Burnham's Joint Model

For many years there has been interest in a model that could make a joint (combined) analysis of band recovery data on dead animals **and** open capture-recapture data on animals released alive. In the band recovery models one can estimate survival (S), reporting (r) or recovery (f) probabilities, while in the open capture-recapture models one can estimate apparent survival (ϕ) and recapture probabilities (p). Then, $S = \phi/F$, where F is a fidelity probability, the probability of remaining or returning to the C-R site and being available for capture. If $F = 1$, then $S = \phi$. Fidelity is the complement of emigration (E), thus $F = 1 - E$.

Given these relationships one can envision the collection of both recovery data and C-R data on a species, over the same time period (*a la* Anderson and Sterling 1974). Then a model is required to allow a joint analysis of the two types of data. Burnham (1993) provided the (complicated) theory for the joint model (also see a series of preceding papers by Burnham). Szymczak and Rexstad (1991) provide a nice example using gadwall ducks (*Anas strepera*).

Some data sets have substantial amounts of data of both types, thus a combined analysis should provide increased precision and the ability to identify additional “effects." The primary advantage of the joint model is the exploitation of the difference in meaning between S and ϕ and be able to estimate a fidelity parameter, F. This is a difficult model with several parameters that are non-identifiable under certain conditions. Parsimonious modeling is possible in *MARK*, thus models might range from {*S(.) p(.) r(.) F(.)*} to {*S(t) p(t) r(t) F(t)*} based primarily on what seems to be known about the biology underlying the study. These parameters can be modeled as functions of covariates using the concept of link functions. Additive models and trend models can be developed using the design matrix capability in *MARK*. Plus, individual covariates are possible using the design matrix capability.

The Encounter History Matrix and Model

A major difference in using the joint model compared to the live recapture models is in the basic format of the encounter history (EH) matrix. Here it is necessary to include *pairs* of 0s or 1s for each occasion (program *MARK* uses the paired version of the capture histories for all the models). This allows one to denote (live) capture (the first value in the pair) vs. death (the second value of the pair). For short, these pairings are called live (l) and dead (d). Thus, a capture history might be

10 00 10 00 10 but written and coded as simply 1000100010.

Encounter histories can only include one d term, i.e., the animal can only die once. However, more than one l term is possible, because the animal can be recaptured alive multiple times. This is in contrast to the dead recovery model because that model only allows one l capture, and one d recovery.

The joint live and dead encounter model is a product multinomial model as in the band recovery and open C-R models considered previously. The “data" are the individual capture histories. The probability structure of this joint model can only be described as “nasty" as there are many possible histories because of the numerous possible paths that an animal can take. Burnham's (1993) paper is difficult to follow because because terms such as the γ_{ij} are, by necessity, recursive.

The joint model can be considered as a Cormack-Jolly-Seber model, then treating animals that were reported dead at the terminal occasion as a recovery. Thus, we start with the C-J-S model where the expected probability under model $\{\phi_t,\ p_t\}$ for the capture history **{10101010101010}** is (ignoring the last d term):

$$(\phi_1 p_2)(\phi_2 p_3)(\phi_3 p_4)(\phi_4 p_5)(\phi_5 p_6)(\phi_6 p_7).$$

Now, imagine the substitution of $F_i S_i$ for ϕ_i in the capture history. In C-R data it is essential to formally include the fact that animals alive at the beginning of occasion *i* were captured (*p*) or not (1–*p* = *q*).

Writing the EH in pairs is useful for the conceptualization and writing down the expected values. For example,

ld ld ld ld ld
10 00 10 00 10

makes the interpretation easier. The key to understanding the joint model is understanding how the parameter F interacts with S and p to model the probability that an animal is available for live recapture. Consider the capture history

ld ld ld ld ld
10 10 10 00 11

which results in the probability

$$S_1 F_1 p_2\, S_2 F_2 p_3\, S_3 F_3 (1-p_4) S_4 F_4 p_5\, (1-S_5) r_5 \ \ .$$

The animal remained on the live trapping area during the 4th live capture occasion because the animal was captured alive on the 5th interval.

Had the animal not been reported dead on the last occasion, the encounter history would be

ld ld ld ld ld
10 10 10 00 10

with probability

$$S_1F_1p_2\,S_2F_2p_3\,S_3F_3(1-p_4)S_4F_4p_5[\,(1-S_5)(1-r_5)+S_5]\ \ ,$$

i.e., the probabiity of not seeing the animal dead during the last interval is now the probability that it died and was not reported, plus the probability that it is still alive. The other key point to recognize about these 2 examples is that we know the animal did not emigrate from the study area because it was caught on the last live capture occasion. Suppose the following capture history was observed

ld ld ld ld ld
10 10 10 10 01 ,

which has probability

$$S_1F_1p_2\,S_2F_2p_3\,S_3F_3p_4S_4[(1-F_4)+F_4(1-p_5)]\,(1-S_5)r_5\ \ .$$

In this example, we know the animal survived the 4th interval, because the animal was found dead during the 5th interval. However, it was not captured during the 5th live capture occasion. Two possibilities explain this non-capture: the animal emigrated from the study area, or it was present on the study area and was not captured.

Additonal possibilities arise concerning the fate of the animal if it is not captured live on the 4th occasion. For example, the encounter history

ld ld ld ld ld
10 10 10 00 01

results in

$$S_1F_1p_2\,S_2F_2p_3\,S_3S_4\{(1-F_3)+F_3(1-p_4)[(1-F_4)+F_4(1-p_5)]\}\,(1-S_5)r_5\ \ .$$

That is, the probability that the animal is not seen during the 4th live capture occasion is the probability that it emigrated from the study area, plus the probability that it stayed on the study area during the interval, but was not captured. If the animal stayed on the study area and was available for capture during the 4th live capture occasion, it could have emigrated during the 4th interval, or stayed on the study area and not have been captured during the 5th live capture occasion. However, we know the animal was still alive during this period when it was not captured, because it died during the 5th interval.

The complexity that results when the animal is not captured during live recapture occasions, but is known to be alive, results in the parameter γ_{ij} of Burnham (1993). The parameter has the lengthy definition;

γ_{ij} is the probability that an animal released at capture occasion *i* and still alive at occasion *j* will never have been removed from the i^{th} released cohort by being recaptured on one of occasions *i*+1 to, and including, occasion *j*.

Basically, γ_{ij} is the probability of not capturing the animal alive during the interval i to j, given that the animal is alive.

Additional complications arise when the animal is not seen again after a last live capture. For example,

ld ld ld ld ld
10 10 10 10 00

has the probability

$$S_1F_1p_2\,S_2F_2p_3\,S_3F_3p_4\{(1-S_4)(1-r_4)+$$

$$S_4[(1-F_4)+F_4(1-p_5)][S_5+(1-S_5)(1-r_5)]\}\ .$$

This nastiness is encompassed in the λ_l and λ_d of Burnham (1993). We will let program *MARK* worry about the intricacies of the parameterization of the capture histories, but the following relations are useful in gaining rough insights into this model:

ϕ_i = F_iS_i and if fidelity is perfect (i.e., no emigration, so that $E=0$, giving $F=1-E=1$), then ϕ_i = S_i.

In addition, for the model {*S(t) p(t) r(t) F(t)*}, the last p and the last F are confounded as a product, $F_{t-1}p_t$. Likewise, the last S and the last r are confounded just as in the dead recoveries model as $(1-S_t)r_t$.

For t = 5, the PIM windows of MARK will have 5 columns for each of S and r, and 4 columns for F and p, giving a total parameter count of 18. However, only 16 of these parameters are estimable given the confounding discussed above.