Known Fate Models

This class of models is important because they provide a theory for estimation of survival probability and other parameters from radio-tagged animals. The focus of known fate models is the estimation of survival probability *S*, the probability of surviving an interval between sampling occasions. These are models where it can be assumed that the sampling probabilities are 1. That is, the status of each tagged animal is known at each sampling occasion. For this reason, precision is typically quite high, even in cases where sample size is often fairly small. The only disadvantages might be the cost of radios and possible effects of the radio on the animal or its behavior. The model is a product of simple binomial likelihoods.

Studies of egg mortality in nests and studies of sessile organisms (mollusks) have also be treated as known fate data. PIT (passive integrated transponders) tags can be used to provide known fate data and have been very widely used in fisheries studies on the Columbia River system. Smith et al. (1994) provide additional details on these models.

The Kaplan-Meier Method

The Kaplan-Meier (1958) estimate is based on observed data at a series of time points, where animals are marked and released only at time 1. The K-M estimator is

$$\hat{S}_t = \prod \left(\frac{n_i - d_i}{n_i} \right) ,$$

where n_i is the number of animals alive and at risk at time i, d_i is the number known dead at time i, and the summation is over i up to the t^{th} time period. Critical here is that n_i is the number known alive at time i minus those individuals known dead or censored during the interval. It is rare that a survival study will observe the time of death of every individual in the study. Animals are "lost" (i.e., censored) due to radio failure or other reasons. The treatment of such censored animals is often important. These K-M estimates produce a survival function (see White and Garrott 1990).

If there are no animals that are censored, then the survival function (empirical survival function or ESF) is merely,

$$\hat{S}_t = \frac{\text{Number of observation} \geq t}{n} \qquad \text{for } t \geq 0.$$

This is the same as the intuitive estimator where not censoring is occurring;

The K-M method is an estimate of this survival function in the presence of censoring. Expressions for the variance of these estimates can be found in White and Garrott (1990).

A simple example of this method can be illustrated using the data from Conroy et al. (1989) on 48 radio-tagged black ducks. The data are

Week	1	2	3	4	5	6	7	8
Number alive at start	48	47	45	39	34	28	25	24
Number dying	1	2	2	5	3	3	1	0
Number alive at end	47	45	39	34	28	25	24	24
Number censored	0	0	4	0	2	0	0	0

Thus,

S_1 = 47/48 = 0.979

S_2 = 45/47 = 0.957

S_3 = 39/41 = 0.951 (note, only 41 because 4 were censored)

S_4 = 34/39 = 0.872

S_5 = 28/32 = 0.875 (note, only 32 because 2 were censored)

S_6 = 25/28 = 0.893

S_7 = 24/25 = 0.960

S_8 = 24/24 = 1.000.

Here one estimates 8 parameters (call this model $S(t)$); one could seek a more parsimonious model in several ways. First, perhaps all the parameters were nearly constant; thus a model with a single survival probability might suffice (i.e., $S(.)$) If something was known about the time intervals (similar to the flood years for the European dipper data) one could model these with one parameter and denote the other periods with a second survival parameter. Finally, one might consider fitting some smooth function across the time periods and, thus, have perhaps only one intercept and one slope parameters (instead of 8 parameters). Still other possibilities exist for both parsimonious modeling and probable heterogeneity of survival probability across animals.

Pollock's Staggered Entry Design

The Kaplan-Meier method assumes that all animals are released at time 1 and they are followed during the study until they die or are censored. Often new animals are released at each time period (say, weekly); we say this entry is “staggered" (Pollock et al. 1989). Assume, as before, that animals are fitted with radios and that these do not affect the animals survival probability. This staggered entry fits easily into the K-M framework by merely redefining the n_i to include the number of new animals released at time i Therefore, conceptually, the addition of new animals into the marked population causes no difficulties in data analysis.

The Binomial Model

Smith et al. (1994) note that there are 3 possible scenarios under the known fate model: for each tagged animal it

1. survives to end of study and is detected at each sampling occasion after its release

2. dies sometime during the study and its carcass is found on the first sampling occasion after its death

3. survives up to the point at which time it is censored.

Note, for purposes of estimating survival probabilities, there is no difference between an animal seen alive and then removed from the population at occasion *k* vs. an animal alive at occasion *k* and then censored due to radio failure or whatever.

The binomial/multinomial model assumes the capture histories are mutually exclusive and exhaustive, that animals are independent, and all animals have the same underlying parameters (homogeneity across individuals).

Radio tagging data can be modeled by a product of binomials. In the black duck example, $n_1 = 48$ and $n_2 = 47$ and the likelihood is

$$\mathcal{L}(S_1 \mid \underline{n}) = \binom{n_1}{n_2} S_1^{n_2} (1 - S_1)^{n_1 - n_2} .$$

Clearly, one could find the MLE, $\hat{S}_1$, for this expression. Of course, the other binomial terms are multiplicative; e.g., survival during the following interval is based on $n_2 = 47$ and $n_3 = 45$,

$$\mathcal{L}(S_2 \mid \underline{n}) = \binom{n_2}{n_3} S_2^{n_3} (1 - S_2)^{n_2 - n_3} .$$

The likelihood function for the entire set of black duck data is the product of these individual likelihoods. The log-likelihood the the sum of terms such as

$$\log(\mathcal{L}(S_i \mid \underline{n})) = \sum \underline{n}_i \log(\text{Prob.}).$$

Consider the following (paired, live (*l*) and dead (*d*) encounter histories:

History	Probability	Number Observed
10 10 10 10	$S_1 S_2 S_3 S_4$	17
Tagged at time 1 and survived until the end of the study		
10 10 11 00	$S_1 S_2 (1 - S_3)$	21
Tagged at time 1 and died during the third interval		
10 11 00 00	$S_1 (1 - S_2)$	24
Tagged at time 1 and died during the second interval		
11 00 00 00	$(1 - S_1)$	43
Tagged at time 1 and died during the first interval		

Estimation of survival probabilities is based on a release (1) and a death (1); if the animal then was censored, it does not provide information about S_i.

More on Binomial Likelihood Functions

Before we move into models for individual covariates, some quick review of the binomial likelihood might be helpful. Consider the usual *n* flips of a coin where,

$p =$ probability the coin lands heads;
$q = 1 - p =$ probability the coin lands tails.

Let *n* = 16 flips (trials). We often write the likelihood in a compact form as

$$\mathcal{L}(p \mid n, y) \;=\; (p)^y \, (1 - p)^{16 - y} \,,$$

where $y =$ number of heads. If we observe *y* = 5, then

$$\mathcal{L}(p \mid 16, 5) \;=\; (p)^5 \, (1 - p)^{16 - 5} \,.$$

Alternatively, we could write the likelihood for each individual outcome and take the product of these terms as the likelihood function. One alternative is to merely write the likelihood as (using the convention that $q = (1 - p)$),

$$\mathcal{L}(p \mid 16, 5) = pppppp \cdot qqqqqqqqqq.$$

Alternatively, we could write this as,

$$\mathcal{L}(p \mid 16, 5) = \prod_{i=1}^{5} p_i \cdot \prod_{i=6}^{16} (1 - p_i).$$

Finally, we could define an indicator variable to denote head or tail; let $y = 1$ if heads, 0 if tails. Then the likelihood can be written for the i^{th} flip as

$$\mathcal{L}\left(p \mid n, \{y_1, y_2, \ldots, y_{16}\}\right) = \prod_{i=1}^{16} \left[(p)^{y_i} (1 - p)^{1-y_i}\right].$$

It these last three forms, each outcome has a term in the likelihood. The likelihood is the product of these individual terms. These formulations ares useful in understanding the modeling of individual covariates.

Individual Covariates

A number of people have suggested modeling of the individual animals, allowing covariates that vary by individual (e.g., White and Garrott, Smith et al. 1994, Pollock ___). This approach is very useful in the biological sciences. In the black duck example, $n_1 = 48$ and $n_2 = 47$ and the binomial likelihood for the survival probability during the first week (i.e., S_1) can be written as

$$\mathcal{L}(S_1 \mid \underline{n}) = \binom{n_1}{n_2} S_1^{n_2} (1 - S_1)^{n_1 - n_2}$$

This can be expressed (omitting the multinomial coefficient) as

$$\mathcal{L}(S_1 \mid \underline{n}) = \prod_{i=1}^{n_2} S_i \cdot \prod_{i=n_2+1}^{n_1} (1 - S_i),$$

where the subscript *i* is over ducks (48 ducks in the study). Thus, the first term in the likelihood is the product of the survival probability over 47 ducks, while the second term in the product of (1–*S*), the ducks that died during the first week (in this examples only a single duck died). So, a final expression of this likelihood is

$$\mathcal{L}(S_1 \mid \underline{n}) = \prod_{i=1}^{47} S_i \cdot (1 - S_i).$$

Now we consider *modeling* the survival probability of these individuals as a nonlinear function of some covariate that varies for each individual animal. The natural choice is the logistic model

$$S_i = \frac{1}{1 + \exp(-[\beta_0 + \beta_1(X_i)])}$$

with link function

$$\log_e\left(S_i/(1\text{-}S_i)\right) = \beta_0 + \beta_1(X_i)$$

where X_i is the value of the covariate for the i^{th} individual. Of course, other functions could be used (log, log-log, complementary log-log, etc.). More than one covariate can also be measured and used with this general approach. If we substitute the logistic submodel and its individual covariate into the likelihood above, the expression *looks* messy, but is conceptually familiar,

$$\mathcal{L}(\beta_0, \beta_1 \mid \underline{n}, X_i) = \prod_{i=1}^{n_2} \left(\frac{1}{1 + \exp(-[\beta_0 + \beta_1(X_i)])} \right) \cdot \prod_{i=n_2+1}^{n_1} \left(1 - \left(\frac{1}{1 + \exp(-[\beta_0 + \beta_1(X_i)])} \right) \right)$$

or

$$\mathcal{L}(\beta_0, \beta_1 \mid \underline{n}, \quad X_i) \quad = \quad \prod_{i=1}^{47} \quad \left(\frac{1}{1+\exp(-[\beta_0 + \beta_1(X_i)])} \right) \quad \cdot \quad \left(1- \left(\frac{1}{1+\exp(-[\beta_0 + \beta_1(X_{48})])} \right) \right)$$

Thus, the MLEs for β_0 and β_1 (the intercept and slope, respectively) are the focus of the estimation. Of course, additional binomial terms could be multiplied for the parameters S_2, S_3, ..., S_8 in the black duck example.

There are two notions to think clearly about:

(1) the survival probabilities are replaced by a logistic (sub)model of the individual covariate X_i. Conceptually, every animal *i* has its own survival probability and this may be related to the covariate X_i.

(2) during the analysis, the covariate of the i^{th} animal must correspond to the survival probability of that animal. Program *MARK* handles this detail. Note, in the last expression of the likelihood (above) we assume it is the 48^{th} duck that died and this corresponds to the 48^{th} covariate, X_{48}.

An Example of the Application

Assume a biologist has found 88 active nests of the red-cockaded woodpecker in which nest initiation occurred on the same day. She selects a single nestling from each of the 88 nests and measures 3 covariates on each of these 88 nestlings. The covariates measured are the number of ectoparasites found, the number of hatchlings in the nest, and the weight at hatching. The “first" occasion is actually on day 3 following hatching. Each bird is tagged uniquely with a colored leg band to allow it to be identified and its fate determined visually be daily inspection of the nest. Birds are followed for 12 days (while they are still in the nest; they typically start to leave the nest after 15 days) and their fate is determined daily. Thus, the data follow the known fate scenario, even though animals are not fitted with radios. Overdispersion should not be a factor as only one bird in each nest is the subject of the study. Sample size is 88 (no staggered entry) and there are 12 occasions (days 3, 4, ..., 15).

If the data were modeled without an occasion effect (i.e., without model S_t), one might potentially include the model with all three covariates for each individual (*i*), as

where,

E_i is the number of ectoparasites on day 3 (= occasion 1)
H_i is the number of nest mates on day 3
W_i is the weight of the nesting on day 3.

Note the interaction term between weight and number of nest mates on day 3. Of course, other *a priori* models would be considered in making inferences from these data, this is just an example.

The estimation would focus on the β parameters, but the *interpretation* would be interesting. For example, one might look at the mean of the S_i given that all the covariates were held at their average values. Then this mean might be compared with means for low vs. high values of weight and the number of nest mates. One could compute the values of *S* for a range of ecoparasites, while holding the other covariates at their mean values. Other possibilities exist and could be explored for the selected model.

Now, one can see that individual covariates can be used in the band recovery models and the open capture-recapture models. *Too few biologists are taking full advantage of the information contained in individual covariates.*

Note, there are problems if the covariate changes through time in the band recovery and open C-R models. For example, if weight changes throughout the study period, one only has weights for those animals recaptured at various times. Thus, when animals are not captured (e.g., the "never recaptured" animals) then the value of their covariate at that time is not known!