Biometrics 60, 100–107
March 2004

Loglinear Models for the Robust Design in Mark–Recapture Experiments

Louis-Paul Rivest* **and Gaétan Daigle**

Département de Mathématiques et de Statistique, Université Laval,
Ste-Foy, Québec G1K 7P4, Canada
**email:* lpr@mat.ulaval.ca

Summary. The robust design is a method for implementing a mark–recapture experiment featuring a nested sampling structure. The first level consists of primary sampling sessions; the population experiences mortality and immigration between primary sessions so that open population models apply at this level. The second level of sampling has a short mark–recapture study within each primary session. Closed population models are used at this stage to estimate the animal abundance at each primary session. This article suggests a loglinear technique to fit the robust design. Loglinear models for the analysis of mark–recapture data from closed and open populations are first reviewed. These two types of models are then combined to analyze the data from a robust design. The proposed loglinear approach to the robust design allows incorporating parameters for a heterogeneity in the capture probabilities of the units within each primary session. Temporary emigration out of the study area can also be accounted for in the loglinear framework. The analysis is relatively simple; it relies on a large Poisson regression with the vector of frequencies of the capture histories as dependent variable. An example concerned with the estimation of abundance and survival of the red-back vole in an area of southeastern Québec is presented.

Key words: Generalized linear model; Mark–recapture study; Multinomial distribution; Poisson regression.

1. Introduction

The robust design was introduced by Pollock (1982) as a combination of models for closed and open populations in mark–recapture studies. This design comprises a series of short-term studies, each one amenable to an analysis with closed population models. The animals are possibly captured at several of these sessions. From one short-term study to the next, the population is experiencing immigration and death; open population models apply. By pooling the data from all short-term studies, the robust design improves the estimation of the demographic characteristics of the population.

A robust design features numerous parameters, for the survival probabilities and the immigration between short-term studies, and for the capture of the units. Pollock et al. (1990) present ad hoc methods for estimating these parameters. Likelihood methods are developed by Kendall, Pollock, and Brownie (1995). Temporary emigration and breeding proportions are incorporated in robust designs by Kendall, Nichols, and Hines (1997), Schwarz and Stobo (1997), and Kendall and Bjorklan (2001). Recent developments are reviewed in Chapter 19 of Williams, Nichols, and Conroy (2002). The program MARK of White and Burnham (1999) implements some of these methods.

This article investigates loglinear models for estimating the parameters in a robust design. In mark–recapture studies, this approach has been pioneered by Cormack (1985, 1989). Recently, Rivest and Lévesque (2001) have shown that loglinear models yield simple estimators for the size of a closed population for models M_0, M_t, M_h, M_b, M_{bh}, and M_{th}. These estimators are all based on simple Poisson regression analyses. A Poisson regression denotes a generalized linear model with Poisson errors and a logarithmic link function. Rivest and Lévesque (2001) also suggested generalizations of Chapman's (1951) correction to improve the small sample properties of loglinear abundance estimators.

In this article, loglinear models and Poisson regressions are used to analyze data collected in a robust design, under several specifications for the capture of the animals. Two models featuring a heterogeneity of the capture probabilities within primary sessions are proposed. This complements MARK nicely since this software deals only with individual heterogeneity accounted for by observed explanatory variables.

It is now convenient to define the parameters and the variables that enter a robust design.

- Subscript i, for $i = 1, \ldots, I$ denotes the primary sampling period and I is the number of primary sampling periods. Subscript j denotes the capture occasions within the primary sessions; $j = 1, \ldots, \ell_i$ where ℓ_i is the number of capture occasions within primary period i.
- The $(\sum \ell_i) \times 1$ capture history vector $\boldsymbol{\omega}$ has entry 1 for capture occasion (i, j) $(\omega_{ij} = 1)$ if the unit has been caught on that capture occasion and 0 if not, $\boldsymbol{\omega}_i$ denotes the vector of the ℓ_i components of $\boldsymbol{\omega}$ pertaining to primary sampling period i.

- The $I \times 1$ vector $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\omega})$ of primary sampling period capture histories has entry 1 for the ith sampling period ($\delta_i = 1$) for a unit caught at least once in that period, that is if $\sum_j \omega_{ij} > 0$, and 0 if not, $p_i^* = \Pr(\delta_i = 1)$ denotes the probability of being captured in the ith session.
- The frequency of capture history $\boldsymbol{\omega}$ among all the animals that have been caught at least once is n_ω, $\boldsymbol{n}$ denotes the vector of the n_ω's, and μ_ω denotes the predicted value of n_ω under a model for the capture of the units.
- N_i denotes the expected size of the population at the start of the ith primary sampling period, for $i = 1, \ldots, I$.
- The survival probability between primary sampling periods i and $i + 1$ is ϕ_i for all animals in the population at primary period i.
- The expected number of new animals entering the population between primary sampling period i and $i + 1$ is B_i and $N_{i+1} = N_i\phi_i + B_i$.

Standard assumptions made in this article are (1) the independence of the fate of each animal from that of the others, (2) the population is closed within a primary sampling period, (3) there is no loss of marks and no trap death, and (4) there is no temporary emigration out of the population. This fourth assumption is relaxed in Section 4.2.

Loglinear models for closed and open populations are reviewed in Sections 2 and 3. Section 4 shows how to combine the models of Sections 2 and 3 to analyze mark–recapture data collected in a robust design. Section 5 presents the analysis of data on the survival of the red-back vole in southeastern Québec.

2. Loglinear Models for Closed Populations

In this section, there are $I = 1$ sampling sessions; the aim of the analysis is to estimate $N_1 = N$, the size of the population. There are $\ell = \ell_1$ capture occasions and a capture history is denoted by $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_\ell)$. A loglinear model for the mark–recapture data expresses the predicted frequency μ_ω as

$$\log \mu_\omega = \gamma + \mathbf{X}_\omega \boldsymbol{\beta}, \tag{1}$$

where $\mathbf{X}_\omega = (X_{\omega 1}, \ldots, X_{\omega p})$ is the row vector of explanatory variables for the capture history ω, and $\gamma, \boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ are unknown parameters. It is convenient to use a parametrization for which $\mathbf{X}_\omega = 0$ when no units are caught (i.e., $\boldsymbol{\omega} = 0$). With this convention, $e^\gamma = \mu_0$ is the predicted frequency of the animals that were missed in the study.

Many models have the form (1). M_t has $p = \ell$, $X_{\omega_j} = \omega_j$, $j = 1, \ldots, p$, and β_j is the logit of the capture probability at occasion j. M_0 is a special case of M_t with $\beta_1 = \beta_2 = \cdots = \beta_p$.

Darroch et al. (1993), Agresti (1994), and Coull and Agresti (1999) assume that the probability that an animal is caught at occasion j is $\exp(\beta_j + h)/\{1 + \exp(\beta_j + h)\}$, where h is a latent variable related to the catchability of an animal. As shown in the Appendix, assuming that the posterior distribution of h given that an animal is not caught has a normal distribution with variance $\beta_{\ell+1}$ leads to $M_{\mathrm{D}th}$, a loglinear model, with $p = \ell + 1$ explanatory variables, $\mathbf{X}_\omega = (\omega_1 \ldots, \omega_\ell, (\sum_j \omega_j)^2/2)$. The posterior variance $\beta_{\ell+1}$ of the latent variable measures the extent of the heterogeneity in catchability.

Heterogeneity in the capture probabilities can be handled as in Chao (1989), by leaving out, when estimating γ, the units caught more than twice as unrepresentative of the units that were never caught. This is implemented in a loglinear framework by adding, to the $\mathbf{X}$ matrix for M_t, one parameter for each capture history featuring more than two captures. This proposal, labeled M_{Cth}, is therefore to have (1) given by

$$\log \mu_\omega = \gamma + \sum_{j=1}^{\ell} \omega_j \beta_j + \sum_{\varpi : \sum \varpi_i > 2} 1_\varpi(\omega)\beta_\varpi, \tag{2}$$

where the second summation is on the $2^\ell - 1 - \ell(\ell + 1)/2$ capture histories ϖ featuring more than two captures and $1_\varpi(\omega) = 1$ if $\omega = \varpi$ and 0 otherwise. It is shown in the Appendix that the animals captured more than twice do not contribute to the estimation of γ. Furthermore, the estimator of abundance for model M_{Ch}, with $\beta_1 = \cdots = \beta_\ell$, coincides with Chao's (1989) estimator when ℓ is large. The column space of the design matrix for $M_{\mathrm{D}th}$ is included in that for M_{Cth}. This suggests carrying out the following sequence of likelihood ratio tests to investigate the presence of heterogeneity: (i) compare $M_{\mathrm{D}th}$ and M_t to test for heterogeneity and (ii) compare M_{Cth} and $M_{\mathrm{D}th}$ to test whether this heterogeneity can be described by Darroch et al. (1993) normal distribution.

Table 1 summarizes the design matrices for several loglinear models for closed populations. They can all be fitted using a Poisson regression. In each case $\hat{N} = \sum n_\omega + \exp(\hat{\gamma})$. Sandland and Cormack (1984), Cormack (1993), and Rivest and Lévesque (2001) discuss the estimation of the variance of $\hat{N}$.

The probability p^* that a unit is caught at least once in the study can be expressed in terms of the parameter β of (1) as

$$p^* = \frac{\sum_\omega \exp(\mathbf{X}_\omega \boldsymbol{\beta})}{1 + \sum_\omega \exp(\mathbf{X}_\omega \boldsymbol{\beta})}, \tag{3}$$

where $\sum_\omega$ denotes a sum on the $2^\ell - 1$ observable capture histories. For model M_t, equation (3) factors nicely into $p^* = 1 - \prod (1 + \exp \beta_j)^{-1}$.

Table 1

The number of explanatory variables, p, and the design matrix, $\mathbf{X}$, corresponding to the loglinear proposals for M_t and M_{th}. Replacing in each one the first ℓ columns of $\mathbf{X}$ by $\sum_{j=1}^{\ell} \omega_j$ yields loglinear models for M_0 and M_h.

Model	M_t, $p = \ell$	$M_{\mathrm{D}th}$, $p = \ell + 1$	M_{Cth}, $p = 2^\ell - 1 - \ell(\ell - 1)/2$
X_ω	$(\omega_1, \ldots, \omega_\ell)$	$(\omega_1, \ldots, \omega_\ell, (\sum \omega_i)^2/2)$	$(\omega_1, \ldots, \omega_\ell, 1_\omega, \text{s.t.} \sum \omega_i > 2)$

3. Loglinear Models for Open Populations

This section reviews the loglinear approach to the Jolly–Seber model for open populations as presented in Cormack (1985, 1989). In the notation of Section 1, open population models are constrained by $\ell_1 = \ell_2 = \cdots = \ell_I = 1$. Thus the capture histories satisfy $\boldsymbol{\delta} = \boldsymbol{\omega}$.

Let μ_δ denote the predicted frequency for capture history $\boldsymbol{\delta}$. It can be expressed as a loglinear function of the population sizes N_i, of the survival probabilities ϕ_i, and of the capture probabilities p_i^* as

$$\log \mu_\delta = \mathbf{Z}_\delta \boldsymbol{\gamma} + \mathbf{X}_\delta \boldsymbol{\beta}, \tag{4}$$

where $\mathbf{X}_\delta$, the row of the $\mathbf{X}$ matrix for capture history $\boldsymbol{\delta}$, has I entries and $\mathbf{Z}_\delta$, the row of the $\mathbf{Z}$ matrix for $\boldsymbol{\delta}$, has $2I - 1$ entries. One has $\mathbf{X}_\delta = \boldsymbol{\delta}$ and $\beta_i = \log\{p_i^*/(1 - p_i^*)\}$, $i = 1, \ldots, I$. Some additional notations for $\mathbf{Z}_\delta\boldsymbol{\gamma}$ are:

U_i is the expected number of unmarked animals in the population before the ith capture occasion: $U_i = U_{i-1}(1 - p_{i-1}^*)\phi_{i-1} + B_{i-1}$ for $i = 1, \ldots, I - 1$, with $U_1 = N_1$.

χ_i is the probability of not being captured after sampling occasion i: $\chi_i = 1 - \phi_i + \phi_i(1 - p_{i+1}^*)\chi_{i+1}$ for $i = 1, \ldots, I - 1$, with $\chi_I = 1$.

For an arbitrary capture history $\boldsymbol{\delta}$, $\exp(\mathbf{Z}_\delta\boldsymbol{\gamma})$ depends on the smallest index j and the largest index k such that $\delta_j = 1$ and $\delta_k = 1$. Thus,

$$\mathbf{Z}_\delta = \Bigg(\underbrace{1, \bar{\delta}_1, \bar{\delta}_1\bar{\delta}_2, \ldots, \prod_{j=1}^{I-1}\bar{\delta}_j}_{I-1}, \underbrace{\bar{\delta}_I, \bar{\delta}_I\bar{\delta}_{I-1}, \ldots, \prod_{j=0}^{I-2}\bar{\delta}_{I-j}}_{I-1}\Bigg), \tag{5}$$

where $\bar{\delta}_j = 1 - \delta_j$, for $j = 1, 2, \ldots, I$. The first entry of $\mathbf{Z}_\delta$ is 1; the corresponding intercept parameter is

$$\gamma_0 = \log\left\{N_1\big(1 - p_I^*\big)\prod_{j=1}^{I-1}\big(1 - p_j^*\big)\phi_j\right\}.$$

Entries 2 to I of $\mathbf{Z}_\delta$ is a sequence of 1 followed by a sequence of 0; the change occurs at the first capture of the unit. Entries $I + 1$ to $2I - 1$ have a similar form, with the switch from 1 to 0 occurring on the last capture. The loglinear parameters for columns 2 to I of $\mathbf{Z}$ all have the same form

$$\gamma_i = \log\left(1 + \frac{B_i}{\big(1 - p_i^*\big)\phi_i U_i}\right) = \log\left(\frac{U_{i+1}}{\big(1 - p_i^*\big)\phi_i U_i}\right)$$

for $i = 1, \ldots, I - 1$ while for $i = I, \ldots, 2I - 2$,

$$\gamma_i = \log\left(1 + \frac{1 - \phi_{2I-i-1}}{\phi_{2I-i-1}\big(1 - p_{2I-i}^*\big)\chi_{2I-i}}\right) = \log\left(\frac{\chi_{2I-i-1}}{\phi_{2I-i-1}\big(1 - p_{2I-i}^*\big)\chi_{2I-i}}\right).$$

The survival probabilities ϕ_i can be evaluated from $\boldsymbol{\gamma}$ by noticing that, for $i = 1, \ldots, I - 1$,

$$\frac{1 - \phi_i}{\phi_i} = \frac{u_{I-i}}{\sum_{k=0}^{I-i-1} u_k} \tag{6}$$

where $u_0 = 1$ and for $i = 1, \ldots, I - 1$

$$u_i = \left\{\prod_{k=1}^{i} e^{\gamma_{I+k-1}}\big(1 - p_{I-k+1}^*\big)\right\}(1 - e^{-\gamma_{I+i-1}}).$$

To recover the N_i's from the loglinear parameters let $v_0 = 1$ and, for $i = 1, \ldots, I - 1$, define

$$v_i = \left\{\prod_{k=1}^{i} e^{\gamma_k}\big(1 - p_k^*\big)\right\}(1 - e^{-\gamma_i}),$$

then

$$\frac{N_{i+1}}{\phi_i N_i} - 1 = \frac{v_i}{\sum_{k=0}^{i-1} v_k}. \tag{7}$$

The whole design matrix featuring both the $\mathbf{X}$ and the $\mathbf{Z}$ matrices in the linear component of (4) is not of full rank. Indeed $(\gamma_0, \gamma_1, \gamma_I, \beta_1, \beta_I)$ is not estimable; only $(\gamma_0 + \beta_1 + \beta_I, \gamma_1 - \beta_1, \gamma_I - \beta_I)$ is. Written in terms of the demographic parameters, these identifiability constraints mean that only $N_1 p_1$ and $\phi_{I-1}p_I$, and $N_I p_I$ are estimable; this lack of identifiability is well known for the Jolly–Seber model (see Pollock et al., 1990).

The Jolly–Seber model is easily fitted by using a Poisson regression on the observed frequencies n_δ. By construction $\gamma_i \geq 0$ for $i = 1, \ldots, 2I - 2$. When an estimate for γ_i is negative, one can set this parameter to 0 and refit the model to estimate the remaining parameters. If $\gamma_i = 0$, for $i \leq I - 1$ then $v_i = 0$ in (7) and $B_i = 0$. Similarly $\gamma_{2I-i-1} = 0$ corresponds in (6) to $\phi_i = 1$ for $i = 1, \ldots, I - 1$.

4. Loglinear Models for the Robust Design

Our proposal for analyzing the data of a robust design is to use a Poisson regression with a design matrix constructed by juxtaposing a $\mathbf{Z}$ component for the open population part of the model and $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_I]$, where $\mathbf{X}_i$ is the design matrix for modeling the capture of the units within the ith primary sampling period. The predicted frequency μ_ω for global capture history $\boldsymbol{\omega}$ is

$$\log \mu_\omega = \mathbf{Z}_\delta\boldsymbol{\gamma} + \mathbf{X}_\omega\boldsymbol{\beta} = \mathbf{Z}_\delta\boldsymbol{\gamma} + \sum_{i=1}^{I}\mathbf{X}_{\omega i}\boldsymbol{\beta}_i, \tag{8}$$

where $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{\omega})$ is the between primary sampling period capture history corresponding to $\boldsymbol{\omega}$, $\boldsymbol{\gamma}$ is a $(2I - 1) \times 1$ vector of loglinear parameters associated with the demographic information, and $\boldsymbol{\beta}_i$ parameterizes the capture probabilities of the units within the ith primary sampling session. The vector $\boldsymbol{\gamma}$ is the same as that presented in Section 3 for open population models. The design matrix $\mathbf{X}_i$ can be specified according to one of the models of Table 1. Seen in the light of (8), the robust design improves on the open population model presented in Section 3 by allowing a finer modeling of the capture mechanism within the primary sampling periods.

Model (8) contains as special cases models (1) and (4) for closed and open populations considered in Sections 2 and 3. The predicted value for the number of units having $\boldsymbol{\delta}$ as between session capture history is the summation of the μ_ω's

corresponding to the capture histories $\boldsymbol{\omega}$ having $\boldsymbol{\delta}$ as a common between session capture history. In mathematical terms, this is expressed as

$$\sum_{\boldsymbol{\omega}:\boldsymbol{\delta}(\boldsymbol{\omega})=\boldsymbol{\delta}} \mu_\omega = \sum_{j_1}\sum_{j_2}\cdots\sum_{j_I} \exp\left(\mathbf{Z}_\delta\boldsymbol{\gamma} + \sum_{i=1}^{I}\mathbf{X}_{\omega i}\boldsymbol{\beta_i}\right)$$
$$= \exp(\mathbf{Z}_\delta\boldsymbol{\gamma})\sum_{j_1}\exp(\mathbf{X}_{\omega 1}\boldsymbol{\beta_1})$$
$$\times \sum_{j_2}\exp(\mathbf{X}_{\omega 2}\boldsymbol{\beta_2})\cdots\sum_{j_I}\exp(\mathbf{X}_{\omega I}\boldsymbol{\beta_I}), \qquad (9)$$

where $\sum_{j_i}$ features only one term, with $\mathbf{X}_{\omega i} = 0$ if $\delta_i = 0$. This sum has $2^{\ell_i} - 1$ terms if $\delta_i = 1$. Furthermore, when $\delta_i = 1$, from (3),

$$\sum_{j_i}\exp(\mathbf{X}_{\omega i}\boldsymbol{\beta_i}) = \frac{p_i^*}{1-p_i^*},$$

where p_i^* is the probability of being caught at least once in primary session i. This shows that (9) has the same form as (4). In a similar way, if $\boldsymbol{\varpi}$ denotes a $\ell_i \times 1$ vector of 0s and 1s giving a possible capture history for the ith primary sampling session, then $\sum_{\boldsymbol{\omega}:\boldsymbol{\omega}_i=\boldsymbol{\varpi}}\mu_\omega$ reduces to the loglinear model (1) for the closed population of the ith sampling session.

Fitting a robust design involves the following four steps:

(1) Calculate $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\beta}}$ the parameter estimates for (8) using a Poisson regression.
(2) For $i = 1, \ldots, I$, evaluate $\hat{p}_i^*$ by applying (3) to the component $\hat{\boldsymbol{\beta}}_i$ of $\hat{\boldsymbol{\beta}}$ for the ith primary session.
(3) Estimate the demographic parameters using $\hat{\boldsymbol{\gamma}}$ and formulae (6) and (7), with the $\hat{p}_i^*$ as calculated at step 2.
(4) Calculate the standard errors with the parametric bootstrap. For each capture history $\boldsymbol{\omega}$, a Poisson random variable with mean $\hat{\mu}_\omega$ is simulated. The parameters for this sample are estimated by going through steps 1–3 of this procedure. This is repeated L times. The variance of a parameter estimate is estimated by the variance of the L bootstrap estimates for this parameter.

The assumption underlying this fitting procedure is that the observed frequencies n_ω follow independent Poisson distributions. This differs from Kendall et al. (1995), who assume that the number of unmarked animals before primary sampling period i, U_i, is fixed. This usually has a small impact on the results of the analysis.

4.1. *The Special Case $I = 2$*

Robust designs with $I = 2$ are interesting alternatives to closed population models. They permit investigating the closure assumption. The $\mathbf{Z}$ component in (8) has a simple form, the row for capture history $\boldsymbol{\omega}$ is $\mathbf{Z}_{\delta(\omega)} = (1,\ 1 - \delta_1(\omega),\ 1 - \delta_2(\omega))$. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ denote the components of the $\mathbf{X}$ matrix for the two primary sampling periods. These matrices have $2^{\ell_1+\ell_2} - 1$ rows; their columns depend on the specification of the mark–recapture model within each period. This section shows how the parameter estimates for the robust design are related to the estimates obtained by fitting the closed population models with design matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ to the data for the two primary periods. The following notation is used in this section, $\mathbf{1}$, $\boldsymbol{\delta}_{[1]}$, $\boldsymbol{\delta}_{[2]}$, $\mathbf{n}$, and $\boldsymbol{\mu}$ denote $(2^{\ell_1+\ell_2} - 1) \times 1$ vectors, with respective components 1, $\delta_1(\omega)$, $\delta_2(\omega)$, n_ω, and μ_ω.

In the Poisson regression for (8), the estimating equations for the parameters are $(\mathbf{Z},\ \mathbf{X})'(\mathbf{n} - \boldsymbol{\mu}) = 0$. Simple manipulations show that this system of equations is equivalent to $(\mathbf{1},\ \boldsymbol{\delta}_{[1]},\ \mathbf{X}_1,\ \boldsymbol{\delta}_{[2]},\ \mathbf{X}_2)'(\mathbf{n} - \boldsymbol{\mu}) = 0$. Now, $(\boldsymbol{\delta}_{[1]},\ \mathbf{X}_1)'(\mathbf{n} - \boldsymbol{\mu}) = 0$ and $(\boldsymbol{\delta}_{[2]},\ \mathbf{X}_2)'(\mathbf{n} - \boldsymbol{\mu}) = 0$ are the Poisson estimating equations for fitting the closed population models corresponding to $\mathbf{X}_1$ and $\mathbf{X}_2$ to the mark–recapture data for the first and the second primary sampling session, respectively. Thus $\hat{N}_1$ and $\hat{N}_2$ calculated with the robust design coincide with the closed population estimates for the two primary sampling sessions. A simple estimate for ϕ_1 is obtained by noting that $(\boldsymbol{\delta}_{[1]} + \boldsymbol{\delta}_{[2]} - \mathbf{1})'\boldsymbol{\mu} = N_1 p_1^* \phi_1 p_2^*$ is the predicted frequency for the animals caught in the two primary sampling sessions. This suggests taking $\hat{\phi}_1 = (\boldsymbol{\delta}_{[1]} + \boldsymbol{\delta}_{[2]} - \mathbf{1})'\mathbf{n}/(\boldsymbol{\delta}'_{[1]}\mathbf{n}\hat{p}_2^*)$.

When $I > 2$, the robust design and closed population model estimates of N_i and p_i^* coincide provided that $\boldsymbol{\delta}_{[i]}$, the vector of the ith components of $\boldsymbol{\delta}$, is in the column space of $\mathbf{Z}$. For (8), it does so for $i = 1$ and $i = I$ only. This agrees with the observation made in Section 3 that the capture probabilities in the first and the last sessions are not estimable using between primary session information only. For periods 2 to $I - 1$, the estimators for p^* combine within and between primary sampling period information; they differ from the closed population model estimators which are based on within primary period information only.

4.2. *Accounting for Temporary Emigration*

Temporary emigration occurs when a unit available for capture leaves the study area temporarily and comes back later in the experiment. Let τ_i^* denote the probability that a unit alive in the survey area at the start of the ith primary sampling period is available for capture for that sampling period. Formula (3) estimates the probability that a unit is captured given that it is present. However, as noted by Burnham (1993), in formulae (6) and (7) the probability p_i^* should in this context be replaced by the probability that a unit is available and is captured, i.e., by $\tau_i^* p_i^*$.

Temporary emigration makes the within session capture probabilities (p_i^*) larger that the between session capture probabilities ($\tau_i^* p_i^*$). These discrepancies can be included in the Poisson regression for the robust design by adding $I - 2$ columns to the $\mathbf{Z}$ component of the design matrix. These columns are $\boldsymbol{\delta}_{[i]}$, the vector of the ith components of capture history $\boldsymbol{\delta}$, for $i = 2, \ldots, I - 1$. The additional loglinear parameters are the differences, on the logit scale, of the two capture probabilities

$$\gamma_{2I+i-3} = \log\left(\frac{\tau_i^*(1-p_i^*)}{1-\tau_i^* p_i^*}\right)$$

for $i = 2, \ldots, I - 1$. Observe that $\gamma_{2I+i-3} < 0$.

The sufficient statistics for the Poisson regression featuring a temporary emigration split nicely into within and between primary period components. Sufficient statistics $\boldsymbol{\delta}'_{[i]}\mathbf{n}$ and $\mathbf{X}'_i\mathbf{n}$ are equal to those for fitting the closed population model corresponding to $\mathbf{X}_i$ to the data for the ith primary

session. Also $\mathbf{Z}'\mathbf{n}$, $\boldsymbol{\delta}'_{[2]}\mathbf{n}, \ldots, \boldsymbol{\delta}'_{[I-1]}\mathbf{n}$, are the sufficient statistics for the Cormack–Jolly–Seber model of Section 3 fitted to the observed frequencies for the between primary sessions capture histories, $\boldsymbol{\delta}$. The parameter estimates for the robust design can be obtained by fitting separate closed population models to the data for the individual sampling sessions and a Cormack–Jolly–Seber model to the between session data. The ratio of the closed population model estimate of the population size for the ith session to the corresponding Cormack–Jolly–Seber estimate provides an estimate for τ_i^* (see Schwarz and Stobo [1997]).

Fitting the robust design in a single Poisson regression can be useful if some of the $\boldsymbol{\gamma}$-parameter values are out of range. Parameters with inadmissible estimates are set to 0, and the model is refitted with a smaller set of parameters. The transformation of the loglinear parameters into demographic parameters proceeds in the following three steps: (1) estimate $\hat{p}_i^*$ the within primary period capture probability using (3), (2) estimate the between primary period capture probability $\hat{\tau}_i^*\hat{p}_i^*$ by $\hat{p}_i^* \exp(-\hat{\gamma}_{2I+i-3})/\{1-\hat{p}_i^*+\hat{p}_i^* \exp(-\hat{\gamma}_{2I+i-3})\}$ for $i = 2, \ldots, I-1$, and (3), estimate the demographic parameters using $(\hat{\gamma}_0, \ldots, \hat{\gamma}_{2I-2})$ and formulae (6) and (7), with $\hat{p}_i^*$ replaced by $\hat{\tau}_i^*\hat{p}_i^*$ calculated at step (2).

The model for temporary emigration can be enlarged to account for a possible Markovian dependency. It suffices to add to the $\mathbf{Z}$ component of the model the $I-1$ interaction terms $\boldsymbol{\delta}_i\boldsymbol{\delta}_{i+1}$. In the context of open population models, Cormack (1989) suggested adding similar terms to the loglinear model design matrix to account for trap dependence.

4.3. *Reducing the Computational Burden*

The example of the next section has six primary sessions, with three capture occasions for each session. There are $2^{18} - 1 = 262{,}143$ capture histories. This large sample size in the Poisson regression may create computation problems. This section suggests an alternative formulation for the Poisson regression, when either M_0 or M_h is used for the within session capture mechanisms.

Suppose for now that there is only one primary session ($I = 1$), with $\ell_1 = \ell$ capture occasions. Model M_0 can be expressed in terms of the frequencies f_h, representing the number of units captured exactly h times as

$$E(f_h) = \binom{\ell}{h} \exp(\gamma + h\beta), \qquad h = 1, \ldots, \ell.$$

An alternative to the Poisson regression of Section 2 for fitting M_0 is a Poisson regression with the $\ell \times 1$ vector of the f_h's as dependent variable, an $\ell \times 2$ design matrix, and a vector of offsets given by $(-\log\binom{\ell}{h} : h = 1, \ldots, \ell)$. This brings the size of the dependent vector to ℓ, down from $2^\ell - 1$. Models $M_{\mathrm{D}h}$ and $M_{\mathrm{C}h}$ can also be fitted in this way.

This approach generalizes to the robust design, where either M_0 or M_h is used within all the sessions. The capture histories $\boldsymbol{\omega}$ can be replaced by $I \times 1$ vector $\mathbf{h} = (h_1, h_2, \ldots, h_I)$ where h_i gives the number of times a unit is captured during session i. In this context f_h is the frequency of the units captured h_i times in primary period i, for $i = 1, \ldots, I$. Observe that $E(f_h)$ has a loglinear structure similar to (8), with an offset given by $-\log\prod\binom{\ell_i}{h_i}$. Thus, with this offset variable, (8) can be fitted by using a Poisson regression with the vector of f_h's as dependent variable. In the example of Section 5, this reduces the size of the dependent vector to $4^6 - 1 = 4095$ entries.

5. Data Analysis: The Estimation of Abundance and Survival of Red-Back Voles in Southeastern Québec

This section analyzes the data from a two-year study of the demographics of the red-back vole (*Clethrionomys gapperi*) in the Duchénier conservation area in southeastern Québec. Data collection was carried out by Pierre Etcheverry, Michel Crête, and Jean-Pierre Ouellet. Duchénier covers 271 km^2 so that mark–recapture was carried out at a sample of 11 locations, randomly selected in the study area. The sampling points were observed during six primary sampling periods, in May, July, and August of 1999 and 2000. One sampling period consisted of three successive nights of mark–recapture. Grids of 10×10 Sherman live traps covering 1 ha were used to capture voles at the sampling points. Over the two years of the study, each sampling point was visited 18 times. The data from the 11 sampling points have been pooled for the analysis.

5.1. *Analysis within Primary Periods Using Closed Population Models*

Table 2 presents the mark–recapture data for the second primary sampling session. The reduced captures on the first night suggests a time effect and the relatively large number of voles caught three times supports the presence of heterogeneity. The deviances of M_0, M_t, M_h, and M_{th} are equal to 21.8, 15.8, 9.3, and 2.9 with respectively 5, 3, 4, and 2 degrees of freedom. Thus M_{th} provides the best fit.

Models $M_{\mathrm{D}th}$ and $M_{\mathrm{C}th}$ have the same deviance of 2.9 since the column spaces of their design matrices are identical, as can be seen in Table 2. These two models have the same heterogeneity parameter β_4; however the intercept for Darroch's model is equal to that for Chao's plus β_4. Thus Darroch's model gives larger estimates of abundance than Chao's in the presence of heterogeneity.

In a small mark–recapture experiment such as that reported in Table 2, $\hat{\beta}_4$ is usually highly variable and Darroch's estimate of abundance is not reliable. Indeed for the data of Table 2, Darroch's estimate is $\hat{N}_{\mathrm{D}} = 905$ (SE 681) while for Chao $\hat{N}_{\mathrm{C}} = 228$ (SE 42); both estimates were calculated

Table 2
Capture–recapture data for the second sampling session and the column of the design matrix for heterogeneity, $\mathbf{X}_4$, of $M_{\mathrm{D}th}$ and $M_{\mathrm{C}th}$

Capt. hist.			$\mathbf{X}_4$		
ω_1	ω_2	ω_3	$M_{\mathrm{D}th}$	$M_{\mathrm{C}th}$	n_ω
0	0	1	1/2	0	33
0	1	0	1/2	0	32
0	1	1	2	0	5
1	0	0	1/2	0	15
1	0	1	2	0	4
1	1	0	2	0	7
1	1	1	9/2	1	9

Table 3
Number of voles caught on the 18 *days of the experiment (D1 to D18) and the number of recaptures after* 1 *day (Recap1),* 2 *days (Recap2), and* 3 *days or more (Recap3$^+$)*

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18
No. caught	30	37	59	35	53	51	59	65	51	29	26	25	47	79	74	36	33	36
Recap1	9	14	3	16	14	2	15	21	0	12	14	3	23	35	2	13	12	NA
Recap2	8	3	1	4	2	1	8	1	0	7	1	1	7	1	2	6	NA	NA
Recap3$^+$	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA

with Rivest and Levesque (2001) frequency corrections. The heterogeneity parameter cannot be estimated accurately with only three capture occasions; the conservative Chao estimate is then preferred. In the robust design for the vole data, six primary sessions are available to estimate β_4. If the model with a common heterogeneity parameter for the six periods fits well, Darroch's model could be used. On the other hand, if the heterogeneity parameters differ significantly between primary sessions, Chao's approach of minimizing the impact of the heterogeneity is better.

5.2. *Between Primary Periods Analysis*

Table 3 provides a summary of the number of voles caught on each day and of their eventual recapture. Less than 30 voles are captured at more than one primary session. This suggests a high mortality between sessions.

The robust design models were fitted using SAS's PROC GENMOD. Table 4 gives the deviances for six models. The index t common to all models means that the parameters for the capture of the units within the sessions vary from one session to the next. The subscript refers to the modeling of the capture mechanism within a session. Following the discussion in Section 5.1, Darroch's model has one common heterogeneity parameter for all the sessions; this parameter is allowed to vary between sessions with Chao's model.

Comparing the deviances for $M^t_{\mathrm{D}th}$ and M^t_t tests for the presence of heterogeneity. The χ^2_1 observed value is $20\,(p < .001)$, and the null hypothesis of no heterogeneity is rejected. To test whether this heterogeneity can be described by a common parameter for the six sessions, the χ^2_5 statistics obtained by comparing the deviances of $M^t_{\mathrm{C}th}$ and $M^t_{\mathrm{D}th}$ is $15\,(p = .01)$. The null hypothesis is rejected and the heterogeneity parameters appear to vary, from -0.61 (SE 0.65) in session 4 to 2.09 (SE 0.6) in session 2. Thus the second primary session has a strong heterogeneity, as noted in Section 5.1, while the fourth session does not exhibit any individual heterogeneity. The best fitting model is $M^t_{\mathrm{C}th}$. It has 18 capture probabilities, 6 heterogeneity parameters, and 11 demographic parameters for a total of 35 parameters.

Table 5 gives the parameter estimates for three models derived using (6) and (7) as proposed in Section 4. The estimates for M^t_t are almost identical to MARK's estimates. Model M^t_t gives estimates that are generally smaller than those obtained with $M^t_{\mathrm{C}ht}$. The survival probability estimates $\hat\phi_i$ for $M^t_{\mathrm{C}ht}$ are 10 to 20% larger than for M^t_t. Thus M^t_t appears to underestimate ϕ_i in the presence of heterogeneity. Under $M^t_{\mathrm{D}ht}$, the common heterogeneity parameter is estimated by 1.02 (SE 0.22). This parameter has a substantial impact on the $\hat\phi_i$'s, increasing them by about 40%, when compared with $M^t_{\mathrm{C}ht}$'s estimates. They have to be considered with caution since the data do not support a constant heterogeneity.

In Table 5 the estimates of survival probabilities have coefficients of variation around 40%. They have been calculated using 200 parametric bootstrap replicates of the dependent vector, featuring more than 250,000 entries. The poor precision of the survival probabilities was to be expected given the limited number of voles captured at more than one session. In this context, the robust design is useful since it allows all the parameters to be estimated. Indeed, an open population model cannot be fitted to the between primary session capture histories because there are too many zeros in the dependent vector. The scarce between session data do not allow to test for temporary emigration as proposed in Section 4.2. Also the $\hat p^*_i$'s reflect mostly within primary session information. This makes the $\hat N_i$'s of Table 5 very close to the estimates obtained with M_t, $M_{\mathrm{C}ht}$, and $M_{\mathrm{D}ht}$ for the individual sessions. The estimated probability of surviving the 1999–2000 winter, $\hat\phi_3$, is, as expected, small.

It is interesting to compare the fits $M^t_{\mathrm{C}h}$ and $M^t_{\mathrm{C}ht}$ since, as shown in Section 4.3, the size of the dependent vector can be

Table 4
Deviances for six robust design models fitted to the vole data

Model	M^t_0	$M^t_{\mathrm{D}h}$	$M^t_{\mathrm{C}h}$	M^t_t	$M^t_{\mathrm{D}th}$	$M^t_{\mathrm{C}th}$
No. of parameters	17	18	23	29	30	35
Deviance	232	213	198	192	172	157

Table 5
Demographic parameter estimates, and their coefficients of variation in percentage, for three loglinear models for the robust design

Parameter	M^t_t	$M^t_{\mathrm{C}ht}$	$M^t_{\mathrm{D}ht}$
$\hat\phi_1$	0.147 (34)	0.191 (38)	0.266 (37)
$\hat\phi_2$	0.101 (38)	0.106 (43)	0.183 (36)
$\hat\phi_3$	0.008 (45)	0.008 (47)	0.011 (41)
$\hat\phi_4$	0.137 (43)	0.153 (43)	0.224 (32)
$\hat\phi_5$	0.052 (44)	0.065 (47)	0.095 (45)
$\hat N_1$	149 (13)	156 (16)	292 (28)
$\hat N_2$	168 (13)	225 (21)	328 (25)
$\hat N_3$	210 (12)	227 (15)	421 (23)
$\hat N_4$	54 (15)	51 (17)	79 (23)
$\hat N_5$	176 (9)	196 (13)	300 (20)
$\hat N_6$	105 (14)	131 (22)	197 (23)

reduced drastically for models without a time effect. Table 4 shows a relatively large deviance difference of 41 on 12 degrees of freedom between these two models. Still their parameter estimates are remarkably similar. The differences are less than 2% for the 11 parameters of Table 5. The estimated standard errors are also of similar magnitude. This suggests that M^t_{Ch} is a useful alternative to M^t_{Cht} when the number of capture occasions is large.

6. Discussion

This article has shown how to use loglinear models to estimate the parameters of a robust design. Two methods for coping with a possible heterogeneity in the capture probabilities of the primary sessions have been proposed. A loglinear analysis of the robust design can be implemented on standard packages for statistical data treatment.

The loglinear approach to the robust design has its limitations. It works well for M^t_0, M^t_t, M^t_h, and M^t_{ht}. However it cannot handle M^t_b and M^t_{bh}. Rivest and Lévesque (2001) present loglinear models for M_b and M_{bh}; unfortunately these models cannot be combined with (4) in a simple loglinear way. Thus the fitting method presented by Kendall et al. (1995) is the only technique available for M^t_b.

Acknowledgements

We are grateful to Pierre Etcheverry, Michel Crête, and Jean-Pierre Ouellet for providing the data on the vole study presented in Section 5. The financial support of the Natural Science and Engineering Research Council of Canada and of the Fonds québecois de la recherche sur la nature et les technologies are gratefully acknowledged.

Résumé

Le "dispositif robuste" est un type d'étude de capture–recapture ayant un plan d'échantilonnage emboîté. Le premier degré comporte des sessions primaires d'échantillonnage; la population subit de la mortalité et reçoit des immigrants entre ces sessions primaires ce qui fait que les modèles pour populations ouvertes s'appliquent. Le deuxième degré d'échantillonnage comporte de courtes études de marquage-récupération dans chaque session primaire. Les modèles pour populations fermées sont utiles à ce niveau pour estimer le nombre d'animaux pour chaque session primaire. Cet article suggère une approche loglinéaire à l'ajustement d'un dispositif robuste. Les modèles loglinéaires utilisés pour traiter les données de marquage-récupération récoltées lors d'études de populations ouvertes et fermées sont d'abord passés en revue. Ces deux types de modèles sont ensuite combinés pour analyser les données obtenues grâce à un dispositif robuste. La méthode loglinéaire mise de l'avant permet d'inclure des paramètres pour l'hétérogénéité des probabilités de capture dans chaque session primaire. L'émigration temporaire hors de l'aire d'étude peut également être prise en compte dans un modèle loglinéaire. L'analyse statistique est relativement simple; elle s'appuie sur une grosse régression Poisson avec les vecteurs des fréquences observées pour toutes les histoires de capture possibles commme variable dépendante. Un exemple portant sur l'estimation de l'abondance et de la survie du campagnol à dos roux dans une région du sud-est du Québec est présenté en guise d'illustration.

References

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

Burnham, K. P. (1993). A theory for combined analysis of ring recovery and recapture data. In *Marked Individuals in the Study of Bird Populations,* J. P. Lebreton and P. North (eds), 199–214. Basel: Birkhauser Verlag.

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

Chapman, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological censuses. *University of California Publications in Statistics* **1,** 131–160.

Cormack, R. M. (1985). Example of the use of GLIM to analyze capture-recapture studies. In *Statistics in Ornithology,* B. J. T. Morgan and P. M. North (eds), *Lecture Notes in Statistics,* Volume 29, 242–274. New York: Springer-Verlag.

Cormack, R. M. (1989). Loglinear models for capture–recapture. *Biometrics* **45,** 395–413.

Cormack, R. M. (1993). Variance of mark–recapture estimates. *Biometrics* **49,** 1188–1193.

Coull, B. A. and 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., and Junker, B. (1993). A three sample multiple capture–recapture approach to the census population estimation with heterogeneous catchability. *Journal of the American Statistical Association* **88,** 1137–1148.

Kendall, W. L. and Bjorklan, R. (2001). Using open robust design models to estimate temporary migration from capture–recapture data. *Biometrics* **57,** 1113–1122.

Kendall, W. L., Pollock, K. H., and Brownie, C. (1995). A likelihood-based approach to capture–recapture estimation of demographic parameters under the robust design. *Biometrics* **51,** 293–308.

Kendall, W. L., Nichols, J. D., and Hines, J. E. (1997). Estimating temporary emigration using capture–recapture data with Pollock's robust design. *Ecology* **78,** 563–578.

Pollock, K. H. (1982). A capture–recapture design robust to unequal probabilities of capture. *Journal of Wildlife Management* **46,** 757–760.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). *Statistical Inference in Capture–Recapture Experiments. Wildlife Society Monographs*, Volume 107.

Rivest, L.-P. and Lévesque, T. (2001). Improved loglinear model estimators for abundance in capture–recapture experiments. *Canadian Journal of Statistics* **29,** 555–572.

Sandland, R. L. and Cormack, R. M. (1984). Statistical inference for Poisson and multinomial models for capture-recapture experiments. *Biometrika* **71,** 27–33.

Schwarz, C. J. and Stobo, W. T. (1997). Estimating temporary migration using the robust design. *Biometrics* **53,** 178–194.

White, G. C. and Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study* **46**(Suppl.), 120–138.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and Management of Animal Populations*. San Diego: Academic Press.

Received December 2002. *Revised August* 2003. *Accepted September* 2003.

Appendix

Derivation of Darroch et al. (1993) model. Let $f(h)$ be the density of the latent variable; the marginal probability of capture history ω is

$$\Pr(\omega) = \int \left\{ \frac{\exp\left\{ \sum \omega_j \beta_j + h\Big(\sum \omega_j \Big) \right\}}{\prod \{1 + \exp(\beta_j + h)\}} \right\} f(h)\, dh.$$

The posterior density of h for an animal that is not caught is proportional to $f(h)/\prod\{1 + \exp(\beta_j + h)\}$. Assuming that this is a $N(0,\, \beta_{\ell+1})$ density gives

$$\begin{aligned} \Pr(\omega) &= C^{-1} \exp\Big(\sum \omega_j \beta_j \Big) \\ &\quad \times \int_{-\infty}^{\infty} \frac{\exp\Big\{ h\Big(\sum \omega_j \Big) - h^2 \Big/ (2\beta_{\ell+1}) \Big\}}{\sqrt{2\pi\beta_{\ell+1}}}\, dh \\ &= C^{-1} \exp\Big(\sum \omega_j \beta_j + \Big(\sum \omega_j \Big)^2 \beta_{\ell+1} \Big/ 2 \Big), \end{aligned}$$

where C is a normalizing constant.

On the estimation of the parameters for M_{Cth}. A reparametrization of (2) in terms of $\gamma,\, \beta_1, \ldots, \beta_\ell$ and $\beta_{\omega^*} = \beta_\omega - \gamma - \sum_i \omega_i \beta_i, \sum \omega_i > 2$ is

$$\log \mu_\omega = \begin{cases} \gamma + \displaystyle\sum_{j=1}^{\ell} \omega_j \beta_j, & \displaystyle\sum_{j=1}^{\ell} \omega_j \le 2, \\ \beta_{\omega^*}, & \displaystyle\sum_{j=1}^{\ell} \omega_j > 2. \end{cases}$$

The Poisson estimating equations for $\gamma,\, \beta_i,\, i = 1, \ldots, \ell$ are

$$\sum_i n_{1;i} + \sum_{j>i} n_{2;i,j} = \exp(\gamma) \left\{ \sum_i \exp \beta_i + \sum_{j>i} \exp(\beta_i + \beta_j) \right\},$$

$$n_{1;i} + \sum_{j \ne i} n_{2;i,j} = \exp(\gamma + \beta_i) \left(1 + \sum_{j \ne i} \exp \beta_j \right), \qquad i = 1, \ldots, \ell,$$

where $n_{1;i}$ (resp. $n_{2;i,j}$) denotes the number of animals caught once (twice), at occasion i (i and j). The solutions to these equations depend only on the counts of the animals captured once or twice. When $\beta_1 = \beta_2 = \cdots = \beta_\ell = \beta$ the Poisson estimating equations for $(\gamma,\, \beta)$ become

$$f_1 + f_2 = \ell \exp(\gamma + \beta) \left\{ 1 + \frac{\ell - 1}{2} \exp \beta \right\},$$

$$f_1 + 2f_2 = \ell \exp(\gamma + \beta) \{ 1 + (\ell - 1) \exp \beta \},$$

where f_1 (resp. f_2) denotes the frequency of animals caught exactly once (resp. twice). These equations have closed form solutions, in particular $\hat{\gamma} = \log\{(\ell - 1) f_1^2 / (2\ell f_2)\}$. When ℓ is large $\exp(\hat{\gamma})$ is equal to Chao's (1989) estimators for the number of missed animals.