Yet another general theory for the analysis of catch at age data: applying survival analysis to fisheries research

Marco Kienzle*and Nicole White†

September 22, 2014

1 Introduction

One of the purposes of stock assessment is to estimate mortalities affecting fish stocks. The central mortality model used in fisheries research was proposed by Baranov [Quinn and Deriso, 1999] to describe the variation of the number of fish belonging to a cohort through time. This deterministic exponential model has a statistical counterpart in the form of the exponential probability distribution function [Cowan, 1998]. Adoptiong a statistical view of this problem allows to develop maximum likelihood estimators [Burnham and Anderson, 2003] of parameters of importance to stock assessment. The branch of statistics focused on survival analysis has extended and refined methods to estimate mortality rates [Cox and Oakes, 1984]. This document describes an application of survival analysis to fisheries catch at age data.

2 Survival analysis

2.1 Abundance of individual belonging to a cohort

Let F be the instantaneous fishing mortality; M be the instantaneous natural mortality; C be the catch (removals due to mortality *i.e.* in number of fishes); D be natural deaths (removals due to natural causes) and T (=C+D) the total removal of fishes.
We consider only one year-class at that stage of life where instantaneous natural mortality (M) (expressed in number per unit time *e.g.* $years^{-1}$) and instantaneous fishing mortality (F) are constant and assumed to operate concurrently and continuously, so that $Z = F + M$ (fig. 1). The year-class vary as follow :

$$N(t) = N_0 e^{-Zt} = N_0 e^{-(F+M)t} \tag{1}$$

The corresponding differential equation is

$$\frac{dN}{dt} = -FN - MN \tag{2}$$

*DAFF biometry, Brisbane, Queensland, Australia
†QUT, Brisbane, Queensland, Australia

is made of term FN which represent the rate of fishes removes by fishing per unit time and MN the rate due to natural mortality. The catch rate is then

$$\frac{dC}{dt} = FN \tag{3}$$

Similarly, the death due to natural causes are

$$\frac{dD}{dt} = MN \tag{4}$$

Integrating catch rate over the interval of time $t = 0$ to $t = \tau$ provides catch over that period of time

$$\int_0^\tau dC\ dt = \int_0^\tau FNdt = \int_0^\tau F \times N_0 e^{-Zt}dt = FN_0 \int_0^\tau e^{-Zt}dt = FN_0[\frac{1}{-Z}e^{-Zt}]_0^\tau$$

$$C_{0\to\tau} = \frac{F}{Z}N_0(1 - e^{-Z\tau}) \tag{5}$$

Equation 5 is known as the Baranov catch equation after the Russian scientist who derived it in the 1910s. It states that the proportion of fish that died in the interval $[0;\tau]$ which have been caught by fishing gear is $\frac{F}{Z}$.

Simplification of equation 5 is possible if we consider the average abundance of fishes between time $[0;\tau]$. Since the definition of the average value of the integral of a function $X(t)$ over age t_1 and t_2 as

$$\bar{X} = \int_{t_1}^{t_2} \frac{X(t)dt}{t_2 - t_1} \tag{6}$$

If apply this to $N(t)$

$$\bar{N} = \frac{N_0(1 - e^{-Z\tau})}{Z\tau} \tag{7}$$

Comparing equation 5 and 7

$$\frac{C_{0\to\tau}}{\tau} = F\bar{N} \tag{8}$$

showing that the catch over time period τ is proportional to instantaneous fishing mortality and the average abundance during the interval of fishing. The term τ is necessary to reconcile the dimension of numbers per time in the equation. Using an interval of time of one year, we get

$$C_{0\to 1} = F\bar{N} \tag{9}$$

Numerical application Having $N_0 = 1000$ individual, $F = 0.4\ years^{-1}$, $M = 0.2\ years^{-1}$ (Fig. 1), we have after a period of 1 year:
$N_1 = 549$, which mean that 451 have died. *i.e.* survival rate is 0.549 and death fraction 0.451 $C = 301$ fishes and $D = 150$ fishes.
Note that the mean number of fishes is equal to 751.98 and NOT TO $\frac{1000+549}{2} = 774.5$ only because the variation of N_t is exponential and not linear.

Now, suppose that the mortality rates are in days^{-1}, 0.4/365 and 0.2/365 for fishing and natural mortality resp., then

$$C_{0\to 365} = F \times 365 \times \bar{N} \tag{10}$$

where $\bar{N}$ now denotes an average number per day.

2.2 Survival analysis approach to single cohort

The exponential decrease in abundance of a fish cohort is described in survival analysis [Cox and Oakes, 1984] using a constant harzard function of time (t) and parameters θ

$$h(t;\theta) = M + F \tag{11}$$

It follows that the density function is:

$$f(t;\theta) = (M + F)\ e^{-(M+F)t} \tag{12}$$

The survivor function gives the proportion of cohort surviving longer than t [Kleinbaum and Klein, 2005]

$$P(T > t) = S(t;\theta) = e^{-(M+F)t} \tag{13}$$

Finally, the cumulative distribution function $F(t)$ with density $f(t)$ gives the proportion of the cohort that died until time $T = t$

$$F(t) = 1 - S(t) \tag{14}$$

The probability of dying between t_1 and t_2 is

$$P(E_{[t_1-t_2]}) = \int_{t_1}^{t_2} f(t;\theta) = F(t_2) - F(t_1) = S(t_2) - S(t_1) \tag{15}$$

The total number of individuals that die in an interval ($E_{[t_1-t_2]}$) is made of those that were caught in the interval ($C_{[t_1-t_2]}$) plus the number of individual dying of natural causes ($D_{[t_1-t_2]}$): $E_{[t_1-t_2]} = C_{[t_1-t_2]} + D_{[t_1-t_2]}$

So, survival analysis provides the tools to build a likelihood estimator of mortality rates. Let's take the same example as above, supposing that we have access to the number of individual dying in each interval. In the R script below, we implicitly assume a knowledge of likelihood for binned data [Cowan, 1998]

```
# Simple illustration of using survival analysis concepts
# to estimate mortality rates

N0 <- 1e3
F <- 0.4 # per year
M <- 0.2 # per year
time <- seq(0,20)
nb.alive.at.time <- N0 * exp(- (M+F) * time)
nb.dying.in.interval <- nb.alive.at.time[-length(time)] - nb.alive.at.time[-1]

llfunc <- function(Z){
    -sum( nb.dying.in.interval * log(exp(-Z * time[-length(time)]) - exp(-Z*time[-1])))
}

result <- optim(par = 0.5, llfunc, method = "L-BFGS-B")
```

This simplistic example above illustrates the usage of survival analysis. It suffers from several shortcomings that we will address one by one to apply survival analysis to fisheries catch at age data. In principle, survival analysis offers the tools to represent any mortality schedule and the likelihood approach provides a method to compare them and determine which is most supported by the data.

First, we introduce the concept of truncated distributions to address the case where some age-groups are not samples, for example in Torres Strait lobster where age-group 2+ migrates outside the fishing grounds to Papua New Guinea to spawn.

```
# survival analysis to estimate mortality rates
# using truncated distributions

N0 <- 1e3
F <- 0.4 # per year
M <- 0.2 # per year
time <- seq(0,20)
nb.alive.at.time <- N0 * exp(- (M+F) * time)
nb.dying.in.interval <- nb.alive.at.time[-length(time)] - nb.alive.at.time[-1]

# Suppose you have only the first 3 age-groups
x <- 1:3

llfunc <- function(Z){

    P <- (exp(-Z * time[-length(time)]) - exp(-Z*time[-1]))[x]
    -sum( nb.dying.in.interval[x] * log(P/sum(P)))
}

result <- optim(par = 0.5, llfunc, method = "L-BFGS-B")

# or only 3 age-groups in the middle of the distribution
x <- 4:7
result2 <- optim(par = 0.5, llfunc, method = "L-BFGS-B")
```

Second, total death data are not available to fisheries scientist. Instead, they are often provided with catch from a fishery.

```
# Simple illustration of using survival analysis concepts
# to estimate mortality rates

N0 <- 1e3
F <- runif(1, min = 0, max = 2) # per year
M <- runif(1, min = 0.1, max = 0.5) # per year
time <- seq(0,20)
nb.alive.at.time <- N0 * exp(- (M+F) * time)
nb.dying.in.interval <- nb.alive.at.time[-length(time)] - nb.alive.at.time[-1]
catch <- F/(F+M) * nb.dying.in.interval

llfunc <- function(f){
    # Here you assume that you know M or you could just estimate total mortality
```

```
    Z <- f+M
    P <- f/(Z) * (exp(-Z * time[-length(time)]) - exp(-Z*time[-1]))
    -sum( catch * log(P/sum(P)))
}

result <- optim(par = 0.5, llfunc, method = "L-BFGS-B")
print(paste("Simulated F is", round(F,2)))
print(paste("Estimated F is", round(result$par,2)))
```

Third, we would also to estimate natural mortality from catch at age data. This is possible if you have a measure of effort (E) and that it relates to fishing mortality via catchability (q): $F = qE$

Fourth, often gear properties interfere with the selection of fish caught.

```
# CREATION DATE     13 Nov. 2008
# MODIFICATION DATE  5 Sep. 2014

# AUTHOR marco.kienzle@gmail.com

# BACKGROUND we are looking for a method to estimate M and F with data similar to the Torres Strai
#

# STATUS works wonderful

# PURPOSE express a likelihood function to estimate cohort-specific mortality
#         using catch data (and not survey)

# METHOD
# ASSUMPTION

# COMMENT if you have only catches, you can't estimate both fishing and natural mortality, you wi
#         BUT if you can provide the ratio of catch over total abundance (basically you have a ca
#             then the ratio of catch over total death (i.e. N(t+1) - N(t)) is equal to F/Z

# Suppose age varies between 0 and 10
age = seq(0,10)

# Suppose you have a 60% (M+F = 0.51) chance survival from both fishing (F=0.405) and natural mor
M <- runif(1, min = 0.1, max =0.5)
x <- runif(1, min = 0.1, max = 1)
s <- runif(1, min = 0.1, max = 1/3)
effort <- runif(length(age), min = 1e3, max = 2e3)
catchability <- runif(1, min = 1/2e4, max = 1/1.5e4)
F <-  catchability * effort * c(s * age[1:4], rep(1,length(age[5:11])))


print(paste("Simulated q is", round(catchability * 1e4,3), "10^(-4)"))
print(paste("Simulated slope of selectivity is ", round(s,3)))
```

```
print(paste("Simulated natural mortality is", round(M,3)))

N0 <- runif(1, min = 1e3, max = 1e6)
nb.at.age = cbind(age, N0 * exp(-(M + F)) ^ age) # this correspond to what you would see
                                                  # with an exhaustive survey
# Calculate the total number of individual dying at age
total.death = N0 * (exp(-(M+F) * age) - exp(-(M+F) * (age+1)))

# According to stock assessment books, catch = F / Z
catch = F/(M+F) * total.death

# Plot the data
par(mfrow=c(1,2))
plot(age, nb.at.age[,2], main = "Surviving both nature and fishing", las = 1)
plot(age, catch, main = "Catch", type = "h", las=1)


# Maximum likelihood function to estimate F
llfunc = function(par){ # a function of only 1 parameter (F)

    F <- par[1] * 1e-4 * effort
    alpha <- par[2]
    M <- par[3]

# And total mortality
    selectivity <- c(alpha * age[1:4], rep(1, length(age[5:11])))
  Z = M + F * selectivity

  # Calculating the probability of surviving until certain age
  prob1 = F * selectivity / Z * (1 - exp(-Z * age))
  prob2 = F * selectivity / Z * (1 - exp(-Z * (age+1)))

  # Finally the likelihood
  P = prob2-prob1

#    print(P)
#    print(sum(P))

    index <- which(P != 0)
    -sum( catch[index]  * log( P[index] / sum(P[index]) ))


} # End of function

# Estimate cohort-specific mortality rates

```

```
result <- optim(par = c(10,0.5, 0.1), fn = llfunc, method = c("L-BFGS-B"),
      lower = c(1e-2,1e-2,1e-2), upper = c(1e2,1,1), hessian = TRUE)
errors <- sqrt(diag(solve(result$hessian)))

print(" ***** ")
print(paste("Estimated catchability is", round(result$par[1],3), "+-", round(errors[1],3)))
print(paste("Estimated s is", round(result$par[2],3), "+-", round(errors[2],3)))
print(paste("Estimated M is", round(result$par[3],3), "+-", round(errors[3],3)))
```

Finally, we can process data from several cohorts at the same time using the standard format of a catch at age matrix (year x age-groups). Multiplicity of data to estimate natural mortality, catchability and selectivity (using the assumption of separability) improves our capability substantially.

```
# CREATED   3 Sep. 2014
# MODIFIED  8 Sep. 2014

# AUTHOR marco.kienzle@gmail.com;

# STATUS works

rm(list=ls())

source("UsefulFunctions.R")

#### Simulate some data

# Similar to sand whiting
max.age <- 13
sim <- GenerateData2(max.age = max.age, nb.of.cohort = 17) # Generate catch using gear selectivit

# idealised
#max.age <- 10
#sim <- GenerateData2(max.age = max.age, nb.of.cohort = 40) # Generate catch using gear selectivi


# The method is fairly robust to pretty large random errors
# sim$catch <- sim$catch * matrix(runif( nrow(sim$catch) * ncol(sim$catch), min = 0.7 , max = 1.3

# The method is robust to mild random errors
sim$catch <- sim$catch * matrix(runif( nrow(sim$catch) * ncol(sim$catch), min = 0.9 , max = 1.1),

plot.catch.by.cohort(sim$catch)

# log-likelihood function
llfunc <- function(par){

    # Re-arrange input data into cohorts
```

```
    catch.by.cohort <- Caaa2Coaa(sim$catch)
    effort.by.cohort <- Caaa2Coaa(sim$effort)

    # optim works best on scaled parameters
    catchability.mf <- 1e-4

    # Allocate param to readable variable names
    M <- par[1]
    q <- par[2] * catchability.mf
    # Simulated selectivity
    #selectivity.at.age <- c(0,0,seq(1/(max.age - 3), (max.age - 4)/(max.age - 3), length = max.ag
    alpha <- par[3]
    selectivity.at.age <- c(0,0, alpha * seq(1,max.age-4), 1, 1)

    # matrix of fishing mortality
    F <- q * effort.by.cohort * outer(rep(1, nrow(effort.by.cohort)), selectivity.at.age)

    # total mortality
    Z <- M + F

    # cumulative mortality
    cum.Z <- my.cumsum(Z)

    # Calculate the probability of observation in each interval
    prob1 <- F/Z * (1 - exp(-cum.Z))
    prob2 <- F/Z * (1 - exp(-(cum.Z-Z)))
    P <- prob1-prob2

    # discard zeroes and NA from sum of logs
    index <- which(!is.na(catch.by.cohort) & catch.by.cohort!=0)


    # Negative log-likelihood
    -sum(catch.by.cohort[index] * log( P[index] / total.over.lines(P)[index]))
}

result <- optim(par = c(0.2,1, 0.2), fn = llfunc, method = c("L-BFGS-B"),
      lower = c(5e-2,5e-2,1e-2), upper = c(0.5,10,0.5), hessian = TRUE)
errors <- sqrt(diag(solve(result$hessian)))

print("")
print(paste("Estimated catchability is", round(result$par[2],3), "+-", round(errors[2],3), " x 10
print(paste("Estimated M is", round(result$par[1],3), "+-", round(errors[1],3)))
print(paste("Estimated alpha is", round(result$par[3],3), "+-", round(errors[3],3)))
```

References

K. P. Burnham and D. Anderson. *Model Selection and Multi-Model Inference*. Springer-Verlag, second edition, 2003.

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

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

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

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

Figures

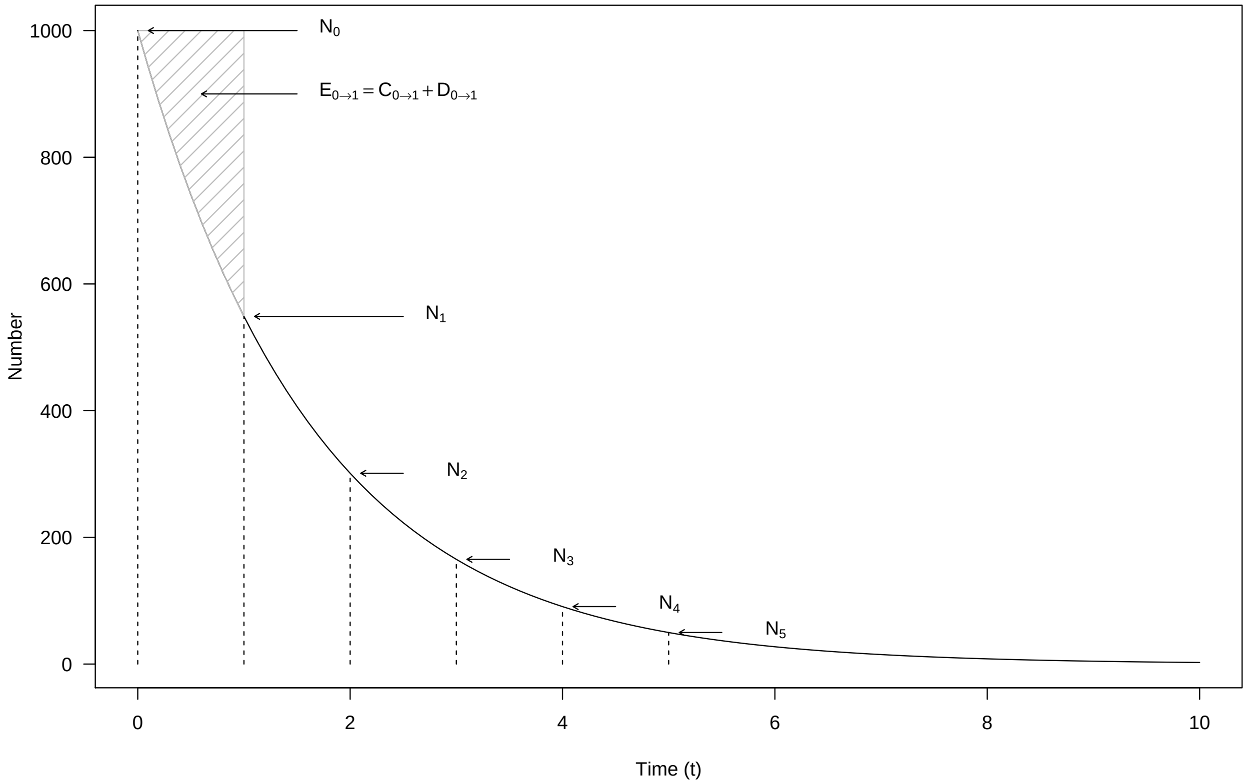

Figure 1: An illustration of the decreasing number of individual in a cohort, assuming $Z = 0.6$ years^{-1} and $\mathrm{N}_0 = 1000$.