RECEIVED BY TIC AUG 11 1972

BNWL-SA- 43 52

CONF-720643--1

Partitioning of the Bias of the Jolly-Seber
Capture-Recapture Estimate of Population
Size When Unequal Catchability is Present

By

Richard O. Gilbert

MASTER

Presented at the Western Regional
Meeting of the WNAR, IMS and ASA
University of Washington
Seattle, Washington
June 19-21, 1972

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

Battelle Memorial Institute
Pacific Northwest Laboratories
Richland, Washington 99352

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

peg

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Introduction

We are concerned here with partitioning the bias in estimates of population size obtained from capture-recapture field studies conducted on wildlife populations. Results have been obtained for the Jolly-Seber estimates $\hat{N}_i$ and $\hat{N}_i^*$ (defined below) of the population size N (Jolly (1965), Seber (1965)) in closed populations for which all animals are not equally catchable at a given time, i.e., heterogeneity is present. This partitioning of the bias of $\hat{N}_i$ and $\hat{N}_i^*$ is an extension of the results of Gilbert (1970) who showed that the bias of these estimators could in most cases be closely approximated using a Taylor series approximation. The results of Gilbert (1970) and those in the present paper are obtained under the assumption that all animals captured at time i are subsequently released before time $i + 1$.

Carothers (1972) points out that $\hat{N}_i$ (and hence $\hat{N}_i^*$) is "subject to various sources of bias which can be categorized either as arising from failures of one or more of the underlying assumptions of the model or as intrinsic to the method of estimation and present even when all the assumptions of the model are fulfilled". This latter case he calls "small sample bias". He considers large populations of indeterminate size to approximate the relative bias (bias divided by population size) which arises when the assumption of equal catchability is violated. He terms this discrepancy (which in fact measures the lack of robustness of $\hat{N}_i$ in the presence of heterogeneity) the asymptotic relative bias of $\hat{N}_i$. In the present paper, the total relative biases (TRB) of $\hat{N}_i$ and $\hat{N}_i^*$ obtained using Taylor series for the distributions in Table 1 are partitioned into relative small sample bias (RSSB) (that bias which can arise when R_i or m_i are small) and the asymptotic relative bias (ARB) arising from the failure of all animals

to be equally catchable. This partitioning helps clarify the relationship between population size (over which the experimenter has no control), the distribution of probability of capture (which the experimenter can alter by changing his experimental design and sampling intensity), and the total relative biases of $\hat{N}_i$ and $\hat{N}_i^*$ resulting from the combined effect of these two factors.

If all of the assumptions underlying the Jolly-Seber model are fulfilled then the only bias present is "small sample bias" which goes to zero as the probability of capture of all animals approaches 1. If heterogeneity is present the discrepancy arising from this unequal catchability is in addition to any small sample bias which may be present and it does not disappear even though the number of animals captured may be very large.

Under the assumption that all animals captured at time i are subsequently released before time $i + 1$, the Jolly-Seber estimates $\hat{N}_i$ and $\hat{N}_i^*$ of the true population size N are given by

$$\hat{N}_i = \frac{n_i^2 Z_i}{R_i m_i} + n_i \quad (1)$$

and

$$\hat{N}_i^* = \frac{(n_i + 1)^2 Z_i}{(R_i + 1)(m_i + 1)} + n_i, \quad (2)$$

($i = 2, 3, \dots, L-1$), respectively, where n_i = number of animals captured on the i th trap day, m_i = number out of n_i that were previously captured, R_i = number out of n_i which are subsequently recaptured at any time i' for $i < i' \leq L$, Z_i = number captured before, after, but not at time i , and L = number of trap days in the trapping experiment. $\hat{N}_i^* + 1$ was proposed by Seber (1965) as an appropriate modification of $\hat{N}_i$ when R_i and/or m_i are zero.

Taylor Series Expansions of $\text{TRB}(\hat{N}_i)$ and $\text{TRB}(\hat{N}_i^*)$

If the total population of N animals is divided into J distinct groups whose sizes are N_j , ($j = 1, 2, \dots, J$), where $N = \sum_{j=1}^J N_j$, and P_j is the probability of capture of each animal in group j on each of the L trap days, then Gilbert (1970) showed that when heterogeneity is present, i.e. $P_j \neq P_{j'}$ for some $j \neq j'$, the Taylor series approximation (expanding through second partial derivatives) to $\text{TRB}(\hat{N}_i^*)$ is given by

$$\begin{aligned} \text{TRB}(\hat{N}_i^*) \approx & \frac{1}{N} \left[\frac{(n_i + 1)^2 Z_i}{(R_i + 1)(m_i + 1)} + n_i - N \right] \\ & + \frac{1}{N} \left[\frac{(n_i + 1)^2 Z_i}{(R_i + 1)(m_i + 1)} \right] \left[\frac{1}{(R_i + 1)^2} \sum_{j=1}^J R_{ij} (N_j - R_{ij})/N_j \right. \\ & \left. + \frac{1}{(m_i + 1)^2} \sum_{j=1}^J m_{ij} (N_j - m_{ij})/N_j \right] \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{(\bar{n}_i + 1)^2} \sum_{j=1}^J \bar{n}_{ij} (N_j - \bar{n}_{ij})/N_j \\
& - \frac{2}{(\bar{n}_i + 1)\bar{z}_i} \sum_{j=1}^J \bar{n}_{ij} \bar{z}_{ij}/N_j \\
& - \frac{2}{(\bar{R}_i + 1)(\bar{n}_i + 1)} \sum_{j=1}^J \bar{R}_{ij} (N_j - \bar{n}_{ij})/N_j \\
& - \frac{2}{(\bar{n}_i + 1)(\bar{m}_i + 1)} \sum_{j=1}^J \bar{m}_{ij} (N_j - \bar{n}_{ij})/N_j \\
& + \frac{1}{(\bar{R}_i + 1)\bar{z}_i} \sum_{j=1}^J \bar{z}_{ij} \bar{R}_{ij}/N_j \\
& + \frac{1}{(\bar{m}_i + 1)\bar{z}_i} \sum_{j=1}^J \bar{z}_{ij} \bar{m}_{ij}/N_j \\
& + \frac{1}{(\bar{R}_i + 1)(\bar{m}_i + 1)} \sum_{j=1}^J \bar{m}_{ij} \bar{R}_{ij} (N_j - \bar{n}_{ij})/\bar{n}_{ij} N_j \Big] \quad (3)
\end{aligned}$$

for $i = 2, 3, \dots, L-1$, where bold faced letters indicate expected values and

$$n_i = \sum_{j=1}^J N_j P_j, m_i = \sum_{j=1}^J N_j P_j P_{1ij}$$

$$z_i = \sum_{j=1}^J N_j P_{1ij} (1-P_j) P_{2ij}, R_i = \sum_{j=1}^J N_j P_j P_{2ij}$$

$$P_{1ij} = 1 - (1-P_j)^{i-1} \text{ and } P_{2ij} = 1 - (1-P_j)^{L-i}.$$

P_{1ij} and P_{2ij} are the probabilities of capture at least once before and at least once after time i , respectively, in group j . The approximation to $TRB(\hat{N}_i)$ is given by equation (3) after substituting n_i, m_i and R_i for $n_i + 1, m_i + 1$ and $R_i + 1$, respectively.

All relative biases obtained here are multiplied by 100 to put them on a percent basis. For comparative purposes we also compute the percent mean total relative bias (MTRB) over all times i , defined e.g. for $\hat{N}^*$ as

$$MTRB(\hat{N}^*) = (L-2)^{-1} \sum_{i=2}^{L-1} 100 TRB(\hat{N}_i^*).$$

The percent mean RSSB and ARB over all trap days i are also computed and are denoted by MRSSB and MARB, respectively.

Partitioning of the Total Relative Bias

When heterogeneity is present in a capture-recapture experiment for which sample sizes are small, then $TRB(\hat{N}_i)$ is a function of both heterogeneity and small sample size. If $J > 1$ so that heterogeneity is present, we compute $ARB(\hat{N}_i)$, i.e. that lack of robustness of $\hat{N}_i$ due to heterogeneity alone, by using the definition

$$ARB(\hat{N}_i) = \lim_{N \rightarrow \infty} TRB(\hat{N}_i) \quad (4)$$

If we take the limit of equation (3) as $N \rightarrow \infty$ after substituting n_i, m_i and R_i for $n_i + 1, m_i + 1$ and $R_i + 1$, respectively, we find that

$$ARB(\hat{N}_i) = \frac{(\sum_{j=1}^J N_j P_j)^2 \sum_{j=1}^J N_j P_{1ij} (1-P_j) P_{2ij}}{N \sum_{j=1}^J N_j P_j P_{2ij} \sum_{j=1}^J N_j P_j P_{1ij}} + \frac{\sum_{j=1}^J N_j P_j}{N} - 1 \quad (5)$$

which is a special case of the more general form of $ARB(\hat{N}_i)$ obtained by Carothers (1972) for open populations. Equation (5) is simply the first term in the Taylor series approximation to $TRB(\hat{N}_i)$ the numerical value of which does not depend on N since the proportions $N_1/N, N_2/N, \dots, N_J/N$ remain constant as $N \rightarrow \infty$. The remaining terms in $TRB(\hat{N}_i)$ go to zero as $N \rightarrow \infty$ (Carothers (1972)) and may be interpreted as the small sample bias of $\hat{N}_i$ which is present when N is not large. Thus we may write $TRB(\hat{N}_i)$ as

$$TRB(\hat{N}_i) \simeq ARB(\hat{N}_i) + RSSB(\hat{N}_i) \quad (6)$$

where $ARB(\hat{N}_i)$ is given by equation (5) and $RSSB(\hat{N}_i)$ is approximated by the 2nd order terms of the Taylor series expansion.

Whereas the small sample bias of $\hat{N}_i$ is entirely a function of 2nd and higher order terms of the Taylor series expansion, this is not the case for $\hat{N}_i^*$ since the first term of $TRB(\hat{N}_i^*)$ (equation (3)) is not independent of N . However, if we divide both numerator and denominator of $(n_i + 1)^2 Z_i / (R_i + 1)(m_i + 1)$ in the first term of equation (3) by N^2 and take the limit of the resulting equation as $N \rightarrow \infty$, we find that

$$ARB(\hat{N}_i^*) = \lim_{N \rightarrow \infty} TRB(\hat{N}_i^*) = ARB(\hat{N}_i) \quad (7)$$

which is given by equation (5). Now, since $n_i > R_i$ and $n_i > m_i$ which implies that $(n_i + 1)^2 / (R_i + 1)(m_i + 1) < n_i^2 / R_i m_i$, it follows that

$$\frac{1}{N} \left[\frac{(n_i + 1)^2 Z_i}{(R_i + 1)(m_i + 1)} + n_i - N \right] - ARB(\hat{N}_i^*) < 0, \quad (8)$$

i.e., $[(n_i + 1)^2 Z_i / (R_i + 1)(m_i + 1) + n_i - N] / N$ is a larger negative fraction than $ARB(\hat{N}_i^*)$. Moreover, in light of equation (7), this difference approaches zero as $N \rightarrow \infty$. Hence, equation (8) is the contribution from the first term of equation (3) to the relative small sample bias of $\hat{N}_i^*$ for given N .

We may compute $RSSB(\hat{N}_i^*)$ for given N , L and $(N_j, P_j), (j = 1, 2, \dots, J)$, as

$$RSSB(\hat{N}_i^*) \cong TRB(\hat{N}_i^*) - ARB(\hat{N}_i^*) \quad (9)$$

where $TRB(\hat{N}_i^*)$ is given by equation (3) and $ARB(\hat{N}_i^*)$ by equation (5).

A more precise approximation to $RSSB(\hat{N}_i^*)$ could be obtained by including higher order terms in the Taylor series expansion, but simulation results obtained to check on the accuracy of equation (3) when $N \geq 50$ indicate the 2nd order terms give a sufficiently accurate approximation for all practical purposes.

Results

The primary results of Gilbert (1970) were

- (i) Serious negative bias can result in both $\hat{N}_i$ and $\hat{N}_i^*$ for distributions of probability of capture clustered primarily below $P = .5$.
- (ii) The bias of $\hat{N}_i$ and $\hat{N}_i^*$ is symmetric with respect to trap day with the least bias occurring on the mid trap day.
- (iii) Substantial reduction in the total biases of $\hat{N}_i$ and $\hat{N}_i^*$ can be obtained by increasing the number of trap days L , particularly if the probability of capture of most animals is low.

By partitioning the total bias of $\hat{N}_i$ and $\hat{N}_i^*$ using equations (6) and (9), respectively, we obtain new insight into the robustness of these estimates when heterogeneity is present. In Table 2 for $N = 50$, $L = 10, 25$ and 50 , and the distributions in Table I we have partitioned $MTRB(\hat{N}^*)$ into its two component parts $MRSSB$ and $MARB$ which measure the average bias over all trap days due to small sample size and the lack of robustness arising from heterogeneity, respectively. We see that for these values of N and L the $MRSSB(\hat{N}^*)$ is a very small proportion of the mean total relative bias $MTRB(\hat{N}^*)$ for all distributions in Table I except VII and VIII. We also note that

$MARB(\hat{N}^*)$ is negative for all distributions considered (see Marten (1970) for a discussion of this point).

In Table 3 we have examined the change in $MRSSB(\hat{N}^*)$ as N decreases from 50 to 10 for trapping experiments of length $L = 5$. These small sample results are relevant to many trapping studies today since the cost of continuing to trap animals beyond 5 or 10 days may be prohibitively expensive. We see that for the small population case of $N = 10$ when $L = 5$ the contribution to $MTRB(\hat{N}^*)$ from $MRSSB(\hat{N}^*)$ becomes important for distributions V, IX and XIII in addition to VII and VIII mentioned above. The remaining distributions continue to have small values of $MRSSB(\hat{N}^*)$ even for these very small population sizes. We note, however, that the values of $MRSSB(\hat{N}^*)$ given in Table 3 for $N < 50$ are somewhat tentative since their accuracy has not been substantiated by computer simulations.

Insight into the types of distributions which result in large asymptotic biases versus those yielding large small sample biases may be gained by an examination of the partitioning of the total bias of $\hat{N}_i^*$ under distribution I with that under a distribution identical to I (denoted by I(a)) except that the two subgroups have probabilities of capture $P_1 = .05$, $P_2 = .07$ rather than $P_1 = .05$ and $P_2 = .95$. For distribution I, $MRSSB(\hat{N}^*)$ comprises 0.35 percent of $MTRB(\hat{N}^*)$, while for distribution I(a) $MRSSB(\hat{N}^*)$ contributes 95.12 percent of the total ($L = 25$ in both cases). The values of $MTRB(\hat{N}^*)$ for I and I(a) are -62.18 and -42.39 percent, respectively. Thus, while the small sample bias of $\hat{N}_i^*$ is nearly nonexistent for distribution I because 20 percent of the population have a high probability of capture, the heterogeneity of I is so great as to produce a mean total bias of greater size that exists for distribution I(a) in which heterogeneity is nearly nonexistent, but small sample bias is of considerable size due to

few animals being captured.

Results for $\hat{N}_i$ (not shown) indicate that $\text{MTRB}(\hat{N})$ is nearly identical to $\text{MTRB}(\hat{N}^*)$ for all distributions in Table I except VII, VIII and XIII which are characterized by low probabilities of capture for most animals. In contrast to $\hat{N}_i^*$, the $\text{RSSB}(\hat{N}_i)$ was found to be positive for nearly all distributions which is a manifestation of the infinite positive small sample bias of N_i . (The very small negative values of $\text{RSSB}(\hat{N}_i)$ obtained for some distributions ($< - .017\%$) may be due to truncation error of the Taylor series).

The partitioning of the total bias into small sample bias and the lack of robustness resulting from heterogeneity offers some guidance concerning field procedures. It is clear that the presence of heterogeneity per se is not as important as the particular pattern of heterogeneity and whether the heterogeneity is near zero or 1. If the heterogeneity is distributed above $P = .5$, then little bias due to either small sample size or unequal catchability will result. This would seem to suggest that an experimenter need not attempt to design an experiment so that all animals have the same probability of capture, but only that the distribution be shifted nearer 1 so that nearly all animals have probabilities of capture, say $\geq .50$.

It is informative to note that the large total bias obtained here for distributions VII, VIII and XIII when $N = 50$ is substantially reduced when $N = 300$ (particularly for $L = 10$ trap days). This occurs since for these distributions a relatively large proportion of the total bias is due to small sample bias which goes to zero as N becomes large. However a large N does not necessarily imply a small total bias. This is clear since the total bias in distribution I, II, III, etc., for which the bias due to heterogeneity is large, will not be appreciable less even if the population size N is very

large. Thus, if the population should happen to be very large, small sample bias will tend to be negligible even if P is near zero, but that bias due to heterogeneity could still be large resulting in a large total bias.

REFERENCES

Carothers, A. D. (1972). The Effects of Unequal Catchability on Jolly-Seber Estimates. Biometrics (issue to be notified).

Gilbert, R. O. (1970). Approximations of the Bias in Jolly's Stochastic Capture-Recapture Model. Presented at the Joint Statistical Meetings of the American Statistical Association, Detroit, Michigan, December 27-30, 1970.

Jolly, G. M. (1965). Explicit Estimates from Capture-Recapture Data with Both Death and Immigration - Stochastic Model, Biometrika 52, 225-247.

Marten, G. G. (1970). A Regression Method for Mark-Recapture Estimation of Population Size with Unequal Catchability, Ecology 51, 291-295.

Seber, G. A. F. (1965). A Note on the Multiple-Recapture Census, Biometrika 52, 249-259.

TABLE 1

DISTRIBUTIONS[†] OF PROBABILITY OF CAPTURE FOR WHICH ESTIMATES
OF % RELATIVE BIASES OF $\hat{N}_i$ AND $\hat{N}_i^*$ HAVE BEEN OBTAINED

Distribution	Probability									
	.05	.15	.25	.35	.45	.55	.65	.75	.85	.95
I	80									20
II	80							8	6	6
III	50									50
IV	50					50				
V	50	20	10	4	4	4	4	4		
VI	34				32					34
VII	50	40	10							
VIII	40	32	20	8						
IX	24	20	16	12	8	8	8	4		
X	20	12	8	6	4	4	6	8	12	20
XI	10	10	10	10	10	10	10	10	10	10
XII	6	10	18	10	6	6	10	18	10	6
XIII	16	28	24	18	10	4				
XIV		6	10	14	20	20	14	10	6	
XV		6	8	10	10	12	12	12	14	16
XVI	4	4	4	4	4	4	8	8	20	40
XVII					4	10	18	24	28	16
XVIII						10	16	20	24	30

† Numbers in table are percentages.

TABLE 2
 PARTITIONING OF $MTRB(\hat{N}^*)$ INTO $MRSSB(\hat{N}^*)$ AND
 $MARB(\hat{N}^*)$ USING TAYLOR SERIES EXPANSIONS WHEN $N = 50^{\dagger}$

Distri- bution	L = 50		L = 25		L = 10	
	MRSSB	MARB	MRSSB	MARB	MRSSB	MARB
I	- .311	-45.243	- .223	-61.959	- .066	-72.398
II	- .351	-44.440	- .253	-60.992	- .081	-71.373
III	- .040	-29.870	- .026	-39.918	- .005	-45.652
IV	- .083	-27.950	- .063	-37.190	- .075	-42.131
V	- .411	-28.326	- .563	-40.136	- .829	-51.791
VI	- .042	-20.616	- .049	-27.872	- .088	-34.058
VII	-2.972	-18.076	-5.548	-23.902	-13.216	-27.589
VIII	-1.222	-18.658	-2.226	-25.505	- 5.338	-30.830
IX	- .153	-16.491	- .247	-24.239	- .441	-33.587
X	- .042	-14.831	- .061	-21.796	- .088	-30.304
XI	- .039	- 8.771	- .067	-13.637	- .127	-21.173
XII	- .039	- 6.868	- .073	-11.441	- .134	-19.667
XIII	- .295	-11.323	- .547	-16.989	- 1.232	-23.769
XIV	- .031	- 2.138	- .062	- 4.316	- .140	- 9.277
XV	- .020	- 2.319	- .040	- 4.687	- .087	-10.204
XVI	- .009	- 3.848	- .016	- 6.097	- .032	- 9.835
XVII	- .004	- .239	- .008	- .499	- .024	- 1.420
XVIII	- .003	- .188	- .007	- .392	- .019	- 1.124

$\dagger$ Data are percents

Table 3

Partitioning $MTRB(\hat{N}^*)$ Into $MRSS(\hat{N}^*)$
And $MARB(\hat{N}^*)$ Using Taylor Series When $L = 5^+$

<u>Distribution</u>	<u>MRSSB</u>			<u>MARB</u>
	<u>N = 50</u>	<u>N = 25</u>	<u>N = 10</u>	
I	- .014	- .037	- .121	-74.72
II	- .042	- .135	- .474	-73.77
III	- .001	- .002	- .010	-46.83
IV	- .232	- .852	- 3.200	-43.01
V	- 1.198	- 3.02	- 7.38	-57.34
VI	- .141	- .283	- .703	-39.07
VII	-24.68	-37.49	-48.999	-28.58
VIII	-11.069	-21.15	-34.44	-32.54
IX	- .706	- 1.98	- 6.01	-38.84
X	- .102	- .217	- .60	-35.56
XI	- .181	- .403	- 1.18	-27.05
XII	- .169	- .390	- 1.195	-25.86
XIII	- 2.571	- 6.77	-16.88	-27.01
XIV	- .232	- .613	- 2.14	-13.53
XV	- .132	- .284	- .80	-15.55
XVI	- .046	- .095	- .25	-13.01
XVII	- .051	- .111	- .33	- 3.16
XVIII	- .043	- .090	- .25	- 2.65

† Data are percents