**SAMPLING FOR SPATIAL CAPTURE-RECAPTURE STUDIES**

Many biologists have experience with the design of natural resource surveys, a key feature of which involves sampling space. That is, we identify a sample frame comprised of spatial units and we sample randomly (or by some other method, such as GRTS) those units and measure some attribute. The resulting inference applies to the attribute of the sample frame. There are some distinct aspects of the design of SCR studies which many people struggle with in their attempts to reconcile SCR design with classical survey design problems.

The main thing is that SCR models clearly involve a type of spatial sampling. However, unlike conventional natural resource sampling the attribute we measure is not relevant to the sample location and, indeed, the sample frame is not comprised of spatial units. Instead, the sample frame is a list of N individuals where N is unknown. In order to make this a well-defined sample frame we have to prescribe a spatial region within which those N individuals are located - we’ll call this the “study area” which we will think of (even define) as a formal subset of the state-space we use in SCR studies.

The relevance of spatial sampling in SCR studies is as a device for accumulating individuals in the sample. That is, we’re not interested in any sample unit attribute directly but, rather, we use spatial units as a means for sampling individuals and obtaining individual level encounter histories that indicate the different sample locations at which each individual is encountered. It makes sense in this context that we should want to choose a set of sample units that provides an adequate sample size. But we need to be careful because arbitrary selection of sample units has a side-effect that it induces unequal probabilities of inclusion into the sample.

The fact that we have unequal probabilities of sampling is consistent with the classical sampling ideas of Horvitz-Thompson estimation (see http://en.wikipedia.org/wiki/Horvitz%E2%80%93Thompson\_estimator) which has motivated capture-recapture models similar to SCR (Huggins 1989; Alho 1990). In the H-T framework, the sample inclusion probabilities are fixed, known, even prescribed a priori. However, in all real animal sampling problems they are unknown because we never know precisely the area over which each individual lives and therefore cannot characterize its availability to capture. Therefore, we have to estimate the sampling inclusion probabilities using a model. This was the idea of Huggins (1989) and Alho (1990). The basic SCR model that achieves this objective (i.e., a characterization of individual-specific encounter probabilities) is described elsewhere. In formalizing this model for individual encounter probability, in terms of where individuals live and trap locations, allows for the estimation of the total size of the sample frame, N, which is the potential number of individuals that exist in the study area.

Operationally we proceed as follows: We define our study area a priori. This is the region where the animals live that we care about. As a conceptual device, we might think of this as the area that, given an infinite amount of resources, we might wall-off so that we can study a real closed population. This “study area” should exist independent of any model or estimator of some population quantity. i.e., the subject-matter context should determine what is the study area. Given a well-defined study area, we use some method to arrange data collecting devices (“detectors” is the term from Efford XXXX) within this study area. The method of arrangement can be completely arbitrary but, naturally, we want to choose arrangements of traps that produce higher values of n (number of sampled units) and better precision for estimating the parameters of the SCR model.

As a practical matter, some animals within the study area might have vanishingly small probability of being included in the sample. i.e., p = 0. This is historically described as their being “holes” in the state-space. Are these guys included in our sample frame, or not? My thinking is this: The individuals in the study area are potentially included in the sample, prior to the distribution of traps, as long as they have some non-zero probability of having a trap located in their vicinity. Therefore, they are elements of the sample frame. This is a design-based argument but we don’t evaluate estimators by that consideration instead using the model from which we define the sample inclusion parameters.

The other twist on this is that once we lay out our sample locations we can wind up having individuals exposed to sampling that live outside of our study area. I don’t see a problem with this for any possible inference objective.

The other way to think about having p = 0 individuals is that extending inference from the sample to individuals that live in these "holes" represents an extrapolation (prediction of the model outside the range of the data). That is, we're saying our model applies to any area within the study area even to areas where we can't capture guys because we happened to not put a trap there. So this probably makes sense as long as holes are small relative to the state-space, or that contain few individuals relative to the rest of the state-space, and/or that is homogeneous with respect to the state-space.

Random sampling: Suppose we take a random sample of locations within the study-area for setting up our sampling devices. In a design-based sense, this forces the probability of sample inclusion to be equal for every individual. This is *unconditional* on the trap locations, however. Conditional on the realized trap locations, individuals do not have a constant sample inclusion probability. This is not to say that random sampling is lacking in virtue. Clearly if we were measuring a trap-specific variable then random sampling would be reasonable.

Reconciling random sampling of space vs. random sampling of individuals. Define a quantity y(x) = number of encounters at site x.

Misc comments –

Moreover, the propertieis of SCR estimators are not a function of the spatial sampling mechanism in any formal way (rather, they are purely model-based) and so sampling in SCR is not really like classical survey sampling.

As such, the spatial sampling inherent in SCR should more properly be called “spatial data collection” although we stick with the word sampling just out of convention.