Work Book

This document lists a number of exercises on solving various issues related to estimating selectivity from single hauls.

The exercises make use of the "SOLVER" function of the EXCEL spread-sheet. This function basically implements a general minimizer (or optimizer); i.e. it minimizes (or maximizes) the value of a target cell, with respect to some specified dependent cells.

In the context of estimating selectivity, the general approach is to define the log-likelihood function. This is described as sum over all length-classes and over all compartments of some values that depend on

- the observed data: number of measurements by length and by compartment and potentially the sub-sampling ratios
- the unknown parameters: e.g. (α, β) or $(L_{50\%}, SR)$

Exercise 1:
Consider the log-likelihood function for a covered codend experiment with no (or equal) sub-sampling:

$$l(\pi_\ell; n_{\ell,\text{codend}}, n_{\ell,\text{cover}}) = \sum_\ell n_{\ell,\text{codend}} \cdot \log(r(\ell;\alpha,\beta)) + n_{\ell,\text{cover}} \cdot \log(1 - r(\ell;\alpha,\beta))$$

If r is the logistic function this simplifies to

$$l(\pi_\ell; n_{\ell,\text{codend}}, n_{\ell,\text{cover}}) = \sum_\ell n_{\ell,\text{codend}} \cdot (\alpha + \beta \cdot \ell) - n_{\ell,+} \cdot \log(1 + \exp(\alpha + \beta \cdot \ell))$$

Now let us take the "Clark-20-1" data set and do the following steps:

1. Open a new spread sheet in Excel and save it as "CovCodend-1.xls"
2. Import the catch data from the "Clark-20-1.txt." text file into cell A13 and downwards and transform the text to data.
3. Label cell A12, B12, C12 and D12 as "Length", "Codend", "Cover" and "Total" respectively
4. Make D13 the sum of B12 and C12 and copy this cell throughout the range of length classes
5. Remove all rows with zero total catch
6. Allocate the cells B10 and C10 to hold the α and β parameters respectively and enter some numbers; f.x. $\alpha = -3$ and $\beta = 0.5$ (Note $\alpha < 0$ and $\beta > 0$)
7. Label cell E12 "Eta", make E13=$B$10+$C$10*A13 (i.e. the linear predictor $\alpha + \beta \cdot \ell$) and copy this cell throughout the range of length classes. Note the absolute references to the α and β cells.
8. Label cell F12 "r(l)", make F13=1/(1+EXP(E13) (i.e. the retention probability for the first length-class) and copy this cell throughout the range of length classes
9. Label cell G12 "LL", make G13= B13*E13+D13*LN(1-F13) (i.e. the log-likelihood contribution for the first length-class) and copy this cell throughout the range of length classes

10. Make cell B11 the sum of all the log-likelihood values calculated in column G
11. Apply the solver to maximize the log-likelihood value in B11 (target cell) by varying the values of B10 and C10 (α and β)

Exercise 2:

Add a column that produces the *observed* proportions $\frac{n_{\ell,\text{codend}}}{n_{\ell,\text{codend}} + n_{\ell,\text{cover}}}$ and produce a plot that draws the *expected* proportions $r(\ell)$ as a curve and the observed proportions as points. Does the fit seem reasonable?

Exercise 3:

Amend the spread sheet developed in exercise 1 to

- calculate the selectivity parameters (*L50,SR)* (See the formula in the table in the Single Haul 1 document)
- calculate deviance residuals (See the Single Haul 2 document)
- Produce a plot to illustrate the residuals. Are there any problems concerning high residuals and patterns?
- Use the deviance residuals to calculate the deviance statistic and the dof. Refer the deviance statistic to $\chi^2(dof)$ distribution to calculate the p-value of the fit. What is the conclusion?

Exercise 4:

The selectivity parameters (*L50,SR)* are less correlated than the generic parameters (α, β). These provide therefore a numerical more stable estimation (less prone to lack of convergence). Reconstruct the spread sheet to estimate the selectivity using a *selectivity* parametrisation rather than a *generic* parametrisation for the logit selectivity curve. (See the Single Haul 2 document).

Exercise 5:

Reconstruct the spread sheet to estimate the selectivity using a C-Log-Log curve instead of the Logit curve. Are there any differences in

- parameter estimates?
- goodness of fit?

Exercise 6:

When an estimation results in a poor fit it can be the model which is inadequate in describing the variation in the data or it can be a problem with the data.

The text file Clark-60-Total.txt contains so-called pooled data. These are data aggregated from the hauls represented in Clark-60-1.txt, Clark-60-2.txt and Clark-60-3.txt.
Analyze these data. Comment on the fit. Is there a problem with

- the model?

- the data?

Exercise 7:

In the data considered so far all fish caught have been measured. The catch in one or more compartments are however often too big to allow for measurements of all fish. In these cases the statistical analysis is based on a sub-sample of the total catch. There are two approaches to deal with sub-sample catches:

- Data-modification: The measured sub-sample is raised (i.e. scaled by the inverse of the sampling ratios).
- Model-Modification: The sampling ratios are treated as effort parameters.

See the Single Haul 2 document for more details.

Implement the analysis of sub-sampled data for the covered codend case using the data listed in the subsampled.txt file and using the logit selectivity curve with a selectivity parameterisation.

The text file raised.txt contains the corresponding raised data. Analyse these and compare the results with the previous results. What are the differences and how do you explain it?

Exercise 8:

Use the spreadsheet from exercise 1 to calculate the variance estimates of (α, β). You can find details in the ICES manual p.45.
Next convert these to variance estimates for the generic parameters (*L50,SR)* using the delta-theorem (see the ICES manual p.46).
Now amend the spread sheet from exercise 4 (which uses the same data) to produce variance estimates of (*L50,SR)* directly, using the formulas in the Single Haul 2 document.
How do the two sets of variances compare ?

Exercise 9:

Raw data of sub-sampled catches are not susceptible for demonstrating the observed proportions retained in the codend. For that purpose we will instead use raised data.

Compare the observed proportions in the data from exercise 7 with the estimated selectivity based on the model-modification approach. How do you explain the difference?

Exercise 10:

In some cases the catch is divided into two or more categories (e.g. "small" and "large") prior to the sub-sampling. This may lead to different sub-sampling ratios for different length-classes.

Analyse the data in the had-8.txt file using a length based sub-sampling model.

Exercise 11:

Produce a spread-sheet that estimates the selectivity of a paired gear experiment using the data from the TT_1.txt file.
Amend the spread sheet to produce graphics of the selectivity curve and the observed proportions. How do they compare?. What is wrong here?

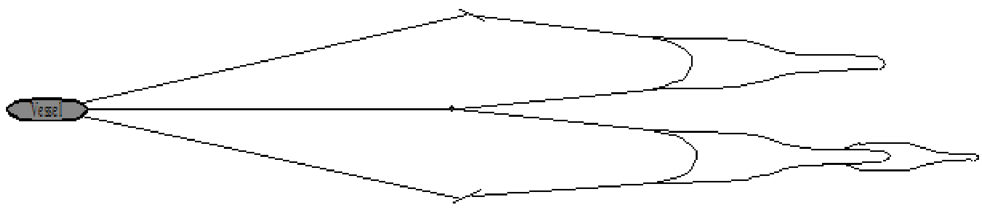
Amend the spread sheet to produce deviance residuals, the deviance statistic and *dof.*

Calculate the variance estimates using the formulas from the ICES manual (p.45)

Exercise 12:

Many new experiments with fishing gear aims at testing more sophisticated gear with multiple selective devices. The complexity of these devices obviously propagates to the analytical aspect of the research. Luckily the SELECT method is general and versatile approach that can accommodate such designs.

In this exercise we will consider a twin trawl design with two identical (standard) codends, one of which is covered with a kite-cover



The catch is consequently collected in three compartments:

I: Codend without cover
II: Codend with cover
III: Cover

Compartment II and III constitutes a regular covered codend experiment (ignoring the catch in compartment I). Similarly the experiment can be considered as a regular twin trawl experiment by aggregating the catch in compartment II and III (This compares to the small mesh codend) and comparing that with the catch in compartment I. This way it is possible to obtain estimates of the selectivity of the two identical codends. The disadvantage of this approach is however that it does not allow for assessing the interaction between the two sets. This can be crucial when dealing with multiple hauls.

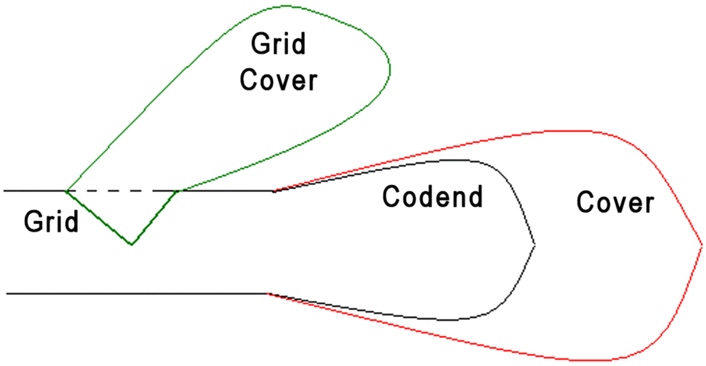
It is however possible obtain simultaneous estimates of the two sets of selectivity parameters along with the split parameter. This model carries a total of 5 parameters.

Construct a spread sheet that accomplish the simultaneous estimation.

Hint: Start by assigning theoretical retentions to each of the three compartments. Use this to calculate the expected proportions caught in each of the three compartments. Plug in these proportions in the log-likelihood function.

Exercise 13:

The figure below presents an experimental gear with two selective devices in it: a grid and a codend.



Describe the model for this design.

How would you assess the effective selectivity of this gear; i.e. the probability for a length ℓ fish to be retained by the gear, given it has entered it ?.