A DISTRIBUTION FREE NESTING SURVIVAL MODEL

K. H. Pollock and W. L. Cornelius

Department of Statistics, North Carolina State University
Box 8203, Raleigh, North Carolina 27695-8203

SUMMARY

In nesting survival studies nests are not found until after incubation has begun and therefore nests which fail are of unknown age while nests which succeed can be aged afterwards. Typically ornithologists have considered the survival time to be from first encounter and have used the Mayfield method (Mayfield 1961, 1975) or variations of it. Here we consider a new discrete survival model which allows estimation of the survival distribution from the time of nest initiation. Nests which succeed are used to estimate the proportions of nests being found (encountered) in each time interval. Assuming that these encounter probabilities also apply to the nests which fail it is possible to estimate a series of discrete failure time probabilities for all the nests. We present maximum likelihood estimators and computer programs to calculate them. Our methodology is illustrated by a real example on mourning doves. We conclude with some suggestions for future research.

Key Words: Survival analysis; Nesting survival models; Mayfield model; Distribution free procedure.

1. Introduction

In nesting survival studies an ornithologist chooses a study area and searches for active nests to observe. Once a nest has been found it is visited regularly until it either fails (gets destroyed) or until it succeeds (the young first hatch and later fledge). Typically nests are not found until after incubation has begun and therefore nests which fail are of unknown age. Of course for nests which succeed the age at first encounter can be established because the total incubation and fledging time is approximately constant.

If the survival time from first encounter rather than initiation is considered then the Mayfield method (Mayfield 1961, 1975; Miller and Johnson (1978); Johnson (1979); Hensler and Nichols (1981); Bart and Robson (1981) can be used. This method assumes that each nest day of survival is independent and has constant probability. It does not use the information on age at discovery of successful nests and lumps nests of very different ages. (The method of analyzing radio-telemetry survival data proposed by Trent and Rongstad (1974) is very similar to the Mayfield method.)

Here we consider a new discrete survival model which allows estimation of the survival distribution from the time of nest initiation rather than the time of first encounter used in earlier models. Nests which succeed are used to estimate "first encounter probabilities". Assuming these first encounter probabilities are the same for the nests which fail it is possible to estimate a series of discrete failure time probabilities for all the nests. A shorter version of our approach appeared in Pollock (1984).

2. Model Development

2.1 Model Structure

In this model we consider a discrete distribution free approach. We denote the age at death of the ith nest as T_i and this random variable can take on the integer values 1, 2, ..., J because J is the fixed known number of units of time for incubation and fledging.

Model Parameters

Failure Parameters

$$q_1 = P(T = 1);\ q_2 = P(T = 2);\ \ldots\ldots;\ q_J = P(T = J)$$

where q_i is the probability a nest fails at age i units of time. Note that the probability of a nest succeeding is $P = 1 - q_1 - q_2 - \ldots - q_J$. Therefore there are J distinct failure parameters in the model. Note also that the time units could be days, two days, weeks or whatever is most appropriate.

Encounter Parameters

There are a set of J nuisance parameters $\delta_1, \delta_2, \ldots, \delta_J$ where δ_i is the probability of an intact nest being first encountered at age i.

A critical assumption to the model we develop is that these encounter probabilities are independent of the survival probabilities given above. In other words, nests encountered early or late in the cycle are not more or less likely to survive future time units.

Data Available

Successful Nests

$n_{1H}, n_{2H}, \ldots, n_{JH}$

n_{iH} is the number of nests found of age i which later succeed.

$n_H = \sum_{i=1}^{J} n_{iH}$ is the total number of successful nests encountered.

Failed Nests

$n_{1F}, n_{2F}, \ldots, n_{JF}$

n_{iF} is the number of nests of unknown age which are observed for i units of time and then fail.

$n_F = \sum_{i=1}^{J} n_{iF}$ is the total number of unsuccessful nests encountered.

2.2 The Likelihood

We consider the likelihood as the product of three conditional multinomial distributions

(i) $$P(n_{1H}, n_{2H}, \ldots, n_{JH} | n_H) = \binom{n_H}{n_{1H}, n_{2H}, \ldots, n_{JH}}$$

$$\cdot \left[\frac{\delta_1}{\sum_{i=1}^{J} \delta_i} \right]^{n_{1H}} \cdot \left[\frac{\delta_2}{\sum_{i=1}^{J} \delta_i} \right]^{n_{2H}} \cdots \left[\frac{\delta_J}{\sum_{i=1}^{J} \delta_i} \right]^{n_{JH}}$$

(ii) $$P(n_{1F}, n_{2F}, \ldots, n_{JF} | n\ ) = \binom{n_F}{n_{1F}, n_{2F}, \ldots, n_{JF}}$$

$$\cdot \left[\frac{\delta_1 q_1 + \delta_2 q_2 + \ldots + \delta_J q_J}{\delta_1 q_1 + (\delta_1 + \delta_2) q_2 + \ldots + (\delta_1 + \delta_2 + \ldots + \delta_J) q_J} \right]^{n_{1F}}$$

$$\cdot \left[\frac{\delta_1 q_2 + \delta_2 q_3 + \ldots + \delta_{J-1} q_J}{\delta_1 q_1 + (\delta_1 + \delta_2) q_2 + \ldots + (\delta_1 + \delta_2 + \ldots + \delta_J) q_J} \right]^{n_{2F}}$$

$$\vdots$$

$$\cdot \left[\frac{\delta_1 q_J}{\delta_1 q_1 + (\delta_1 + \delta_2) q_2 + \ldots + (\delta_1 + \delta_2 + \ldots + \delta_J) q_J} \right]^{n_{JF}}$$

$$\text{(iii)}\quad P\,(n_H | n_H + n_F) = \binom{n_H + n_F}{n_H}$$

$$\cdot \left[\frac{(1 - \sum_{i=1}^{J} q_i)\ \sum_{i=1}^{J} \delta_i}{(1 - \sum_{i=1}^{J} q_i) \sum_{i=1}^{J} \delta_i + \delta_1 q_1 + (\delta_1 + \delta_2) q_2 + \ldots + (\delta_1 + \delta_2 + \ldots + \delta_J) q_J} \right]^{n_H}$$

$$\cdot \left[\frac{\delta_1 q_1 + (\delta_1 + \delta_2) q_2 + \ldots + (\delta_1 + \delta_2 + \ldots + \delta_J) q_J}{(1 - \sum_{i=1}^{J} q_i) \sum_{i=1}^{J} \delta_i + \delta_1 q_1 + (\delta_1 + \delta_2) q_2 + \ldots + (\delta_1 + \delta_2 + \ldots + \delta_J) q_J} \right]^{n_F}$$

The first multinomial component of the likelihood is the conditional distribution of the numbers of successful nests of each age at encounter (n_{iH} i = 1, 2, ..., J) given the total number of successful nests (n_H). The distribution only involves the encounter parameters in the cell probabilities.

The second multinomial component of the likelihood is the conditionsl distribution of the numbers of nests of unknown age which fail after i units of time (n_{iF} i = 1, 2, ..., J) given the total number of nests which fail (n_F). This distribution is easy to derive when one realizes that an encountered nest which fails after 1 time unit has age at death between 1 and J units ($1 \leq T \leq J$) and this event has probability $(\delta_1 q_1 + \delta_2 q_2 + \ldots + \delta_J q_J)$. Similarly an encountered nest which fails after 2 units of time has age at death between 2 and J units ($2 \leq T \leq J$) and this event has probability $(\delta_1 q_2 + \delta_2 q_3 + \ldots + \delta_{J-1}\, q_J)$. Finally an encountered nest which fails after J units of time has age at death T = J and this event has probability $\delta_1 q_J$ because it must have been encountered on the first day.

The denominator of each cell probability is the total probability of an encountered nest failing because this is a conditional distribution.

The third component of the likelihood is a conditional binomial distribution. It is the conditional distribution of the number of successful nests encountered at any time (n_H) given the total number of nests encountered ($n_H + n_F$).

2.3 Maximum Likelihood Estimation

From the first component we can estimate the encounter parameters explicitly by imposing a constraint $\sum_{i=1}^{J} \delta_i = 1$. The estimators take the very simple form

$$\hat{\delta}_1 = \frac{n_{1H}}{n_H}, \ \hat{\delta}_2 = \frac{n_{2H}}{n_H}, \ \ldots, \ \hat{\delta}_J = \frac{n_{JH}}{n_H}$$

Note that $\hat{\delta}_J = 1 - \hat{\delta}_1 - \hat{\delta}_2 \ldots - \hat{\delta}_{J-1}$ due to the constraint.

From the second component we obtain the following set of J equations of which only J - 1 are independent.

$$\frac{\hat{\delta}_1\hat{q}_1 + \hat{\delta}_2\hat{q}_2 + \ldots + \hat{\delta}_J\hat{q}_J}{\hat{\delta}_1\hat{q}_1 + (\hat{\delta}_1 + \hat{\delta}_2)\hat{q}_2 + \ldots + (\hat{\delta}_1 + \hat{\delta}_2 + \ldots + \hat{\delta}_J)\hat{q}_J} = \frac{n_{1F}}{n_F}$$

$$\vdots$$

$$\frac{\hat{\delta}_1\hat{q}_J}{\hat{\delta}_1\hat{q}_1 + (\hat{\delta}_1 + \hat{\delta}_2)\hat{q}_2 + \ldots + (\hat{\delta}_1 + \hat{\delta}_2 + \ldots + \hat{\delta}_J)\hat{q}_J} = \frac{n_{JF}}{n_F}$$

From the third component we obtain an additional independent equation

$$\frac{(1 - \sum_{i=1}^{J}\hat{q}_J)}{(1 - \sum_{i=1}^{J}\hat{q}_J) + \hat{\delta}_1\hat{q}_1 + (\hat{\delta}_1 + \hat{\delta}_2)\hat{q}_2 + \cdots + (\hat{\delta}_1 + \hat{\delta}_2 + \ldots + \hat{\delta}_J)\hat{q}_J} = \frac{n_H}{n_H + n_F}$$

Therefore we have a set of J independent linear equations in J unknowns (the $\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_J$) which can be solved if we substitute in the estimates of the $\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_J$ from the first component.

In Section 3 we illustrate this estimation procedure with an example using program SURVIV (White 1983).

3. Example

The following analysis of mourning dove (*Zenaida macroura*) nest data recorded in 1978 and 1979 at Patuxent Wildlife Research Station, Laurel, Maryland by James D. Nichols.

Field workers located 59 ($n = n_H + n_F$) nests containing 1 or 2 eggs each; 34 (n_F) of the nests failed to produce fledglings, 16 (n_{1F}) of those within 1 week of discovery, 9 (n_{2F}) within the 2nd week after discovery, and 9 (n_{3F}) within the 3rd week. The remaining 25 (n_H) nests did produce nestlings, and of these, 11 (n_{1H}) were nests first found no more than 1 week after the eggs were laid, 7 (n_{2H}) were found when eggs were in the 2nd week of development, and the remaining 7 (n_{3H}) were found in the 3rd week.

Encounter probability estimates for this example are

$$\hat{\delta}_1 = n_{1H}/n_H = 11/25 = 0.44 \text{ (s.e. = 0.0993)}$$

$$\hat{\delta}_2 = n_{2H}/n_H = 7/25 = 0.28 \text{ (s.e. = 0.0898)}$$

$$\hat{\delta}_3 = n_{3H}/n_H = 7/25 = 0.28 \text{ (s.e. = 0.0898)}$$

Associated failure probabilities are

$$\hat{q}_1 = 0.26 \text{ (s.e. = 0.1345)}$$

$$\hat{q}_2 = 0.10 \text{ (s.e. = 0.1330)}$$

$$\hat{q}_3 = 0.29 \text{ (s.e. = 0.1171)}$$

The asymtotic variance-covariance matrix of the $\underline{\delta}$ and $\underline{q}$ vectors is given in Table 1. Notice that while the standard errors are large here there is evidence that the failure probabilities are not constant over the nesting period. In the first week of incubation the failure probability is high ($\hat{q}_1 = 0.26$). This might be attributed to predators finding it easy to locate nests and nests of inexperienced breeders being abandoned. In the second week of incubation the failure probability is much lower ($\hat{q}_2 = 0.10$). In the first week of nestlings the failure probability appears to rise again ($\hat{q}_3 = 0.29$). This might be attributed to predators being attracted to the nests by the parents feeding the nestlings.

The probability of a nest succeeding is given by $\hat{P} = 1 - \hat{q}_1 - \hat{q}_2 - \hat{q}_3 = 0.35$. Notice that the naive estimate of a nest succeeding ignoring the different encounter times would be $n_H/n = 25/59 = 0.42$ which is positively biased due to nests already having survived some period before being found.

Program SURVIV (White 1983) is useful for this estimation problem, if input specifications are properly made, and especially if the parameter J is small. (We found with J = 6 the distribution IBM PC version of SURVIV would not accept the full expression for n_{1F} conditional on n_F. Presumably this can be overcome by recompiling SURVIV with a larger character expression size limit.)

The 3 likelihood component systems of equations are expressed to SURVIV as 3 "cohorts" in the PROC MODEL specification. We recommend that SURVIV be asked to solve for 3J, not 2J parameters. This is because the required procedure is to solve for the $\underline{\delta}$ vector first, and use that result to solve for the q_i parameters. SURVIV, however, solves all equations in PROC MODEL simultaneously. Thus our 2-stage solution must be effected by a PROC MODEL step in which the $\underline{\delta}$ vector parameters are indexed by different variables in the first COHORT than in the 2nd and 3rd COHORT. The 2-stage solution can be implemented with 2 PROC ESTIMATE steps, the 1st solving for all 3J parameters simultaneously, and the 2nd taking the output of the 1st as initial values but constraining the 2 duplicate δ vectors to be elementwise equal.

For our example, the input to SURVIV should appear as in Figure 1. Note that only 2 J - 1 equations need to be specified, because the COHORT cell probabilities always sum to 1, and SURVIV by default creates an additional cell in each cohort whose probability is assigned the difference between 1 and the sum of the other, explicit cell probabilities. This means that one of the J equations in each of the 1st 2 COHORTs can be dropped from the specification. In our example, we omitted the equation defining δ_3; we introduced it once in the 2nd COHORT as the expression $1 - \delta_1 - \delta_2$.

Another computer program to solve this estimation problem, NESTING, is available from the authors. Our program can handle larger failure vectors than SURVIV. NESTING solves the likelihood equations explicitly for point estimates and uses simulation for variance estimates, while SURVIV uses numerical optimization methods to obtain all estimates.

4. Model Assumptions

The first assumption of the model is that the nests observed constitute a random sample with respect to survival. This same assumption is required of all sampling procedures to estimate survival rates (radio-tagging method, capture-recapture method). If the nests which are easier to find also have higher predation rates it could be violated.

There is no seasonal component in this model. All nests are treated as a cohort from time of initiation. If the sample were large enough one could stratify the sample into early, medium and late nests and compare the survival distributions.

This model requires the assumption that visiting the nest does not influence its survival. Bart and Robson (1982) state that, "In some cases, visiting the subject may temporarily depress its chance of survival. --- observers may lead predators to the nest or cause nest abandonment". They present an extension of the Mayfield type model to allow a temporary visitor induced increase in mortality just after the visit.

Bart and Robson (1982) also emphasize the importance of a regular schedule of visits (daily, weekly). Sometimes biologists are tempted to change the schedule if they think a nest is just about to fail.

A very critical assumption to this model is that encounter probabilities are not related to the subsequent success of nests. Take for example nests encountered close to the beginning of the nesting cycle. It may be that these nests are highly visible and more likely to suffer predation. This assumption will need to be investigated in more detail using simulation in a future article.

ACKNOWLEDGEMENTS

We would like to thank James D. Nichols for providing the data and for many helpful discussions on nesting survival experiments.

REFERENCES

Bart, J. and Robson, D. S. (1982). Estimating survivorship when the subjects are visited periodically. Ecology 63, 1078-1090.

Hensler, G. L. and Nichols, J. S. (1981). The Mayfield method of estimating nesting success: a model, estimators and simulation results. Wilson Bulletin 93, 42-53.

Johnson, D. H. (1979). Estimating nest success: the Mayfield method and an alternative. Auk 96, 651-661.

Mayfield, H. (1961). Nesting success calculated from exposure. Wilson Bulletin 73, 255-261.

Mayfield, H. (1975). Suggestions for calculating nest success. Wilson Bulletin 87, 456-466.

Miller, H. W. and Johnson, D. H. (1978). Interpreting the results of nesting studies. J. Wildl. Manage. 42, 471-476.

Pollock, K. H. (1984). Estimation of survival distributions in ecology. Proceedings of 12th International Biometrics Conference. Tokyo, Japan.

Trent, T. T. and Rongstad, O. J. (1974). Home range and survival of cottontail rabbits in Southwestern Wisconsin. J. Wildl. Manage. 38, 459-472.

White, G. C. (1983). Numerical estimation of survival rates from band-recovery and biotelemetry data. J. Wildl. Manage. 47, 716-728.

Table 1. Variance-Covariance Matrix for $(\hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3, \hat{q}_1, \hat{q}_2, \hat{q}_3)$

.009856	-.004928	-.004928	-.000017	0.005158	-.005838
-.004928	.008064	-.003136	.004384	-.006323	.002635
-.004928	-.003136	.008064	.004384	-.006323	.002635
-.000017	-.004384	-.004384	.018093	-.008869	-.004901
.005158	-.006323	-.006323	-.008869	.017690	-.008836
-.005838	.002635	.002635	-.004901	-.008836	.013709

FIGURE 1. Input specification for Nesting survival example analysis by SURVIV (White 1983).

```
PROC TITLE Nesting Survival Example (diskfilename NESTEX.WLC);

PROC MODEL npar=7;

    COHORT = 25          /*  total successful nests    */;
       11: s(1)          /*  seen period 1             */;
        7: s(2)          /*  seen period 2             */;

    COHORT = 34          /*  fail nests                */;
       16: s(6)*s(3)+s(7)*s(4)+(1-s(1)-s(2))*s(5))/
              (s(6)*s(3)+(s(6)+s(7))*s(4)+s(5))/*FAIL AFTER 1 PERIOD*/;
        9: s(6)*s(4)+s(7)*s(5))/
              (s(6)*s(3)+(s(6)+s(4)+s(5))/*FAIL AFTER 2 PERIODS*/;

    COHORT = 59         /*  total nests                */;
       25: (1 -s(3) -s(4) -s(5))  /
           (1 -s(3) -s(4) -s(5)   +  s(6)*s(3)+(s(6)+s(7))*s(4) + s(5))
                                     /*TOTAL SUCCESSFUL NESTS*/;

    LABELS;
       s(1) = First encounter in period 1, delta(sub1);
       s(2) = First encounter in period 2, delta(sub2);
       s(3) = Prob(failure at age 1), q(sub1);
       s(4) = Prob(failure at age 2), q(sub2);
       s(5) = Prob(failure at age 3), q(sub3);
       s(6) = First encounter in period 1, delta(sub1);
       s(7) = First encounter in period 2, delta(sub2);

PROC ESTIMATE nsig=5 maxfn=1000 name=DELTA;
     INITIAL;
       ALL = 0.3;

PROC ESTIMATE nsig=5 maxfn=1000 name=QPROB;
     INITIAL;
        RETAIN = DELTA;
     CONSTRAINTS;
        s(6) = s(1);
        s(7) = s(2);
```