Natural Resources Data Analysis – Lecture Notes
Brian R. Mitchell

VII. Week 7:

A. Goodness of Fit Testing – Mark-Recapture models

1. ***Testing goodness of fit gets difficult*** as models become more complex. Mark-recapture models are especially resistant to simple tests of fit. Evan Cooch's "Using MARK – A Gentle Introduction" contains a good introduction to goodness of fit testing in Chapter 5. The book is available for download at http://www.phidot.org/software/mark/docs/book/.

2. In most cases, ***even mark-recapture data that meets the model assumptions will have a high $\hat{c}$***, because the deviance for mark-recapture models is not normally χ^2 distributed. It then becomes important to look at how the model $\hat{c}$ compares to the $\hat{c}$ you would expect for a model with data that meets all assumptions. Model $\hat{c}$ divided by the expected $\hat{c}$ is your true estimate of model fit, and this value (calculated for your most general model) is the value you should use for QAIC and $QAIC_c$.

3. When can ***statistical methods*** be used?

   a) Certain models can be tested using statistical methods and existing computer programs (e.g. RELEASE and U-CARE).

   b) ***No individual covariates***

   c) ***CJS and multi-strata models*** that are 1) fully time dependent, 2) transience models (two age classes for survival), or 3) immediate trap-dependence models.

   d) Note that these programs ***assume*** that the deviance is χ^2 distributed, which is generally not true for mark-recapture data. These programs also typically ***lack statistical power*** because of the amount of data pooling required.

   e) See Evan Cooch's Chapter 5 if you are interested in these methods.

4. Some models can be tested using MARK's median $\hat{c}$ or bootstrapped GOF simulations.

   a) ***No individual covariates***

   b) ***Known fate, live recapture, and recovery models*** only

   c) ***CJS models should use the median $\hat{c}$ simulation***, other models can use bootstrapped GOF. The bootstrap is biased for CJS models.

d) How does MARK's ***bootstrapped GOF*** work [demonstrate on Terri's CJS spreadsheet]?

(1) ***Estimate*** your data set. For a CJS model, you will get estimates of ϕ (survival) and p (chance of resighting, given survival). You also need to know the number of animals you marked.
(2) ***Simulate*** a data set based on these parameters. Start with the first individual, and generate as many random numbers as you have parameters.
(3) Create a ***history*** for the individual based on the parameter estimates and random numbers.
(4) ***Repeat*** (2) and (3) until you reach the number of released individuals.
(5) Summarize the capture histories, and ***fit the model***. Estimate the deviance and $\hat{c}$ from the model.
(6) ***Repeat*** (2) – (5) the desired number of times (usually about 100 or 1,000).

e) How do you ***run a simulation in MARK***?

(1) Fit your model.
(2) Select Tests… Bootstrap GOF.
(3) Choose your summary variables, results location, number of simulations, and random number seed.
(4) After the simulation finishes, look at the results using Simulations… View Simulation Results.

f) ***How do you use the results*** from the bootstrap simulation?

(1) The p-value for a ***bootstrapped GOF test*** is the proportion of bootstraps that exceeds the deviance for the actual model (so if 12 models in 100 had a higher deviance, the bootstrapped GOF probability is 0.12).
(2) ***Estimate $\hat{c}$*** as the original model deviance divided by the mean bootstrap deviance.
(3) Estimate $\hat{c}$ as the original model $\hat{c}$ divided by the mean bootstrapped $\hat{c}$.
(4) Use the larger value of $\hat{c}$ obtained using (2) and (3).
(5) Note that the two methods of estimating $\hat{c}$ will be very similar unless your data is sparse, in which case the number of estimable parameters (and hence the degrees of freedom) will vary, causing the two methods to give different results.

g) What does the ***median $\hat{c}$*** test do?

(1) For a range of true $\hat{c}$ values, MARK simulates multiple data sets and estimates $\hat{c}$. Then MARK plugs these values into a logistic regression, and estimates the value of the true $\hat{c}$ at which the observed $\hat{c}$ would be the median value.

(2) This procedure is a bit mysterious, and is ***not really documented*** anywhere other than the cryptic entry in the MARK help files.
(3) I am not sure what procedure is used to simulate a population with a known ĉ; I suppose this can be done by forcing some dependence between individuals, but I don't know what level of dependence would be needed to generate a specific value for ĉ.
(4) If you are using this procedure, it is important to ***use a large number of replicates*** at each ĉ value; with small numbers of replicates the results can be unpredictable. I recommend at least 50 replicates at each point. You should probably conduct at least 1,000 total simulations.
(5) Gary White suggests doing a small number of replicates over a large range of ĉ values to start with, and then zeroing in with more intensive simulations near the expected final ĉ value.

5. ***Otherwise***…

a) Look at the ***estimate of ĉ for your global model***. Note that ĉ is currently not estimated correctly for closed capture or robust designs.

b) ***Examine plot of model residuals***. Gary White has told me that he does not think that the residuals are useful for examining goodness of fit for models with individual covariates. This function also does not work properly in MARK because the expected values are calculated incorrectly. I originally thought that the expected values needed to be divided by the number of observations per capture history. If that is the case, then the following directions will generate a residual plot.

(1) To examine residuals, you need to ***export your residuals*** to Excel (Output… Specific Model Output… Residuals… List Observed and Expected in Excel).
(2) Add a new column, and ***create corrected expected value*** by dividing the expected value by the number of records with the same capture history.
(a) Sort the data by capture history.
(b) Create a list of possible capture histories and the number of time each occurred; make sure this is also sorted.
(c) Use the LOOKUP function in Excel to divide the reported expected value by the number of observations.
(3) Calculate the ***Pearson residual***:

$$\frac{\text{observed} - \text{expected}}{\sqrt{\text{expected}}}$$

(4) ***Plot the residuals*** on a scatterplot.
(5) I was unable to find the proper formula for calculating deviance residuals.

c) ***Unfortunately***, I think that the expected values in MARK are wrong in some other way than just the number of observations per capture history. This is because there are odd patterns visible for residuals of simulated data. For fun, order your input file by capture history, and then fit the model and plot the residuals.

6. ***Beyond MARK?*** Bootstrap GOF for individual covariates and closed capture models.

a) ***Individual Covariates***

(1) I'm not sure why MARK does not include the possibility of calculating a bootstrapped GOF and $\hat{c}$ for models with individual covariates; my understanding is that this procedure is essentially the same as the procedure without covariates, with a slight twist and a heavier computational load.
(2) There may be some ***theoretical issues*** that I am not aware of; my impression is that the bootstrapped GOF is probably biased low in some way. ***My response*** to this is that I would rather have a biased GOF (in which case if the fit is poor or borderline I know that my model doesn't fit) than no test at all, and no way to estimate $\hat{c}$.
(3) The heavier computational load is because the simulation must consider each individual, rather than a summary of captures.
(4) The first part of the twist is that at least one mark-recapture parameter (i.e. ϕ or p) must be altered based on the value of the individual covariate (or covariates) via a logistic link.
(5) The second part of the twist is that the covariate values used in the simulation should be sampled (with replacement) from the covariate values in the actual data set. This is standard practice for bootstrapping procedures.
(6) Luckily for me, Terri Donovan has created a spreadsheet that takes care of simulating the appropriate data; I just needed to write a macro to take care of the repeated simulations.
(7) When generating the bootstrapped data set for multiple covariates, the appropriate procedure is to select a row of data (a set of covariate measurements) rather than independently selecting covariates.

b) ***Closed Capture Models***

(1) Again, the reason why MARK doesn't implement bootstrap GOF appears to be that ***the method is biased*** in some way, and Gary White has not conducted the simulations to determine the amount of bias to his satisfaction. Gary and Evan Cooch seem to feel that a biased estimate of $\hat{c}$ could be worse than no estimate of $\hat{c}$. I'm not sure that I agree, but you should be aware that

bootstrapped GOF tests are not generally done for this type of model.
(2) Part of the problem is that estimating the ***deviance*** of a closed capture model without individual covariates is mathematically complicated; the deviance calculation requires knowing the -2ln(L) of the saturated model (because D = -2ln$(L)_{fit}$ minus -2ln$(L)_{sat}$), which contains a term equal to ln(n!), where n is the number of individuals captured. Note that for models with individual covariates, the saturated likelihood is zero, so the deviance is just -2ln$(L)_{fit}$.
(3) There is ***no technical reason that prevents constructing a bootstrap GOF*** for a closed capture model; the concept is straightforward as long as you come up with a way to evaluate the ln(n!) function. Excel can't do this for numbers greater than about 170, so I resorted to using Mathematica to create a lookup table for my Excel calculations (in a Mathematica window, type: "N[log[n!],x]", where n = the number and x = decimal places of the result).
(4) Another consideration is that the population size should be allowed to vary for each simulation, according to the estimated N and the standard deviation of this estimate.
(5) In addition, closed capture GOF simulation (as well as simulations for models with individual covariates) might be less biased if the actual parameter estimates (ϕ and p) were also allowed to vary for each population according to their estimated standard deviations.
(6) My initial attempt at bootstrap simulation of a closed capture model did not work well; so there are clearly some bugs to be worked out. I may not be calculating deviance correctly (at least my calculations do not agree with MARK's), so that may be a big part of the problem.

B. Presentations

C. Class Evaluations