Maximum likelihood estimates of mortality rates from catch at age data using survival analysis

Marco Kienzle*, Jason McGilvray† and You-Gan Wang‡

January 9, 2015

Abstract

Survival analysis was applied to fisheries catch at age data to develop maximum likelihood estimators for stock assessment. This new method estimated natural mortality, fishing mortality and catchability from typical catch at age matrices. Monte Carlo simulations suggested estimates were un-biased and provided a better fit than the traditional multinomial likelihood. Application to a dataset from Queensland's Sea Mullet fishery (Australia) estimated natural mortality to be equal to 0.319 ± 0.165 year^{-1}.

1 Introduction

One purpose of stock assessment is to estimate mortalities affecting fish stocks. This estimation problem is easier to solve for species that can be aged as opposed to those for which age can't be determined, as for example crustaceans. The reason is that mortality and longevity are inversely related, hence age is a measure of mortality. The central mortality model in fisheries research relating catch to the number of fish belonging to a cohort through time was proposed by Baranov [Quinn and Deriso, 1999]. Given recruitment and mortalities, the proportions of individuals at age in the catch can be calculated and used in a multinomial likelihood [Fournier and Archibald, 1982]. This method has become by far the most common likelihood to integrate age data into modern stock assessment models [Francis, 2014, Maunder and Punt, 2013].

The deterministic exponential model in Baranov's catch equation has a statistical counterpart in the form of the exponential probability distribution function which first and second moments quantify the relationship between longevity (age) and mortality rate [Cowan, 1998]: the mean age of a cohort which abundance declines at a constant rate is the inverse of that rate. Adopting such a statistical view of the exponential decay of individual belonging to a cohort allows the development of a set of maximum likelihood functions to estimate parameters of importance when assessing stocks. The field of survival analysis in statistics has created both a conceptual framework and refined methods to estimate mortality rates [Kleinbaum and Klein, 2005, Cox and Oakes, 1984] which are widely applied in the fields of medical research and engineering.

*Queensland Dept of Agriculture, Fisheries and Forestry, Ecosciences Precinct, Joe Baker St, Dutton Park, Brisbane, QLD 4102, Australia;
University of Queensland, School of Agriculture and Food Sciences, St. Lucia, QLD 4072, Australia
†Queensland Dept of Agriculture, Fisheries and Forestry, Ecosciences Precinct, Joe Baker St, Dutton Park, Brisbane, QLD 4102, Australia
‡University of Queensland, Centre for Applications in Natural Resource Mathematics, School of Mathematics and Physics, St. Lucia, QLD 4072, Australia

Despite the commonalities between survival analysis for medical and fisheries research, this theory has seldom been applied to animal ecology [Pollock et al., 1989]: to our knowledge, there hasn't been any application to age data for the purpose of fisheries stock assessment. In this manuscript, we describe how to apply survival analysis to create likelihood functions of age data for the purpose of estimating natural and fishing mortalities as well as gear selectivity. We started with a simplistic example using constant natural and fishing mortalities to introduce fundamental concepts from survival analysis before moving to more sophisticated cases leading to its application to real data from the sea mullet fishery in Queensland (Australia). The proposed methods were tested with simulated data to characterize some of their properties and their capacity to estimate population dynamic parameters of interest. Finally, the application to the mullet fishery case study provided specific estimates of natural mortality, catchability and selectivity.

2 Materials and methods

Fish can be assigned an age by examining its otolith, which is found just below its brain. Fish otoliths deposit calcium carbonate through time, thus increasing in size each year of their life. Microscopic observation of otolith sections often reveal alternate opaque and translucent zones, which can be used to assign individual fish to a particular age group.

Sampling programs in fisheries research centers around the world aim to collect a representative sample of fish each year to determine the distribution of age of any species of interest. In most cases, the data are binned into age-groups of width 1 year. For this reason, we split the lifespan of cohorts from their birth ($t \in [0; \infty]$) into n yearly intervals from $a_1 = 0$ to the maximum age of a_{n+1} years. While the theory presented in this document used that particular subdivision of time (t), un-equal ones also applies. In fact, an un-equal subdivision of time was used for the sea mullet case study.

2.1 The likelihood for constant natural and fishing mortalities

The exponential decrease in abundance of individuals belonging to a single cohort due to constant natural (M) and fishing (F) mortalities was described from a survival analysis point of view [Cox and Oakes, 1984] using a constant hazard function of time (t) and parameters θ

$$h(t;\theta) = M + F \tag{1}$$

The probability density function (pdf) describing survival from natural and fishing mortality is

$$f(t;\theta) = (M+F)\ e^{-(M+F)t} = \underbrace{M \times e^{-(M+F)t}}_{=f_1(t;\theta)} + \underbrace{F \times e^{-(M+F)t}}_{=f_2(t;\theta)} \tag{2}$$

Since age data belonging to individuals dying from natural causes are generally not available to fisheries scientists, we used only the component of the pdf that relates to fishing mortality ($f_2(t;\theta)$). This component of $f(t;\theta)$ integrates over the entire range of t to

$$\int_{t=0}^{t=\infty} f_2(t;\theta)\ dt = \frac{F}{M+F}. \tag{3}$$

Hence, the probability density of being caught at age t is (by normalizing $f_2(t;\theta)$),

$$\begin{aligned} g(t;\theta) &= \frac{M+F}{F}\, f_2(t;\theta) && (4)\\ &= \frac{M+F}{F}\, F \times e^{-(M+F)t} && (5)\\ &= f(t;\theta), && (6) \end{aligned}$$

which is the same as the original pdf.

The probability of being caught during age (a_i, a_{i+1}) is

$$P_i = \int_{t=a_i}^{t=a_{i+1}} f(t;\theta)\ dt = \exp\{M+F)a_i)\} H_{a_i}(a_{i+1}-a_i).$$

where $H_t(s)$ is the conditional probability of surviving up to age A conditional on being alive at age a $(A > a)$, specifically, $H_a(A) = 1 - \exp(A-a)$.

Suppose a total of S fish being caught with S_i being the number of fish at age between a_i and a_{i+1}. The overall marginal likelihood of this sample

$$\begin{aligned} \mathcal{L} &= \prod_{i=1}^{n} \Big(\int_{t=a_i}^{t=a_{i+1}} f(t;\theta)\ dt\Big)^{S_i} && (7)\\ &= \prod_{i=1}^{n} P_i^{S_i} && (8) \end{aligned}$$

This is often referred to as the likelihood of a multinomial probability (P_i) where $P_i = \int_{t=a_i}^{t=a_{i+1}} f(t;\theta)\ dt$.

This is not multinominal, which requires $\sum P_i = 1$

Because the fish younger than a_1 is not considered are not taken into account (they exist theoretically in the population). The likelihood is not complete and hence incorrect!

The way of getting around this is to look at the relative distribution between these age groups. The younger fish may live somewhere else.

Let $p_i = P_i / \sum_{j=1}^{n} P_j$, the relative proportions among those age groups of interest. Conditional on the total catch S with age between age a_1 and a_n, the age frequency of the total sample S follows the following multinominal distribution (up to a constant $S!/s_1!...S_n!$) (see Wang, 1999).

$$\prod_{i=1}^{n} p_i^{S_i}.$$

The logarithm of the likelihood was

$$\begin{aligned} \log(\mathcal{L}) &= \sum_{i=1}^{n} S_i\ \log\Big(\int_{t=a_i}^{t=a_{i+1}} f(t;\theta)\ dt\Big) && (9)\\ &= \sum_{i=1}^{n} S_i\ \log\Big(\int_{t=a_i}^{t=a_{i+1}} (M+F)\ e^{-(M+F)t}\ dt\Big) && (10)\\ &= \sum_{i=1}^{n} S_i\ \log\big(e^{-(M+F)\times a_i} - e^{-(M+F)\times a_{i+1}}\big) && (11)\\ & && (12) \end{aligned}$$

This development illustrated an application of survival analysis to estimate mortality rates affecting a cohort of fish by maximum-likelihood using a sample of catch at age. This method was implemented in R [R Core Team, 2013] in the package Survival Analysis for Fisheries Research (SAFR) provided as supplement material. Numerical application were made available using the following commands: **library(SAFR); example(llfunc1)**.

Natural and fishing mortality cannot be disentangled with catch data only but the next section will show that the provision of **varying** effort data allowed to estimate both catchability (q) and natural mortality.

2.2 Estimating catchability and natural mortality

In this section, we assumed that a time series of effort (E_i) associated with a sample of catch at age (S_i) was available to the researcher. And the assumption that fishing mortality varied according to fishing effort through constant catchability (q) held: $F(t) = q\ E(t)$. In this situation, the hazard function was written as

$$h(t,\theta) = M + q\ E(t) \tag{13}$$

And the pdf

$$\begin{aligned} f(t,\theta) &= (M + q\ E(t))\ e^{-Mt - q\int_0^t E(t)\ dt} && (14)\\ &= \underbrace{M \times e^{-Mt - q\int_0^t E(t)\ dt}}_{=f_1(t;\theta)} + \underbrace{q\ E(t) \times e^{-Mt - q\int_0^t E(t)\ dt}}_{=f_2(t;\theta)} && (15) \end{aligned}$$

As in the previous section, we had

$$\int_{t=0}^{t=\infty} f_2(t;\theta)\ dt = 1 - \int_{t=0}^{t=\infty} M \times e^{-Mt - q\int_0^t E(t)\ dt}\ dt \tag{16}$$

But we did not know an analytic solution to the integral since the function $E(t)$ was not specified. Nevertheless, as we knew the value of effort in any given interval ($\int_{t=a_i}^{t=a_{i+1}} E(t)\ dt = E_i = \int_{t=0}^{t=a_{i+1}} E(t)\ dt - \int_{t=0}^{t=a_i} E(t)\ dt, \forall i \in [1;n]$), we could calculate the value of $\int_{t=0}^{t=\infty} f_2(t;\theta)\ dt$ assuming $E(t)$ was constant over each interval i

$$\begin{aligned} \int_{t=0}^{t=\infty} f_2(t;\theta) &= 1 - \sum_{i=1}^{n} \Big[-\frac{M}{M + q\ E_i} e^{-Mt - q\int_0^t E(t)\ dt} \Big]_{t=a_i}^{t=a_{i+1}} && (17)\\ &= 1 - \sum_{i=1}^{n} \frac{M}{M + q\ E_i} \big(e^{-M\ a_i - q\int_0^{a_i} E(t)\ dt} - e^{-M\ a_{i+1} - q\int_0^{a_{i+1}} E(t)\ dt} \big) && (18)\\ &= \sum_{i=1}^{n} \frac{q\ E_i}{M + q\ E_i} \big(e^{-M\ a_i - q\int_0^{a_i} E(t)\ dt} - e^{-M\ a_{i+1} - q\int_0^{a_{i+1}} E(t)\ dt} \big) && (19)\\ & && (20) \end{aligned}$$

In practice, $0 \leq \int_{t=0}^{t=\infty} f_2(t;\theta) \leq 1$ and took a specific value depending on the values of M, q and E_i. Naming this constant value K, we could write the pdf of catch at age given that effort data are available as

$$g(t;\theta) = \frac{1}{K}\ f_2(t;\theta) \tag{21}$$

And the log-likelihood:

$$\log(\mathcal{L}) = \sum_{i=1}^{n} S_i\ \log\Big(\int_{t=a_i}^{t=a_{i+1}} g(t;\theta)\ dt \Big) \tag{22}$$

Numerical application of this method were made available using the following commands: **library(SAFR); example(llfunc2)**.

Accounting for age-specific gear selectivity ($s(t)$) effects on fishing mortality ($F(t) = q\ s(t)\ E(t)$) was included in a similar way into the likelihood using constant value for selectivity at age. In practice, it is difficult to estimate n additional selectivity parameters using only the data from a single cohort but processing several cohorts at the same time and assuming separability of fishing mortality rendered estimation of catchability, natural mortality and selectivity possible.

2.3 Estimates from catch at age matrix using fishing mortality separability

This section describes an application of survival analysis to matrices of catch at age, developed for the purpose of estimating catchability (q), selectivity at age ($s(t)$) and constant natural mortality (M). The matrix $(S_{i,j})$ containing a sample of fishes aged to belong to a particular age-group j in year i contains $n + p - 1$ cohorts. These cohorts were indexed by convention using k ($k \in [1, n + p - 1]$) and an increasing number r_k ($1 \leq r_k \leq \min(n, p)$) identifying incrementally each age-group (see appendix p. 21 for more information). Each matrix $S_{i,j}$ has two cohorts with only 1 age-group representing them.

The derivation for a single cohort were the same as those presented in the previous section, here reproduced with indexations relative to a single cohort and accounting for selectivity

$$g_k(t;\theta) = \frac{q\ s(t)\ E(t) \times e^{-Mt - q\int_0^t s(t)\ E(t)\ dt}}{\sum_{l=1}^{r_k} \frac{q\ s_{k,l}\ E_{k,l}}{M+q\ s_{k,l}\ E_{k,l}} \left(e^{-M\ a_{k,l} - q\int_0^{a_{k,l}} s(t)\ E(t)\ dt} - e^{-M\ a_{k,l} - q\int_0^{a_{k,l+1}} s(t)\ E(t)\ dt}\right)} \tag{23}$$

This expression needs more justification. $qs(t)E(t)$ should be $qs(t)E(t)/(M + qs(t)E(t))$?

Why not define

$$P_l = \frac{q\ s_{k,l}\ E_{k,l}}{M + q\ s_{k,l}\ E_{k,l}} \left(e^{-M\ a_{k,l} - q\int_0^{a_{k,l}} s(t)\ E(t)\ dt} - e^{-M\ a_{k,l} - q\int_0^{a_{k,l+1}} s(t)\ E(t)\ dt}\right)$$

and $g_k = P_l / \sum_{l=1}^{r_k} P_l$.

The following likelihood $\mathcal{L}$ is in fact the conditional likelihood as mentioned in the previous section. To highligh the novelty, it is better to put more details, and it will be easier to read and appreciate by others.

The likelihood function of a catch at age matrix was build using each pdf specific to each cohort ($g_k(t;\theta)$):

$$\mathcal{L} = \prod_{k=1}^{n+p-1} \prod_{l=1}^{r_k} \left(\int_{t=a_{k,l}}^{t=a_{k,l+1}} g_k(t;\theta)\ dt\right)^{S_{k,l}} \tag{24}$$

The expression above is equivalent to

$$\mathcal{L} = \prod_{i,j} P_{i,j}^{S_{i,j}} \tag{25}$$

where the $P_{i,j}$ are the probabilities of observing a fish of a given age j in year i given by the hazard model. In this likelihood, the $P_{i,j}$ sum to 1 along the cohort instead of summing to 1 for each year as described for the multinomial likelihood in Fournier and Archibald [1982].

This method was implemented in R [R Core Team, 2013] in the package Survival Analysis for Fisheries Research (SAFR). Numerical application of this method are available using the following commands: **library(SAFR); example(llfunc3); example(llfunc4); example(llfunc5);**.

2.4 Testing methods by simulation

Methods to estimate mortality and selectivity from a matrix containing a sample of number at age were tested with simulated datasets to characterize their performance. Variable number of cohorts ($n+p-1 = 25$, 35 or 45); maximum age ($p = 8$, 12 or 16 years) and sample size of age measurement in each year varying from 125 to 2000 increasing successively by a factor 2. The simulated datasets were created by generating an age-structure population using random recruitment for each cohort, random constant natural mortality, random catchability and random fishing effort in each year (Tab. 1). A catch at age matrix was calculated using a logistic gear selectivity with 2 parameters:

$$s_{a_i} = \frac{1}{1 + \exp(\alpha - \beta \times a_i)} \tag{26}$$

Several sampling strategies were implemented to assess how it affected mortality estimates. To test estimators derived from survival analysis, one would like to draw randomly from the probability distribution. This is obviously impossible in the real world because field biologists never have in front of them a entire cohort to chose from. Nevertheless, we implemented a sampling strategy (sampling strategy 1) that randomly selected from the entire simulated catch at age dataset as a benchmark. In the real world, samples can be drawn by accessing only a single year-class of every cohort every year, so the second strategy implemented was to simulate a random selection of a fixed number of sample (N) each year (sampling strategy 2). Finally, the third strategy investigated was to apply a weighting by the estimated total catch at age ($\hat{C}_{i,j}$) to the sample of number at age in the sample ($S_{i,j}$) – sampling strategy with weighting :

$$\hat{C}_{i,j} = p_{i,j} \odot C_i \otimes v(j) \tag{27}$$

where $p_{i,j}$ is the proportion at age (see appendix p. 21), C_i is a column vector containing the total number of fish caught in each year i and $v(j)$ is a row vector of 1's. And a weighted sample ($S^*_{i,j}$) was obtained using the fraction of total catch sampled

$$S^*_{i,j} = \hat{C}_{i,j} \times \frac{\sum_{i,j} S_{i,j}}{\sum_i C_i} \tag{28}$$

Note that $\sum_{i,j} S_{i,j} = \sum_{i,j} S^*_{i,j}$.

Comparison with the multinomial likelihood proposed by Fournier and Archibald [1982] were made using differences in negative log-likelihood between that method and the survival analysis approach described in the present article. Simulated catch were used to calculate the proportion of individual at age, constraining them to sum to 1 in each year. This method to calculate proportions for the multinomial likelihood was regarded as the best case scenario because we expect any estimation algorithm based on the multinomial likelihood to, at best, match exactly the simulated catch at age. The logarithm of these proportions were then multiplied by the simulated age sample (weighted or not depending on the case) to calculate the log-likelihood as described in Fournier and Archibald [1982]. This quantity was compared to that calculated using the survival analysis approach to determine which model best fitted the simulated data. This comparison ignored the number of parameter used in each model as the Akaike criteria would. The multinomial likelihood requires $n + p - 1$ more parameters to be estimated than the survival analysis because the former requires an estimate of recruitment for each cohort in order to calculate the proportion at age in the catch.

2.5 A case study: Queensland's sea mullet fishery

The straddling Sea Mullet (*Mugil cephalus*) population stretches along the east coast of Australia, with most landings occurring between 19°S (approx. Townsville) and 37°S (roughly the border between New South Wales and Victoria). Following recommendations from Bell et al. [2005], an existing (1999–2004) scientific survey design was modified from 2007 onwards to include both estuarine and ocean habitats in order to provide representative demographic statistics for Queensland component of this fishery. The number of fish

at age obtained by otolithometry (Tab. 2) were analyzed to estimate natural mortality, catchability and gear selectivity.

Sea Mullet are thought to spawn in oceanic waters adjacent to ocean beaches from May to August each year. By convention, the birth date was assumed to be on July 1^{st} each year. Opaque zones are thought to be deposited on the otolith margin during spring through early summer (September to December). Biologists have come to the conclusion that the first identifiable opaque zone is formed 14 to 18 months after birth, and all subsequent opaque zones are then formed at a yearly schedule [Smith and Deguara, 2003]. Each fish in the sample was assigned an age-group based on opaque zone counts and the amount of translucent material at the margin of otolith. Age-group 0–1 comprised fish up to 18 months old ($a_1 = 18$ months) while all subsequent age-groups spanned 12 months ($a_2 = 30$ months, $a_3 = 42$ months, etc ...).

Sensitivity of survival analysis estimates to these data, a matrix containing 7 years and 16 age-groups, were performed by truncating the dataset in 2 ways to assess the robustness of the method to varying number of years and age-groups. The first truncation removed the last and last-two years of data to evaluate the sensitivity of parameters estimates to addition/omission of data in order to anticipate possible effects of future addition of newly available data. The second truncation removed older age-groups from 10–11 to 15–16 to evaluate the importance of few old fish on natural mortality estimates as one could think *a priori* that these longer-lived individuals provided a lot of information on mortality.

3 Results

3.1 Method tests using simulated data

Weighting the numbers of sampled fish each year by total catch (sampling strategy 2 - weighted sample) performed as well as the benchmark sampling strategy 1 (Fig 1 and Fig. 2). By contrast, estimations using a fixed number of fish each year were biased suggesting that weighting by catch is necessary in practical applications of the survival analysis approach.

This weighting of age-data samples considerably reduced the uncertainty on natural mortality estimates (Fig 1) and almost completely removed bias: a small of amount of bias was still noticeable at the extremity of the range of natural mortality (0.1–1.0) tested. Increasing the number of samples reduced uncertainty associated with natural mortality estimates.

Estimates of catchability were much more consistent across the range of values tested (1–10 $\times 10^{-4}$) for all methods (Fig. 2). The bias of the un-weighted approach was often similar to that of the weighted one. But the uncertainty associated with the former approach was much larger than the latter. For both strategy 1 and strategy 2 with weighting, the benefit of increasing sampling size were very noticeable up to a 1000 fish aged but less so beyond that.

The comparison between the likelihood function from survival analysis and the multinomial likelihood (Fig. 3) showed that, apart sampling strategy 2 which provided biased estimates, the approach using survival analysis provided in the majority of cases smaller negative log-likelihood values than the multinomial likelihood. The substantial advantage given the multinomial likelihood in this comparison played an important role at low sampling intensity where the assumption that proportion at age was known perfectly artificially improved its performance in most difficult situations. This artificial advantage faded away as the simulated sample sizes were increased resulting in the survival analysis approach outperforming the multinomial likelihood.

3.2 Mortality estimates for sea mullet

Applying survival analysis to age data from a sample of Sea Mullet weighted by total yearly catch, catchability was estimated to be equal to $7.055 \pm 2.724\ 10^{-5}$ per boat-day (Tab. 4). Natural mortality for Sea Mullet was estimated to 0.319 ± 0.165 year^{-1} using the entire dataset (comprising 2013 and 16 age-groups, Tab. 5). The sensitivity analysis showed consistent estimates with the removal of 1 or 2 years and up to 6 age-groups: catchability estimates varied between [7.054; 7.126] 10^{-5} with mean equal to 7.079 10^{-5} boat.days^{-1} and natural mortality estimates varied between [0.319; 0.382] with mean equal to 0.336 year^{-1}. This sensitivity analysis suggested that the presence of age-groups in the dataset with fewer, sparse observations increased the uncertainty of both catchability and natural mortality estimates.

The maximum likelihood matrix of probabilities ($P_{i,j}$) associated with the weighted observations at age in the sample ($S^*_{i,j}$) were presented in Tab. 3. They illustrate that the construction of the likelihood estimator using this survival analysis relied on probabilities summing to 1 along the cohort instead of summing to one along rows and across cohorts, as previously proposed to develop the multinomial likelihood of age data by Fournier and Archibald [1982]. Note that 2 cohorts in the dataset were described by a single observation (top-right and bottom-left corner of the matrix in Tab. 3) which did not provide any information to estimate mortality rates, as represented by their associated probability equal to 1.

Maximum likelihood estimates of gear selectivity, catchability and natural mortality were slightly affected by weighting the sample of observed number at age by total yearly catch (Tab. 6), suggesting that variation of catch within $\pm$ 12% of the coefficient of varatioin influenced on the outcome of the analysis (Tab. 2).

4 Discussion

This application of survival analysis to fisheries research provided a novel approach to develop maximum likelihood estimators of natural, fishing mortalities and gear selectivity from age data. Monte Carlo simulations showed that it provided un-biased estimates of natural mortality and catchability over a wide range of simulated values.

The comparison between the negative log-likelihood from the survival analysis approach with the multinomial likelihood [Fournier and Archibald, 1982] suggested that the former offered a better model to represent the data. This comparison was made using the best possible outcome for the multinomial likelihood because it used the simulated proportions of individuals at age in place of the probabilities to compute the likelihood. Arguably, this approach gave a substantial advantage to the multinomial likelihood over the survival analysis: no one would reasonably expect any estimation method to systematically provide exactly the proportion at age in the catch using a sample of the data. Therefore, the present comparison really focused on which probabilities to use in the likelihood function, whether they should sum to 1 in each years along age-groups or along cohorts. Despite the strong advantage given to the multinomial likelihood, the results suggested that simulated data according to Baranov's catch equation were fundamentally better fitted by a statistical method that modelled the exponential decay of individuals along cohort rather than by one that assumed the data followed a multinomial probability distribution specific to each year.

Weighting of sample provided un-biased estimates of natural mortality and catchability. Mortality estimates, in particular fishing mortality, depended on the magnitude of catch. The un-realistic sampling strategy which assumed that all catch data would be in front of the experimenter at once for sampling, accounted automatically for variation of catch and effort in each year because the abundance of each age-group

in the catch determined the probability to choose at random an individual belonging to any age-group. In practical application of survival analysis to fishery research, weighting is necessary because one cannot know *a priori* the magnitude of catch in coming years.

The simulations used a logistic gear-selectivity to generate and fit the data although we would have preferred to generate data from a wide range of possible gear-selectivity functions or even using non-parametric procedures. Simulations showed that gear selectivities were the most difficult parameters to estimate. The sea mullet case study was in fact not fitted with a logistic curve but selectivities were estimated through a tedious process to search each proportion retained at age that best fitted the data as measured by the likelihood. This process could not be automatized into the simulation testing framework to provide automatic identification of gear-selectivity. This aspect of the analysis was left out of the present manuscript for future work. Criticisms that this somewhat simplified the problem would be correct. But the current article was designed as an introduction to the application of survival analysis to fisheries research not one that solves all problems at once. As such, the likelihood approach presented in this manuscript provides a method to identify the gear selectivity that best fit the data, just not an automatic one.

The estimations of natural mortality and catchability using data from a fixed number of fish every year were biased probably because data were simulated with large variations of recruitment and fishing effort, resulting in large variation of catches between years. Hence weighting number at age samples by total catch probably introduced large correction to the datasets in many simulations. The effect of weighting on parameter estimates was noticeable also in the case of the mullet fishery where the coefficient of variation of catch was 12.2%.

It was surprising that this analysis of Sea Mullet data from the QLD fishery estimated similar values, in particular gear selectivity estimates, to the most recent stock assessment performed on a much larger and diverse dataset that included data from New South Wales (NSW) [Bell et al., 2005]. Lester et al. [2009] suggested, using parasites, that the bulk of Sea Mullet caught in Queensland fishery is based on local fish populations and not migrating from NSW. While genetic analyses could not identify differences in single nucleotide polymorphism between samples from south QLD and NSW [Krück et al., 2013]. A clarification of the boundaries of stock of Sea Mullet on the Australian east coast should precede further data analysis and development of management strategies for this fishery.

The sensitivity analysis to data truncation showed a weak trend in increasing uncertainty associated with natural mortality. Intuition would have suggested that old, rare, individuals provided valuable information about mortality hence increasing our knowledge on survival. The results of data truncation suggested the opposite, that inclusion of older age-groups containing few or no observations increased our uncertainties on mortality estimates. Possibly this lack of data induced large uncertainties on gear selectivities for older age group because lack of observations in those could be the result of high mortality or low selectivity. Uncertainties in that aspect of the model might have propagated into other components, increasing uncertainty about natural mortality.

This likelihood method might find its place naturally into integrated stock assessment [Maunder and Punt, 2013] as it provided an efficient method to deal with samples of age data. Applications of survival analysis to fishery data could be expanded further, a particular area of interest for future development would be to use this method to derive recruitment estimates based on the probabilities estimated from survival analysis and total catch data from the fishery to generate the most likely time series of recruitment.

Acknowledgements

We are greatful to both W.N. Venables from the CSIRO and Nicole White from Queensland University of Technology for the many discussions on the topic of applying survival analysis to fisheries data.

References

P.A. Bell, M.F. O'Neill, G.M. Leigh, A.J. Courtney, and S.L. Peel. Stock assessment of the Queensland-New South Wales sea mullet fishery (*Mugil cephalus*). Technical Report QI05033, Queensland Government, 2005.

G. Cowan. *Statistical Data Analysis*. Oxford Science Publications, 1998.

D.R. Cox and D. Oakes. *Analysis of survival data*. Chapman and Hall, 1984.

D. Fournier and C.P. Archibald. A General Theory for Analyzing Catch at Age Data. *Canadian Journal of Fisheries and Aquatic Sciences*, 39(8):1195–1207, 1982.

R.I.C.C. Francis. Replacing the multinomial in stock assessment models: A first step. *Fisheries Research*, 151(0):70–84, 2014.

D.G. Kleinbaum and M. Klein. *Survival Analysis: A Self-Learning Text*. Springer, 2005. ISBN 9780387239187.

N.C. Krück, D.I. Innes, and J.R. Ovenden. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern australian sea mullet (*Mugil cephalus*). *Molecular Ecology Resources*, 13(4):715–725, 2013.

R.J.G. Lester, S.E. Rawlinson, and L.C. Weaver. Movement of sea mullet mugil cephalus as indicated by a parasite. *Fisheries Research*, 96(23):129 – 132, 2009.

M.N. Maunder and A.E. Punt. A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142:61–74, 2013.

K.H. Pollock, S.R. Winterstein, and M.J. Conroy. Estimation and analysis of survival distributions for radio-tagged animals. *Biometrics*, 45(1):pp. 99–109, 1989.

T. J. Quinn and R. B. Deriso. *Quantitative Fish Dynamics*. Oxford University Press, 1999.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`.

K. Smith and K. Deguara. Formation and annual periodicity of opaque zones in sagittal otoliths of *Mugil cephalus* (Pisces: Mugilidae). *Marine & Freshwater Research*, 54:57–67, 2003.

Figures

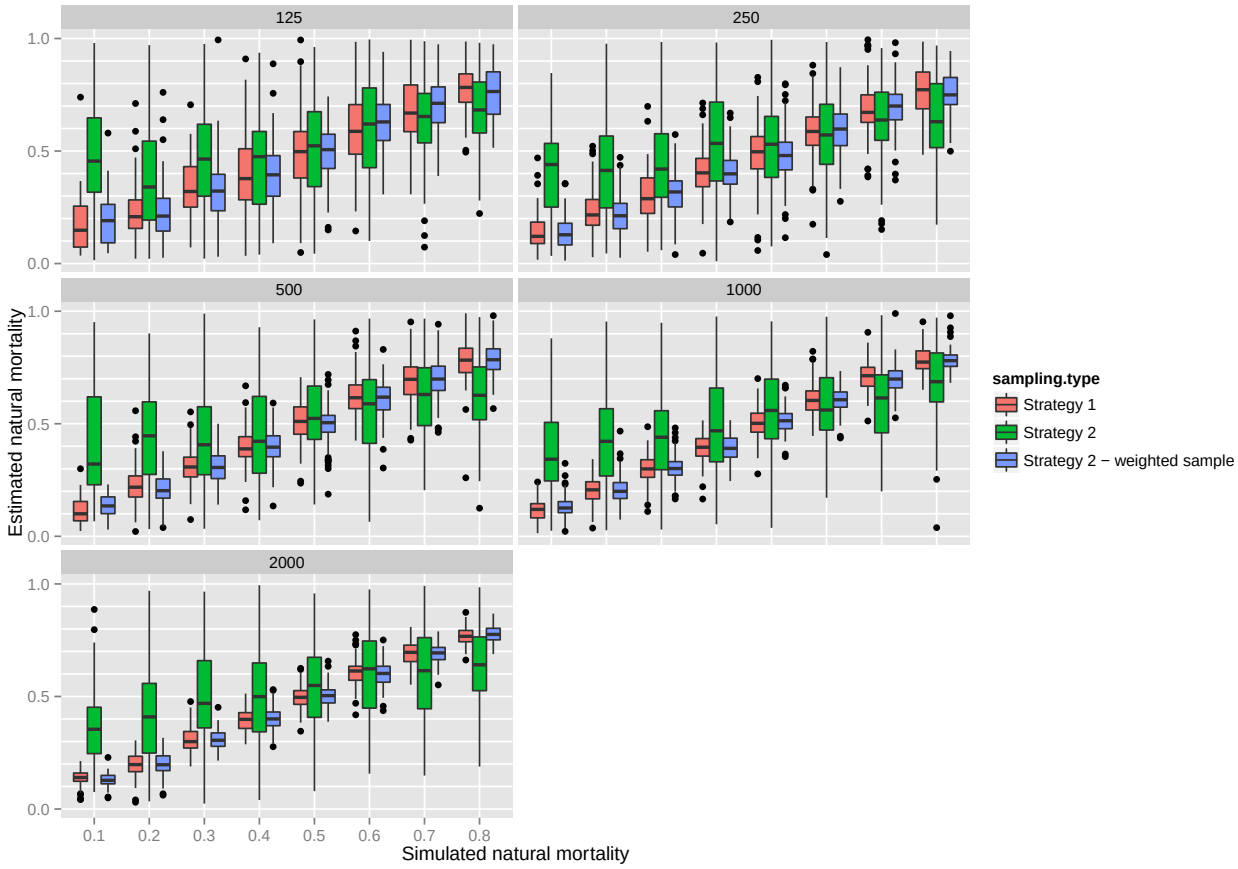

Figure 1: Comparison between simulated natural mortality (x-axis) and estimated using (a) a fixed number of sample each year or (b) data weighted by catch. Each panel correspond to an increasing number of samples per year varying from 125 to 2000.

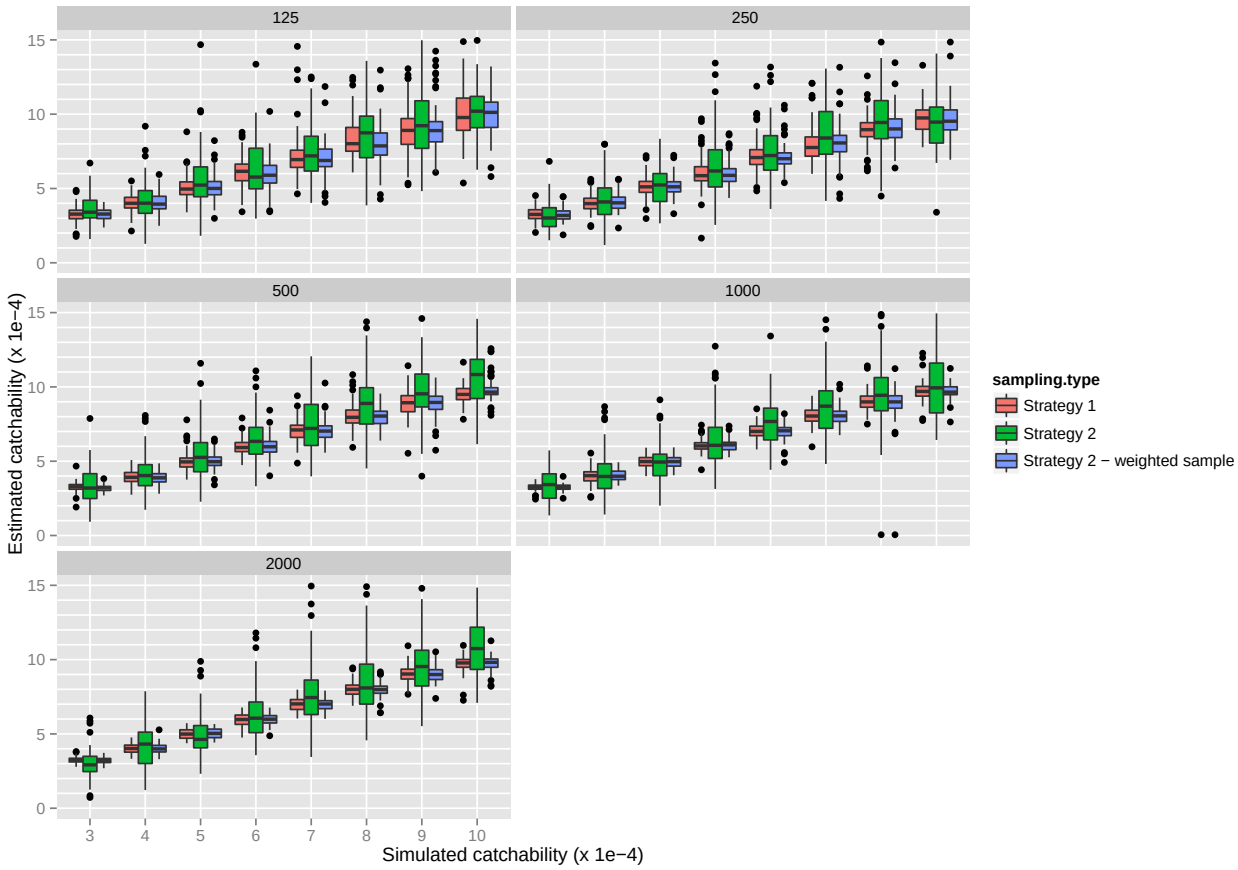

Figure 2: Comparison between simulated catchability (x-axis) and estimated using (a) a fixed number of sample each year or (b) data weighted by catch. Each panel correspond to an increasing number of samples per year varying from 125 to 2000.

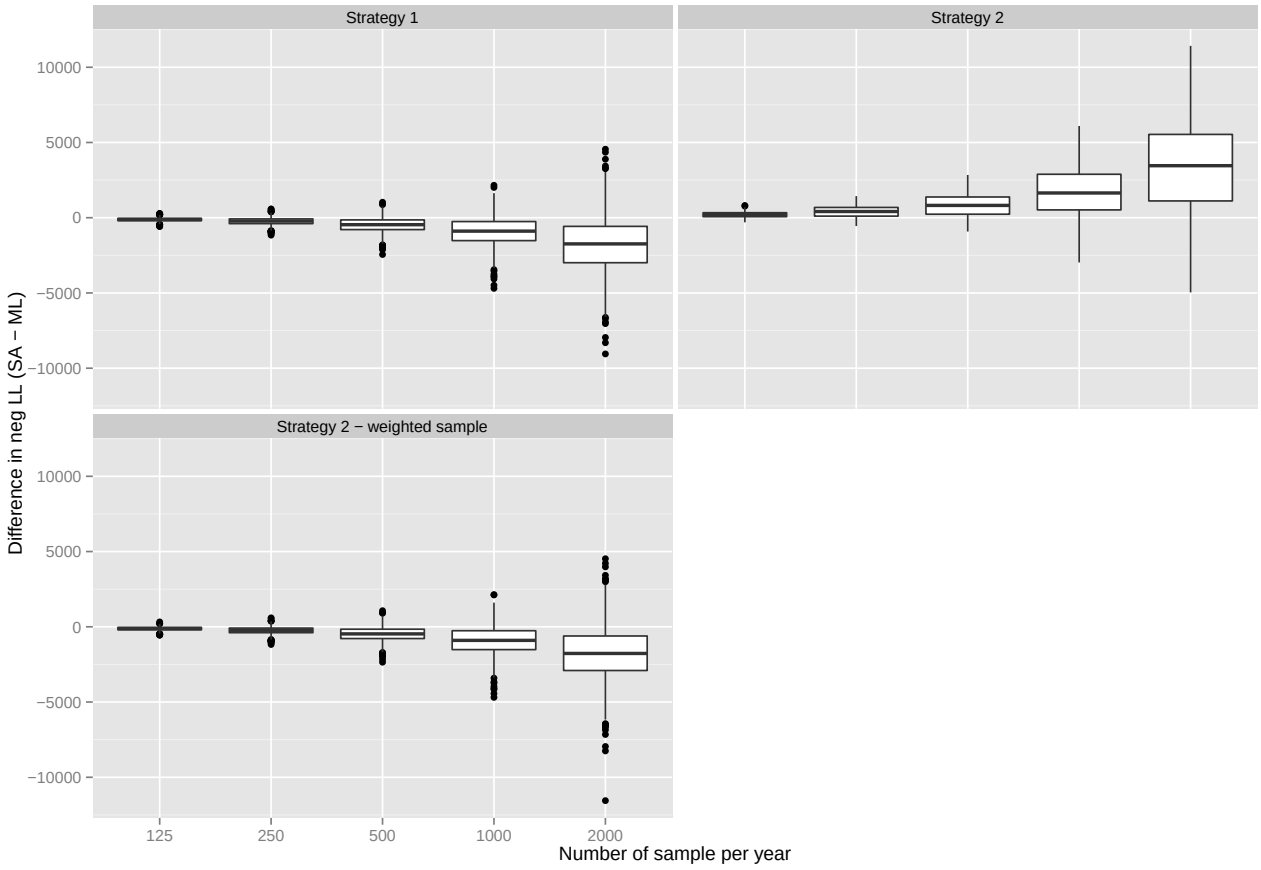

Figure 3: Difference between the negative log-likelihood (negLL) from survival analysis (SA) and multinomial (ML) as a function of the number of sample per year. Each panel represents a particular sampling strategy.

Tables

Variable type	Distribution	Parameters
recruitment	uniform	min=1e6, max=1e7
natural mortality	uniform	min=0.1, max=0.8
catchability	uniform	min=3e-4, max=1e-3
fishing effort	uniform	min=1e3, max=5e3
gear selectivity α	uniform	min=8, max=12
gear selectivity β	uniform	min=1, max=3

Table 1: Distribution and range of value taken by different type of random variable in simulations.

	0–1	1–2	2–3	3–4	4–5	5–6	6–7	7–8	8–9	9–10	10–11	11–12	12–13	13–14	14–15	15–16	Catch	Effort
2007	11	180	517	561	118	105	45	24	11	3							1350	7400
2008		42	468	618	409	100	57	21	10	8	2	2		2			1795	7875
2009	1	110	280	679	251	151	29	17	6	1	3	2	1	1			1815	6529
2010	2	239	541	250	200	97	50	11	9	2							1757	6109
2011	6	244	598	500	115	71	35	10	2	4		1	1		1	1	1542	6412
2012	1	99	633	563	298	57	32	15	11								1649	6993
2013		89	405	955	532	183	25	24	5					1			1993	6667

Table 2: Distribution of yearly samples (in rows) of sea mullet into age-groups of width 1 year (in columns); catch in tonnes and effort in boat-days.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.0025	0.0968	0.3290	0.5339	0.5739	0.5750	0.5691	0.5690	0.5690	0.5689	0.5712	0.5766	0.5874	0.6192	0.6886	1.0000
2	0.0028	0.1071	0.3233	0.3812	0.2833	0.2595	0.2563	0.2599	0.2599	0.2598	0.2598	0.2609	0.2633	0.2721	0.2800	0.3114
3	0.0025	0.0932	0.2835	0.3003	0.1615	0.1020	0.0920	0.0932	0.0945	0.0945	0.0945	0.0945	0.0948	0.0972	0.0980	0.1008
4	0.0029	0.0938	0.2811	0.3072	0.1525	0.0702	0.0437	0.0403	0.0408	0.0414	0.0414	0.0414	0.0414	0.0422	0.0422	0.0426
5	0.0057	0.1212	0.3167	0.3412	0.1759	0.0749	0.0339	0.0216	0.0200	0.0202	0.0205	0.0205	0.0205	0.0208	0.0207	0.0207
6	0.0258	0.2458	0.4212	0.3912	0.1973	0.0871	0.0365	0.0169	0.0108	0.0100	0.0101	0.0102	0.0102	0.0104	0.0103	0.0103
7	1.0000	0.9742	0.7486	0.4547	0.1957	0.0843	0.0367	0.0157	0.0073	0.0047	0.0043	0.0044	0.0044	0.0045	0.0044	0.0044

Table 3: Maximum likelihood probabilities ($P_{i,j}$) of the observed mullet sample age dataset weighted by total catch.

	10	11	12	13	14	15	16
2012	7.093 ± 3.254	7.101 ± 0.973	7.126 ± 1.029	7.095 ± 3.195	7.116 ± 1.636	7.095 ± 7.221	7.095 ± 6.491
2013	7.054 ± 1.176	7.055 ± 1.237	7.056 ± 1.367	7.056 ± 1.384	7.056 ± 2.393	7.056 ± 2.131	7.055 ± 2.724

Table 4: Sensitivity of catchability estimates (x10^{-5} boat-day^{-1}) to data truncations. Rows indicate the most recent year of data and columns the maximum age-group included in the analysis.

	10	11	12	13	14	15	16
2012	0.334 ± 0.117	0.36 ± 0.052	0.382 ± 0.05	0.339 ± 0.066	0.382 ± 0.046	0.337 ± 0.248	0.337 ± 0.239
2013	0.319 ± 0.075	0.319 ± 0.083	0.32 ± 0.091	0.32 ± 0.091	0.32 ± 0.154	0.319 ± 0.132	0.319 ± 0.165

Table 5: Sensitivity of natural mortality estimates (in year^{-1}) to data truncations. Rows indicate the most recent year of data and columns the maximum age-group included in the analysis.

	un-weighted	weighted
q	6.153 ± 1.504	7.055 ± 2.724
M	0.396 ± 0.038	0.319 ± 0.165
s_1	0.002 ± 0	0.002 ± 0
s_2	0.074 ± 0.008	0.085 ± 0.015
s_3	0.399 ± 0.02	0.419 ± 0.04
s_4	0.909 ± 0.047	0.882 ± 0.074
s_5	1 ± 0.033	1 ± 0.078
s_6	1 ± 0.053	1 ± 0.076
s_7	0.96 ± 0.091	0.981 ± 0.098
s_8	0.96 ± 0.116	0.981 ± 0.115
s_9	0.961 ± 0.131	0.982 ± 0.104
s_{10}	0.961 ± 0.154	0.982 ± 0.162
s_{11}	0.961 ± 0.157	0.982 ± 0.798
s_{12}	0.962 ± 0.322	0.982 ± 0.878
s_{13}	0.962 ± 0.532	0.982 ± 1.925
s_{14}	1 ± 0.505	1 ± 0.741
s_{15}	1 ± 1.06	1 ± 1.555
s_{16}	1 ± 1.039	1 ± 1.199

Table 6: Comparison of maximum likelihood parameters estimates for the mullet fishery using un-weighted or weighted samples of age data.

Appendices

Definitions of some mathematical symbols

This appendice contains definitions of some of the mathematical symbols used in previous sections

- $S_{i,j}$: a matrix of dimensions $n \times p$ ($i \in [1, n]$ and $j \in [1, p]$) containing a number of fishes that were aged and found to belong to specific age-groups j in a particular year i. This matrix contains data belonging to $n + p - 1$ cohorts, which by convention were labeled using k varying from 1 on the top-right corner of the matrix to $n + p - 1$ on the bottom-left (Tab. 7).

	1		...			p
1	...	...	...	3	2	1
	...	...	...	...	3	2
⋮	...	...	...	...	4	3
	...	...	k	...	...	...
	...	...	...	...	...	...
n	$n+p-1$	...	...	...	...	...

Table 7: Convention used to associate each element of the catch at age matrix ($C_{i,j}$) with particular cohort referred to as with the number given in this table.

The number of data in $S_{i,j}$ belonging to each cohort (r_k) varies from 1 to $\min(n, p)$ and was determined as follow:

$$r_k = \begin{cases} i - j + p \text{ if } k < \min(n,p) \\ \min(n,p) \text{ if } \min(n,p) \leq k < \max(n,p) \\ j - i + n \text{ if } k \geq \max(n,p) \end{cases} \tag{29}$$

Each element of the $S_{i,j}$ matrix is uniquely identified using indices i and j ($1 \leq i \leq n$ and $1 \leq j \leq p$) or indices k and l ($1 \leq k \leq n + p - 1$ and $1 \leq l \leq r_k$), so for example

$$\sum_{i,j} S_{i,j} = \sum_{k,l} S_{k,l} \tag{30}$$

- $p_{i,j}$: a matrix of dimensions $n \times p$ ($i \in [1, n]$ and $j \in [1, p]$) containing the proportion at age in the sample ($S_{i,j}$). Rows of this matrix sum to 1.

$$p_{i,j} = \frac{S_{i,j}}{\sum_j S_{i,j}} \tag{31}$$

- $F_{i,j}$ a matrix of fishing mortality with dimension $n \times p$ ($i \in [1, n]$ and $j \in [1, p]$). This matrix was constructed as the outer product of year specific fishing mortalities ($q\ E_i$) and selectivity at age (s_j):

$$F_{i,j} = q\ E_i \otimes s_j \tag{32}$$