An extravariation model for improving confidence intervals of population size estimates from removal data

You-Gan Wang and Neil R. Loneragan

Abstract: We propose a new model for estimating the size of a population from successive catches taken during a removal experiment. The data from these experiments often have excessive variation, known as overdispersion, as compared with that predicted by the multinomial model. The new model allows catchability to vary randomly among samplings, which accounts for overdispersion. When the catchability is assumed to have a beta distribution, the likelihood function, which is referred to as beta-multinomial, is derived, and hence the maximum likelihood estimates can be evaluated. Simulations show that in the presence of extravariation in the data, the confidence intervals have been substantially underestimated in previous models (Leslie–DeLury, Moran) and that the new model provides more reliable confidence intervals. The performance of these methods was also demonstrated using two real data sets: one with overdispersion, from smallmouth bass (Micropterus dolomieu), and the other without overdispersion, from rat (Rattus rattus).

Résumé: Nous proposons un nouveau modèle pour estimer une population à partir de prises successives dans le cadre d’une expérience d’enlèvement. Les données générées par ces expériences sont souvent entachées de variations très marquées, décrites par le terme de surdispersion, par rapport aux données obtenues à partir du modèle multinomial. Le nouveau modèle permet de laisser varier au hasard la capturabilité d’un prélèvement à l’autre, ce qui règle le problème de la surdispersion. Lorsqu’on prend comme hypothèse que la capturabilité suit la distribution bêta, la fonction de probabilité du type bêta-multinomiale, est dérivée et les évaluations du maximum de vraisemblance peuvent être calculées. On montre, par simulation, que lorsqu’il y avait trop de variation dans les données, les intervalles de confiance ont été substantiellement sous-estimés dans les modèles précédents (Lesley–DeLury, Moran), et que le nouveau modèle en fournit qui sont plus fiables. L’efficacité de cette méthode a été démontrée aussi au moyen de deux ensembles de données réelles : l’un avec surdispersion, dans le cas de l’achigan à petite bouche (Micropterus dolomieu), et l’autre sans surdispersion, dans le cas du rat (Rattus rattus).

[Traduit par la Rédaction]

Introduction

Removal experiments are often used in biological research to estimate the abundance of animals in a population (Seber 1973; Ricker 1975). In these types of experiments, an area is sampled repeatedly over a relatively short period of time and the population size is estimated from the change in catch rates. The data generated from a removal experiment consist of a sequence of catch numbers and the corresponding efforts. The population is assumed to be closed in the sense that no emigration, immigration, or natural mortality occurs during the experiment, and all catches are removed from the population (Otis et al. 1978; Schnute 1983).

Typically, the catches from a removal experiment can be expressed by a sequence \( C_1, C_2, \ldots, C_k \), where \( C_i \) is the \( i \)th catch and \( k \) is the number of samples taken. For the convenience of further discussion, we list below the notation used in this paper:

- \( N \): initial population size before sampling
- \( k \): number of samples
- \( q_i \): capture efficiency at time \( i \)
- \( p_i \): capture proportion of the \( i \)th sampling
- \( n_i \): number of animals caught from the \( i \)th sampling
- \( T_i \): cumulative capture up to (including) time \( i \)
- \( N_i \): population size after \( i \)th sampling
- \( E(C_i | N_{i-1}) \): expected value of \( C_i \)

We are interested in estimating the parameter \( N \), and possibly the parameter \( q \), as well. If we assume that the capture efficiencies are independent of the population size:

\[
E(C_i | N_{i-1}) = q_iN_{i-1}, \quad N_i = N_{i-1} - C_i, \quad i = 1, 2, \ldots, k.
\]

In this study, we assume, for simplicity, that effort is constant during each sampling interval, although the methods discussed here can easily be extended to allow for variable effort if the relationship between catchability and effort is specified.

Many methods already exist to analyse data from removal
experiments. The performance of these methods has been tested in many ways (e.g., Seber 1973; Ricker 1975; Otis et al. 1978; Schnute 1983; Crittenden 1983). Two fundamental methods for analysing removal data are the regression method, which will be referred to as the Leslie–DeLury (L–D) method (Leslie and Davis 1939; DeLury 1947; Ricker 1975), and the multinomial likelihood (ML) method (Moran 1951; Schnute 1983; Loneragan et al. 1995).

A major complication in the estimation arises when the variation in the data exceeds the nominal level of variation from the multinomial models, a phenomenon known as overdispersion. A simple way to address this situation is to assume that the catchabilities (capture efficiencies) \(q_1, q_2, \ldots, q_k\) are different parameters (Otis et al. 1978). However, this will create an oversaturated model (with \(k + 1\) parameters and \(k\) data points) and makes estimation impossible. In general, too many parameters in the model often result in unreliable estimates of \(N\) with very wide confidence intervals (cf. Otis et al. 1978). To reduce the number of parameters, Schnute (1983) suggested a parametric function to govern the \(q_i\)'s.

However, in some cases, catchability cannot be specified in a deterministic way. This paper suggests a way of modelling extrapolation that provides an alternative to the model of deterministic catchability outlined by Schnute (1983). Our “extravariation” model assumes that catchabilities vary randomly over occasions, and it improves the estimates of the confidence intervals when there is overdispersion in the data.

The model’s performance was compared with that of two commonly used models: the L–D model and the multinomial model of constant catchability. Real data from smallmouth bass (Micropterus dolomieu) (Ricker 1975) and rat (Rattus rattus) (Schnute 1983) were also used to test the performance of the various models.

**Two existing fundamental models**

**L–D model**

If the proportion of the population removed during each sampling period is assumed to be have the same expectation, i.e., \(q_i = q\) for all \(i\), we have

\[
E(C_i | T_{i-1}) = q(N - T_{i-1})
\]

(2)  

Here, \(T_0\) is defined as 0. Parameter estimates can be obtained by regression of the catch against the cumulative catch in the form of \(Y = a + bX\) with \(a = qN\) and \(b = -q\) (Ricker 1975, p. 151). Calculation of confidence intervals is given by DeLury (1951) and Ricker (1975).

A weighted regression may result in a better estimate of \(N\), and \(1/N_i\) can be used as the weighting function where a preliminary estimate of \(N\) is obtained by eye (Zippin 1956). We suggest using the unweighted regression analysis to obtain a preliminary estimate of \(N\) first for the weighting function. The formulae for calculating confidence limits using weighted regression are given in Appendix A.

**Multinomial model**

If we assume that each catch is sampled from a binomial distribution with a probability parameter \(q_i\) and a population parameter \(N - T_{i-1}\), the catch vector \(C\) will have a multinomial distribution with parameters \((N, p_1, p_2, \ldots, p_k)\), where

\[
p_i = q_i \sum_{j=1}^{i-1} (1 - q_j).
\]

The probability function of \(C\) under this general multinomial model can be written as (2.4) in Schnute (1983):

\[
p(C|q_1, q_2, \ldots, q_k) = \frac{N!}{(N - T_k)!} \prod_{i=1}^{k} \frac{q_i^C_i (1 - q_i)^{N - T_i}}{C!}.
\]

(3)  

A special and important case, considered by Moran (1951), is when the \(q_i\)'s are assumed to be a constant, \(q\), and the log-likelihood becomes

\[
l_{MULT}(N, q) = \log \frac{N!}{(N - T_k)!} + T_k \log(q)
\]

\[
+ \sum_i (N - T_i) \log(1 - q) - d
\]

where \(d\) is a constant \(\sum_i \log(C_i)\), independent of the parameters. Maximum likelihood estimates can be obtained from the estimating Eq. (C.2) (see Appendix C). For a given \(N\), let \(l_{MULT}(N, q)\) be the maximum value of \(l_{MULT}(N, q)\) evaluated at the optimal \(q\). The function \(l_{MULT}(N)\) is known as a profile likelihood function of \(N\) (McCullagh and Nelder 1989, p. 254). A confidence interval for \(N\) can be obtained based on the fact that \(-2 \log(l(N)/l(\hat{N}))\) is approximately \(\chi^2(1)\) distributed (Schnute 1983; Bedrick 1994).

**Extravariation model**

To account for the extramultinomial variation in the data, we assume that the capture efficiencies \(q_i\) are random variables, independent and identically distributed.

**Likelihood function**

The joint likelihood of \(C\) can be written as

\[
p(C) = p(C_1)p(C_2|C_1) \cdots p(C_k|C_1, C_2, \ldots, C_{k-1}).
\]

(5)  

If we assume that the distribution of \(q_i, f(q)\), is beta with parameters \((\alpha, \beta)\), this allows considerable flexibility (Johnson and Kotz 1970):

\[
f(q) = \frac{q^{\alpha-1}(1 - q)^{\beta-1}}{B(\alpha, \beta)}
\]

where \(B(\alpha, \beta)\) is the beta function \(\int_0^1 x^{\alpha-1}(1 - x)^{\beta-1}dx\). Therefore:

\[
p(C_i|C_1, C_2, \ldots, C_{i-1}) = \int p(C_i|q, C_1, C_2, \ldots, C_{i-1})f(q)dq
\]

\[
= \left(\frac{N - T_{i-1}}{C_i}\right) \int q^{C_i}(1 - q)^{N - T_i} \frac{q^{\alpha-1}(1 - q)^{\beta-1}}{B(\alpha, \beta)} dq
\]

\[
= \left(\frac{N - T_{i-1}}{C_i}\right) B(\alpha_2, \beta_2) B(\alpha, \beta)
\]

© 1996 NRC Canada
where \( \alpha_i = \alpha + C_i \) and \( \beta_i = \beta + N - T_i \). From Eqs. (5) and (7) we obtain the log-likelihood function of \( C \) as
\[
I_{\text{NEW}}(N, \alpha, \beta) = \log \frac{N!}{(N - T_k)!} + \sum_{i=1}^{k} \log \left( \frac{B(\alpha_i, \beta_i)}{B(\alpha, \beta)} \right) - d.
\]

In the case of fixed catchability, i.e., \( q_i = q \) for all \( i \), or when \( \alpha + \beta \to \infty \) and \( \alpha/(\alpha + \beta) = q \), the likelihood function (Eq. (8)) becomes that of the multinomial model given by Eq. (4).

If \( I_{\text{NEW}}(N) \) is the profile likelihood function (Eq. (8)) evaluated at optimum values of \( (\alpha, \beta) \), the approximate 95% confidence limits can be obtained by solving the equation 
\[-2 \log(I(N)/I(\hat{N})) = 3.84.\]

**Variance components**

Under the general multinomial model (Eq. (3)), the variance of \( C_i \) is \( Np_i(1 - p_i) \). Therefore, for Moran’s multinomial model, the variance of \( C_i \) is
\[
V_1(C_i) = N\gamma_i(1 - \gamma_i)
\]
in which \( \gamma_i = (1 - q)^{i-1} q \).

In the extrapolation model, \( q_i \) is assumed to be a random variable with a mean \( q \) and a variance \( \sigma^2 \). Recall that \( p_i = \frac{q_i}{\sum_{j=1}^{i} (1 - q_j)} \). The variance of \( C_i \) is
\[
V_2(C_i) = E\{\text{var}(C_i|p_i)\} + \text{var}(E(C_i|p_i))
\]
\[
= V_1(C_i) + N(N - 1)\text{var}(p_i)
\]
in which
\[
\text{var}(p_i) = E(p_i^2) - [E(p_i)]^2
\]
\[
= \{(1 - q)^2 + \sigma^2\}^{i-1}(q^2 + \sigma^2) - \gamma_i^2.
\]

If \( q_i \) is assumed to have a beta distribution with parameters \( (\alpha, \beta) \), we have \( q = \alpha/\alpha + \beta \) and \( \sigma^2 = q(1-q)/(\alpha + \beta + 1) \).

Note that \( V_2(C_i) \) consists of two components: the variance by the multinomial model \( V_1(C_i) \) and the extrapolation.

**Simulation studies**

To investigate the performance of the different models, catch data were generated from Monte Carlo simulations. The population \( N \) was assumed to be 1000 and three different values of \( q \) were chosen: 0.1, 0.2, and 0.5. Three sweep numbers \( (k = 10, 5, \text{and} \ 3) \) were used. The values of \( k \) were chosen to be smaller for larger values of \( q \) as is the case in typical removal experiments. Estimates of \( N \) were obtained by three methods: the weighted L-D method, the ML method based on Moran’s (1951) multinomial model, and the new method based on the extrapolation model.

If we rewrite \( \text{var}(C_i) \) as \( (1 + \tau)Nq(1 - q) \), where \( \tau = (N - 1)\sigma^2/(q(1-q)) \), \( \tau \) can be used as an index of overdispersion. The values of \( \tau \) were chosen to be 0.0, 0.5, and 1.0. The case of \( \tau = 0 \) corresponds to Moran’s multinomial model.

The catchability (capture efficiency) was sampled each time from a beta distribution with parameters calculated from the values of \( (q, \tau) \). A catch was then obtained by sampling from the remaining population with the catchability obtained from the beta distribution.

The estimates, together with corresponding 95% confidence intervals, were evaluated by the three methods for each simulated catch sequence of 1000 simulations. To assess the reliability of the confidence intervals, we calculated 95% coverage probability (the proportion of the number of simulations in which the 95% confidence interval covers the true value of \( N = 1000 \)).

The performance of the three methods was quite similar for the point estimate of \( N \) (Table 1). Although the L-D method performs well when \( q \) is small (i.e., 0.1, 0.2 in our simulations), the confidence intervals are substantially underestimated when \( q = 0.5 \). It is interesting to note that the L-D method does not seem to be affected by the extent of overdispersion (\( \tau \)) in the data, while Moran’s method does, as one might expect (Table 1). The 95% coverage probabilities also clearly show that, in the case of overdispersion, the confidence limits’ of Moran’s model are too optimistic and can be substantially underestimated (Table 1). The standard er-

<table>
<thead>
<tr>
<th>( q )</th>
<th>( k )</th>
<th>( \tau )</th>
<th>var(( q )) \times 1000</th>
<th>Model</th>
<th>( \hat{N} )</th>
<th>(SE)</th>
<th>( p(N) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>10</td>
<td>0.0</td>
<td>0.00</td>
<td>L-D</td>
<td>1007.0</td>
<td>(87.1)</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>1005.0</td>
<td>(86.4)</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>1004.7</td>
<td>(86.7)</td>
<td>0.952</td>
</tr>
<tr>
<td>0.1</td>
<td>10</td>
<td>0.5</td>
<td>0.045</td>
<td>L-D</td>
<td>1006.4</td>
<td>(94.7)</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>1005.8</td>
<td>(94.9)</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>1003.7</td>
<td>(93.9)</td>
<td>0.942</td>
</tr>
<tr>
<td>0.1</td>
<td>10</td>
<td>1.0</td>
<td>0.090</td>
<td>L-D</td>
<td>1019.3</td>
<td>(117.2)</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>1019.6</td>
<td>(117.6)</td>
<td>0.891</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>1016.6</td>
<td>(116.3)</td>
<td>0.934</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>0.0</td>
<td>0.000</td>
<td>L-D</td>
<td>1007.9</td>
<td>(77.8)</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>1005.0</td>
<td>(77.0)</td>
<td>0.953</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>1004.3</td>
<td>(77.5)</td>
<td>0.965</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>0.5</td>
<td>0.080</td>
<td>L-D</td>
<td>1008.9</td>
<td>(89.5)</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>1006.9</td>
<td>(88.9)</td>
<td>0.919</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>1005.5</td>
<td>(88.5)</td>
<td>0.944</td>
</tr>
<tr>
<td>0.2</td>
<td>5</td>
<td>1.0</td>
<td>0.160</td>
<td>L-D</td>
<td>1015.2</td>
<td>(102.6)</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>1014.1</td>
<td>(102.0)</td>
<td>0.872</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>1011.7</td>
<td>(101.6)</td>
<td>0.925</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>0.0</td>
<td>0.000</td>
<td>L-D</td>
<td>999.3</td>
<td>(22.6)</td>
<td>0.840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>997.9</td>
<td>(22.4)</td>
<td>0.948</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>997.7</td>
<td>(22.6)</td>
<td>0.958</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>0.5</td>
<td>0.125</td>
<td>L-D</td>
<td>1001.1</td>
<td>(25.5)</td>
<td>0.809</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>999.8</td>
<td>(25.3)</td>
<td>0.921</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>999.6</td>
<td>(25.3)</td>
<td>0.937</td>
</tr>
<tr>
<td>0.5</td>
<td>3</td>
<td>1.0</td>
<td>0.250</td>
<td>L-D</td>
<td>1003.3</td>
<td>(28.8)</td>
<td>0.757</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Moran</td>
<td>1003.6</td>
<td>(28.6)</td>
<td>0.887</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New</td>
<td>1001.4</td>
<td>(28.5)</td>
<td>0.922</td>
</tr>
</tbody>
</table>

© 1996 NRC Canada
Fig. 1. Observed and predicted catches of smallmouth bass. The Pearson $\chi^2$ test shows overdispersion in the catch data.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\hat{N}$</th>
<th>(95% CI)</th>
<th>$q$</th>
<th>DR</th>
<th>df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-D (Ricker)</td>
<td>1078</td>
<td>(814, 2507)</td>
<td>0.1068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-D (weighted)</td>
<td>1112</td>
<td>(852, 2016)</td>
<td>0.0966</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moran</td>
<td>1124</td>
<td>(979, 1357)</td>
<td>0.0958</td>
<td>0</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Otis</td>
<td>$\infty$</td>
<td>(1018, $\infty$)</td>
<td>0.0000</td>
<td>2.15</td>
<td>1</td>
<td>0.142</td>
</tr>
<tr>
<td>New</td>
<td>1134</td>
<td>(917, 1785)</td>
<td>0.0935</td>
<td>9.25</td>
<td>1</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Fig. 2. Scaled deviance functions of three different models. The horizontal lines indicate 95% confidence intervals.

Errors produced by the three methods are similar because they were calculated from the 1000 point estimates. In practice, it is unlikely that the same experiment could be repeated many times, and therefore, one may rely on the likelihood ratio statistics to evaluate the uncertainty of the estimate and its confidence intervals. The performance of the new method was reliable in all aspects.

Applications

Smallmouth bass data

Ricker (1975, p.151) described a removal experiment which estimated the population size of smallmouth bass in Little Silver Lake, Lanark County, Ontario. The successive catch numbers were 131, 69, 99, 78, 56, 76, 49, 42, 63, and 47, making a total catch of 710. Using the unweighted L–D regression method, Ricker estimated $N$ as 1078, with a 95% confidence interval of (814, 2507).

To find out whether there is an overdispersion problem in the data, we calculated the Pearson $\chi^2$ statistic using the estimates from Moran’s model of constant catchability. It was 29.1 with 8 degrees of freedom, corresponding to a $p$-value <0.001 (Fig. 1). This strongly indicates that Moran’s multinomial model is inappropriate and that the variation in the data is significantly in excess of that predicted by the multinomial model (Fig. 1). Therefore, the extravariation model was used. Although the estimates of $N$ by these methods differ slightly, there were substantial differences in the widths of the confidence intervals (Table 2). The variances of $C_i$ were evaluated under the multinomial model and the new model. The ratios of the two variances for the 10 catches were found to be (3.83, 3.58, 3.35, 3.14, 2.95, 2.78, 2.63, 2.49 2.37, 2.25). Therefore the inflation factor for the first catch $\tau$ is 2.83, i.e., 283% higher than the nominal variance.

We further analysed the data using the model of Otis et al. (1978), which requires the same number of parameters as the extravariation model (see Appendices B and C). Unfortunately, it gave the estimate of $N$ as $\infty$. The deviance functions $\sim$ log (profile likelihood) of $N$ from different models were evaluated (Fig. 2). The corresponding 95% confidence intervals were obtained using likelihood ratio statistics (Schnute 1983) (Table 2).

The reduction of the deviance by the extravariation model with one extra degree of freedom is 9.25 ($p < 0.002$), a substantial improvement over the other methods. On the other hand, there is little reduction by the Otis model (Table 2).

Rat data

Schnute (1983) reanalysed the rat data in Leslie and Davis (1939). The 18 successive catches are $C = (49, 32, 31, 34, 16, 33, 22, 27, 17, 19, 18, 16, 18, 12, 14, 12, 17, 7)$. Schnute (1983) obtained the maximum likelihood estimate of $N$ as 522.3 using Moran’s multinomial model (see also Moran 1951). The corresponding Pearson $\chi^2$ value is 17.8 with 16 degrees of freedom ($p = 0.34$). These results suggest that the variation in the data is within the range predicted by
the multinomial model with constant capture efficiency and therefore there is no overdispersion problem. Using the new extraveration model, we obtained the same estimate of $N$ as using Moran’s model. Based on the profile log-likelihood function of the catch, the 95% confidence interval was found to be (468, 607).

The programs are written in GAUSS and are available from the first author.

**Discussion**

The catches at successive times during a removal experiment are always correlated because the catch at time $i$ depends on how many were being caught before time $i$. As pointed out by Schnute (1983), the ML method allows for this correlation between the catches.

Otis et al. (1978) also proposed a series of models to account for the individual heterogeneity of catchability, which appears to have great similarity to the model proposed here. However, their models are actually different versions of the multinominal models, and cannot be used to explain extraveration in the catches (Appendix B). Their models were also found to be unreliable (Appendix B) and did not provide a sensible estimate of $N$ for the smallmouth bass data.

Carle and Strub (1978) also proposed a method to make Moran’s ML more robust. However, their model assumes constant capture probability at different times, and is therefore inappropriate for overdispersed data.

Schnute (1983) realized that the assumption of constant capture efficiency over time might not be valid in practice, and proposed that a steadily decreasing or increasing function for the capture efficiency at time $i$ be

$$q_i = q_1 + (q - q_1)(1 - d^{i-1})$$

where $q_1$ is the initial capture efficiency, $q$ is the limit of the efficiency as $i$ becomes large, and $a$ is a parameter measuring the change rate from $q_1$ to $q$. He referred to this as Model 3, and it included his Model 1 (the same as Moran’s model of constant capture efficiency: $a = 1$) and Model 2 (constant capture efficiency except the initial one: $a = 0$) as special cases. He further evaluated the likelihood function based on the binominal model of capture for estimating the parameters. In Schnute’s model, the capture efficiency is determined by a parametric function. In our model, it is governed by a stochastic law which has a parametric distribution.

When we are interested in estimating the catch efficiency from multiple experiments, we can analyse the pooled data by multiplying the likelihoods together. But it is not obvious for the L–D regression method, which is based on the catch versus the cumulative catch. Crittenden (1983) examined the effect of variability in $q$ and found that the variance of the catch was a quadratic function of the expected catch. He therefore used a quadratic weighting function in the regression and found that his method was more efficient than the conventional weighting one, but still showing some positive bias in its estimate of the initial population. This may be due to the violation of the assumptions in the regression. As pointed out by Pollock (1991, p. 232), this approach is not recommended because the likelihood can be very right-skewed, and the $\chi^2$ approximation for the confidence intervals can be unsatisfactory, necessitating an adjustment.

Overdispersion is a phenomenon with many causes, each of which should be modelled in its own way to obtain the most appropriate analysis. We have developed a model that allows for the catchability to vary randomly; the model is natural and easy to interpret and it is a counterpart of the deterministic model of Schnute (1983). If there is evidence that the catchability changes systematically, a deterministic model may be more appropriate. A more realistic model is to assume that catchability is random and that its expectation varies among sweeps. However, parameter estimation for this model has to rely on multiple removal experiments.

In practice, the Pearson $\chi^2$ test can be used first to see if the multinominal model can adequately describe the data, and if extraveration is found, the beta-multinomial model may be used. It is possible that in some cases the catchability may change in a deterministic way and Schnute’s model would be more appropriate.

The likelihood ratio statistic would then be useful to determine the most appropriate model for the estimation (Schnute 1983). However, likelihood statistics are available only for likelihood-based methods, which makes the regression methods less attractive for the analysis of removal data. On the other hand, it may be difficult to justify the distribution used, although the distribution we used for $q_i$ (beta distribution) is quite flexible.

Existing methodology for overdispersion is to inflate the variance by a dispersion parameter and to apply a quasi-likelihood approach that requires specification of a mean–variance relationship (Wang 1996). However, such a relationship cannot be uniquely specified when both the population parameter ($N$) and probabilities ($q$’s) are unknown. Therefore, it is of great interest to develop distribution-free methods to account for overdispersion in removal data.

Not much attention has been paid to the effect of overdispersion in fisheries research, although the effect of overdispersion in general is well known in statistical literature (e.g., Cox 1983; Roberts et al. 1987). In estimating the recapture rate of marked fish, Bayley (1993) found that the standard errors of the parameters were substantially lower when overdispersion is ignored. Our simulation results are consistent with his findings, albeit in a different context.

Risk evaluation of fisheries management strategies is based on the uncertainties associated with parameter estimates. Incorrect confidence intervals will result in inappropriate estimates of the risk associated with management decisions. The proposed extraveration model will therefore prove useful because of its ability to account for overdispersion in the removal data when estimating a population size.

**Acknowledgments**

We are grateful to Dr. Mervyn R. Thomas for his support and encouragement on this work and Drs. David Die and Yongshun Xiao for helpful comments on an early version of the manuscript. We also thank Dr. Nick Ellis for his assistance in some programming for this work.

**References**


**Appendix A: Confidence interval of \( \tilde{N} \) for the weighted L–D method**

For the convenience of readers, we present general expressions of the L–D estimates and corresponding standard errors when a weighting function is used. The estimation can be written in the regression form of \( C = X\beta + \epsilon \), where \( C \) is the vector of catch \((k \times 1)\), \( X \) is a \((k \times 2)\) matrix with the \( i \)th row being \((1, T_{i-1})\), and \( \beta \) is a vector of two parameters \( qN \) and \( -q \). The estimate of \( \beta \) is \( \hat{\beta} = (X'X)^{-1}X'Y \) under the assumption that the error vector \( \epsilon \) is independently distributed with mean 0 and a common variance \( \sigma^2 \).

DeLury (1951) obtained the confidence limits based on the approximation of \( \beta \)-distribution to \( b'\hat{\beta}/\hat{\sigma} \), where \( b = (1, N) \) and \( \hat{\sigma} \) is the standard error of the numerator (see also Ricker 1975). If the regression is weighted by a diagonal matrix \( W \):

\[
\hat{\beta} = (X'WX)^{-1}X'WC \quad \text{and} \quad \hat{\sigma}^2 = b'(X'WX)^{-1}b \hat{\sigma}^2
\]

where \( \hat{\sigma}^2 = \{C'W(C - X\hat{\beta})\}/(k - 2) \). Assuming \( \hat{\beta} \) can be approximated by a binormal distribution, the \((1 - \alpha)\) confidence limits can be expressed as the two roots of the quadratic function

\[
(A.2) \quad b'\hat{\beta}/\hat{\sigma} = z^2 t_{\alpha, k-2} + t_{\alpha, k-2}
\]

where \( t_{\alpha, k-2} \) is the critical value of a \( t \)-distribution with \( k - 2 \) degrees of freedom at significance level \((1 - \alpha)\). If \( W \) is the identity matrix, the expressions derived here become identical to those given in Ricker (1975, p. 150).

**Appendix B. Approach of Otis et al. (1978)**

The above multinomial model assumes that each animal has the same constant catchability. To account for the heterogeneity of individuals, one may assume that the individual catchability is a random variable with a distribution function \( \pi(q) \) (Otis et al. 1978). In this case, the capture efficiency at time \( i \) becomes the conditional expectation given that the animal has not been captured before:

\[
(B.1) \quad q_i = \int_0^1 x(1 - x)^{i-1}\pi(x)dx = \frac{\int_0^1 x^{i-1}\pi(x)dx}{\int_0^1 (1 - x)^{i-1}\pi(x)dx}
\]

From a Bayesian point of view, \( q_i \) is a posterior mean when the prior distribution is \( \pi(x) \).

In the model of Otis et al. (1978), \( C \) has a multinomial distribution with marginal probabilities \( \lambda_i = E[q(1 - q)^{i-1}] \) \((1 \leq i \leq k)\). The variance of \( C_i \) is thus \( N\lambda_i(1 - \lambda_i) \), which does not differ from \( V_i(C_i) \) substantially. If \( \pi(x) \) is taken to be beta(a,b), we have

\[
(B.2) \quad \lambda_i = ab(b + 1) \ldots (b + i - 2)/(a + b)(a + b + 1)(a + b + 2) \ldots (a + b + i - 1)
\]

To make use of the ML method, one needs to specify the function form of \( \pi(\cdot) \) to obtain the likelihood of the observed catches. If \( \pi(x) \) is assumed to be a beta distribution with parameters \((a,b)\), we have \( q_i = a/(a + b + i - 1) \), and from Eq. (3) the likelihood function of the catch \( C \) is

\[
(B.3) \quad l_{\text{OTIS}} = \log \frac{N!}{(N - T_i)!} + T_k \log(a) + \sum_{i=1}^{k}(N - T_i)\log(b + i - 1) - \sum_{i=1}^{k}(N - T_{i-1})\log(a + b + i - 1) - \frac{1}{2}d
\]

**Appendix C. Estimating functions of Moran’s and Otis’ models**

If \( l(N, q) \) is the log-likelihood function of the catch \( C \) which is assumed to follow some distribution, the ML estimates of \((N, q)\) satisfy the following equations:

\[
(C.1) \begin{cases} 
I(N, q) - I(N - 1, q) = 0 \\
\frac{\partial l}{\partial q} = 0.
\end{cases}
\]

© 1996 NRC Canada
From Eq. (B.3), we obtain the estimating equations for Moran’s model:

\[
\begin{align*}
N &= \frac{T_k}{1 - (1 - q)^k} \\
q &= \frac{k}{kN - \sum_{i=1}^{k-1} T_i}
\end{align*}
\] (C.2)

From Eq. (B.3), we obtain the estimating equations for Otis’ model:

\[
\begin{align*}
N &= \frac{T_k}{1 - \prod_{i=1}^{k} \frac{b+i-1}{a+b+i-1}} \\
a &= \frac{T_k}{\sum_{i=1}^{k} \frac{N - T_{i-1}}{a+b+i-1}} \\
= \sum_{i=1}^{k} \frac{N - T_i}{b+i-1}
\end{align*}
\] (C.3)