Why a time effect often has a limited impact on capture-recapture estimates in closed populations

Louis-Paul RIVEST

Key words and phrases: Heterogeneity in capture probabilities, log-convexity, mixture models, multinomial distribution, overdispersion, Poisson regression.

Abstract: This paper is concerned with loglinear estimators of $N$, the size of the population, in a capture-recapture experiment featuring an heterogeneity in the individual capture probabilities and a time effect. Models where the first capture influences the probability of subsequent captures, are also considered briefly. Several results are derived from a new inequality associated with a dispersive ordering for discrete random variables. In a loglinear model with heterogeneity, the estimator $\hat{N}$ is shown to increase with the heterogeneity parameter. Including a time effect decreases $\hat{N}$ in models without heterogeneity. Finally a model featuring heterogeneity can accommodate a time effect through a small change of its heterogeneity parameter. This is demonstrated using an inequality for the estimators of the heterogeneity parameters and in a Monte Carlo experiment.

Pourquoi un effet temporel a souvent peu d’impact sur les estimateurs de capture-recapture dans des populations fermées

Résumé : Cet article traite des estimateurs loglinéaires de la taille $N$ d’une population dans une expérience de capture-recapture où les probabilités de capture varient dans le temps et entre individus. Des modèles où la première capture change la probabilité de captures ultérieures sont aussi considérés brièvement. De nombreux résultats sont démontrés à l’aide d’une nouvelle inégalité associée à un ordre de dispersion pour des variables aléatoires discrètes. On montre d’abord que l’estimateur $\hat{N}$ produit par un modèle loglinéaire avec hétérogénéité inter individu est une fonction croissante du paramètre pour l’hétérogénéité. Inclure des variations temporelles des probabilités de capture diminue $\hat{N}$ dans un modèle sans hétérogénéité. Finalement un modèle avec hétérogénéité peut s’accomoder d’un effet temporel en changeant son paramètre d’hétérogénéité. Ce résultat est démontré par le biais d’une inégalité pour les estimateurs des paramètres d’hétérogénéité obtenus avec les deux modèles et dans une étude par simulations.

1. INTRODUCTION

The estimation of the size $N$ of a closed population using mark recapture techniques is an important topic in population management. The data is collected over $t \geq 2$ capture occasions. In experiments involving small mammals, a grid of live traps is installed over the target area and the capture occasions are often successive days. A capture history is a vector $\omega = (\omega_1, \omega_2, \ldots, \omega_t)$ where $\omega_i = 1$ for a capture at occasion $i$ and 0 for a miss. When an animal is caught for the first time, it is given a unique tag that allows its identification at subsequent captures. Thus each animal captured at
least once has an observable capture history \( \omega \). The aim of the analysis is to estimate \( n_0 \), the number of animals that are never caught in the experiment, using the frequencies \( n_{\omega} \) of the \( 2^t - 1 \) observable capture histories.

In experiments involving small mammals, the daily capture probabilities often vary with the weather conditions. Cold temperatures may make the animals passive and decrease the probability that they are captured. Model \( M_t \) applies in this case, where the subscript \( t \) denotes a time effect. Active animals might also have larger capture probabilities than animals that are passive. This is an instance of model \( M_h \), where the subscript \( h \) is associated with heterogeneous capture probabilities. Model \( M_{th} \) featuring both a time effect and an heterogeneity in capture probabilities is considered in this work.

Model \( M_h \) has been extensively studied and several estimation procedures are available. Burnham & Overton (1978) and Pollock & Otto (1983) propose a jackknife estimator, Chao, Lee & Jeng (1993), Agresti (1994) and Coull & Agresti (1999) investigate continuous mixture models, Norris & Pollock (1996) and Pledger (2000) consider finite mixtures while Dorazio & Royle (2003) use a beta-binomial model. Selecting a model for \( M_h \) on the basis of a goodness of fit statistic and calculating its estimator for \( N \) is not a good strategy since models fitting the data equally well can give drastically different estimators for \( N \) (Huggins, 2001; Link, 2003).

There is no consensus on a unique method to deal with heterogeneous capture probabilities so providing a lower bound estimator (Chao, 1989, Rivest & Baillargeon 2007) is advisable. One may also investigate how including a particular feature in the model, such as a time effect or an heterogeneity in capture probabilities, influences the resulting abundance estimator. This is the approach taken in this paper. Its main finding is that a significant time effect often has a small impact on the value of \( N \). This is especially true when an heterogeneity is present since it is shown that a time effect can be incorporated in a loglinear model for \( M_h \) by decreasing its heterogeneity parameter. Mathematical results explaining why loglinear models for \( M_h \) and \( M_{th} \) often give similar estimators of abundance are provided.

The main mathematical tool for the derivations is a dispersive ordering for discrete random variables. The key inequality is established in Section 2 following a brief review of log-convex orderings. This inequality applies to estimators of abundance derived from loglinear models for \( M_{th} \). These estimators are reviewed in Section 3 where the main results are presented. Models \( M_h \) and \( M_{th} \) having a behavioural effect are also considered briefly; subscript \( b \) means that the probability of capture changes after the first capture. A simulation study is presented in Section 4 to validate some of the findings.

2. A NEW INEQUALITY FOR LOG-CONVEX SEQUENCES

This section reviews a log-convex ordering considered by Whitt (1985); it derives a new crossing property for ordered sequences.

**Definition 1.** (Whitt, 1985) Let \( a_k \) and \( b_k \) be for \( k = 0, 1, \ldots, t \) two sequences of non-negative numbers. The sequence \( \{a_k\} \) is log-convex with respect to \( \{a_k\} \) if \( (b_{k}/a_k)^2 \leq (b_{k-1}/a_{k-1})(b_{k+1}/a_{k+1}) \) for \( k = 1, \ldots, t - 1 \). If \( \{b_k\} \) is log-convex with respect to \( \{a_k\} \) then \( \{a_k\} \) is said to be log-concave with respect to \( \{b_k\} \).

If the probability mass function of \( Y \) is log-convex with respect to that of \( X \) one says that \( Y \) is “more dispersed” than \( X \). Let \( X \) be a random variable having a binomial distribution with parameters \( t \) and \( p \). Heterogeneous capture probabilities can be modelled using a binomial mixture whose probability mass function is given by \( p_{th}(k) = \binom{t}{k} \int_0^1 p^k(1-p)^{t-k} dF(p) \) for \( k = 0, \ldots, t \), where \( F(p) \) is the mixture distribution. Several authors, including Whitt (1985), Lindsay (1986) and Gelfand & Dalal (1990), have shown that \( p_{th}(k)/P(X = k) \) is log-convex, that is \( p_{th}(k)/\binom{t}{k} \) is
log-convex, for any mixing distribution $F$. Thus binomial mixtures are more dispersed than the binomial distribution.

Let $p_t(k)$, $k = 0, \ldots, t$ denote the probability function of the sum of $t$ independent nonidentically distributed Bernoulli random variables. Hardy, Littlewood & Pólya (1952, p. 52) and Marshall & Olkin (1979, p. 93) prove that $p_t(k)/P(X = k)$ is log-concave so that $p_t(k)$ is less dispersed than the binomial distribution with parameters $(t, p)$. The next proposition gives a crossing property and a moment inequality for random variables that are ordered according to the log-convex ordering.

**Proposition 1.** (Shaked, 1980) Let $X$ and $Y$ be two discrete random variables such that the probability mass function of $Y$ is log-convex with respect to that of $X$. If $E(X) = E(Y)$, then

1. $E\{\psi(Y)\} \geq E\{\psi(X)\}$, for any convex function $\psi(\cdot)$, the inequality is strict if $\psi(\cdot)$ is strictly convex, that is if $\psi(k - 1) + \psi(k + 1) > 2\psi(k)$ for all $k > 0$, and if $X$ and $Y$ do not have the same probability mass function;

2. As $k$ goes from 0 to $t$, $P(Y = k) - P(X = k)$ has exactly two sign changes and the sign sequence is $(+,-,+)$.

In a capture-recapture experiment, let $Y$ represent the number of captures for an animal. Model $M_0$ has no time effect and no heterogeneity in capture probabilities, thus $Y$ has a binomial distribution with variance $\text{Var}_0(Y) = tp(1 - p)$ where $p$ is the capture probability at one occasion. For $M_t$, $Y$ is distributed as a sum of independent non-identically distributed Bernoulli random variables and Proposition 1 implies that $\text{Var}_t(Y) \leq tp(1 - p)$ where $p = \sum p_i/t$ and $p_i$ is the capture probability at occasion $i$. Under model $M_h$, $p$ is a random variable and $\text{Var}_h(Y) \geq t E(p)(1 - E(p))$, by Proposition 1. Thus $\text{Var}_h(Y) \geq \text{Var}_0(Y) \geq \text{Var}_1(Y)$; Section 3.2 shows that this ordering also holds for the loglinear estimators of $N$ derived under these three models. The following proposition is applied repeatedly in the next section; its proof is given in the Appendix.

**Proposition 2.** Let $\{a_k; k = 0, \ldots, t\}$ and $\{b_k; k = 0, \ldots, t\}$ be two sequences of positive numbers such that

1. $\{b_k\}$ is log-convex with respect to $\{a_k\}$;
2. $\sum_1^t b_k = \sum_1^t a_k$ and $\sum_1^t kb_k = \sum_1^t ka_k$,

then $b_0 \geq a_0$.

If the sequences $\{a_k; k = 1, \ldots, t + 1\}$ and $\{b_k; k = 1, \ldots, t + 1\}$ satisfy assumptions $i)$ and $ii)$ of Proposition 2 then $b_{t+1} \geq a_{t+1}$. The proof of this result mimics that of Proposition 2.

3. LOGLINEAR MODELS FOR ESTIMATING THE SIZE OF A CLOSED POPULATION

3.1 A review of some well known models

For model $M_t$, Cormack (1989) observed that the predicted value for capture history $\omega$ is

$$\mu_\omega = E(n_\omega) = N \prod_{j=1}^t p_j^{\omega_j} (1 - p_j)^{1-\omega_j}.$$ 

This has a loglinear form, $\log \mu_\omega = \gamma + \sum_{j=1}^t \omega_j \beta_j$, where $\beta_j = \log\{p_j/(1 - p_j)\}$ and $\gamma = \log\{N \prod (1 - p_j)\}$. If $\beta_1 = \ldots = \beta_t = \beta$, one gets a loglinear model for $M_0$. Following Darroch et al. (1993) and Agresti (1994), Rivest & Baillargeon (2007) discuss how to extend this
model to account for heterogeneous capture probabilities. Their proposal for model $M_{th}$ is

$$\log \mu_\omega = \gamma + \sum_{j=1}^{t} \omega_j \beta_j + \tau \psi(\sum \omega_j), \quad (1)$$

where $\psi(k)$ is a convex function of $k$ and $\tau$ is the heterogeneity parameter. Taking $\psi(k) = \exp(ak) - 1$ for some $a > 1$ yields a model where the logits of the individual capture probabilities have a scaled Poisson mixture distribution with shape parameter $\tau$. The functions $\psi(k) = k^2/2$ and $\psi(k) = -\log(\lambda + k) + \log(\lambda)$, for some $\lambda > 0$, give mixing distributions equal to a normal mixture and to a mixture of the negative of a gamma random variable. In these three cases, the variance of the logits of the capture probabilities increases with the heterogeneity parameter $\tau$. The bias of Chao’s lower bound for $N$ under model $M_{th}$ is also an increasing function of the heterogeneity parameter $\tau$, see Table 1 of Rivest & Baillargeon (2007) which provides a closed form expression for this bias. The abundance estimator obtained with these three $\psi$-functions are typically ordered $\text{Poisson} < \text{normal} < \text{gamma}$. The mixing distributions underlying (1) depend on the the number of occasions $t$; thus the heterogeneity parameters $\tau$ for different values of $t$ are not really comparable. This point is discussed by McCullagh (1994) who argues that for $\psi(k) = k^2/2$, $\tau/t$ is nearly invariant with respect to $t$.

In (1) the intercept $\gamma$ stands for the logpredicted frequency for the units that were missed in the experiment. A simple method to estimate the parameters of (1) is to use a generalized linear model, with a Poisson distribution and a logarithm link function (see Sandland and Cormack, 1984, and Cormack and Jupp, 1991). An efficient estimator of the populations size is

$$\hat{N} = n + \exp \hat{\gamma}, \quad (2)$$

where $n = \sum n_\omega$ is the number of animals caught at least once. Its variance is estimated by

$$v(\hat{N}) = \exp \hat{\gamma} + \exp(2\hat{\gamma})v(\hat{\gamma}),$$

where $v(\hat{\gamma})$ is the variance estimate for the intercept in the Poisson regression, see Rivest and Lévesque (2001) for details.

Let $f_k$ and $\hat{\phi}_k$ be the observed and estimated predicted frequencies for the number of animals caught $k$ times,

$$f_k = \sum_{\omega_i = k} n_\omega \quad \text{and} \quad \hat{\phi}_k = \sum_{\omega_i = k} \hat{\mu}_\omega, \quad k = 1, \ldots, t$$

where $\hat{\mu}_\omega$ is the estimated predicted frequency for capture history $\omega$. Note that $\hat{\phi}_0 = \exp \hat{\gamma}$ estimates the number of animals that were missed. For all the possible submodels of (1), the $\hat{\phi}_k$ satisfy the following Poisson estimating equations,

$$n = \sum_{k=1}^{t} f_k = \sum_{k=1}^{t} \hat{\phi}_k \quad \text{and} \quad (3)$$

$$\sum_{k=1}^{t} kf_k = \sum_{k=1}^{t} k\hat{\phi}_k. \quad (4)$$

Equations (3) and (4) come from the score functions for $\gamma$ and $\beta$, when there is no time effect. In the presence of a time effect, (4) can be derived by summing the $t$ Poisson estimating equations for the $\beta_j$’s. Observe that (3) and (4) mean that the predicted frequencies $\{\hat{\phi}_k : k = 0, \ldots, t\}$ for two different submodels of (1) satisfy condition ii) of Proposition 2. The estimating equation for the heterogeneity parameter $\tau$ is

$$\sum f_k \psi(k) = \sum \hat{\phi}_k \psi(k). \quad (5)$$

### 3.2 Some applications of Proposition 2 to loglinear estimators of abundance
First consider model $M_h$ with $\beta_1 = \ldots = \beta_t = \beta$ in (1). The estimator of abundance obtained for a fixed value of $\tau$ is $\hat{N_h}(\tau) = n + \exp \hat{\gamma}_\tau$, where $\hat{\gamma}_\tau$ is the estimated intercept in a Poisson regression featuring $\tau \psi(\sum \omega_i)$ as an offset. Let $\tau_2 > \tau_1$ be two values for the heterogeneity parameter and $\hat{\phi}_{k1}$ and $\hat{\phi}_{k2}$ be, for $k = 0, \ldots, t$, the corresponding predicted frequencies for the number of units caught $k$ times. That is $\hat{\phi}_{kj} = \binom{n}{j} \exp \{ \hat{\gamma}_0 + k \hat{\beta} + \tau_j \psi(k) \}$ for $j = 1, 2$. These predicted values satisfy equations (3) and (4) and $\hat{\phi}_{k2}$ is log-convex with respect to $\hat{\phi}_{k1}$. The assumptions of Proposition 4 are met, thus $\exp(\hat{\gamma}_2) \geq \exp(\hat{\gamma}_1)$. We have proved the following

**Proposition 3.** Let $\hat{N}_h(\tau_2)$ and $\hat{N}_h(\tau_1)$ be two loglinear estimators of abundance obtained under model $M_h$ when the heterogeneity parameter is set to respectively $\tau_2$ and $\tau_1$. If $\infty > \tau_2 \geq \tau_1 > -\infty$, then $\hat{N}_h(\tau_2) \geq \hat{N}_h(\tau_1)$.

If $\tau > 0$, then $\hat{N}_h(\tau) \geq \hat{N}_h(0) = \hat{N}_0$. Thus when the heterogeneity parameter is positive, the estimator of abundance under $M_h$ is larger than $\hat{N}_0$, that under $M_0$. It may also happen that $\hat{\tau} < 0$. One has $\hat{N}_h < \hat{N}_0$ in this case. This could be associated with an under binomial variation of the number of captures for one animal, as discussed in Section 2. As $\tau \to \infty$, $\hat{N}_h(\tau)$ diverges to $\infty$ so that accounting for an heterogeneity can produce arbitrarily large estimators. This is exemplified by some of the numerical examples of Dorazio & Royle (2003).

As stated in Section 2, if the number of captures has a binomial mixture distribution then $E(f_k) = \binom{n}{k} \exp \psi(k)$, $k = 0, \ldots, t$ for some convex function $\psi(k)$. Fitting $M_0$ to $\{E(f_k) : k = 1, \ldots, t\}$ yields loglinear parameters $\gamma_0$ and $\beta_0$. $E(f_k)$ and the predicted frequencies under $M_0$, $\phi_{k0} = \binom{n}{k} \exp \{ \gamma_0 + k \beta_0 \}$, $k = 0, \ldots, t$, fulfill the assumptions of Proposition 2. Thus $E(f_0) \geq \exp \gamma_0$; this highlights that fitting $M_0$ when $M_h$ is true underestimates $N$, see Hwang & Huggins (2005) for an alternative derivation of this result.

Now we turn to situations where there is a time difference in the capture probabilities. Let $\{\hat{\phi}_{0k}, k = 0, \ldots, t\}$ and $\{\hat{\phi}_{tk}, k = 0, \ldots, t\}$ denote the predicted frequencies under $M_0$ and $M_t$. Observe that $\{\hat{\phi}_{tk}/\hat{N}_t : k = 0, \ldots, t\}$ is the probability function of a sum of independent Bernoulli random variables with varying probabilities, where $\hat{N}_t = \sum_k \hat{\phi}_{tk}$. The inequality of Hardy, Littlewood & Pólya (1952) presented in Section 2 implies that $\hat{\phi}_{0k}$ is log-convex with respect to $\hat{\phi}_{tk}$. In addition, these two sequences satisfy (3) and (4) so that Proposition 2 holds; we have proved the following proposition.

**Proposition 4.** Let $\hat{N}_0$ and $\hat{N}_t$ denote the estimators of $N$ obtained under $M_0$ and $M_t$ then $\hat{N}_0 \geq \hat{N}_t$.

In Proposition 4, $\hat{N}_0$ and $\hat{N}_t$ are the classical estimators for $M_0$ and $M_t$ respectively defined by estimating equations (4.26) p. 164 and (4.4) p. 131 in Seber (1982). When $t = 2$, $\hat{N}_0 = n + (n_{10} + n_{01})^2/(4n_{11})$ and $\hat{N}_t = n + n_{10}n_{01}/n_{11}$; Proposition 4 can be demonstrated using elementary algebra.

In the absence of heterogeneity, $M_0$ is a worst case scenario since it leads to a larger correction for missed animals than $M_t$. More importantly, this suggests that the difference $\hat{N}_t - \hat{N}_0$ is small in many situations where $M_t$ fits better than $M_0$, provided that the time differences in the capture probabilities are small. Under $M_t$ the sufficient statistics are $n = \sum f_k$ and $n_j$, the total number of animals caught at the $j$th occasion for $j = 1, \ldots, t$. One can write $\hat{N}_t = N(n_1, \ldots, n_t)$ and $\hat{N}_0 = N(\bar{n}, \ldots, \bar{n})$, where $\bar{n} = \sum n_j/t$. Proposition 4 means that the function $\hat{N}(\cdot)$ has a maximum at $n_1 = \ldots = n_t = \bar{n}$. In addition $\hat{N}(\cdot)$ does not vary much around this maximum since its gradient is approximately equal to 0. This agrees with many simulation studies which noted that $\hat{N}_0$ is robust to moderate differences in the capture probabilities, see for instance Williams, Nichols, & Conroy (2002), p. 310.

Consider now the loglinear proposals for $M_h$ and $M_{th}$ in (1). The corresponding estimators
of abundance are not necessarily ordered: Proposition 4 does not generalize to \( \hat{N}_h \) and \( \hat{N}_{th} \). The estimated predicted frequencies under \( M_h \) and \( M_{th} \), \( \{ \hat{\phi}_{hk} \} \) and \( \{ \hat{\phi}_{thk} \} \), satisfy (3), (4), and (5). If \( \hat{\tau}_h > \hat{\tau}_{th} \) then \( \{ \hat{\phi}_{hk} \} \) would be log convex with respect to \( \{ \hat{\phi}_{thk} \} \). Proposition 1 would then imply that \( \sum_{k=1}^{t} \psi(k) \hat{\phi}_{hk} > \sum_{k=1}^{t} \psi(k) \hat{\phi}_{thk} \) provided that \( \psi(k) \) is strictly convex. This contradicts (5); we have proved the following result.

**Proposition 5.** Let \( \hat{\tau}_h \) and \( \hat{\tau}_{th} \) be the estimators of the heterogeneity parameter under \( M_h \) and \( M_{th} \), then \( \hat{\tau}_h \geq \hat{\tau}_{th} \) provided that \( \psi(k) \) is strictly convex.

Proposition 5 shows that removing the time effect in \( M_{th} \) decreases its heterogeneity parameter. Thus an unrecorded time effect is a possible explanation for a negative value of the heterogeneity parameter when fitting \( M_h \). Model \( M_h \) accommodates, in an indirect way, time differences in the capture probabilities. One may therefore expect that estimators of abundance obtained with \( M_{th} \) and \( M_h \) should be relatively close, even in the presence of a time effect.

### 3.3 Models with a behavioural effect

Under models \( M_h \) and \( M_{th} \), the capture probability changes after the first capture. This section investigates how leaving aside such a behavioural effect influences the abundance estimator obtained under \( M_0 \). The impact of an heterogeneity in the capture probabilities on the estimator obtained under \( M_0 \) is also investigated.

At a given capture occasion, the capture probabilities for \( M_h \) are \( p \) if the animal has never been caught, and \( r \) otherwise. The sufficient statistics for \( N \) are \( \{ u_i : i = 1, \ldots, t \} \), where \( u_i \) is the number of units first captured on occasion \( i \). Under this model the time of the first capture has a geometric distribution. The predicted value \( \mu_i = E(u_i) \) satisfy \( \log \mu_i = \alpha + (i - 1)\beta \) where \( \alpha = \log(Np) \) and \( \beta = \log(1 - p) \). How does a behavioural effect affects the abundance estimator obtained under \( M_0 \)? Note that the expected numbers of animals caught and missed are the same, \( N(1 - (1 - p)^t) \) and \( N(1 - p)^t \) respectively, for the two models. When \( r \neq p \), the estimators of \( p \) and \( N \) obtained under \( M_0 \) are biased. If \( r > p \), \( M_0 \) underestimates \( N \) while the opposite is true if \( r < p \). This is proved by considering the estimating equation for \( p \). One can conjecture that fitting \( M_h \), see (1), when the true model is \( M_{th} \) produces a negative bias if the first capture increases the probability of a subsequent capture.

Under \( M_{th} \), the probability \( p \) of first capture varies between animals. Thus the time of the first capture has a geometric mixture distribution for this model. From Lindsay (1986), \( E(u_k) = \exp \psi(k) \) where \( \psi(k) \) is a convex function. Solving the Poisson estimating functions for \( M_h \),

\[
n = \sum_{i=1}^{t} \hat{\mu}_i \quad \text{and} \quad \sum_{i=1}^{t} iu_i = \sum_{i=1}^{t} i\hat{\mu}_i.
\]

with \( u_i = \exp \psi(i) \), \( i = 1, \ldots, t \), yields parameter values \( \alpha_\psi \) and \( \beta_\psi \). Consider the sequences \( a_k = \exp\{\alpha_\psi + \beta_\psi(k - 1)\} \) and \( b_k = \exp \psi(k) \) for \( k = 1, \ldots, t \), \( m \) for any \( m > t \). The assumptions of Proposition 2 are met so that \( a_m \leq b_m \) for any \( m > t \). Thus the number of animals not caught \( \sum_{k=t+1}^{\infty} E(u_k) \) is larger than its predicted value under \( M_h \), \( \sum_{k=t+1}^{\infty} \exp\{\alpha_\psi + \beta_\psi(k - 1)\} \). In other words, fitting \( M_h \) underestimates the abundance when the data is generated according to \( M_{th} \).

Cormack (1989) showed that \( \mu_\omega \) has a loglinear form under \( M_h \); Baillargeon & Rivest (2007) suggest a loglinear model for \( M_{th} \) where the first occasion has its own probability of first capture, possibly different from that of the other capture occasions. For these models (1) does not hold; the intercept is not the logpredicted frequency for the animals that were missed in the experiment.

### 3.4 An example using the snowshoe hare data set of Cormack (1989)


Exploratory Heterogeneity Graphs

Figure 1: Heterogeneity plots for the hare data.

To get a numerical illustration of the results of this section, consider the snowshoe hare data set of Cormack (1989), see also Baillargeon & Rivest (2007). This data set has \( t = 6 \) capture occasions and \( n = 68 \). Sections 3.2 and 3.3 have highlighted that \( \{ (f_k, u_k) : i = 1, \ldots, t \} \) are useful descriptive statistics to assist the selection of a capture-recapture model. Plots of \( \{ (k, \log f_k/(t_k)) : i = 1, \ldots, t \} \) and of \( \{ (k, \log u_k) : i = 1, \ldots, t \} \) can help to detect an heterogeneity in capture probabilities and a time effect. Figure 1 presents the plots of \( (k, \log f_k/(t_k)) \) and \( (k, \log u_k) \) produced by the package \texttt{Rcapture} of Baillargeon & Rivest (2007) for the snowshoe hare data set. The first one suggests an heterogeneity in capture probabilities; \( M_h \) and \( M_{th} \) should be considered for this data set.

The deviance difference between \( M_0 \) and \( M_t \) is 10.2 on 6 degrees of freedom (\( p \text{-value} = .08 \)). According to the inequality of Hardy, Littlewood & Pólya (1952) presented in Section 2, this time effect should make the graph of \( (k, \log f_k/(t_k)) \) concave. This is however hidden by the heterogeneity effect since its shape is convex. The method of Mao & Lindsay (2003) could be used to check whether this effect is significant. One has \( \hat{N}_0 = 75.4, \ s.e. = 3.5 \) and \( \hat{N}_t = 75.1, \ s.e. = 3.4 \). Thus \( \hat{N}_0 > \hat{N}_t \), in agreement with Proposition 4. Even if \( M_t \) gives a better fit, \( \hat{N}_0 \approx \hat{N}_t \) showing that \( \hat{N}_0 \) is robust to a time effect.

Adding an heterogeneity component to either \( M_0 \) or \( M_t \) yields a significant deviance reduction as was expected considering the first plot of Figure 1. Taking \( \psi(k) = \exp(2k) - 1 \) provides a good fit. It yields \( \hat{N}_{th} = 81.1, \ s.e. = 5.6 \) while \( \hat{N}_h = 81.5, \ s.e. = 5.7 \). The two estimates are similar even if the deviance difference between the two models is 10.97 for 5 degrees of freedom (\( p \text{-value} = .052 \)). The estimates of the heterogeneity parameters are \( \hat{\tau}_{th} = 0.0876 \) and \( \hat{\tau}_h = 0.0834 \); they are ordered in agreement with Proposition 5.
Table 1: Comparison of the estimators \( \hat{N}_{th} \) and \( \hat{N}_h \) when the true model is \( M_{th} \) with \( \psi(k) = k^2/2 \).

<table>
<thead>
<tr>
<th>( t )</th>
<th>( N )</th>
<th>( \mu )</th>
<th>( B(\hat{N}_{th}) )</th>
<th>( B(\hat{N}_h) )</th>
<th>( RMSE(\hat{N}_{th}) )</th>
<th>( RMSE(\hat{N}_h) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>100</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>35</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>3</td>
<td>22</td>
<td>-6</td>
<td>75</td>
<td>41</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>1</td>
<td>0</td>
<td>-3</td>
<td>57</td>
<td>56</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>2</td>
<td>7</td>
<td>-6</td>
<td>62</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>500</td>
<td>3</td>
<td>13</td>
<td>-73</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>2</td>
<td>0</td>
<td>-1</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>500</td>
<td>1</td>
<td>-2</td>
<td>-3</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>500</td>
<td>2</td>
<td>0</td>
<td>-1</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>10</td>
<td>500</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>44</td>
<td>44</td>
</tr>
</tbody>
</table>

The loglinear models considered in this paper are flexible modelling tools. Discrete covariates, such as a sex or weight, can be incorporated by stratification. A loglinear model can be set up in each stratum; likelihood ratio tests of homogeneity of the parameters are easily carried out. Fienberg (1972) fitted loglinear models to mark recapture data by adding interactions between capture occasion. The \( t \)-variable interaction term was not estimable and needed to be set to 0.

The loglinear models considered in this work use a different parametrization for the \( 2^t-1 \) estimable parameters that allows a non null \( t \)-variable interaction. For instance under (1), this \( t \)-variable loglinear interaction is \( \tau \sum_{k=0}^{t} (-1)^k \binom{t}{t-k} \psi(k) \). It is equal to \( \tau \{ \exp(-\mu) - 1 \}^t \) for the Poisson mixture model.

4. A MONTE CARLO COMPARISON OF ESTIMATORS OF ABUNDANCE OBTAINED UNDER \( M_h \) AND \( M_{th} \)

Proposition 5 suggests that \( M_h \) can accommodate a time effect through a simple change of its heterogeneity parameter. This was investigated in a Monte Carlo study where the observed frequencies were simulated using a multinomial distribution whose predicted values \( \mu \) satisfy (1), with \( \psi(k) = k^2/2 \). The population size \( N \) was estimated by \( \hat{N}_h \) and \( \hat{N}_{th} \), in (2), obtained by fitting (1) respectively with and without the constraint \( \beta_1 = \ldots = \beta_t \). Thus models \( M_h \) and \( M_{th} \) were considered in the simulations. They used \( t = 5 \) and \( t = 10 \) capture occasions. Heterogeneity was introduced by setting the variance of the number of captures for one animal equal to \( \mu \sum p_i (1-p_i) \), where \( p_i \) is the marginal probability of capture at occasion \( i \) and \( \mu \), the over-dispersion parameter, is either 1, 2, or 3; \( \mu = 1 \) gives model \( M_t \). For \( t = 5 \), the simulations use \( (p_1, \ldots, p_5) = (0.1, 0.1, 0.2, 0.3, 0.3) \), while for \( t = 10 \), \( (p_1, \ldots, p_{10}) = (.05, .05, .05, .05, .1, .1, .1, .2, .2, .2) \). The proportion of animals not caught ranged between 30% when \( \mu = 1 \) to more than 50% when \( \mu = 3 \). The loglinear parameters for a particular combination of \( (p_1, \ldots, p_t) \) and \( \mu \) were evaluated numerically.

In Table 1, \( \hat{N}_h \) has a smaller root mean squared error than \( \hat{N}_{th} \), especially when the parameter for over-dispersion is small. In these simulations \( M_{th} \) fits better than \( M_h \) since it accounts for the time effect in the data. Still \( \hat{N}_h \) is in general more precise than \( \hat{N}_{th} \). This is so because model \( M_h \) approximates \( E(f_k) \) well even when \( \{f_k\} \) is obtained under \( M_{th} \). These simulations
strengthen Proposition 5; lowering the heterogeneity parameter of $M_h$ accounts for the time effect of $M_{th}$. Proceeding as in McCullagh (1994) one can show that, as $t$ goes to infinity, the number of captures for an individual, properly normalized, has the same limiting distribution under the two models, $M_h$ and $M_{th}$. Thus when $t$ is large, $M_h$ and $M_{th}$ lead to the same distribution for $\{E(f_k)/N : k = 0, \ldots, t\}$. Both models should then yield similar estimators for $N$; this is exemplified in the simulations reported in Table 1.

The larger root mean squared errors for $\hat{N}_{th}$ in Table 1 agrees with formula (6) of Rivest and Lévesque (2001). This formula shows that if two nested models give the same estimator for $\mu$, the largest model leads to the estimator with the largest variance. Observe also that fitting $M_h$ is an inefficient way of modelling a time effect when there is no heterogeneity. For example, for the first line of Table 1 ($t = 5$, $N = 100$, $\mu = 1$), simulations show that $\hat{N}_t$ has a bias of 1 and a RMSE of 11; it is much more accurate than $\hat{N}_{th}$. It would be interesting to investigate whether the robustness property for $\hat{N}_{th}$ highlighted in Proposition 5 and Table 1 holds true for other specifications of $M_h$ and $M_{th}$, such as the finite mixture model considered in Pledger (2000).

5.DISCUSSION

This paper has shown that loglinear estimators of the size of a closed population accounting for an heterogeneity in the capture probabilities are robust to the presence of a time effect. The estimator of abundance for $M_0$ is not sensitive to small deviations in the capture probabilities between capture occasions. Thus leaving aside a significant time effect has, in many instances, a negligible impact on the outcome of an analysis.

When there is no time effect $\{f_k : k = 1, \ldots, t\}$, the frequencies of the number of animals caught $k$ times, are sufficient for $M_h$. Loglinear models for $M_h$ can be fitted through a Poisson regression using $\{f_k : k = 1, \ldots, t\}$ as the dependent vector provided that the offset $\log \left( \binom{t}{k} \right)$ is included in the regression. This is important in large experiments featuring several capture occasions. For instance Rivest & Daigle (2004) use a Poisson regression to analyze a robust design with six closed population models having three capture occasions each; the number of observable capture histories is $2^{18} - 1 = 262,143$. If the analysis is restricted to $M_h$ for the six closed population models, then the capture histories can be coded in terms of $(f^{(1)}, \ldots, f^{(6)})$ where $f^{(j)} = 0, 1, 2, 3$ is the number of captures in the $j$th experiment. This reduces the size of the dependent vector in the Poisson regression, to $4^6 - 1 = 4,095$. Such manipulations of capture histories can be carried out with Rcapture, see Baillargeon & Rivest (2007).

APPENDIX

Proof of Proposition 2. Let $X$ and $Y$ be discrete random variables defined on $\{1, \ldots, t\}$ such that

$$P(X = k) = \frac{a_k}{\sum_{\ell=1}^t a_\ell} \quad \text{and} \quad P(Y = k) = \frac{b_k}{\sum_{\ell=1}^t b_\ell}.$$

Since $P(Y = k)$ is log-convex with respect to $P(X = k)$ and $E(X) = E(Y)$, by Proposition 1, $P(Y = k) - P(X = k)$ has two sign changes in $\{1, \ldots, t\}$. Thus, if the first one occurs at $k \geq 1$,

$$b_1 \geq a_1 \quad \text{and} \quad b_k \geq a_k, b_{k+1} \leq a_{k+1},$$

for some $k > 1$. The second set of inequalities implies that $b_{k+1}a_k/(a_{k+1}b_k) \leq 1$. The logconvexity assumption implies that $b_{k+1}a_\ell/(a_{k+1}b_\ell)$ is increasing. Thus $b_1a_0/(a_1b_0) \leq 1$ or $b_0 \geq b_1a_0/a_1 \geq a_0$.

ACKNOWLEDGEMENTS

The financial support of the Natural Science and Engineering Research Council of Canada and of the Fonds québécois de la recherche sur la nature et les technologies are gratefully acknowledged.
REFERENCES


Received October 30 2006
Accepted April 16 2007

Louis-Paul RIVEST: lpr@mat.ulaval.ca
Département de mathématiques et de statistique
Université Laval
Québec, Québec
Canada, G1K 7P4