**Models for Search-Encounter Data**

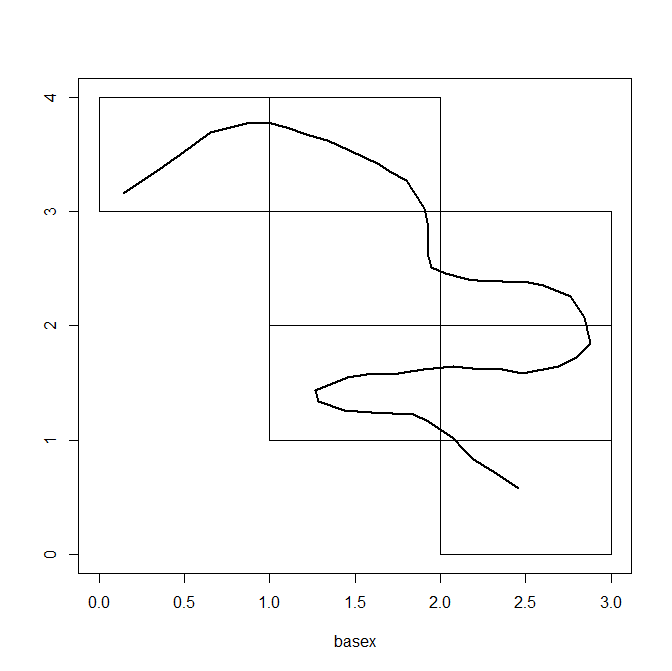
The focus of this chapter should be these search-encounter models. What do we mean by this? Probably better to say “models for search-encounter data”. These are models that arise where you get actual location data of individuals not biased by trap locations but rather by searching space in some fashion. In most cases both detection probability and movement parameters are resolvable. i.e., models that preserve the “movement outcomes” – the u[i,t] variables – the locations. The models we differentiate here depend on a number of things related to data structure or protocol – basically whether or not we record the exact location and how we record it. How exactly are these different from models for data from fixed arrays? (1) sample units are either continuous space polygons or lines, not points; (2) we have location information that is not biased by trap locations (but is biased by the observation device somehow); (3) because we have direct observations of location that exist independent of traps we can often build an explicit model of space usage or an explicit movement model.

Cite Royle and Young, Royle et al. 2011; Efford 2011/2012??

They also provide something of a bridge between the standard trap models and the types of models by Chandler. [where does SCR/DS fit in?]

We define 4 basic sampling designs:

1. Design 1 is when we have a continuous search-path or lines, or multiple such lines, in some region (Fig. 1). This is the type of problem described by Royle et al. (2011 MEE). We assume the path or lines are laid out a priori in some manner that is done independent of the activity centers of individuals and the collection of data does not affect the lines. Sometimes the lines are within well-defined polygons but the polygon boundaries are not meaningful in terms of the observation process. A number of variations of the data collection protocol are possible:
   1. Protocol (1a) has us just record the locations of individuals
   2. Protocol (1b) has us record location of individuals AND location on the transect where we observed the individual
   3. Protocol (1c) has us record neither of those things, instead we record the closest perpendicular distance. This is a typical distance sampling situation which produces exactly a DS type of a model (or a CR-DS model). We don’t recommend recording closest perpendicular distance and we don’t discuss these models too much here
   4. Protocol (1d) . In this case, observations are restricted to the line itself. We imagine that the line is evolving in response to search activity. It is not quite like the other ones so let’s call it “ad hoc”. In this case we use small bins as traps and the length of the line in each grid as a covariate. Thompson et al. and Russell et al.



Analysis of protocols 1a is that of Royle et al. (2011). Analysis of 1b is a distance-sampling like model but with an additional hierarchical structure the describes the individual location scatter about the home range center. This is precisely a type of DS with measurement error. Analysis of 1c is a similar idea except it represents an explicit model misspecification since one is approximating the observation process by the nearest perpendicular to the line. Analysis of 1d is the “unstructured survey data” like from Thompson et al. or Russell et al. Note also that the capcrap paper is a version of this – grids or polygons were sampled but no information on the search path is available. This could be a Design 3 problem but that is excess computation I think.

Hierarchical model for this situation:

u[I,t] ~ bvn( s[i], sigma^2)

Log(h(u[I,t],x)) = beta0 + beta1\*dist(u[I,t],x)

Total hazard is this:

H(u[I,t]) = exp(beta0)\* sum\_{j} exp(beta1\*dist(u[I,t],x[j])

x[j] = point on line

Then: p[I,t] = 1-exp(-H(u[I,t]))

This is the hazard model of Hayes and Buckland 1983 and see Borchers and Efford 2008 also

1. Design 2 is the uniform search intensity model from Royle and Young (2008). In this case we have a well defined search area, such as a quadrat or a transect, and we imagine that the area is uniformly searched so that p = p0 constant within the search area. Sampling produces locations of individuals within well-defined boundaries. The polygon of search is important because it tells us that p=0 by design outside of the boundary.

Using the example from the Figure above, we could imagine that each quadrat was uniformly searched. The individual quadrat boundaries are irrelevant and we only need to be concerned about the “total” boundary of the composite polygon (the intersection of all little ones). That said for analysis in BUGS it is easier to work with square polygons.

We show a simulation example here and we analyze it either using a bivariate normal movement model or else a 2-d random walk type of model. But we don’t provide a real example as Royle and Dorazio 2008 did a reanalysis of the lizard data and see also Efford (XXXX).

1. Design 3 is based on a bad implementation of Design 1 or 2. We set up search polygons (e.g., the grid cells of above figure) and record locations of encountered individuals but we do not do a uniform search of quadrats and we forgot to record the GPS path.

Analysis of this design: We imagine that we can assume a uniform search intensity here and maybe it won’t be so bad. We should do a simulation study of this somehow. I am working on methods to lay down some standard sets of lines for simulating data, and then ignoring the lines in doing an analysis.

Alternative 2 for analysis: We could map each location to the CENTER of the grid cell and pretend this is trap array (traps at center of each grid). This was the idea of Kery et al. and some other papers.

1. Design 4: In this case we screw up even further and forget to record the locations of individuals within a bunch of quadrats. I believe Richard has been thinking along these lines – using the underlying movement model as a latent model.
   1. Protocol (4a) – we imagine that you could have counts BY individual identity within each quadrat. Not sure what analysis model this would be
   2. Protocol (4b) – we DON’T have individual identities but just total counts. This is Chandler and Royle (2012/13).

For the book chapter I think my plan is to lay out these various data types or designs (a typology of models I guess), provide some simulation code, and a few limited analyses (not of every type) either of simulated data or real examples. Perhaps the MHB example from the MEE paper and the lizard data from Royle and Young and possibly the Fisher data from Thompson et al. or lion data from Robin Russell. That is too much material.

Example 1: Analysis of the Swiss MHB survey using Design 1

Example 1b: Lizard data. No need to analyze this as it was done in RD book. Mention polygon detectors in secr.

Example 2: Fisher data possibly – lion data or --- or Capricaillie data?

Inference for Search-Encounter Designs

We cover the basic Design 1 here which also is relevant to Design 2 as a special case.

Analysis of protocols 1a is that of Royle et al. (2011).

The model:

##

#u[I,t] ~ bvn( s[i], sigma^2)

#Log(h(u[I,t],x)) = beta0 + beta1\*dist(u[I,t],x)

#Total hazard is this:

#H(u[I,t]) = exp(beta0)\* sum\_{j} exp(beta1\*dist(u[I,t],x[j])

#x[j] = point on line

#Then: p[I,t] = 1-exp(-H(u[I,t]))

1.2. Example: Simulator and WinBUGS code for this example [in repo]

1.3. Hard quadrat boundaries: Quadrat boundaries might be relevant or might not be. If they are then, for Bayesian analysis, a value of u outside the boundary has p forced to 0, its as simple as that.

In general, we define p[I,t] = p[I,t]\*I(u in X).

We see how this relates to the uniform search intensity model. P[I,t] = p0 then defines precisely the model of Royle and Young (2008).

1.4. Analysis of other protocols of Design 1:

Analysis of 1b is a distance-sampling like model but with an additional hierarchical structure the describes the individual location scatter about the home range center. This is precisely a type of DS with measurement error. Analysis of 1c is a similar idea except it represents an explicit model misspecification since one is approximating the observation process by the nearest perpendicular to the line. Analysis of 1d is the “unstructured survey data” like from Thompson et al. or Russell et al. Note also that the capcrap paper is a version of this – grids or polygons were sampled but no information on the search path is available. This could be a Design 3 problem but that is excess computation I think.

Protocol (1b) has us record location of individuals AND location on the transect where we observed the individual. This is an easier problem I think, but you have to account for “not seen” prior to x0 so maybe some kind of a cumulative hazard model or something.

Protocol (1c) has us record neither of those things, instead we record the closest perpendicular distance. This is a typical distance sampling situation which produces exactly a DS type of a model (or a CR-DS model). We don’t recommend recording closest perpendicular distance and we don’t discuss these models too much here

Protocol (1d) . In this case, observations are restricted to the line itself. We imagine that the line is evolving in response to search activity. It is not quite like the other ones so let’s call it “ad hoc”. In this case we use small bins as traps and the length of the line in each grid as a covariate. Thompson et al. and Russell et al.

Simulation results -----------------

|

> wbout

Inference for Bugs model at "model0.txt", fit using jags,

3 chains, each with 5000 iterations (first 1000 discarded)

n.sims = 12000 iterations saved

mu.vect sd.vect 2.5% 25% 50% 75% 97.5% Rhat n.eff

N 94.448 5.237 81.000 92.000 96.000 98.000 100.000 1.006 520

beta0 -0.539 0.743 -1.714 -1.072 -0.637 -0.102 1.258 1.042 54

beta1 -11.943 2.196 -17.229 -13.236 -11.665 -10.375 -8.378 1.035 62

psi 0.936 0.056 0.792 0.908 0.951 0.979 0.998 1.005 650

sigma 0.340 0.043 0.268 0.309 0.337 0.366 0.436 1.004 820

deviance 206.987 25.474 160.405 189.069 205.779 224.080 259.491 1.017 130

For each parameter, n.eff is a crude measure of effective sample size,

and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)

pD = 319.4 and DIC = 526.3

DIC is an estimate of expected predictive error (lower deviance is better).

>

Section: Design 3: Ad hoc implementation of Design 1.

We don’t do anything new in terms of modeling here but we look at how bad do we do if we don’t have the search path and use the USI model? We consider 4 cases: Case 1 and 2: regular searching of low and high intensity. E.g., for a 1 unit block, then we can have a sinusoid track through each block of length 1.5 or 2 and then 4 or 5 km. For case 3 and 4 we use heterogeneity in search intensity.

Section: Design 4 – no location info

1. Imagine we have a bunch of quadrats or segments that are contiguous and we do the surveys like above and record counts PER individual but no other sampling information.
2. Design 4b: We don’t record individual ID.

Basic concept is that correlation among quadrats is induced. Like Chandler and Royle….. This ties us to that paper.