Summer School in Quantitative Fisheries Stock Assessment

Day 4: An introduction to Statistical Catch-at-Age models

Cóilín Minto, Chato Osio, Alessandro Orio, Alessandro Ligas

July 20, 2017

Outline

Goal

Build a fully statistical model to estimate the parameters of the stock assessment

Cohort

Cohort progression

$$
\begin{array}{cccccc}
 & \multicolumn{5}{c}{\text{Age} \longrightarrow} \\
 & N_{a1,t1} & N_{a2,t1} & N_{a3,t1} & N_{a4,t1} & N_{a5,t1} \\
 & N_{a1,t2} & N_{a2,t2} & N_{a3,t2} & N_{a4,t2} & N_{a5,t2} \\
\text{Time} \longrightarrow & N_{a1,t3} & N_{a2,t3} & N_{a3,t3} & N_{a4,t3} & N_{a5,t3} \\
 & N_{a1,t4} & N_{a2,t4} & N_{a3,t4} & N_{a4,t4} & N_{a5,t4} \\
 & N_{a1,t5} & N_{a2,t5} & N_{a3,t5} & N_{a4,t5} & N_{a5,t5}
\end{array}
$$

Cohort

Cohort progression

$$\begin{array}{c|ccccc}
 & \multicolumn{5}{c}{\text{Age} \longrightarrow} \\
\hline
 & N_{a1,t1} & N_{a2,t1} & N_{a3,t1} & N_{a4,t1} & N_{a5,t1} \\
 & N_{a1,t2} & N_{a2,t2} & N_{a3,t2} & N_{a4,t2} & N_{a5,t2} \\
\text{Time} \downarrow & N_{a1,t3} & N_{a2,t3} & N_{a3,t3} & N_{a4,t3} & N_{a5,t3} \\
 & N_{a1,t4} & N_{a2,t4} & N_{a3,t4} & N_{a4,t4} & N_{a5,t4} \\
 & N_{a1,t5} & N_{a2,t5} & N_{a3,t5} & N_{a4,t5} & N_{a5,t5}
\end{array}$$

Cohort

Cohort progression

$$
\begin{array}{cccccc}
 & & & \text{Age} \longrightarrow & & \\
 & N_{a1,t1} & N_{a2,t1} & N_{a3,t1} & N_{a4,t1} & N_{a5,t1} \\
 & N_{a1,t2} & N_{a2,t2} & N_{a3,t2} & N_{a4,t2} & N_{a5,t2} \\
\text{Time} \downarrow & N_{a1,t3} & N_{a2,t3} & N_{a3,t3} & N_{a4,t3} & N_{a5,t3} \\
 & N_{a1,t4} & N_{a2,t4} & N_{a3,t4} & N_{a4,t4} & N_{a5,t4} \\
 & N_{a1,t5} & N_{a2,t5} & N_{a3,t5} & N_{a4,t5} & N_{a5,t5}
\end{array}
$$

$$N_{a+1,t+1} = N_{a,t}e^{-(F_{a,t}+M_{a,t})}$$

Starting a cohort

But how do we start them?

$$
\begin{array}{c|ccccc}
 & \multicolumn{5}{c}{\text{Age} \longrightarrow} \\
\hline
 & N_{a1,t1} & N_{a2,t1} & N_{a3,t1} & N_{a4,t1} & N_{a5,t1} \\
 & N_{a1,t2} & N_{a2,t2} & N_{a3,t2} & N_{a4,t2} & N_{a5,t2} \\
\text{Time} \downarrow & N_{a1,t3} & N_{a2,t3} & N_{a3,t3} & N_{a4,t3} & N_{a5,t3} \\
 & N_{a1,t4} & N_{a2,t4} & N_{a3,t4} & N_{a4,t4} & N_{a5,t4} \\
 & N_{a1,t5} & N_{a2,t5} & N_{a3,t5} & N_{a4,t5} & N_{a5,t5}
\end{array}
$$

Year one

We also need to start all ages in the first year

$$
\begin{array}{cc}
 & \text{Age} \longrightarrow \\
\begin{array}{c}\text{Time} \\ \downarrow\end{array} &
\begin{array}{ccccc}
N_{a1,t1} & N_{a2,t1} & N_{a3,t1} & N_{a4,t1} & N_{a5,t1} \\
N_{a1,t2} & N_{a2,t2} & N_{a3,t2} & N_{a4,t2} & N_{a5,t2} \\
N_{a1,t3} & N_{a2,t3} & N_{a3,t3} & N_{a4,t3} & N_{a5,t3} \\
N_{a1,t4} & N_{a2,t4} & N_{a3,t4} & N_{a4,t4} & N_{a5,t4} \\
N_{a1,t5} & N_{a2,t5} & N_{a3,t5} & N_{a4,t5} & N_{a5,t5}
\end{array}
\end{array}
$$

Year one

We also need to start all ages in the first year

$$
\begin{array}{c|ccccc}
 & \multicolumn{5}{c}{\text{Age} \longrightarrow} \\
\hline
\text{Time} \downarrow & N_{a1,t1} & N_{a2,t1} & N_{a3,t1} & N_{a4,t1} & N_{a5,t1} \\
 & N_{a1,t2} & N_{a2,t2} & N_{a3,t2} & N_{a4,t2} & N_{a5,t2} \\
 & N_{a1,t3} & N_{a2,t3} & N_{a3,t3} & N_{a4,t3} & N_{a5,t3} \\
 & N_{a1,t4} & N_{a2,t4} & N_{a3,t4} & N_{a4,t4} & N_{a5,t4} \\
 & N_{a1,t5} & N_{a2,t5} & N_{a3,t5} & N_{a4,t5} & N_{a5,t5}
\end{array}
$$

How about F?

Age

Time $\downarrow$

$F_{a1,t1}$	$F_{a2,t1}$	$F_{a3,t1}$	$F_{a4,t1}$	$F_{a5,t1}$
$F_{a1,t2}$	$F_{a2,t2}$	$F_{a3,t2}$	$F_{a4,t2}$	$F_{a5,t2}$
$F_{a1,t3}$	$F_{a2,t3}$	$F_{a3,t3}$	$F_{a4,t3}$	$F_{a5,t3}$
$F_{a1,t4}$	$F_{a2,t4}$	$F_{a3,t4}$	$F_{a4,t4}$	$F_{a5,t4}$
$F_{a1,t5}$	$F_{a2,t5}$	$F_{a3,t5}$	$F_{a4,t5}$	$F_{a5,t5}$

How about F?

Age

Time $\longrightarrow$

$F_{a1,t1}$	$F_{a2,t1}$	$F_{a3,t1}$	$F_{a4,t1}$	$F_{a5,t1}$
$F_{a1,t2}$	$F_{a2,t2}$	$F_{a3,t2}$	$F_{a4,t2}$	$F_{a5,t2}$
$F_{a1,t3}$	$F_{a2,t3}$	$F_{a3,t3}$	$F_{a4,t3}$	$F_{a5,t3}$
$F_{a1,t4}$	$F_{a2,t4}$	$F_{a3,t4}$	$F_{a4,t4}$	$F_{a5,t4}$
$F_{a1,t5}$	$F_{a2,t5}$	$F_{a3,t5}$	$F_{a4,t5}$	$F_{a5,t5}$

That's a lot of parameters!

How about F?

Need to conserve degrees of freedom in the model but also need the **F** matrix. One approach is a separable model

$$\mathbf{F} = \begin{pmatrix} s_{a1} & s_{a2} & s_{a3} & s_{a4} & s_{a5} \\ s_{a1} & s_{a2} & s_{a3} & s_{a4} & s_{a5} \\ s_{a1} & s_{a2} & s_{a3} & s_{a4} & s_{a5} \\ s_{a1} & s_{a2} & s_{a3} & s_{a4} & s_{a5} \\ s_{a1} & s_{a2} & s_{a3} & s_{a4} & s_{a5} \end{pmatrix} \times \begin{pmatrix} f_{t1} & f_{t1} & f_{t1} & f_{t1} & f_{t1} \\ f_{t2} & f_{t2} & f_{t2} & f_{t2} & f_{t2} \\ f_{t3} & f_{t3} & f_{t3} & f_{t3} & f_{t3} \\ f_{t4} & f_{t4} & f_{t4} & f_{t4} & f_{t4} \\ f_{t5} & f_{t5} & f_{t5} & f_{t5} & f_{t5} \end{pmatrix}$$

How about F?

Need to conserve degrees of freedom in the model but also need the **F** matrix. One approach is a separable model

$$\mathbf{F} = \begin{pmatrix} s_{a1}f_{t1} & s_{a2}f_{t1} & s_{a3}f_{t1} & s_{a4}f_{t1} & s_{a5}f_{t1} \\ s_{a1}f_{t2} & s_{a2}f_{t2} & s_{a3}f_{t2} & s_{a4}f_{t2} & s_{a5}f_{t2} \\ s_{a1}f_{t3} & s_{a2}f_{t3} & s_{a3}f_{t3} & s_{a4}f_{t3} & s_{a5}f_{t3} \\ s_{a1}f_{t4} & s_{a2}f_{t4} & s_{a3}f_{t4} & s_{a4}f_{t4} & s_{a5}f_{t4} \\ s_{a1}f_{t5} & s_{a2}f_{t5} & s_{a3}f_{t5} & s_{a4}f_{t5} & s_{a5}f_{t5} \end{pmatrix}$$

Separable model

So the parameters we have in the model are:

- $N_{1,t=1\ldots T}$ number of recruits
- $N_{a=2\ldots A,t=1}$ numbers in first year
- $s_{a=1\ldots A}$ exploitation pattern (age effects)
- $f_{t=1\ldots T}$ fishing mortality level (year effects)

Predictions

From this set of parameters we can predict $\hat{\mathbf{F}}$ and $\hat{\mathbf{N}}$ from which we can predict catch

$$\hat{C}_{a,t} = \frac{\hat{F}_{a,t}}{\hat{F}_{a,t} + M_{a,t}} \left(1 - e^{-(\hat{F}_{a,t} + M_{a,t})}\right) \hat{N}_{a,t}$$

Or, with the addition of catchability, predict a survey index

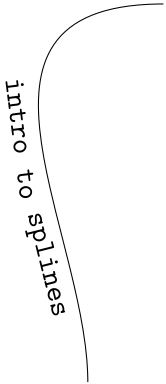
$$\hat{I}_{a,t} = q_{a,t} \hat{N}_{a,t}$$

Estimation

We observe $C_{a,t}$, $I_{a,t}$ so in principle we can estimate the parameters by maximising the likelihood of the parameters given the data

$$
\begin{aligned}
L(\mathbf{N_{a=1}}, \mathbf{N_{t=1}}, \mathbf{s}, \mathbf{f}, \mathbf{q}, \boldsymbol{\sigma} | \mathbf{C}, \mathbf{I}) = \prod_{a=1}^{A} \prod_{t=1}^{T} & (\mathrm{N}(\ln(I_{a,t}) - \ln(\hat{I}_{a,t}), \sigma_i^2) \\
& \times \mathrm{N}(\ln(C_{a,t}) - \ln(\hat{C}_{a,t}), \sigma_c^2))
\end{aligned}
$$

We maximise the (log) likelihood to estimate the best fitting parameters of the model given the data!


intro to splines