Coverpage

```
## Document build date: Tue Dec 20 14:55:52 2016
## Working directory :
##       D:/projects/flr/drafting-doc
## Current contents of .GlobalEnv:
##      .First .Last thm
##
## Session information:
## R version 3.3.2 (2016-10-31)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 10586)
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.1252
## [2] LC_CTYPE=English_United Kingdom.1252
## [3] LC_MONETARY=English_United Kingdom.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] xtable_1.8-2       lattice_0.20-34    RColorBrewer_1.1-2
## [4] formatR_1.4        knitr_1.15.1
##
## loaded via a namespace (and not attached):
## [1] magrittr_1.5  tools_3.3.2   stringi_1.1.2 grid_3.3.2    digest_0.6.10
## [6] stringr_1.1.0 evaluate_0.10
```

Description of Catch at Age model

Colin P. Millar, Ernesto Jardim

Abstract

This document presents the statistical catch-at-age stock assessment model developed in the JRC Assessment For All (`a4a`) initiative. The stock assessment model framework is a non-linear catch-at-age model implemented in R http://www.r-project.org/ / FLR http://www.flr-project.org/ / ADMB http://www.admb-project.org/ that can be applied rapidly to a wide range of situations with low parametrization requirements. The model structure is defined by submodels, which are the different parts that require structural assumptions. There are 5 submodels in operation: a model for F-at-age, a model for the initial age structure, a model for recruitment, a (list) of model(s) for abundance indices catchability-at-age, and a list of models for the observation variance of catch-at-age and abundance indices. The submodels form use linear models. This opens the possibility of using the linear modelling tools available in R: see for example the mgcv http://cran.r-project.org/web/packages/mgcv/index.html gam formulas, or factorial design formulas using. Detailed model formulas, several diagnostic tools and a large set of models are presented in the document. Additionaly, advanced features like external weighting of the likelihood components and MCMC fits are also described. The target audience for this document are readers with some experience in R and some background on stock assessment. The document explains the approach being developed by a4a for fish stock assessment and scientific advice. It presents a mixture of text and code, where the first explains the concepts behind the methods, while the last shows how these can be run with the software provided.

Contents

1 Background

The stock assessment model framework is a non-linear catch-at-age model implemented in R/FLR/ADMB that can be applied rapidly to a wide range of situations with low parametrization requirements.

In the `a4a` assessment model, the model structure is defined by submodels, which are the different parts of a statistical catch at age model that require structural assumptions.

There are 5 submodels in operation:

- a model for F-at-age,
- a (list) of model(s) for abundance indices catchability-at-age,
- a model for recruitment,
- a list of models for the observation variance of catch-at-age and abundance indices,
- a model for the initial age structure,

In practice, we fix the variance models and the initial age structure models, but in theory these can be changed.

The submodels form use linear models. This opens the possibility of using the linear modelling tools available in R: see for example the mgcv gam formulas, or factorial design formulas using `lm()`. In R's linear modelling language, a constant model is coded as ~ 1, while a slope over age would simply be $\sim age$. For example, we can write a traditional year/age separable F model like $\sim factor(age) + factor(year)$.

The 'language' of linear models has been developing within the statistical community for many years, and constitutes an elegant way of defining models without going through the complexity of mathematical representations. This approach makes it also easier to communicate among scientists

- 1965 J. A. Nelder, notation for randomized block design
- 1973 Wilkinson and Rodgers, symbolic description for factorial designs
- 1990 Hastie and Tibshirani, introduced notation for smoothers
- 1991 Chambers and Hastie, further developed for use in S

There are two basic types of assessments available in `a4a` : the management procedure fit and the full assessment fit. The management procedure fit does not compute estimates of covariances and is therefore quicker to execute, while the full assessment fit returns parameter estimates and their covariances at the expense of longer fitting time.

2 The data and simple description of the model

The data are

C_{at} catch at age a and year t

S_{atk} abundance index for age a and year t from the kth survey or CPUE series, $k = 1, 2, \ldots$

The model is an age structure model where the number of fish in a given cohort N at the start of the following year is the number of fish that survived the perils of the current year. We assume that fish die through the year at a constant rate e^{-Z} (Z is positive), and that this rate is solely due to natural causes (M) and fishing (F) so that the total mortality rate is $Z = F + M$. This results in the model

$$N_{a+1,t+1} = N_{at}e^{-Z_{at}}$$

Abundance indices are observations of the relative abundance not of absolute abundance. This is because trawl surveys do not detect every fish but a fixed proportion Q. This proportion depends on age through length and means the index is proportional to abundance

$$S_{at} = Q_a N_{at}$$

If F and M are constant through the year catches arise as a fraction of those fish that died, and is written here as the familiar Baranov catch equation

$$\begin{aligned} C_{at} &= \frac{F_{at}}{Z_{at}} (N_{at} - N_{a+1,t+1}) \\ &= \frac{F_{at}}{Z_{at}} \left(1 - e^{-Z_{at}}\right) N_{at} \end{aligned}$$

These last two equations show that in there own way, catches and abundance indices are both observations of the numbers of fish in the population. Neither is sufficient to estimate the absolute abundances N but together they can be used to estimate both N and F. One way of doing this is using a statistical catch at age approach

3 More detailed description of the stock assessment model

Modelled catches C are defined in terms of the three quantities, natural mortality M, fishing mortality F and recruitment R, using a modified form of the well known Baranov catch equation:

$$C_{ay} = \frac{\boldsymbol{F}_{ay}}{\boldsymbol{F}_{ay} + M_{ay}} \left(1 - e^{-(\boldsymbol{F}_{ay} + M_{ay})}\right) \boldsymbol{R}_y e^{-\sum(\boldsymbol{F}_{ay} + M_{ay})}$$

where a and y denote age and year. Modelled survey indices I are defined in terms of the same three quantities with the addition of survey catchability Q:

$$I_{ays} = \boldsymbol{Q}_{ays} \boldsymbol{R}_y e^{-\sum(\boldsymbol{F}_{ay} + M_{ay})}$$

where s denotes survey or abundance index and allows for multiple surveys to be considered. Observed catches $C^{(obs)}$ and the observed survey indices $I^{(obs)}$ are assumed to be log-normally distributed, or equivalently, normally distributed on the log-scale, with age, year and survey specific observation variance:

$$\log C^{(obs)}_{ay} \sim \text{Normal}\Big(\log C_{ay}, \boldsymbol{\sigma}^2_{ay}\Big) \qquad \log I^{(obs)}_{ays} \sim \text{Normal}\Big(\log I_{ays}, \boldsymbol{\tau}^2_{ays}\Big)$$

The full log-likelihood for the `a4a` statistical catch at age model can now be defined as the sum of the log-likelihood of the observed catches (ℓ_N is the log-likelihood of a normal distribution)

$$\ell_C = \sum_{ay} w^{(c)}_{ay}\ \ell_N\Big(\log C_{ay}, \boldsymbol{\sigma}^2_{ay};\ \log C^{(obs)}_{ay}\Big)$$

and the log-likelihood of the observed survey indices

$$\ell_I = \sum_s \sum_{ay} w^{(s)}_{ays}\ \ell_N\Big(\log I_{ays}, \boldsymbol{\tau}^2_{ays};\ \log I^{(obs)}_{ays}\Big)$$

giving the total log-likelihood

$$\ell = \ell_C + \ell_I$$

which is defined in terms of the strictly positive quantites, M_{ay}, F_{ay}, Q_{ays} and R_y, and the observation variances σ_{ay} and τ_{ays}. As such, the log-likelihood is over-parameterised as there are many more parameters than observations. In order to reduce the number of parameters, M_{ay} is assumed known (as is common), and the remaining parameters are written in terms of a linear combination of covariates x_{ayk}, e.g.

$$\log F_{ay} = \sum_k \beta_k x_{ayk}$$

where k is the number of parameters to be estimated and is sufficiently small. Using this tecnique the quantities $\log F$, $\log Q$, $\log \sigma$ and $\log \tau$ (in bold in the equations above) can be described by a reduced number of parameters. The following section has more discussion on the use of linear models in `a4a` .

Stock recruitment relationships

The `a4a` statistical catch at age model can addiionally allow for a functional relationship to be imposed that links predicted recruitment $\tilde{R}$ based on spawning stock biomass and modelled recruitment R, included as a fixed variance random effect. Options for the relationship are the hard coded models Ricker, Beverton Holt, smooth hockeystick or geometric mean. This is implemented by including a third component in the log-likelihood

$$\ell_{SR} = \sum_y \ell_N\Big(\log \tilde{R}_y(a, b), \phi^2_y;\ \log R_y\Big)$$

giving the total log-likelihood

$$\ell = \ell_C + \ell_I + \ell_{SR}$$

Using the (time varying) Ricker model as an example, predicted recruitment is

$$\tilde{R}_y(a_y, b_y) = a_y S_{y-1} e^{-b_y S_{y-1}}$$

where S is spawning stock biomass derived from the model parameters F and R, and the fixed quantites M and mean weights by year and age. It is assumed that R is log-normally distributed, or equivalently, normally distributed on the log-scale about the (log) recruitment predicted by the SR model $\tilde{R}$, with known variance ϕ^2, i.e.

$$\log R_y \sim \text{Normal}\Big(\log \tilde{R}_y, \phi_y^2\Big)$$

which leads to the definition of ℓ_{SR} given above. In all cases a and b are strictly positive, and with the quantities F, R, etc. linear models are used to parameterise $\log a$ and/or $\log b$, where relevant.

By default, recruitment R as apposed to the reruitment predicted from a stock recruitment model $\tilde{R}$, is specified as a linear model with a parameter for each year, i.e.

$$\log R_y = \gamma_y$$

This is to allow modelled recruitment R_y to be shrunk towards the stock recruitment model. However, if it is considered appropriate that recruitment can be determined exactly by a relationship with covariates, it is possible, to instead define $\log R$ in terms of a linear model in the same way as $\log F$, $\log Q$, $\log \sigma$ and $\log \tau$.

3.1 Model fitting

The parameters that are estimated are log recruitment r_t, survey catchability q_a, the F parameters (MORE ON THESE LATER) and log stock numbers n_{a1} in the first year. The model is written in terms of these parameters,

$$\begin{aligned}
c_{at} &= r_{t-a+1} - \sum_{i=1}^{a-1} Z_{a-i,t-i} + \log\left(\frac{F_{at}}{Z_{at}}\left(1 - e^{-Z_{at}}\right)\right) + \epsilon_{at} \\
s_{at} &= q_a + r_{t-a+1} - \sum_{i=1}^{a-1} Z_{a-i,t-i} + \epsilon'_{at}
\end{aligned}$$

where ϵ denotes the Gaussian observation error. There are some modifications required for the early cohorts as they use n_{a1} rather than recruits and for the plus groups, but these are trivial and not presented. It is possible to write these equations in matrix notation if we combine all catches into a single vector: $\boldsymbol{c} = (c_{11}, c_{21}, \dots, c_{A_+1}, c_{12}, \dots, c_{A_+2}, \dots)^T$ and do similarly for the survey indices $\boldsymbol{s}$, it is also simpler if we define $\boldsymbol{r} = (n_{A1}, n_{A-1,1}, \dots, n_{21}, r_1, \dots)^T$ and combine the F model parameters into a single vector $\boldsymbol{f}$ then we can write the model as

$$\begin{aligned}
\boldsymbol{c} &= \boldsymbol{X}_r\boldsymbol{r} + o_1(\boldsymbol{f}) + o_2(\boldsymbol{f}) + \boldsymbol{\epsilon} \\
\boldsymbol{s} &= \boldsymbol{M}\boldsymbol{X}_q\boldsymbol{q} + \boldsymbol{M}\boldsymbol{X}_r\boldsymbol{r} + \boldsymbol{M}o_1(\boldsymbol{f}) + \boldsymbol{\epsilon}'
\end{aligned}$$

where the functions o_1 and o_2 are nonlinear (vector) functions of the F model parameters and the $\boldsymbol{X}$ matrices are various design matrices based on the full set of ages and years (to be described later). The $\boldsymbol{M}$ matrix maps the correct age and year in the survey to that in the full set of ages and years. These equations can be further combined by stacking the equations

$$\begin{pmatrix}\boldsymbol{c}\\ \boldsymbol{s}\end{pmatrix} = \begin{pmatrix}\boldsymbol{0} & \boldsymbol{X}_r \\ \boldsymbol{M}\boldsymbol{X}_q & \boldsymbol{M}\boldsymbol{X}_r\end{pmatrix}\begin{pmatrix}\boldsymbol{q}\\ \boldsymbol{r}\end{pmatrix} + \begin{pmatrix}\boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{M} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}o_1(\boldsymbol{f})\\ o_2(\boldsymbol{f})\end{pmatrix} + \begin{pmatrix}\boldsymbol{\epsilon}\\ \boldsymbol{\epsilon}'\end{pmatrix}$$

so that the model is of the form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + o(\boldsymbol{f}) + \boldsymbol{\epsilon}$$

where

$$\boldsymbol{\epsilon} \sim N\left(\boldsymbol{0}, \boldsymbol{W}^{-1}\right) \quad \text{where} \quad \boldsymbol{W}^{-1} = \begin{pmatrix}\sigma_c^2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_s^2\boldsymbol{I}\end{pmatrix} = \sigma_c^2\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{\sigma_s^2}{\sigma_c^2}\boldsymbol{I}\end{pmatrix}$$

In other words, this statistical catch at age model can be written as a linear model with an offset due to nonlinear functions of the F model parameters. We use this to reduce the parameters in the fitting process by concentrating the likelihood. This can be done by inserting the maximum likelihood estimates of β conditional on $\boldsymbol{f}$ and $\boldsymbol{W}$ into the likelihood. The maximum likelihood estimate of β conditional on the other parameters is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{y} - o(\boldsymbol{f}))$$

If we decide that the surveys and catches have the same observation variance then we can also estimate the obervation variance conditionally, in this case

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}(\boldsymbol{f}) &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{y} - o(\boldsymbol{f})) \\
\hat{\boldsymbol{y}}(\boldsymbol{f}) &= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{y} - o(\boldsymbol{f})) + o(\boldsymbol{f}) \\
&= \boldsymbol{H}(\boldsymbol{y} - o(\boldsymbol{f})) + o(\boldsymbol{f}) \\
&= \boldsymbol{H}\boldsymbol{y} + (1 - \boldsymbol{H})o(\boldsymbol{f})
\end{aligned}
$$

and the conditional (unbiased) estimate of σ^2 is

$$
\hat{\sigma^2}(\boldsymbol{f}) = \frac{1}{n-p}\log\Big((\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})^T(\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})\Big)
$$

where n is the length of $\boldsymbol{y}$ and p is the number of unique parameters in β. The concentrated log likelihood for $\boldsymbol{f}$ is then

$$
\begin{aligned}
l(\boldsymbol{f}) &\propto -\frac{n}{2}\log\Big((\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})^T(\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})\Big) \\
&= -\frac{n}{2}\log\left(\Big|\Big|(1 - \boldsymbol{H})\boldsymbol{y} + (1 - \boldsymbol{H})o(\boldsymbol{f})\Big|\Big|\right)
\end{aligned}
$$

Since $1 - \boldsymbol{H}$ and $(1 - \boldsymbol{H})\boldsymbol{y}$ only depend on the data these can be calculated outside of an iterative optimization procedure. It is straightforward to give different weights to the survey and catch components without increased computation (this is the same as assuming you know the ratio $\frac{\sigma_c^2}{\sigma_s^2}$). However, if you want to estimate both variances then the estimate of β is

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}(\boldsymbol{f}) &= (\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{y} - o(\boldsymbol{f})) \\
\hat{\boldsymbol{y}}(\boldsymbol{f}) &= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(\boldsymbol{y} - o(\boldsymbol{f})) + o(\boldsymbol{f})
\end{aligned}
$$

and as $\boldsymbol{W}$ is not known before hand the inverse $(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})^{-1}$ must be computed at every iteration. The concentrated log likelihood in this case is

$$
\begin{aligned}
l(\boldsymbol{f}) &= \frac{1}{2}\log|\boldsymbol{W}| - \frac{1}{2}\Big((\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})^T\boldsymbol{W}(\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})\Big) \\
&= -\frac{n_c}{2}\log\sigma_c^2 - \frac{n_s}{2}\log\sigma_s^2 - \frac{1}{2}\Big((\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})^T\boldsymbol{W}(\hat{\boldsymbol{y}}(\boldsymbol{f}) - \boldsymbol{y})\Big)
\end{aligned}
$$

4 Extending the model (brief)

First potential models for q_a. Linear forms can be included my simply changing the design matrices. By linear forms i mean spline smoothers with fixed degrees of freedom. An interesting addition would be to use penalised splines or better (i think) structured random effects

The model for F has not been considered so far. Options for this are a simple separable model, a model with several separable periods all these models can be expressed as linear models on the log link. Within the same framework is simple then to include smoothers (splines) with fixed degrees of freedom. More interesting but perhaps out with the scope of this project are structured random effect models for F. These include seasonal models (treating the number of ages as the season length) and correlated random walks.

5 Summary

At its simplest the model is a non-linear fixed effects regression, fitting the F parameters to the data but this requires that the survey variances are known relative to the catch variance. This model can include splines, or random effects with known parameters (variances, degrees of freedom, autocorrelation) in the q and r models. The reason this is being considered is that computational speed is important and allowing some fixed variability may reduce bias in model outputs by allowing some flexibility, it is acknowledged that such assumptions make statistical testing a bit dodgy.

The next level of complexity is where we want to estimate the survey variance. This means we have to recalculate the hat matrix at every iteration which involves a matrix inversion. Since the matrix inversion is being done already, including structured random effects for q and recruitment into this only adds the variance parameters to the objective function.

If a stock recruit function were to be added, recruits could not be integrated out and would have to be estimated in the objective function.

If more complicated F models were used, such as structured random effects (random walks, correlated random walks, seasonal models) the number of parameters would increase in the objective function. Therefore a model with few parameters is one with a highly parametrised F model and no SRR relationship.

6 Implementation (sketch)

We require two functions that return an objective

A inputs are F at age and observation error and arguments are the data and the hat matrix

B inputs are F at age and any variance parameters taking as arguments the design matrix H, the weight matrix W and the structural prior matrix (not mentioned yet but lets call it Q)

The full objective function is then

1. take input parameters (F pars, variances, recruitments (if SRR model being used))
2. convert F pars into F at age
3. calculate objective value using one of the two functions A or B above
4. add on SRR density and prior densities for variances if necessary

$$N_{at} = \begin{cases} R_t & \text{if} \quad a = 1 \\ R_{t-a+1} e^{-\sum_{i=1}^{a-1} Z_{a-i,t-i}} & \text{if} \quad a > 1 \end{cases} \tag{1}$$

EJ: We need to separate the model(s) from their fitting. I'd like to be more "fisheries science"-like in the general description. We need to work on it, but I like the idea that we have a single model based on the common dynamics and we increase the complexity by adding S/R, Selectivity or growth models, depending on the information available. That means using the usual equations for the dynamics. Afterwards we can add a section on fitting which shows all the technical details, including how it links with the model.

The data are assumed to be gaussian observations of numbers at age in the population, which itself is generated by the common age structured model

$$n_{at} = \begin{cases} r_t & \text{if} \quad a = 1 \\ r_t - \sum_{i=1}^{a-1} Z_{a-i,t-i} & \text{if} \quad a > 1 \end{cases} \tag{2}$$

where $Z_{at} = F_{at} + M_{at}$, where M_{at} is known and F_{at} is modelled as a seperable function

$$\log F_{at} = \gamma_a + \delta_t \tag{3}$$

with suitable constraints. The observation equations are

$$c_{at} \sim N\left(\log \left(\frac{F_{at}}{Z_{at}} \left(1 - e^{-Z_{at}} \right) \right) + n_{at}, \quad \kappa \right) \tag{4}$$

and

$$s_{atk} \sim N\left(q_{ak} + n_{at}, \quad \tau_k \right) \tag{5}$$

The parameters to be estimated in this model are: log recruitment r_t, F at age and year, F_{at}, log survey catchability q_{ak}, and the precisions $\theta = (\kappa, \tau_1, \ldots)$.

7 model details / parameterisations

7.1 Almost linear models

The approach taken here is built around the observation that the model can be written in the form

$$\mathbf{y} = f(\alpha) + X\beta + \epsilon \tag{6}$$

So that conditional on α, the model is linear with gaussian errors, so there are nice properties to be exploited. The question then is, how much can we squeeze into β...

If the survey catchability model is linear, the parameters α define F_{at} and the function f is a non-linear function involving F and M. So that recruitment and survey catchabilities are contained in β. A full description of this definition of f and X is given in appendix **??**. This is in a sense the best we can do, as there is no way to (analytically) linearise the model with respect to F_{at}.

Two ways in which structure can be introduced to this model is through parametric forms, for example using a basis smoother $q_{ak} = B_q\beta_q$ or through structural priors, for example by imposing that

$$(q_{2k} - q_{1k}) - (q_{3k} - q_{2k}) \sim N\Big(0, \lambda\Big)$$

That is deviations from a straight line are penalised much like a penalized smoother (see Rue and Held for details). Other priors based on the normal distribution that could be of use are things like: smooth shapes for q, smooth shapes with flat tops or with plataues defined in terms of restrictions to gradients or step sizes, random walk or AR1 process (plus optional smooth trend) for recruitment, an even an evolving pattern for q or F over time (*but this might be pushing it a bit!*). See section 9 for a description of these priors.

Since F_{at} is always non-linear there is no reason to prefer structural priors over parameteric forms (such as logistic, double normal etc.) unless one wishes to take advantage of the flexiblilty of the structural prior models.

7.2 Not so linear models

Aside from nice linear priors, there is also the possibility of a recruitment model being imposed. These are of the form

$$E\big[r_{t+1}\big] = f\left(\sum_a \omega_{at} e^{n_{at}}\right) \tag{7}$$

and so in general if there is a recruitment model specified the only parameters that are linear are the survey catchabilities.

7.3 Totally non-linear models

Finaly if there is a non-linear function specified for the survey catchabilities the model is completely non-linear.

8 Fitting strategies

8.1 Almost linear models

In this case the model, in full generality, is

$$\mathbf{y} = f(\alpha) + X\beta + \epsilon$$

where

$$\alpha \sim N\big(0, Q_a(\theta)\big)$$
$$\beta \sim N\big(0, Q_b(\theta)\big)$$
$$\epsilon \sim N\big(0, H(\theta)\big)$$

and

$$\theta \sim \text{uninformatively}$$

where the Q matrices are structural priors, the H matrix is diagonal with different variances for each component and all the variance parameters are contained in θ which are given uninformative priors. For precisions these will be gamma priors and for AR1 parameters these will be ...

As all the priors are gaussian, we can integrate out β leaving us with

$$\log \pi(\alpha, \theta|y) \propto -\frac{1}{2}\Big[(y - f(\alpha))^T(H - HX(X^THX + Q_b)^{-1}X^TH)(y - f(\alpha)) + \alpha^TQ_a\alpha - \log\frac{|Q_a||Q_b|}{|X^THX + Q_b|}\Big] + \log \pi(\theta)$$

the joint distribution of α and θ to maximise. Note that the matrices Q_a, Q_b and H all depend on θ.

An additional possibility...

Conditional on α (and the components of θ that define Q_a), the joint distribution above becomes a kind of multivariate chi-square distribution, and i suspect this has an analytically tractable maximum, so that we could plug in the best estimates of the remaining θ...

The strategy therefore is to maximise over α and θ. If we can plug in values of θ that maximise the distribution $\pi(\theta|\alpha, y)$, then we could propose α, plug in best θ, $\tilde{\theta}$ say, and return the value of the joint density at the point $(\alpha, \tilde{\theta})$. Proposals would be calculated by the optimiser - nlminb for example.

8.2 Not so linear models

These models can be treated the same as in the previous section, however the size of α is now much larger as it contains all the recruitments r_t and only the catchabilities remain in the β vector.

8.3 Totally non-linear models

With these models we are resigned to maximising $\log \pi(\alpha, \beta, \theta|y)$. I am not aware of any tricks here.

9 The structure in the model

or

What the parameters will look like

9.1 log Catchabilities: q_a (and possibly F at age)

A simple non parametric model is an approximation to a thin plate spline with 3 or 4 degrees of freedom. The problem with these smoothers is that the smaller the degrees of freedom the poorer the approximation - these smoothers usually perform better in a penalised setting where the model space contains more degrees of freedom than required. This problem can be cast into a random effects setting, but it is simpler to set a prior that penalizes deviations form a straight line (this is pretty much the discrete analogue of the thin plate spline anyway). Such a prior looks like this:

$$(x_{i-1} - x_{i-2}) - (x_i - x_{i-1}) \sim N\Big(0, \lambda\Big)$$

or

$$Dx \sim N\Big(0, \lambda I\Big)$$

in other words x is gaussian with zero mean and precision matrix $Q = D^TD$. This also gives an insight into why this prior is a smoother. The second difference matrix D enters into the posterior as λx^TD^TDx i.e. the sum of the squared (discrete) second derivatives of x and weighted by the variance of the gradient changes - this is analagous to the smoothing penalties, typically written $\lambda\int f''(x)^2$.

Related priors are the random walk defined by

$$x_i - x_{i-1} \sim N\Big(0, \lambda\Big)$$

Combinations of these priors can be used, in particular to impose a flat (or flatish) curve beyond a given point, one can specify a weight, w_i on the variances so that the differences are forced to be effectively zero. In adition to this you could also make the curve smoother as this point is approached forcing somthing like a flat top. The details of this have still to be finalised.

Figure 1: Basis functions for a thin plate spline

Figure 2: Some selectivity functions