

# Age\_Structure

December 17, 2023

When dealing with age structures the recipe is pretty straightforward:

1. Group fish into bins by length
2. Age a sample of fish in each bin
3. Extrapolate that age key across all the fish in each bin
4. Get an age structure

However usually those bins are equally sized. Let's see if we can do better using what we know about the Fisher Information.

First we need to start with a probability function. Specifically let's look at  $P(t|b)$  - the probability of getting a specific age, given a specific bin. Note that both  $t$  and  $b$  are discrete in this situation so we're looking at a probability function as opposed to a probability distribution function.

Now we don't actually know  $P(t|b)$  (if we did we wouldn't be asking this question in the first place). But we do have a model for the other way around:

$$P(b|t) = \int_b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{L-L_\infty(1-e^{-k(t-t_0)})}{\sigma}\right)^2} dL$$

which we'll go ahead and assume is readily computable (i.e. we've already fit  $L_\infty$ ,  $k$ , and  $\sigma$ ). Furthermore given we're assuming we're taking loads of length samples we also know  $P(b)$ . So by Bayes' Theorem:

$$P(t|b) = \frac{P(b|t)P(t)}{P(b)}$$

However this seems to present a problem - aren't we trying to figure out what  $P(t)$  is? Certainly, but for now let's assume it is a parameter of our model -  $P(t) = \theta_t$ . Therefore our model is:

$$P(t|b) = \frac{P(b|t)\theta_t}{P(b)}$$

Alright so our log likelihood is:

$$l = \ln(P(b|t)) - \ln(P(b)) + \ln \theta_t$$

and:

$$\partial_t l = \frac{1}{\theta_t}$$

and:

$$\partial_t^2 l = -\frac{1}{\theta_t^2}$$

Given none of the other derivatives exist (technically the  $\theta_t$  are related in the fact that they must sum to 1, but that doesn't give them meaningful derivatives with respect to each other as we're more or less free to choose all of them but one).

Now note that we know for every  $\tau \neq t$  that

$$\partial_t^2 l = 0$$

. Therefore:

$$I_{t,t} = -E[\partial_t^2 l] = -\sum_{\tau} \partial_t^2 l \bullet P(\tau|b) = \frac{1}{\theta_t^2} P(t|b) = \frac{1}{\theta_t^2} \frac{P(b|t)\theta_t}{P(b)} = \frac{1}{\theta_t} \frac{P(b|t)}{P(b)}$$

Alright so we now know we have a diagonal matrix made up of these components. Furthermore given we know that a multiplier on a row of our matrix just results in a multiplier on our determinant, we can just remove the  $\theta_t^{-1}$  components as they won't actually contribute to the maximization of our information (they'll just represent a collective constant multiplier). What we are interested in then is the matrix:

$$\begin{pmatrix} \sum_b n \frac{P(b|0)}{P(b)} & 0 & \dots & 0 \\ 0 & \sum_b n \frac{P(b|1)}{P(b)} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sum_b n \frac{P(b|A)}{P(b)} \end{pmatrix}$$

where  $A$  is the maximum practical age of the species and  $n$  is the number of samples to be taken per bin  $b$ .

Now note that for each bin  $b$  there are going to be a series of ages where  $P(b|t)$  is so low that it effectively does not contribute to the information at that age  $t$ . This means that samples in that bin only affect some, but not all, of our diagonal elements. And here we find there is diminishing returns. For each new sample from that bin we add we'll go from having a diagonal element of size  $m$  to size  $m + 1$ . If  $m$  is small this is a big change and definitely ups are information significantly, but if  $m$  is large then this makes no real appreciable difference as compared to sampling from a bin where the number of samples thus far is much smaller. Therefore we're going to end up with a kind of balancing across bins most appropriate to each age.

And this is where we get a divergence from the normal equally distributed sampling. For low ages the bins will be relatively farly spaced because fish grow quickly when they are young. However the grow quite slowly when they are old and so the bins will get progressively placed closer and

closer together. And this means our bins will get narrower and narrower as the lengths they contain increase.

And this makes good intuitive sense - the smaller lengths have relatively fewer ages represented and therefore require fewer samples to get a good idea of the relative distribution of ages. The larger lengths however are muddled with all kinds of ages and so you'd need to sample more fish to get a good sense of the overall distribution. We want non-uniform stratified sampling.

To reverse this it means that if you wanted well defined bounds on each of your  $P(t)$  and used uniform stratified sampling you'd end up sampling the lower lengths *way* more than you actually need to. By using non-uniform stratified sampling you can reduce the amount of wasted effort gathering data you don't really need. Pretty cool!

[ ]: