BIOMETRICS 37, 371–375
June 1981

Maximum Likelihood Applied to a Capture–Recapture Model

D. A. Sprott
Department of Statistics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

SUMMARY

Maximum likelihood is applied to a capture–recapture model in order to estimate the population size, n. The likelihoods to which this model gives rise can be highly skew. It is shown that the use of $n^{-\frac{1}{3}}$ gives an approximately normal likelihood. This results in a function that has an approximately standard normal distribution and is linear in $n^{-\frac{1}{3}}$. Thence complete sets of approximate confidence intervals can be obtained merely by the use of a pocket calculator.

1. Introduction

Darroch and Ratcliff (1980) discussed the estimation of the population size, n, in an interesting capture–recapture model. The model leads to the probability function

$$\operatorname{pr}(r) = \frac{1}{n^s}\binom{n}{r}\Delta^r 0^s, \tag{1}$$

where r, the observed number of individuals caught at least once, is a sufficient statistic for n, the population size, and s is the number of samples, each of size one. Darroch and Ratcliff (1980) compared a new estimator of n with the maximum likelihood estimator, $\hat{n}$, in terms of asymptotic bias and variance. They illustrated their results using two experiments, A and B, which yielded $s = 72$, $r = 69$, giving $\hat{n} = 828$, and $s = 435$, $r = 341$, giving $\hat{n} = 854$, respectively.

In general, however, variances are not very useful for finite samples if the purpose is to make informative statements about n, for instance to specify complete sets of confidence intervals or present tests of significance. In particular, the variance can be completely misleading if the likelihood function of n is skew, since it gives no indication of the skewness. For example, if $r = s$ then $\hat{n} = \infty$, and the likelihood increases from a minimum at $n = r$ to its maximum at $n = \infty$. In Experiment A, for $n = 335$, which is a left deviation of 493 from the maximum likelihood estimate of $\hat{n} = 828$, the relative likelihood is 15%. To obtain such a small relative likelihood to the right of $\hat{n}$ requires n to be as large as 3000, which is a right deviation of 2172 from $\hat{n}$, indicating extreme skewness to the right. The corresponding results for Experiment B are $n = 726$ and 1020, with corresponding deviations 128 and 166, which indicate less skewness.

It is the purpose of this communication to show that (1) provides an example of the method of Sprott (1973, 1975, 1980) in which maximum likelihood can be applied to give confidence intervals by allowing for the possible asymmetry of the likelihood. Theoretical details can be found in the papers by the author cited above. Sprott's approach leads to

Key words: Maximum likelihood; Likelihood function; Capture–recapture; Transformations; Normal likelihood.

the use of (5), below. It should be emphasized that the purpose here is not to compare the efficiency or precision of estimators, as was done by Darroch and Ratcliff (1980), but rather, to show how the full asymptotic efficiency of the maximum likelihood estimator may be utilized on specific finite samples.

2. Use of Maximum Likelihood

Instead of thinking of maximum likelihood in terms of point estimates and their variances, it is more fruitful to think of it as a method of generating quantities u that have an approximate standard normal distribution and that are approximately linear in the parameter. One such candidate is $u=(\hat{n}-n)\{I(\hat{n})\}^{\frac{1}{2}}$, where $I(\hat{n})=-(\partial^2 \log p/\partial \hat{n}^2)$ is the observed Fisher information. This function u is linear in n, but is approximately a standard normal variate only if the likelihood functions of n produced by (1) are approximately normal. As exemplified by Experiment A, this may not be the case in terms of n. However, a transformation $\phi=\phi(n)$ may improve matters. In fact, it will be shown for the model (1) that $\phi=n^{-\frac{1}{3}}$ is often helpful in this respect. This suggests the desirability of using $u=(\hat{n}^{-\frac{1}{3}}-n^{-\frac{1}{3}})\{I(\hat{n}^{-\frac{1}{3}})\}^{\frac{1}{2}}$, which is linear in $n^{-\frac{1}{3}}$, as a standard normal variate to set up confidence intervals for n, leading to (5) below.

Using Stirling's approximation, we find the log likelihood from (1) is approximately

$$\log L(n)=(n-s+\tfrac{1}{2})\log n-(n-r+\tfrac{1}{2})\log(n-r).$$

The maximum likelihood estimate $\hat{n}$ satisfies $\hat{n}-r=\hat{n}/\hat{y}$, where $\hat{y}=(1-1/\hat{n})^{-s}$. Assuming the discontinuity in n can be ignored, we differentiate $\log L(n)$ twice with respect to n and set $\hat{n}-r=\hat{n}/y$; this yields

$$I(\hat{n})=\{2\hat{n}(\hat{y}-1)-\hat{y}^2-2s+1\}/2\hat{n}^2.$$

Differentiating $\log L(n)$ three and four times with respect to n gives the skewness and kurtosis, F_3 and F_4, of the likelihood, as measured by the third and fourth standardized derivatives of $\log L$, respectively. Thence

$$F_3(\hat{n})=\frac{\partial^3 \log L(n)}{\partial \hat{n}^3} I^{-\frac{3}{2}}(\hat{n})=\frac{\{\hat{n}(\hat{y}^2-1)-\hat{y}^3-2s+1\}2^{\frac{3}{2}}}{\{2\hat{n}(\hat{y}-1)-\hat{y}^2-2s+1\}^{\frac{3}{2}}}, \tag{2}$$

and $F_4=(\partial^4 \log L/\partial \hat{n}^4)I^{-2}(\hat{n})$, which can similarly be calculated. The quantity $F_3(n)$ is $2^{\frac{5}{2}}n^{\frac{1}{2}}/\{s(s-1)\}^{\frac{1}{2}}+O(n^{-\frac{1}{2}})$, which is approximately $5.66n^{\frac{1}{2}}/s$. The quantities F_3 and F_4 are the principal deviations from a normal likelihood; the use of a parameter $\phi=\phi(n)$ that reduces these deviations will enhance the normal approximation to the resulting likelihood $L(\phi)$, and also enhance the normal approximation to the distribution of the resulting quantity $u=(\hat{\phi}-\phi)\{I(\hat{\phi})\}^{\frac{1}{2}}$.

For the model (1), use of $\phi=n^{-\frac{1}{3}}$ often substantially reduces F_3 of (2) and F_4. For $\phi=n^{-\frac{1}{3}}$,

$$I(\hat{\phi})=I(\hat{n}^{-\frac{1}{3}})=(9/2)\hat{n}^{\frac{2}{3}}\{2\hat{n}(\hat{y}-1)-\hat{y}^2-2s+1\} \tag{3}$$

and

$$F_3(\hat{\phi})=F_3(\hat{n}^{-\frac{1}{3}})=-2^{\frac{3}{2}}\frac{\{\hat{n}(\hat{y}^2-4\hat{y}+3)+2s-1-\hat{y}^3+2\hat{y}^2\}}{\{2\hat{n}(\hat{y}-1)-\hat{y}^2-2s+1\}^{\frac{3}{2}}} \tag{4}$$

which is $2^{\frac{3}{2}}(2s+\frac{1}{2})/\{9\hat{n}s(s-1)\}^{\frac{1}{2}}+O(\hat{n}^{-\frac{3}{2}})$ or $1.89\ \hat{n}^{-\frac{1}{2}}$, approximately. These expressions can easily be evaluated. For Experiment A, (2) is 2.31, while (4) is -0.7. Also $F_4(\hat{n})$ is -6.02 while $F_4(\hat{n}^{-\frac{1}{3}})$ is $-.11$. For Experiment B, (2) is .42, while (4) is -0.07. The quantities $F_4(\hat{n})$ and $F_4(\hat{n}^{-\frac{1}{3}})$ are $-.20$ and $-.02$, respectively. Thus, use of $n^{-\frac{1}{3}}$ commends itself here also, even though the original likelihood $L(n)$ for Experiment B is much more symmetric, as described in §1. These results could be more fully illustrated graphically.

The foregoing suggests the use of $u=(\hat{n}^{-\frac{1}{3}}-n^{-\frac{1}{3}})\{I(\hat{n}^{-\frac{1}{3}})\}^{\frac{1}{2}}$ as a standard normal variate, as proposed earlier. Use of (3) gives

$$u=3\{1-(\hat{n}/n)^{\frac{1}{3}}\}\{\hat{n}(\hat{y}-1)-\tfrac{1}{2}\hat{y}^2-s+\tfrac{1}{2}\}^{\frac{1}{2}} \tag{5a}$$

as an approximate standard normal variate, where $\hat{y}=(1-1/\hat{n})^{-s}$. This is linear in $n^{-\frac{1}{3}}$ so the data can be summarized approximately by

$$n=\hat{n}\{1-u/A(\hat{n})\}^{-3}, \tag{5b}$$

where

$$A(\hat{n})=3\{\hat{n}(\hat{y}-1)-\tfrac{1}{2}\hat{y}^2-s+\tfrac{1}{2}\}^{\frac{1}{2}} \tag{5c}$$

and u is a standard normal variate.

Application of (5) to Experiment A gives $n=828(1-.187u)^{-3}$; taking $u=\pm 1.96$ gives an approximate 95% confidence interval of $n=(324,\ 3259)$. Experiment B gives $n=854(1-.02904u)^{-3}$; $u=\pm 1.96$ gives the approximate 95% confidence interval $n=(723,\ 1018)$. It can easily be verified, as in §1, that these intervals take into account the skewness of the likelihood.

It should be noted that only the calculation of $\hat{n}$ requires iteration; given $\hat{n}$, then $A(\hat{n})$ from (5c) gives the complete set of approximate confidence intervals in (5b).

3. Discussion

Darroch (1958) has also given a method of obtaining approximate confidence intervals. For (1), this requires the assumption that r is approximately normally distributed about $n\{1-\exp(-s/n)\}$ with variance $n\{\exp(-2s/n)\}\{\exp(s/n)-1-s/n\}$. An iteration is then necessary to find each endpoint of each confidence interval. Also, it is doubtful if the normality assumption for r will lead to more accurate confidence intervals than does (5). This was tested on the small artificial example $s=25$, $r=23$, giving $\hat{n}=142.3$, for which tables of Stirling's numbers are available to calculate (1). From (5), an approximate 95% confidence interval is $n=(48,\ 802)$. From (1), $\mathrm{pr}\,(r\leqslant 23)=.0505$ and $\mathrm{pr}\,(r\geqslant 23)=.0361$ for $n=802$ and 48, respectively, giving a confidence level of 92.34%. Using the above assumed normal distribution of r gives $n=(51,\ 542)$, which does not fully reflect the extreme asymmetry of the likelihood. The exact probabilities are .0998 and .0515 for $n=542$ and 51, respectively, and the corresponding confidence coefficient is 84.87%.

This example, and also Experiment A, are affected by extreme discontinuity since r is close to s. A continuity correction improves the accuracy. This entails replacing r by $r-\frac{1}{2}$ for a lower confidence bound n_L, and by $r+\frac{1}{2}$ for an upper confidence bound n_u. The expression (5b) is then replaced by

$$\begin{aligned} n_u&=\hat{n}_u\{1-u/A(\hat{n}_u)\}^{-3}, \qquad u>0,\\ n_L&=\hat{n}_L\{1-u/A(\hat{n}_L)\}^{-3}, \qquad u<0, \end{aligned} \tag{5d}$$

where $\hat{n}_u$ and $\hat{n}_L$ are the maximum likelihood estimators using $r+\frac{1}{2}$ and $r-\frac{1}{2}$, respectively. Use of (5d) in the above example gives an approximate 95% confidence interval of (42, 1656) which, by (1), has an exact confidence level of 96.8%. For Experiment A, the above continuity correction yields the approximate 95% confidence interval (295, 4702). Note that the upper confidence bound, n_u, is unstable in the sense that, because of extreme skewness, slight changes in the likelihood, or confidence coefficient, produce large changes in n_u. This is because, when r is close to s, (3) is small so the experiment is imprecise and large sets of n_u values have essentially the same likelihood. For Experiment B, on the other hand, r is not so close to s and the continuity correction does not have much affect, yielding an approximate 95% interval (720, 1025).

Thus, (5) appears to give a reasonably accurate and efficient summary of the data. Furthermore, all of the calculations, including the iteration required to obtain $\hat{n}$, are easily performed on a pocket calculator.

4. Extension

The foregoing can be extended to the case of general sample sizes $a_1, a_2, \ldots, a_s$. The conditional distribution of the observations, given the a_i obtained by Darroch (1958), leads to the likelihood function

$$\frac{n!}{(n-r)!}\prod_{i=1}^{s}\frac{(n-a_i)!}{n!}.$$

Using the same procedure as in §2, we obtain $I(n)$, then $F_3(n)$:

$$F_3(n) = 4\left\{n\Big/\sum_{1<j} a_i a_j\right\}^{\frac{1}{2}} + O(n^{-\frac{1}{2}}).$$

Use of $\phi = n^{-\frac{1}{3}}$ reduces F_3 to order $n^{-\frac{1}{2}}$. This helps particularly if $F_3(n)$ is large and the likelihood of n is highly asymmetric. In other cases use of ϕ may not make much difference, and other transformations such as $1/n$, or even $1/n^2$, may be somewhat better. The use of $1/n$ was briefly discussed by Otis *et al.*, (1978, pp. 84–87) who concluded it was of no help. The above results indicate, however, that the use of a transformation depends upon the circumstances and requires a fuller discussion. In fact, if a transformation of the form $\phi = n^{\lambda}$ is assumed, a tentative value of λ can be obtained, using (5) of Sprott (1973), as

$$\hat{\lambda} = 1 - \tfrac{1}{3}\hat{n}F_3(\hat{n})I^{\frac{1}{2}}(\hat{n}).$$

In all cases it is advisable to plot or tabulate the likelihood function of n to examine for skewness, and to compare it with the likelihood function of any proposed $\phi(n)$.

ACKNOWLEDGEMENTS

I should like to thank Dr A. N. Arnason for helpful discussion and for references to additional papers. This work was partially supported by a Natural Science and Engineering Research Council grant.

RÉSUMÉ

Le maximum de vraisemblance est appliqué au modèle de capture–recapture pour estimer la taille, n, d'une population. Pour ce modèle, les vraisemblances obtenues peuvent être fortement biaisées. On montre que l'emploi de $n^{-\frac{1}{3}}$ donne une vraisemblance approximativement normale. On aboutit à une fonction qui suit approximativement une distribution normale et qui est linéaire par rapport à $n^{-\frac{1}{3}}$. Ceci permet de donner des ensembles complets d'intervalles de confiances approchés qui pour être déterminés ne requièrent que l'utilisation d'une calculatrice de poche.

References

Darroch, J. N. (1958). The multiple-recapture census. I. Estimation of a closed population. *Biometrika* **45,** 343–359.

Darroch, J. N. and Ratcliff, D. (1980). A note on capture–recapture estimation. *Biometrics* **36,** 149–153.

Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978). *Statistical Interference from Capture Data on Closed Animal Populations.* Wildlife Monographs 62. Washington, D.C.: The Wildlife Society.

Sprott, D. A. (1973). Normal likelihoods and their relation to large sample theory of estimation. *Biometrika* **60,** 457–465.

Sprott, D. A. (1975). Application of maximum likelihood methods to finite samples. *Sankhyā, Series B* **37,** 259–270.

Sprott, D. A. (1980). Maximum likelihood and small samples: Estimation in the presence of nuisance parameters. *Biometrika* **67,** 515–523.

Received June 1980; *revised September* 1980