n_{i_1, \cdots, i_T}$ be the number of subjects having that capture history. Let $\mu_i = \mathrm{E}(n_i)$. It is easily shown (e.g., Duncan 1985; Darroch and McCloud 1986) that, by averaging over the individuals in the population, model (1) implies the expected frequencies satisfy the quasi-symmetric loglinear model

$$\log(\mu_i) = \beta_0 + \beta_1 I(i_1 = 1) + \cdots + \beta_T I(i_T = 1) + \lambda(i_1, \cdots, i_T), \quad (2)$$

where I is the indicator function, $\lambda(i_1, \cdots, i_T)$ is invariant to permutations of its arguments. This implies interaction terms between the sampling occasions of the same order have equal coefficients. Next we will demonstrate that (2) can be derived without the need to specify the distribution F. From (1), we can get

$$p_{st} = \frac{\exp(\alpha_s + \beta_t)}{1 + \exp(\alpha_s + \beta_t)}.$$

Given subject s, the probability of a particular sequence $i = (i_1, \cdots, i_T)$ is

$$p(i|\alpha_s) = \prod_{t=1}^{T} p_{st}^{i_t} (1 - p_{st})^{1 - i_t} = \frac{\exp(\alpha_s \sum_t i_t + \sum_t \beta_t i_t)}{\prod_t (1 + \exp(\alpha_s + \beta_t))}$$

Assume $\alpha_s \sim F$ with F unspecified, the marginal probability is computed as

$$p(i) = \exp\left(\sum_t \beta_t i_t\right) \int_{\alpha_s} \frac{\exp(\alpha_s \sum_t i_t)}{\prod_t (1 + \exp(\alpha_s + \beta_t))} dF(\alpha_s$$

)

The integral only depends on data through $\sum_t i_t$ and thus

Computer Engineering and Intelligent System Journal
ISSN: 2997-6812 |
Volume 10 Issue 3, July-September, 2022
Journal Homepage: https://ethanpub.online/Journals/index.php/E16
Official Journal of Ethan Publication

$$p\left(\underline{i}\right) = \exp\left(\sum_t \beta_t i_t \quad h\left(\sum_t i_t\right)\right), \quad (\quad)\ (\quad)$$

Where $h(\quad)$ is an unspecified function. Taking logarithms will produce a loglinear model of the form (2). The argument $\sum_t i_t$ of function h implies identical interactions of the same order. For example, among $\{i_1, i_2, \cdots i_T\}$, any two of the T terms taking value 1 with the remaining terms taking value 0 will produce the same argument $\sum_t i_t = 2$, and hence the first-order interaction between any pair of sampling occasions, will be identical. This model also implies higher order interaction terms are identical as long as they are of the same order.

We have to point out though we assume the responses by the same individual are independent, heterogeneity among individuals implies positive associations between each pair of the samples, both conditional and marginal, in the contingency table.

2.2 Partial Quasi-symmetric Model

Now we assume the heterogeneity pattern will change across the sampling occasions. Without loss of generality, assume across the T samples the catchability of a given subject takes one value from sample 1 to k and takes a different value from sample $k+1$ to sample T:

$$\text{logit}(p_{st}) = \begin{cases} \alpha_s + \beta_t, t = 1, \cdots, k, \alpha_s \sim F; \\ \gamma_s + \beta_t, t = k+1, \cdots, T, \gamma_s \sim G; \end{cases}$$

Then we have

$$p_{st} = \begin{cases} \dfrac{\exp(\alpha_s + \beta_t)}{1 + \exp(\alpha_s + \beta_t)}, 1 \le t \le k; \\ \dfrac{\exp(\gamma_s + \beta_t)}{1 + \exp(\gamma_s + \beta_t)}, k+1 \le t \le T; \end{cases}$$

Given subject s, the probability of a particular sequence $\underline{i} = (i_1, \cdots, i_T)$ is

$$p\left(\underline{i}|\alpha_s, \gamma_s\right) = \prod_{t=1}^{T} p_{st}^{i_t}(1 - p_{st})^{1-i_t} = \frac{\exp\left(\alpha_s \sum_{t=1}^{k} i_t + \gamma_s \sum_{t=k+1}^{T} i_t + \sum_{t=1}^{T} \beta_t i_t\right)}{\prod_{t=1}^{k}(1 + \exp(\alpha_s + \beta_t)) \prod_{t=k+1}^{T}(1 + \exp(\gamma_s + \beta_t))}.$$

The marginal probability is

$$p\left(\underline{i}\right) = \exp\left(\sum_t \beta_t i_t\right) \int_{\alpha_s} \frac{\exp\left(\alpha_s \sum_{t=1}^{k} i_t\right)}{\prod_{t=1}^{k}(1 + \exp(\alpha_s + \beta_t))} dF(\alpha_s) \int_{\gamma_s} \frac{\exp\left(\gamma_s \sum_{t=k+1}^{T} i_t\right)}{\prod_{t=k+1}^{T}(1 + \exp(\gamma_s + \beta_t))} dG(\gamma_s).$$

The integrals only depend on data through $\sum_{t=1}^{k} i_t$ and $\sum_{t=k+1}^{T} i_t$, thus the marginal probability satisfies

$$p\left(\underline{i}\right) = \exp\left(\sum_t \beta_t i_t\right) h\left(\sum_{t=1}^{k} i_t, \sum_{t=k+1}^{T} i_t\right), \quad (3)$$

Where h is an unspecified function. Taking logarithms of (3) leads to a partial quasi-symmetry loglinear model

$$\log\ \mu_i = \beta_0 + \beta_1 I(i_1 = 1) + \cdots + \beta_t I(i_t = 1) + \lambda(i_1 + \cdots + i_k, i_{k+1} + \cdots + i_T). \quad (\quad) \quad (4)$$

To characterize the structure of the loglinear model, let us call samples 1 to k subset 1 and samples $k+1$ to T subset 2. Model (4) implies the first-order interaction terms between any pair of sample occasions within the same subset are identical; the first order interaction terms between a randomly selected sample from one subset and a randomly selected sample from the other subset are identical. For second-order interactions, interaction terms among any three samples within the same subset are identical; interaction

Computer Engineering and Intelligent System Journal
ISSN: 2997-6812 |
Volume 10 Issue 3, July-September, 2022
Journal Homepage: https://ethanpub.online/Journals/index.php/E16
Official Journal of Ethan Publication

terms among any two samples from one subset and a randomly selected sample from the other subset are identical.

Take $T = 5$ and $k = 3$ for an example. If we use u to indicate interaction terms, the model implies the first-order interactions satisfy $u_{12} = u_{13} = u_{23}; u_{14} = u_{15} = u_{24} = u_{25} = u_{34} = u_{35}$ and the second-order interactions satisfy $u_{124} = u_{125} = u_{134} = u_{135} = u_{234} = u_{235}; u_{145} = u_{245} = u_{345}$.

3 Numerical Studies

3.1 Simulation Results

In this simulation, we examine the performance of partial quasi-symmetric models under the additive logit specification with varying heterogeneity patterns.

Case 1: We choose 5 total sample occasions, $T = 5$. For $s = 1, \cdots, N$, we generate the capture probability $p_{st} = \frac{\exp(\alpha_s + \beta_t)}{1+\exp(\alpha_s + \beta_t)}$, for $t = 1,2,$ and $p_{st} = \frac{\exp(\gamma_s + \beta_t)}{1+\exp(\gamma_s + \beta_t)}$, for $t = 3,4,5$. We generate each $\alpha_s \sim$ Normal $(\mu = 2, \sigma = 5)$ and each $\gamma_s \sim$ Uniform $(-10,10), s = 1, \cdots, N$. We generate each $\beta_t \sim$ Uniform $(0\ 0\ 5)$ 1 T $(,\ .\ , t =\ , \cdots, .$

Case 2: We choose $T = 4$. We generate $\alpha_s \sim$ Normal $(\mu = 4, \sigma = 1)$ for $t = 1$, and $\gamma_s \sim$ Uniform $(-10,10),$ for $t = 2,3,4$.

We generate $Y_{st} \sim$ Bernoulli $(p_{st}), s = 1, \cdots, N; t = 1, \cdots, T$. We choose $N = 400$ in case 1 and $N = 1000$ in case 2. The frequencies of individuals with the same capture pattern form the cell counts in the contingency table. We pretend the cell count corresponding to capture pattern $(0,0, \cdots 0)$ is missing and use the remaining data to estimate it. Adding the observed table total and the estimate of the missing cell count produces N, the estimate of the population size N.

We compare the performance of the quasi-symmetric model, the partial quasi-symmetric model, the full model and the independence model. We include first-order interaction terms only without including higher order interaction terms. Fitting partial quasi-symmetric loglinear models by R is straightforward. In the model formula, in addition to the regular main effect and interaction terms, we just need to add one or more I() terms, with the sum of the identical interaction terms inside the parenthesis. For example, to fit a partial quasi-symmetric loglinear model with three lists A, B and C including AB interaction and identical AC and BC interactions, the R command is **glm (y~A+B+C+AB+I(AC+BC), family=poisson)**.

For evaluation purpose, we considered the mean squared error $E(\widehat{N} - N)^2$. We generated the data 1000 times and reported the average of the squared errors with its standard error in Table 1. The comparison results showed the quasi-symmetric model performed substantially better than other models under the assumption of varying heterogeneity patterns.

Table 1: Comparison of different models

	Case 1	Case 2
Partial Quasi	16.8(0.8)	356.6(17.1)
Quasi Symmetry	38.5(1.4)	2017(144.0)
Full	120.1(8.7)	7774(48.2)
Independence	68.9(2.0)	8318(49.5)

3.2 Data Examples

We will illustrate the value of partial quasi-symmetric loglinear models with two real-world date sets.

Computer Engineering and Intelligent System Journal
ISSN: 2997-6812 |
Volume 10 Issue 3, July-September, 2022
Journal Homepage: https://ethanpub.online/Journals/index.php/E16
Official Journal of Ethan Publication

The first is a Belgian data set on invasive pneumococcal disease (IPD) cases collected between July 1, 2009 and June 30, 2011 (Braeye et. al 2016).
The IPD cases of adults aged ≥ 50 years were collected on three lists: the hospital list on which adults hospitalized with microbiologically confirmed IPD were eligible for inclusion, the National Reference Centers (NRC) and Sentinel laboratories network (Sentinel). The data is presented in Table 2 with N, H, S indicating the list NRC, Hospital and Sentinel respectively.
Table 2: Belgium IPD data

	H=0	H=1
N=0, S=0	?	232
N=0, S=1	347	34
N=1, S=0	854	188
N=1, S=1	843	281

The data complication suggested possible NH and HS associations as the laboratories of the hospitals were encouraged to send cases to both NRC and the sentinel surveillance. There was possible interaction between the sentinel surveillance and the NRC because their detectors overlapped.
Indeed the full loglinear model $N + H + S + NH + NS + SH$ shows all the three interaction terms NH, NS, SH are significant at 0.05 significance level. This model produces $\widehat{N} = 6364$ with a 95% confidence interval (5114, 8283). How good is this estimate? Since the full model uses all the 7 degrees of freedom, we can not use the deviance G^2 to assess the fit of the model. The relevant outside information on the average IPD incidence rate in Europe in this period (Braeye et. al 2016) suggests possible model overfitting and over-estimation of N.
Next we consider the more concise quasi-symmetric model $N + H + S + I(NH + NS + SH)$. This model fits the data poorly with $G^2 = 166.2$ on 2 df. The estimate $\widehat{N} = 3520$ is likely an under-estimate as it is below the lower bound estimation $N = 4109$ produced by Chao's samp e covera e method (Braeye et. a 2016).
The partial quasi-symmetric model which lies between the two models above could be a reasonable choice. The referral system (from Hospital to NRC and Sentinel) suggests the interaction terms NH and SH may be close to each other and thus it is reasonable to consider the partial quasi-symmetric model $N + H + S + I(NH + SH) + NS$. This model produces $G^2 = 19.02$ on 1 df. Though this G^2 value still indicates a lack of fit, it greatly improves over the quasi-symmetric model with a reduction of G^2 by 147.18. The partial quasi-symmetric model produces $\widehat{N} = 5042$ with a 95% confidence interval (4388, 5965).
Taking into account the available outside information, the lower bound estimate and the goodness-of-fit statistic, the estimate by the partial quasi-symmetric model seems more plausible compared to the estimate by the full model or the quasi-symmetric model. It also produces a narrower confidence interval compared to the full model. Saving 1 degree of freedom by the partial quasi-symmetric model is valuable given the total degrees of freedom is only 7.
The second data set came from a census study in 1988 designed to estimate the undercount of black male renters, a group believed to be seriously undercounted in past census (Darroch et. al 1993). The data in Table 3 compiled black male renters aged 30-44 in St. Louis, Missouri on three lists: preliminary enumeration (list E), post-enumeration survey (list P), and the administrative list supplement (list A) which was compiled from pre-census administrative records of state and federal government agencies.
Table 3: Three source data in 1988 census in St. Louis, Missouri

	A=0	A=1
E=0, P=0	?	43
E=0, P=1	53	13
E=1, P=0	71	7
E=1, P=1	155	72

Fitting the full model $E + P + A + EP + EA + PA$ shows all three interaction terms are significant at 0.10 level of significance. The model produces $\widehat{N} = 1240$ with a 95% confidence interval (698,2815).
The quasi-symmetric model $E + P + A + I(EP + EA + PA)$ produces $G^2 = 73.01$ on 2 df, indicating a serious lack of fit. This model produces $\widehat{N} = 496$ with a 95% confidence interval(455,577).
The poor fit of the model indicates the estimate of N may not be reliable.
The way the data was compiled suggested the choice of the partial quasi-symmetric model, in which the pattern of heterogeneity of individuals was the same in the E and P samples but different in the A sample. Indeed the way the A list was constructed through extensive searches of administrative records of a region covered by the
E or P list suggested it may be reasonable to assume identical PA and EA interactions. Fitting this model $E + P + A + I(PA + EA) + EP$ produces $G^2 = 3.45$ on 1 df, a great improvement over the quasi-symmetric model. This model produces $\widehat{N} = 1182$ and a 95% confidence interval (688,2423). Though we have no outside information to help assess the accuracy of the estimate, the good fit of the partial quasi-symmetric model gives us more confidence in its estimate. Compared to the full model, the partial quasi-symmetric model produces a narrower confidence interval with a decent goodness-of-fit.

4 Conclusions

In this article, we explore the partial quasi-symmetric loglinear model which can be used to model varying heterogeneity patterns across the capture occasions (lists) in capture-recapture studies. Compared to the full model with separate values for different interaction terms, the properly chosen partial-symmetric model can save degrees of freedoms and typically produces shorter confidence intervals by constricting some interactions terms to be identical. Compared to the quasi-symmetric model which constrains all interaction terms of the same order to be identical, the partial quasi-symmetric model provides more flexibility by allowing a subset of interactions of the same order to be identical. This more flexible structure can lead to substantial improvement in the fit of the model without losing too many degrees of freedom. In summary, the partial quasi-symmetric model has potential of offering a good compromise when the full model seems over-fitted and the quasi-symmetric model seems under-fitted. The quasi-symmetric loglinear model is based on the additive logit model which assumes local independence among the sample occasions. When local dependence is present such as in the cases of trap avoidance or trap attraction, we may expand the partial quasi-symmetric model by adding some association terms such as interaction between adjacent sample occasions. If it is reasonable to assume the interaction terms in a certain group do not differ greatly, we can further consider a more concise model by restricting these terms to be identical.

References

Agresti, A. (1994). Simple capture-recapture models permitting unequal catchability and variable sampling effort. Biometrics, 50 (2), 494-500.

Bishop, Y. M. M., Fienberg, S. E. & Holland, P. W. (1977). Discrete Multivariate Analysis: Theory and Practice. Cambridge MA: MIT press.

Braeye, T., Jan Verheagen, J., Mignon, A. ,Flipse, W., Pierard, D., Huygen, K., Schirvel, C., & Hens, N. (2016) Capture-recapture estimators in epidemiology with applications to Pertussis and Pneumococcal invasive disease surveillance. PLoS ONE 11(8).

Chao, A., Lee, S-M., & Jeng, S-L. (1992). Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. Biometrics 48, 201-216.

Chao A., Tsay PK. (1998). A sample coverage approach to multiple-system estimation with application to census undercount. Journal of the American Statistical Association 1998; 93, 283 – 293.

Cormack, R. M. (1989). Log-linear models for capture-recapture. Biometrics 45, 395–413.

Coull, B. A. & Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture-recapture studies. Biometrics 55, 294-301.

Darroch, J.N., Fienberg, S.E., Glonek, G.F.V. & Junker, B. W. (1993). A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability. Journal of the American Statistical Association, 88 (423), 1137-1148.

Duncan, OD. (1985). Some models of response uncertainty for panel analysis. Social Science Research 14, 126-141.

Lloyd, C. J. and Yip, P. (1991). A unification of inference from capture-recapture studies through martingale estimating functions. In Estimating Equations, V. P. Godambe (ed.), 65-88. Oxford: Clarendon Press.

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

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. Proceedings of the 4th Berkeley Symposium in Mathematical Statistics and Probability, 4: 321-333.

Sanathanan, L. (1972a). Estimating the size of a multinomial population. Annals of Mathematical Statistics 43, 142-152.

Sanathanan, L. (1972b). Models and estimation methods in visual scanning experiments. Technometrics 14,813-829.

Seber, G.A.F. (2002). The Estimation of Animal Abundance and Related Parameters. Blackburn Press.