MARK-RECAPTURE MODELS WITH SURVIVAL RATES SPECIFIC TO CAPTURE-HISTORY CLASSES: A SYNTHESIZED NUMERICAL EXAMPLE

BU-826-M November, 1983

by

D. S. Robson

Biometrics Unit, Cornell University, Ithaca, NY 14853

Abstract

A numerical example is constructed to illustrate the notation and methodology contained in Robson's 1969 paper on "Mark-recapture methods of population estimation."

Introduction

The paper [1] entitled "Mark-recapture methods of population estimation" by D. S. Robson contains somewhat awkward notation which is illustrated there only with non-numerical examples. Here we present a numerical example constructed as a supplement to Example III of the paper. The model for this example admits that tagged and untagged individuals may have different survival rates and admits the possibility of a short term stress effect of tagging which may affect survival during the period immediately after tagging and release, and possibly during the next period, as well. Capture probability, conditional upon survival, is assumed to be unaffected by capture history (see Pollock [2] for treatment of this problem with a stress effect on both survival and catchability).

Paper No. BU-826-M in the Biometrics Unit.

Numerical Example

The $k = 5$ sample array of data in Table 1 was constructed to precisely fit the assumed model of Example III with arbitrary amounts of recruitment into the untagged segment of the population between successive sampling occasions, and with arbitrary survival probabilities and sample sizes, and arbitrary numbers released from each sample.

On any given sampling occasion the extant population at the time of release may be classified into four distinct groups with respect to rates of survival to the next sampling occasion. These classes, designated by $H_j^{(\nu)}$ at time t_j, $\nu = 1, 2, 3, 4$, are identified in Table 2 by the capture histories $h_j = (\delta_{1j}, \delta_{2j}, \cdots, \delta_{jj})$ of individuals having identical survival rates between time t_j^+ and time t_{j+1}^-, where

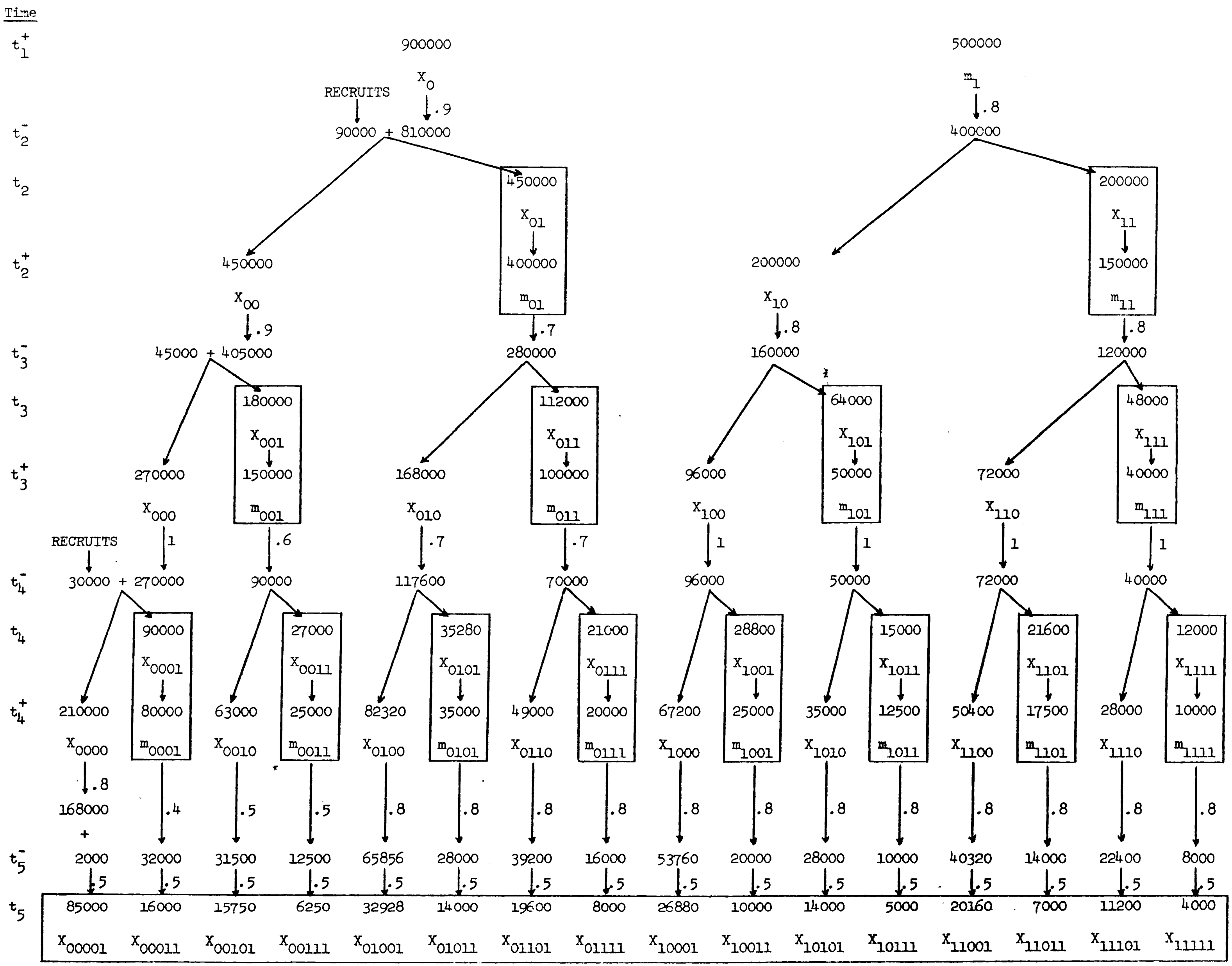
$$\delta_{ij} = \begin{cases} 1 \text{ indicates "captured at time } t_i\text{"} \\ 0 \text{ indicates "not captured at time } t_i\text{"} . \end{cases}$$

Table 2. Capture histories h_j at time t_j grouped according to survival rate from t_j to t_{j+1}

j	$H_j^{(0)}$	$H_j^{(1)}$	$H_j^{(2)}$	$H_j^{(3)}$
1	(0)	(1)	NONE	NONE
2	(00)	(01)	(11)	NONE
3	(000)	(001)	{(010), (011)}	{(100), (101), (110), (111)}
4	(0000)	(0001)	{(0010), (0011)}	{the remaining 12 histories}
5	(00000)	(00001)	{(00010), (00011)}	{the remaining 28 histories}

The numbers X_{h_j} having capture history h_j at time t_j are so indicated in Table 1, where the boxed entries represent observed counts captured (X_h) and released (m_h) while all remaining entries are unobserved and hence represent

Table 1. An array of synthetic data covering 5 samples and conforming exactly to the model of Example III.



unknown quantities to the investigator. Population counts $N_{h_{j-1}}(t_j^-)$ in each capture history category are listed in Table 1 at times t_j^- immediately prior to the j'th sampling, and the identifiable group totals of these population counts are shown boxed in Table 3.

Table 3. Group totals of population counts at times t_j^-. Boxed entries are identifiable from the boxed data in Table 1, while the remaining non-zero counts cannot be estimated.

j	$N_j^{(0)}$	$N_j^{(1)}$	$N_j^{(2)}$	$N_j^{(3)}$	Total = N_j
1	1,400,000	0	0	0	1,400,000
2	900,000	400,000	0	0	1,300,000
3	450,000	280,000	280,000	0	1,010,000
4	300,000	90,000	187,600	258,000	835,600
5	170,000	32,000	44,000	345,536	591,536

A minimal set of summary statistics from Table 1 which are sufficient for estimating the boxed entries in Table 3 are displayed in Table 4A and 4B, along with the (redundant) numbers $R_j^{(\nu)}$ later recaptured (Table 4C) among those $m_j^{(\nu)}$ (Table 4D) in group $H_j^{(\nu)}$ that are released at time t_j^+;

$$R_j^{(\nu)} = \sum_{(h_{j-1},1)\in H_j^{(\nu)}} (X_{h_{j-1}11} + X_{h_{j-1}101} + \cdots + X_{h_{j-1}10\text{-}.01})$$

for example,

$$R_3^{(3)} = (X_{1011} + X_{10101}) + (X_{1111} + X_{11101}) = (15000 + 14000) + (12000 + 11200)$$

since (10) and (11) are the h_2's satisfying $(h_2,1) \in H_3^{(3)}$, as seen in Table 2. The entries $T_j^{(\nu)}$ in Table 4A include the $R_j^{(\nu)}$ plus the number (say $Z_j^{(\nu)}$) of individuals in $H_j^{(\nu)}$ that are seen before t_j and after t_j but not at t_j,

$$Z_j^{(\nu)} = \sum_{(h_{j-1},0)\in H_j^{(\nu)}} (X_{h_{j-1}01} + X_{h_{j-1}001} + \cdots + X_{h_{j-1}0\cdots01})$$

and

$$T_j^{(\nu)} = R_j^{(\nu)} + Z_j^{(\nu)};$$

for example, in Table 4A

$$T_3^{(2)} = R_3^{(2)} + Z_3^{(2)}$$

$$= (X_{0111} + X_{01101}) + (X_{0101} + X_{01001})$$

$$= (21000 + 19600) + (35280 + 32928)$$

$$= 40,600 + 68,208 = 108,808$$

is the number of individuals in $H_3^{(2)}$ at time t_3 that are seen again after t_3. Table 4B lists the summary statistics

$$S_j^{(\nu)} = \sum_{(h_{i-1},0)\in H_j^{(\nu)}} X_{h_{j-1}1}$$

representing the numbers that are caught at t_j <u>and</u> would have entered the class $H_j^{(\nu)}$ if not caught at t_j; for example,

$$S_4^{(3)} = X_{0101} + X_{0111} + X_{1001} + X_{1011} + X_{1101} + X_{1111} = 133,680.$$

Note that these 6 capture histories are also in the class $H_4^{(3)}$; this is not the case, for example, with

$$S_4^{(0)} = X_{0001} = 90,000$$

since the history (0001) is not in $H_4^{(0)} = \{(0000)\}$. Table 4D lists the numbers in class $H_j^{(\nu)}$ that are released at time t_j^+,

Table 4. Summary statistics from Table 1 which are sufficient with respect to the model of Example III.

A

i	$T_i^{(0)}$	$T_i^{(1)}$	$T_i^{(2)}$	$T_i^{(3)}$
1	805,000	319,680	0	0
2	355,000	180,208	209,440	0
3	175,000	42,750	108,808	149,640
4	85,000	16,000	22,000	172,768

B

	$S_i^{(0)}$	$S_i^{(1)}$	$S_i^{(2)}$	$S_i^{(3)}$	n_i
2	450,000	0	200,000	0	650,000
3	180,000	0	112,000	112,000	404,000
4	90,000	0	27,000	133,680	250,680
5	85,000	0	16,000	194,768	295,768

C

	$R_i^{(0)}$	$R_i^{(1)}$	$R_i^{(2)}$	$R_i^{(3)}$
2	0	180,208	89,760	0
3	0	42,750	40,600	52,200
4	0	16,000	6,250	48,000
5	0	0	0	0

D

	$m_i^{(0)}$	$m_i^{(1)}$	$m_i^{(2)}$	$m_i^{(3)}$
1	0	900,000	0	0
2	0	400,000	150,000	0
3	0	150,000	100,000	90,000
4	0	90,000	25,000	120,000

$$m_j^{(\nu)} = \sum_{(h_{j-1},1) \in H_j^{(\nu)}} m_{h_{j-1}1} .$$

For example,

$$m_4^{(2)} = m_{0011} = 25,000.$$

ML Estimation

The likelihood factor depending on the unknown parameters $N_i^{(\nu)}$ is given by Robson's equation (1) as

$$\prod_{i=1}^{k} \frac{\prod_{\nu=0}^{q_i} \binom{N_i^{(\nu)}}{T_{i-1}^{(\nu)}}}{\binom{N_i}{n_i}} \prod_{i=2}^{k} - \frac{1}{\prod_{\nu=0}^{q_i} \binom{N_{i-1,0}^{(\nu)} - S_{i-1}^{(\nu)} + m_{i-1}^{(\nu)}}{T_{i-1}^{(\nu)}}} = L$$

where

$$N_{i-1,0}^{(\nu)} - S_{i-1}^{(\nu)} = \sum_{(h_{i-2}0) \in H_{i-1}^{(\nu)}} (N_{h_{i-2}} - X_{h_{i-2}1})$$

is the number of individuals present at t_{i-1}^{+} that have a capture history in $H_{i-1}^{(\nu)}$ and were not captured at t_{i-1}. In Example III

$$N_{i,0}^{(0)} = N_i^{(0)} \qquad N_{i,0}^{(1)} \equiv 0 \qquad N_{i,0}^{(2)} = N_i^{(1)} \qquad N_{i,0}^{(3)} = N_i^{(2)} + N_i^{(3)}$$

and since $N_1^{(0)} = N_1$ then

$$L = \prod_{2}^{k-1} \frac{\binom{N_i^{(0)}}{T_{i-1}^{(0)}}\binom{N_i^{(1)}}{T_{i-1}^{(1)}}\binom{N_i^{(2)}}{T_{i-1}^{(2)}}\binom{N_i^{(3)}}{T_{i-1}^{(3)}} \Big/ \binom{N_i}{n_i}}{\binom{N_i^{(0)} - S_i^{(0)}}{T_i^{(0)}}\binom{m_i^{(1)}}{T_i^{(1)}}\binom{N_i^{(1)} - S_i^{(2)} + m_i^{(2)}}{T_i^{(2)}}\binom{N_i^{(2)} + N_i^{(3)} - S_i^{(3)} + m_i^{(3)}}{T_i^{(3)}}} .$$

Letting $\Delta_i^{(\nu)}L$ denote the difference $L(N_i^{(\nu)}) - L(N_i^{(\nu)} - 1)$ and setting these differences equal to zero gives the likelihood equations $L(N_i^{(\nu)})/L(N_i^{(\nu)} - 1) = 1$:

(a) $$\frac{(N_i - n_i)N_i^{(0)}}{N_i(N_i^{(0)} - S_i^{(0)})} = 1 \quad \text{for } i = 2, 3, \cdots, k-1 = 4$$

(b) $$\frac{(N_i - n_i)N_i^{(1)}(N_i^{(1)} - S_i^{(2)} - T_i^{(2)} + m_i^{(2)})}{N_i(N_i^{(1)} - S_i^{(2)} - T_i^{(2)} + R_i^{(2)})(N_i^{(1)} - S_i^{(2)} + m_i^{(2)})} = 1 \quad \text{for } i = 2, 3, 4$$

$$\frac{(N_i - n_i)N_i^{(2)}(N_i^{(2)} + N_i^{(3)} - S_i^{(3)} - T_i^{(3)} + m_i^{(3)})}{N_i(N_i^{(2)} - T_{i-1}^{(2)})(N_i^{(2)} + N_i^{(3)} - S_i^{(3)} + m_i^{(3)})} = 1 \quad \text{for } i = 3, 4$$

$$\frac{(N_i - n_i)N_i^{(3)}(N_i^{(2)} + N_i^{(3)} - S_i^{(3)} - T_i^{(3)} + m_i^{(3)})}{N_i(N_i^{(3)} - T_{i-1}^{(3)})(N_i^{(2)} + N_i^{(3)} - S_i^{(3)} + m_i^{(3)})} = 1 \quad \text{for } i = 3, 4.$$

The last two equations may be rewritten as

(c) $$\frac{(N_i - n_i)(N_i^{(2)} + N_i^{(3)})(N_i^{(2)} + N_i^{(3)} - S_i^{(3)} - T_i^{(3)} + m_i^{(3)})}{N_i(N_i^{(2)} + N_i^{(3)} - S_i^{(3)} - T_i^{(3)} + R_i^{(3)})(N_i^{(2)} + N_i^{(3)} - S_i^{(3)} + m_i^{(3)})} = 1 \quad \text{for } i = 3, 4$$

(d) $$\frac{N_i^{(2)}(N_i^{(3)} - T_{i-1}^{(3)})}{N_i^{(3)}(N_i^{(2)} - T_{i-1}^{(2)}} = 1 \quad \text{for } i = 4.$$

Substituting the entries from Tables 3 and 4 into these equations will reveal that all 9 equations are satisfied, indicating that for these synthetic data the ML estimates are exactly correct. Equation (c) for i = 4, for example, gives

$$\left(\frac{584,920}{835,600}\right)\left(\frac{445,600}{187,152}\right)\left(\frac{259,152}{431,920}\right) = 1 \; .$$

For i = 2 the equations (a) and (b) are linear, giving

$$\hat{N}_2^{(1)} = S_2^{(2)} + \frac{m_2^{(2)} T_2^{(2)}}{R_2^{(2)}} \quad \text{and} \quad \hat{N}_2 = \frac{n_2}{S_2^{(2)}} \hat{N}_2^{(1)}$$

but the remaining ML estimates are obtained by solutions of quadratic equations.

Goodness of Fit

The pdf of the sample $\underset{\sim}{X}$ conditional upon the sufficient statistic $\underset{\sim}{U}$ may be expressed as products of the multihypergeometric distributions of the two-rowed tables:

. . .	h_{i-1}	. . .	Total
. . .	$X_{h_{i-1},1}$	. . .	$S_i^{(\nu)}$
	$T_{h_{i-1}}(0) - X_{h_{i-1},1}$		$T_{i-1}^{(\nu)}(0) - S_i^{(\nu)}$
. . .	$T_{h_{i-1}}(0)$	. . .	$T_{i-1}^{(\nu)}(0)$

for $(h_{i-1}0) \in H_i^{(\nu)}$

			Total
	$R_{h_{i-1}1}$		$R_i^{(\nu)}$
	$m_{h_{i-1}1} - R_{h_{i-1}1}$		$m_i^{(\nu)} - R_i^{(\nu)}$
	$m_{h_{i-1}1}$		$m_i^{(\nu)}$

for $(h_{i-1}1) \in H_i^{(\nu)}$

where

$T_{h_{i-1}}(0) = R_{h_{i-1}}$ = number subsequently captured at least once among those individuals present at t_i^- with capture history h_{i-1}.

In Example III only $H_3^{(3)}$ and $H_4^{(3)}$ yield non-degenerate tables; i.e., tables with two or more columns:

$$\frac{\binom{T_{10(0)}}{X_{101}}\binom{T_{11(0)}}{X_{111}}}{\binom{T_{2,0}^{(3)}}{S_3^{(3)}}} = \frac{\binom{119680}{64000}\binom{89760}{48000}}{\binom{209440}{112000}} \quad \text{for } (h_2 0) \in H_3^{(3)}$$

$$\frac{\binom{m_{101}}{R_{101}}\binom{m_{111}}{R_{111}}}{\binom{m_3^{(3)}}{R_3^{(3)}}} = \frac{\binom{50000}{29000}\binom{40000}{23200}}{\binom{90000}{52200}} \quad \text{for } (h_2 1) \in H_3^{(3)}$$

$$\binom{T_{010(0)}}{X_{0101}}\binom{T_{011(0)}}{X_{0111}}\binom{T_{100(0)}}{X_{1001}}\binom{T_{101(0)}}{X_{1011}}\binom{T_{110(0)}}{X_{1101}}\binom{T_{111(0)}}{X_{1111}}\Big/\binom{T_{3(0)}^{(3)}}{S_4^{(3)}}$$

$$\binom{68208}{35280}\binom{30600}{21000}\binom{55680}{28800}\binom{29000}{15000}\binom{41760}{21600}\binom{23200}{12000}\Big/\binom{248448}{133680}$$

for $(h_3 0) \in H_4^{(3)}$, while $(h_3 1) \in H_4^{(3)}$ gives

$$\binom{35000}{14000}\binom{20000}{8000}\binom{25000}{10000}\binom{12500}{5000}\binom{17500}{7000}\binom{10000}{4000}\Big/\binom{120000}{48000} .$$

Note that chi-squares for these 4 tables would give

$$\chi^2_{1df} + \chi^2_{1df} + \chi^2_{5df} + \chi^2_{5df} = \chi^2_{12df} = 0 .$$

References

[1] Robson, D. S. (1969). "Mark-recapture methods of population estimation" in New Developments in Survey Sampling, N. L. Johnson and H. Smith, Jr., Editors, Wiley Interscience, New York, pp. 120-140.

[2] Pollock, K. H. (1975). "A K-sample tag-recapture model allowing for unequal survival and catchability", Biometrika 62, 577-583.