FISH 104 continued

Statistical estimation of number at age from a length distribution

In Fish 104 length distributions were simulated. This time, given a length distribution, the parameters will be estimated using non-linear optimisation.

The parameters to be optimised can be:
mean length at age
standard deviation of length at age
number at age

If you assume the von Bertalanffy equation can be used to model mean length at age this reduces the number of parameters to be estimated for any number of ages only 2 or 3 parameters are estimated rather one for each age.

- Using the length distribution from

  www.hafro.is/∼lorna/ldist1

  estimate the mean length at age and proportion at age of the population. Initially assume that μ and σ are fixed and provide initial values for μ from the von Bertalanffy growth function. If you estimate μ try using the von Bertalanffy equation within the estimation.

  - Start by plotting the length distribution.
  - Estimate mean length at age for the youngest ages from the length distribution and calculate von Bertalanffy parameters.
  - If the proportion at age is being estimated then convert the length distribution into proportions.
  - Write a function which simulates a population (an example is given below) and using `nlm` estimate the parameters by minimising the difference between the simulated length distribution and the data.

Functions are given in the lecture notes for Fish 104 along with a suggested procedure to fit all the parameters.

A function which could be used to estimate the proportion in each group could look like this:

```
le <- ldist$le              # lengths to be estimated

n <- 8                      # number of age groups
a <- 1:n
mu <- vonb(c(100,0.1,0))    # mean length at age
sdev <- rep(1.5,n)          # st dev of length at age

# function to estimate proportion at age
estld.p <- function(gpar){

  p <- gpar
  p[n] <- 1-sum(p[1:(n-1)])  # ensure p sums to 1.

  ld.tmp <- matrix(rep(0, length(le)*n), ncol=length(le))
  for(i in 1:n) {
  ld.tmp[i,] <- (pnorm((le+0.5-mu[i])/sdev[i]) - pnorm((le-0.5- mu[i])/sdev[i]))*p[i]
```

```
  }
  ld.tot <- apply(ld.tmp, 2, sum)

  SSE <- sum((ld.tot-pldist)^2)
  return(SSE)
}
```

This estimates the proportion p and requires μ and σ, an observed length distribution (as proportions) called `pldist` and `n` (the number of ages) to defined in R. The parameters to be estimated are entered as a vector.

In the function `estld.p` it is possible values of `p` to be ¡ 1. One way to prevent this is to increase the `SSE` whenever a value of `p` is ¡ 0. As the aim is to minimise `SSE` these values will not be chosen. To do this include the line:

```
if(min(p)<0) SSE <- SSE*1000
```

after SSE has been calculated and before it is returned.

To estimate the parameters use:

```
fit <- nlm(estld, gpar)
```

and use `typsize`.

This function can be edited to estimate other parameters, or to estimate μ from a von Bertalanffy function.

To estimate the von Bertalanffy parameters, the von Bertalanffy equation needs to be included in the function.

```
estld2 <- function(gpar) {

  K <- gpar[1]
  Linf <- gpar[2]
  t0 <- gpar[3]

  mu <- vonb(c(Linf,K,t0))

...
```

After initially estimating the parameters separately, estimate different groups of parameters at the same time, eg the proportion at age and the growth parameters.

Alternatively ...

Simulate a population length distribution, add some random error, then try to estimate the parameters.

To add lognormal errors to a generated length distribution, multiply the length distribution by

```
exp(rnorm(lel,0,sc^2)-sc^2/2)
```

where `lel` is the length of the length distribution and `sc` is the amount of error eg 0.1 or 0.5.

You can do this simply using the R code from FISH 104 or in more detail using the code from FISH 508 where the selection pattern of the catch is modelled. In this case, the length distribution of the population is modelled, then the length distribution which would be caught using a particular selection pattern can be predicted.

Fitting to real data

There is a dataset `shrimple.txt` in

`/nethome/unu/unu/stockassess/practicals/`

in Linux and in Windows:

`E:\unu\stockassess\practicals`

The dataset contains a subset of 2 years of shrimp survey data. The lengths are in mm. L_∞ is about 35mm and there are about 8 age groups.

- Calculate the length distribution for each year separately — remember to fill in zeros where necessary.
- Estimate the proportion at age for both years (separately) — do you get similar mean lengths and standard deviations for both years?
- If the shrimp length distribution is modelled on 0.5mm length groups (the scale of the original data) the equation to predict the length distribution must be changed to eg:

```
ld.tmp[i,] <- (pnorm((le+0.25-mu[i])/sdev[i]) - pnorm((le-0.25- mu[i])/sdev[i]))*p[i]
```