Appendix S1. Updating of latent encounter history frequencies in `multimark`

Brett T. McClintock[1]

National Marine Mammal Laboratory

Alaska Fisheries Science Center
NOAA National Marine Fisheries Service
Seattle, Washington, U.S.A.
[1] *Email:* brett.mcclintock@noaa.gov

May 11, 2015

The feasible set of latent encounter histories ($\mathbf{Y}$) is explored in `multimark` using an extension of the MCMC algorithms proposed by Bonner & Holmberg (2013) and McClintock *et al.* (2013, 2014) that were originally conceived for a different application by Link *et al.* (2010). The new algorithm conditions on the observed data (thus reducing the dimension of the problem) and only proposes updates with non-negative latent encounter history frequencies, $\mathbf{f} = (f_1, f_2, \ldots, f_{5^T})$. Let $\mathbf{r}$ denote the set of $4^T - 2^{T+1} + 1$ indices for latent encounter histories that spawn >1 observed history, and let $f_{j(1)}$ and $f_{j(2)}$ denote the corresponding frequencies for type 1 and type 2 histories that arise from latent encounter history $j \in \mathbf{r}$. Referring back to Table 2 with $T = 2$, $\mathbf{r} = \{4, 8, 9, 12, 14, 16, 17, 18, 19\}$ and one potential update would involve frequencies for latent history '31' (f_{17}) and its progeny '11' ($f_{17(1)} = f_7$) and '20' ($f_{17(2)} = f_{11}$).

When conditioning on the observed encounter histories, the size of the problem is typically greatly reduced because many of the potential latent histories and corresponding moves are not permissable. For example, with $T = 2$, if encounter history '11' was never observed, then $f_7 = f_9 = f_{17} = f_{19} = 0$ can be ignored and $j \in \{9, 17, 19\}$ can be removed from $\mathbf{r}$ for subsequent computations.

Starting from a permissible $\mathbf{f}$ conditional on the observed encounter histories, the algorithm proceeds as follows:

1. Randomly draw a latent encounter history index $r \in \mathbf{r}^*$, where $\mathbf{r}^*$ is the subset of $\mathbf{r}$ with corresponding frequencies that satisfy $\min(f_1, f_j)+\min(f_{j(1)}, f_{j(2)}) > 0$ for $j \in \mathbf{r}$.

2. Randomly draw c_r from the integer set $\left\{-\min(f_1, f_r), \ldots, -1, 1, \ldots, \min(f_{r(1)}, f_{r(2)})\right\}$.

3. Propose $f_r^* = f_r + c_r$, $f_{r(1)}^* = f_{r(1)} - c_r$, and $f_{r(2)}^* = f_{r(2)} - c_r$.

4. Apportion $\mathbf{f}^*$ to individuals following McClintock *et al.* (2014), and accept proposed move based on the Metropolis-Hastings ratio described therein [pp. 2470-2472, steps 9(b)-9(c)].

Note that because `multimark` uses "semi-complete" data likelihoods that condition on the number of unique individuals encountered at least once (n), the dimension of the data-augmented encounter histories (M) described in McClintock *et al.* (2014) is determined by the number of observed encounter histories (i.e. $M = n_1 + n_2 + n_{known}$), such that $f_1 = M - n$ and $n = \sum_{j=2}^{5^T} f_j$. Any additional constraints, such as those resulting from encounter histories being designated as known with certainty using the *known* argument in *processdata()*, are accounted for by simple modifications to steps 1-2.

In terms of mixing, it can sometimes be advantageous to explore more than one move at a time. At each iteration of the chain, the argument *maxnumbasis* specifies how many times to perform steps 1-3 in sequence before evaluating step 4. The default for *multimarkCJS()* and *multimarkClosed()* is *maxnumbasis=1*.

References

Bonner, S.J. & Holmberg, J. (2013) Mark-recapture with multiple, non-invasive marks. *Biometrics*, **69**, 766–775.

Link, W.A., Yoshizaki, J., Bailey, L.L. & Pollock, K.H. (2010) Uncovering a latent multinomial: analysis of mark-recapture data with misidentification. *Biometrics*, **66**, 178–185.

McClintock, B.T., Bailey, L.L., Dreher, B.P. & Link, W.A. (2014) Probit models for capture-recapture data subject to imperfect detection, individual heterogeneity and misidentification. *The Annals of Applied Statistics*, **8**, 2461–2484.

McClintock, B.T., Conn, P.B., Alonso, R.S. & Crooks, K.R. (2013) Integrated modeling of bilateral photo-identification data in mark-recapture analyses. *Ecology*, **94**, 1464–1471.