Bayesian Catch Curve Analysis
Institute of Statistics Mimeo Series # 2615

Emily H. Griffith[1,5], Sujit K. Ghosh[1], Kenneth H. Pollock[3], and Michael J. Seider[4]

[1]Department of Statistics
North Carolina State University
Raleigh, NC 27695

[3]Department of Zoology
North Carolina State University
Raleigh, NC 27695

[4]Lake Superior Fisheries Team
Wisconsin Department of Natural Resources
P.O. Box 589
141 South Third Street
Bayfield, WI 54814

[5]Corresponding author:
Patuxent Wildlife Research Center
12100 Beech Forest Road, G-2
Laurel, MD 20708
egriffith@usgs.gov

Abstract

Catch curves have been used to estimate survival and instantaneous mortality for fish and wildlife populations for many years. In order to better analyze catch curve data from the Apostle Islands population lake trout *Salvelinus namaycush* in Lake Superior, we develop a Bayesian approach to catch curve analysis. First, the proposed Bayesian approach is illustrated for a single catch curve and then extended to multiple years of data. We also relax the model assumption of a stable age distribution to allow random effects across years. The proposed models are compared with the traditional methods using the focused DIC. There are many potential advantages to the Bayesian approach over the traditional methods such as least squares and maximum likelihood, based on large sample theory. Bayesian estimates are valid for finite samples, and efficient numerical methods can be used to obtain estimates of instantaneous mortality. We conclude that many benefits can be obtained from the Bayesian approach to a single catch curve and to multiple years of data, such as closed-form variance estimates and the ability to both model and estimate the process variation of survival rates.

Key words: Focused DIC; Life table methods; Markov chain Monte Carlo methods; Standing age distribution; Survival estimation.

1   Introduction

Lake trout *Salvelinus namaycush* from the Apostle Islands population in Lake Superior are recovering from both overfishing and predation by sea lampreys

Petromyzon marinus (Linton et al. 2007; Pollock et al. 2007). Due to the importance of lake trout, both economically and ecologically, monitoring the populations' demographic parameters is crucial. Linton et al. (2007) analyze age distribution and catch-per-unit-effort data, while Pollock et al. (2007) analyze mark-recapture data from this population. We obtained a set of catch curves, collected on male lake trout from 1986 through 2005 (unpublished data collected by the Wisconsin Department of Natural Resources). Catch curves tend to be both inexpensive and simple to collect, but their usefulness is limited by strict assumptions regarding population dynamics. Our objective is to utilize Bayesian inference to better use this type of data.

Chapman and Robson (1960) and Robson and Chapman (1961) formally developed statistical methods of analyzing catch curve data. They proposed estimating the survival probability, denoted by S, by using a regression model, a geometric model, and a multinomial model. The regression model is based on the idea that $E(n_x) = NS_1S_2S_3\ldots S_x p$ where S_x denotes the survival probability of an individual at age x, N the total number of individuals, p the capture probability, and n_x is the number of fishes captured from age class x. If capture probability (p) and survival rate (S) are age independent, then it follows that $E(n_x) = NS^x p$ (Seber 1982, p. 426). Thus, in a linear regression relating $\log n_x$ to x, the slope of the line is $\log S$ (Chapman and Robson 1960). This can be used as an approximate model, where the variance being dependent on age is ignored. Both Chapman and Robson (1960) and Seber (1982) recommend truncating the catch curve data at the point where n_x drops below 5 in order to reduce the dependence of σ^2 on x.

Chapman and Robson (1960) also developed methods to use the geometric distribution and the multinomial distribution to obtain a maximum likelihood estimator and then a uniformly minimum variance unbiased estimator (UMVUE) of the survival rate. The assumptions for the Chapman-Robson survival estimate are constant survival for all age classes and all years, a stable age distribution, a stationary population, and equal catchability for all fully vulnerable age classes. A population with a stable age distribution has survival and fecundity rates that have remained constant for a long period of time (Messier 1990; Udevitz and Ballachey 1998). However, a stable age distribution does not necessarily imply that the population size remains constant. A population can increase, decrease, or remain constant in size and still have a stable age distribution (Williams et al. 2002). A stationary population is one where the populations growth rate is one. Let N_x be the number of individuals in the sample of age x. Also let $T = \sum_x x N_x$. Notice that there exists a finite maximum age, x_{max}, such that $N_x = 0$ for all $x > x_{max}$ and hence T is finite with probability 1. The N_x's conditional on $\sum N_x = N$ follow a multinomial distribution, given below.

$$
\begin{aligned}
L(S|N,T) &= P(N_0 = n_0, N_1 = n_1, \ldots | N) \\
&= \frac{N!}{\prod_{x=0}^{\infty} n_x!} \prod_{x=0}^{\infty} (S^x)^{n_x} (1-S)^N \propto S^T (1-S)^N.
\end{aligned}
\tag{1}
$$

The maximum likelihood estimator (MLE) of S based on the above likelihood is

$$
\hat{S}_{MLE} = \bar{X}/(1+\bar{X}),
\tag{2}
$$

where $\bar{X} = T/N$. The MLE is a biased estimator (Chapman and Robson 1960). A uniformly minimum variance unbiased estimator (UMVUE) for S does exist, and is given by the formula

$$\hat{S}_U = \frac{\bar{X}}{1 + \bar{X} - \left(\frac{1}{N}\right)}. \tag{3}$$

The variance for $\hat{S}_U$ cannot be obtained in closed form. The minimum variance unbiased estimator of $\text{var}(\hat{S}_U)$ is

$$\widehat{\text{var}}(\hat{S}_U) = \hat{S}_U \left(\hat{S}_U - \frac{T-1}{N+T-2} \right). \tag{4}$$

Chapman and Robson (1960) also discuss the instantaneous mortality rate, Z, which can be defined by the relationship $Z = -\log S$. For the geometric model, no unbiased estimate of Z exists (Chapman and Robson 1960). Z can be estimated consistently using the fact that $\hat{Z}_{MLE} = \log(\hat{S}_{MLE})$. Z can also be estimated using least-squares estimation. However, Chapman and Robson (1960) do not recommend using the regression estimator due to the potential for bias due to the violation of the assumption of constant variance. Robson and Chapman (1961) conclude that the best estimate for S, and the corresponding standard error, is the UMVUE.

Jensen (1985) compares the UMVUE of S suggested by Chapman and Robson (1960) and Robson and Chapman (1961) to the least-squares estimate. Jensen (1985) concludes that the precision of the UMVUE is a function of the number of animals in the catch curve. However, the precision of the least-squares estimate of S increases with the number of age classes in the sample. Murphy (1997) also

explored the bias in the Chapman-Robson estimator of instantaneous mortality from the UMVUE of S, $\hat{S}_U$, setting $\hat{Z}_U = \log(\hat{S}_U)$ and compared it with the bias for the least-squares regression estimate of the same. Murphy (1997) found that the least-squares regression estimates of Z were biased, underestimating the true instantaneous mortality rate. Small samples and low instantaneous mortality increase this bias (Murphy 1997). We conclude that the UMVUE of S has the most desirable properties out of all these estimators. However, there is not a closed form estimate of the variance of the survival rate parameter under the geometric or multinomial model. Also, the UMVUE does not exist for estimating Z.

In this article we develop a Bayesian approach to estimate S and Z and we show that closed form estimates can be obtained based on a finite sample. In Section 2, we describe our method for a single catch curve analysis. In Section 4, we extend our method to multiple years of catch curves. In Section 6, we extend our method further to allow for the survival rate to have a random effect across years. In Sections 4 and 6, the methods are illustrated with analysis based on real data. In Section 7, we report the results of two simulation studies that we performed to test our random effects method. We conclude with some suggestions for future research, in Section 8.

2 Bayesian Catch Curve with Conjugate Priors

We now present a Bayesian analysis using conjugate priors. The conjugate prior for the multinomial distribution is the Dirichlet distribution (Robert 2001, p. 121).

In the specific case of the catch curve's conditional multinomial distribution, because there are only two multinomial probabilities, S and $1-S$, the Dirichlet distribution reduces to the Beta(α, β) distribution (Robert 2001, p. 521).

Applying Bayes Theorem to the likelihood of the data given in Equation (1) and the Beta(α, β) prior distribution yields the posterior density

$$\pi(S|T,N,\alpha,\beta) = \frac{S^{T+\alpha-1}(1-S)^{N+\beta-1}}{B(T+\alpha, N+\beta)}, \tag{5}$$

meaning that the posterior distribution of S is a Beta$(T+\alpha, N+\beta)$ distribution.

The Bayes estimator of S under squared error loss is the posterior expectation of the Beta$(T+\alpha, N+\beta)$ distribution. Thus, the Bayes estimator of S is given by

$$\hat{S}_B = E[S|N,T] = \frac{T+\alpha}{T+\alpha+N+\beta}, \tag{6}$$

with corresponding variance

$$\widehat{\text{var}}(\hat{S}_B) = \text{var}[S|N,T] = \frac{(T+\alpha)(N+\beta)}{(T+\alpha+N+\beta)^2(T+\alpha+N+\beta+1)}. \tag{7}$$

These results hold true for any α and β that satisfy $T+\alpha > 0$ and $N+\beta > 0$. Using specific α's and β's, we can obtain the other estimators of the survival rate using Bayesian inference. Notice that as $\alpha+\beta \to 0$, $\hat{S}_B \to \hat{S}_{MLE}$. Thus the MLE can be obtained as a limiting case of the Bayes estimator using an improper prior $\pi(S) = 1/(S(1-S))$. This prior has most of its weight at zero and one. Similarly, as $\alpha \to 0$ and $\beta \to -1$, $\hat{S}_B \to \hat{S}_U$ and again we see that the UMVUE is obtained

7

as a limit of the Bayes estimator using an improper prior $\pi(S) = 1/(S(1-S)^2)$. This prior has most of its weight at zero and one. Note that $1/(S(1-S)^2) \to \infty$ as $S \to 0$ or $S \to 1$. Alternatively, if we use the weighted squared error loss, $L_w(S,a) = w(S)(S-a)^2$ with $w(S) = 1/(S(1-S))$ then the Bayes estimator is given by $\hat{S}_{wB} = E(w(S)S|T,N)/E(w(S))|T,N) = \hat{S}_{MLE}$ if $\alpha = \beta = 1$.

Bayesian inference calls for a prior distribution on the parameter S. If we do not have any substantial prior information, we can use a noninformative prior like the Jeffreys prior. Jeffreys prior is $\pi^*(S) \propto |I(S)|^{\frac{1}{2}}$ where $I(S)$ is Fisher's information (Robert 2001, p. 130). Notice that $I(S) = E\left[-\partial^2 \log L(S|N,T)/\partial S^2\right]$ $= E\left[T/S^2 + N/(1-S)^2\right] = N/(S(1-S)^2)$, which leads to Jeffreys prior being $\pi^*(S) \propto [\sqrt{S}(1-S)]^{-1}$, which is the limiting kernel of a Beta$(1/2, \beta)$ distribution when $\beta \to 0$. Thus, in this case the posterior distribution, using Jeffreys prior, is a Beta$(T+1/2, N)$, which is a proper distribution as long as $N > 0$.

Using Jeffreys prior, the posterior mean and variance of S are

$$\hat{S}_J = \frac{T + \frac{1}{2}}{T + N + \frac{1}{2}} \tag{8}$$

and

$$\widehat{\text{var}}(\hat{S}_J) = \frac{(T + \frac{1}{2})(N)}{(T + N + \frac{1}{2})^2 (T + N + \frac{3}{2})}, \tag{9}$$

respectively.

The Bayes estimator of $Z = -\log S$, under the squared error loss, is given by the posterior mean, $E(-\log S|T,N,\alpha,\beta)$. Although this expectation cannot be evaluated analytically, we can numerically integrate it using the `integrate`

8

function available in the software R . Alternatively, we may use Monte Carlo methods to numerically approximate $\hat{Z}_B = E(-\log S | T, N, \alpha, \beta)$.

3 Illustration: Lake Trout Data

We will illustrate the Bayesian approach to a single catch curve using our motivating example of data collected on the Apostle Islands population of lake trout in Lake Superior. The trout were caught and aged annually from 1986 through 2005. The data are sparse, making annual estimates highly variable. The average sample size for male lake trout is 77.25. However, over the 20 years of catches, the small annual catch curves combine to form a large amount of data, indicating the value of a combined analysis. We are using a subset of the data, from 2000 to 2005. We chose to use this subset because the population had been recovering from both overfishing and lamprey predation. Our goal was to select a subset with an approximately stable age distribution, which is almost certainly not the case in the early years. In our subset, the average sample size of male lake trout is 73.3.

We analyzed these data using the UMVUE, recommended by Chapman and Robson (1960); Robson and Chapman (1961); Jensen (1985, 1996); Murphy (1997), and our Bayesian approach using Jeffreys prior. The UMVUE was 0.95, with 95-percent confidence interval limits of 0.947 and 0.957. The Bayesian estimate, using a Jeffreys prior, was 0.952. The 95-percent equal-tailed posterior interval had a lower bound of 0.947 and an upper bound of 0.958. These results are very similar. The regression and adjusted regression estimates were significantly lower

than both the Bayes estimate and the UMVUE. The bias that we found is the same as the bias found in Jensen (1996).

These estimates of the survival rate are all high (above 0.90), indicating low mortality. The data were collected during a spawning survey where larger fish tend to be collected. These larger fish may experience higher survival rates due to an increased ability to survive a lamprey attack and lowered vulnerability to commercial and recreational fishing gears (M. Seider, personal communication, June 20, 2008).

4   Bayesian Catch Curve with Multiple Years of Data

The catch curve estimates by Chapman and Robson (1960) require the assumptions that the age distribution is stable and stationary and survival is constant for all age classes. It is also assumed that the ages are sampled randomly above some minimum age. If we can justify these assumptions, combining catch data over multiple years will result in a more precise estimate of the constant survival rate due to the increase in sample size. The likelihood function for k years of catch-curve data, assuming independence, is given by

$$L(S|N_1,N_2,\ldots,N_k,T_1,T_2,\ldots,T_k) \propto S^{T_1+T_2+\cdots+T_k}(1-S)^{N_1+N_2+\cdots+N_k} \qquad (10)$$

Using the conjugate prior, which remains the Beta(α,β) distribution, the pos-

terior distribution is a Beta(α^*,β^*), where $\alpha^* = T_1 + T_2 + \cdots + T_k + \alpha$ and $\beta^* = N_1 + N_2 + \cdots + N_k + \beta$. The Bayesian estimate of survival under Jeffreys prior is

$$\hat{S}_B = \frac{\alpha^*}{\alpha^* + \beta^*} \tag{11}$$

with $\alpha = 1/2$ and $\beta = 0$. The corresponding variance is

$$\widehat{\text{var}}(\hat{S}_B) = \frac{\alpha^* \beta^*}{(\alpha^* + \beta^*)^2(\alpha^* + \beta^* + 1)}. \tag{12}$$

5 Illustration: Lake Trout Data

We are using the data collected on male lake trout from 2000 to 2005 in order to illustrate the combination of annual data. The individual year estimates and standard deviations are given in Table 1, along with the estimates and standard deviations from combining the data. Combining years involves strict assumptions. The stable age distribution assumption is very important when combining data over many years to form one large sample, as only one parameter will be estimated for all the years of data. As the underlying assumptions of the catch curve analysis rely on the assumption of constant survival over many years, calculating individual survival rate estimates treating each year as independent is not a consistent method for analyzing this data. By grouping the data together, resolution is lost. Table 1 shows that, although the survival estimates vary only slightly between years, there is some variation that may be lost when data are combined.

This leads us to a better model which we develop in the next section.

6 Random Effects Model for Multiple Years of Data

If we consider the possibility that the survival rate is a random effect with a constant mean and some variance, we can use multiple years of catch curve data to estimate the mean and variance of the random effect process. The estimates obtained will be ecologically meaningful, as there is often some natural temporal variation in the survival rates of a population. Now we assume that survival is constant for all age classes but is a function of a random process with a constant mean and variance for all years. Each catch curve will provide a point estimate of survival at that time. Combining years of data will provide estimates of the mean and variance of the random process.

In a Bayesian framework, we can modify the likelihood given in Equation (10) to allow for S to vary with year k. The likelihood will then be a product of the likelihoods for each year, given below.

$$L(S_1,\ldots,S_k|N_1,\ldots,N_k,T_1,\ldots,T_k) = \prod_{j=1}^{k} L(S_j|N_j,T_j) \propto \prod_{j=1}^{k} S_j^{T_j}(1-S_j)^{N_j} \quad (13)$$

If we assign the random effect S_j a $\text{Beta}(\tau S^*, \tau(1-S^*))$ distribution, each S_j will have population mean S^* and population variance $S^*(1-S^*)/(\tau+1)$. Because we do not know S^* or τ, we can use the Bayesian hierarchical framework and assign S^* a $\text{Beta}(\alpha,\beta)$ prior and τ a $\text{Gamma}(a,b)$ prior. We set $\alpha = \beta = .5$

and $a = b = 0.001$. Notice that this results in relatively vague priors for S^* and τ, with prior variances $\text{var}(S^*) = 0.125$ and $\text{var}(\tau) = 1000$.

Our hierarchical framework can be summarized as:

$$
\begin{aligned}
N_j | T_j, S_j &\sim \text{Negative Binomial}(S_j, T_j) \\
S_j | \tau, S^* &\sim \text{Beta}(\tau S^*, \tau(1 - S^*)) \\
\tau | a, b &\sim \text{Gamma}(a.b) \\
S^* | \alpha, \beta &\sim \text{Beta}(\alpha, \beta)
\end{aligned}
$$

We fit this model using the package `R2WinBUGS`, calling WinBUGS from `R`. The data we used are from the male lake trout, between 2000 and 2005. The estimates of survival from our random effect model for each individual year, along with a 95% equal-tailed posterior interval, are given in Table 2. We used the equal-tailed posterior interval instead of the highest posterior density (HPD) interval because of the symmetry of the posterior distributions. Recall that $Z^* = -\log S^*$ represents the population level instantaneous mortality rate.

We used the deviance information criterion (DIC) for model selection. The DIC is based on the deviance, $D(S) = -2\log(f(N, T|S))$, and $\text{DIC} = D(E(S|N, T)) + 2p_D = E(D(S)|N, T) + p_D$ (Spiegelhalter et al. 2002), where the $p_D = E[D(S)|N, T] - D(E[S|N, T])$ represents the number of effective parameters in the model. Spiegelhalter et al. (2002) conclude that DIC is the Bayesian analogue of Akaike's information criterion (AIC). When there is no substantial prior information in the

13

model, the DIC is approximately equal to the AIC. Spiegelhalter et al. (2002) discuss one benefit of the DIC over model selection criteria such as the AIC or the Bayes information criterion (BIC), which is that the BIC and AIC involve the specification of the number of parameters in the model. The number of parameters is not always easily determined in a hierarchical model, especially those that involve random effects.

In Table 2, we present the posterior mean and standard deviation of the random effects (S_j's) and from column 3 of this table it is clear that the variance components are relatively small as all the posterior standard deviations of the S_j's are smaller than 0.01. The deviance information criterion, or DIC, of the random effect model is 46.3, which is larger than the DIC of the constant S model for the same data, which is 43.6 (see Table 1). This implies that the random effect model is not an improvement for this particular dataset. However, the DIC of the random effect model, 46.3, is smaller than that of the individual yearly estimates (fixed effect model), which was 49.2 (see Table 1). The random effect model is an improvement over estimating each year separately for this dataset. Overall, being able to relax the assumption of constant survival and implement a random effects model is an improvement to the analysis of catch curve data, as it adds to the information gained from the analysis, whether that model is the best-fitting model or not.

The traditional DIC focuses on the posterior distribution, not the level of the hierarchy where the difference between the random effects model and the constant survival model lies. The focused DIC (Bob O'Hara, personal website,

http://deepthoughtsandsilliness.blogspot.com/2007/12/focus-on-dic.html, December 5, 2007) looks at the marginal deviance instead of the traditional conditional deviance.

The marginal distribution of **N** and **T** for the random effects model is

$$
\begin{aligned}
m(N, T | S^*, \tau) &= \int_{S_j=0}^{1} \prod_j \binom{N_j + T_j - 1}{T_j} S_j^{T_j} (1 - S_j)^{N_j} \\
&\quad \times \frac{S_j^{\tau S^* - 1} (1 - S_j)^{\tau(1 - S^*) - 1}}{B(\tau S^*, \tau(1 - S^*))} dS_j \\
&= \prod_j \binom{N_j + T_j - 1}{T_j} \frac{B(T_j + \tau S^*, N_j + \tau(1 - S^*))}{B(\tau S^*, \tau(1 - S^*))}.
\end{aligned}
\tag{14}
$$

The focused DIC, or fDIC, uses the focused deviance. The focused deviance is defined as $D^f(S^*, \tau) = -\log(m(N, T | S^*, \tau))$. The fDIC can be defined as

$$
fDIC = D^f(\bar{S}^*, \bar{\tau}) + 2p_D^f
\tag{15}
$$

where $p_D^f = E(D^f(S^*, \tau) | N, T) - D^f(\bar{S}^*, \bar{\tau})$ and $\bar{S}^*$ and $\bar{\tau}$ represent the posterior mean of S^* and τ, respectively. The focused DIC is only appropriate to calculate in the presence of a random effect, but it can be compared with the traditional DIC for model selection between models with and without random effects.

The focused DIC for the random effects model is 75.3, with $p_D^f = 2$. The fixed effects model is clearly selected, since the DIC for the fixed model is 43.6, with $p_D = 1$. The variation between years is small, indicating that the extra parameters

15

to account for such a small random effect are not necessary to explain the yearly variations. In order to more formally explore the use of fDIC as a model selection tool we perform a simulation study.

7 Simulation Studies

As it is likely that survival rates will vary between years in other real fish as wildlife populations, we were interested in seeing what happens to standard estimates of survival when the survival rate is constant for all ages, but varies randomly from year to year. We examined survival rates of $S = 0.6$, 0.75 and 0.9 combined with random effects generated from a Normal$(0, \sigma)$ distribution with $\sigma = 0$, 0.05, and 0.25. We added random variation to them using the formula $\text{logit}(S_i) = \text{logit}(S) + \eta_i$, or $S_i = \frac{\left(\frac{S}{1-S}\right)e^{\eta_i}}{1+\left(\frac{S}{1-S}\right)e^{\eta_i}}$, where η_i is generated from a Normal$(0, \sigma)$ distribution. We know that $n_x \sim \text{Binomial}(N_0, \prod_{i=1}^{x} S_i)$, so we generated catch curve data by randomly generating the number of animals caught in each age class using the Binomial distribution. Using $\sigma = 0.25$ means that the range of values within two standard deviations is 0.24, 0.19, and 0.09 for survival rates of 0.6, 0.75, and 0.9, respectively.

We generated 10,000 single catch curves with 50 age classes for each combination, setting the number of animals in our sample and in the zero age class, N_0, equal to 100. We estimated the survival rate, S, and the corresponding standard deviation, $\hat{\sigma}_{\hat{S}}$, using the UMVUE and the Bayesian estimator with Jeffreys prior. The results are given in Table 3.

16

The results show that, despite the addition of random variation, the point estimate from a single catch curve remains unbiased. These models appear to be very robust against the random effect in the estimation of the survival rate, S, as long as the age distribution is stable. The variance estimates underestimate the true amount of variability in the population. This can be seen by comparing the mean estimates of the standard deviation of S with the standard deviation of the estimates of S.

We also carried out a separate simulation study to judge how well the focused DIC performed at selecting the correct model in the presence of a random effect compared to the traditional DIC.

For this simulation study, we examined two survival rates and four levels of variability. We set $S^* = 0.6$ or 0.8 and $\sqrt{S^*(1-S^*)/(\tau+1)} = 0$, 0.05, 0.10, and 0.20. We used a $\mathrm{Beta}(\tau S^*, \tau(1-S^*))$ distribution to generate the survival rates with the random effect. We generated 6 years of catch curve data 1,000 times and fit both the fixed effect and random effect model to each. The corresponding results are given in Table 4.

The p_D estimate is good under the fixed effect model. The focused DIC selected the fixed effects model correctly much more frequently than the traditional DIC. The focused DIC tended to select the fixed effects model more often than the traditional DIC in all cases, but it did select the random effects model more often as the size of the random effect increased. There was a huge increase in performance between the fDIC and the DIC when the fixed model was the correct model.

The probability of a Type I error, $\Pr[\text{Reject } H_0 | H_0, \sigma = 0]$, when $S = 0.8$ is 0.52 for the fDIC and 0.96 for the DIC. When $S = 0.6$, the probability of a Type I error is 0.37 for the fDIC and 0.82 for the DIC. Table 4 also gives the probabilities of a Type II error, $\Pr[\text{Fail to reject } H_0 | H_A]$, for $\sigma = 0.05$, 0.10, and 0.20 and $S = 0.8$ and 0.6. In terms of total error, the sum of Type I and Type II error, for each level of S and σ, it appears that fDIC outperforms DIC. This suggests that when the focused DIC can be calculated, it is a good improvement over the traditional DIC for model selection involving both fixed and random effect models.

8   Conclusions

For a single catch curve, the MLE and the least-squares estimators of the survival rate are biased, but for the MLE, the bias reduces to zero as the sample size approaches infinity (Chapman and Robson 1960; Jensen 1985, 1996; Murphy 1997; Dunn et al. 2002). For the UMVUE, there is no closed form estimate of the variance of the survival rate parameter(s) under the geometric or multinomial model. Using Bayesian analysis, we can obtain a closed form finite sample estimate of the survival rate(s) as well as a closed form estimate of its variance. It is also straightforward to obtain the Bayesian estimate of the instantaneous mortality rate Z.

We analyzed catch curve data from the Apostle Islands population of lake trout in Lake Superior (Pollock et al. 2007) and found that our closed form Bayesian estimate performed very similar to the UMVUE when Jeffreys prior is used. We also explored analytic relationships between the Bayesian estimator and the UMVUE

and MLE and showed that the latter estimates can be obtained as a limiting case of Bayes estimates.

To extend our model, we considered the situation where multiple years of catch data are collected. We first combined the separate years of data which, under the assumptions of the Chapman and Robson (1960) estimator, results in increased precision. However, information about possible assumption violations is lost, because no information about the individual years is part of this model. We relaxed the assumptions of Chapman and Robson (1960) and fit a random effects model to multiple years of catch data. The random effects model is an important advance because many real populations are likely to have substantial variation in survival rates between years due to environmental perturbations. The focused DIC also appears to be a good advance for model selection between fixed and random effects models.

The Bayesian approach to catch curve data analysis provides a broad and flexible method to extract the most information from the data without having to use marked animals. This is a real benefit for studies of animals that are difficult to mark. Bayesian methods are a very useful tool for data analysis. With software like WinBUGS and R widely available and free to use, we think these methods should become more popular in mainstream fisheries journals and in practice.

Acknowledgments

Schram, for generously sharing their data.

References

Chapman, D. G. and Robson, D. S. (1960), "The analysis of a catch curve," *Biometrics*, 16, 354–368.

Dunn, A., Francis, R., and Doonan, I. (2002), "Comparison of the Chapman-Robson and regression estimator of Z from catch-curve data when non-sampling stochastic error is present," *Fisheries Research*, 59, 149–159.

Jensen, A. L. (1985), "Comparison of catch-curve methods for estimation of mortality," *Transactions of the American Fisheries Society*, 114, 743–747.

— (1996), "Ratio estimation of mortality using catch curves," *Fisheries Research*, 27, 61–67.

Linton, B. C., Hansen, M. J., Schram, S. T., and Sitar, S. P. (2007), "Dynamics of a recovering lake trout population in eastern Wisconsin waters of Lake Superior, 1980-2001," *North American Journal of Fisheries Management*, 27, 940–954.

Messier, F. (1990), "Mammal life histories: Analyses among and within *spermophilus columbianus* life tables–a comment," *Ecology*, 71, 822–824.

Murphy, M. D. (1997), "Bias in Chapman-Robson and least-squares estimators of mortality rates for steady-state populations," *Fishery Bulletin*, 95, 863–868.

Pollock, K. H., Yoshizaki, J., Fabrizio, M. C., and Schram, S. T. (2007), "Factors affecting survival rates of a recovering lake trout population estimated by mark-recapture in Lake Superior, 1969-1996," *Transactions of the American*

Fisheries Society, 136, 185–194.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Robert, C. P. (2001), *The Bayesian choice*, Springer Texts in Statistics, 2nd ed., New York: Springer-Verlag.

Robson, D. S. and Chapman, D. G. (1961), "Catch curves and mortality rates," *Transactions of the American Fisheries Society*, 90, 181–189.

Seber, G. A. F. (1982), *The estimation of animal abundance and related parameters*, 2nd ed., New York: Macmillan Inc.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B*, 64, 583–639.

Udevitz, M. and Ballachey, B. E. (1998), "Estimating survival rates with age-structure data," *Journal of Wildlife Management*, 62, 779–792.

Williams, B., Nichols, J., and Conroy, M. (2002), *Analysis and Management of Animal Populations*, New York: Academic Press.

Table 1: Individual annual and combined posterior estimates (posterior standard deviations) of survival and instantaneous mortality rates from data on the Apostle Island population of male lake trout in Lake Superior, 2000-2005. Estimates are based on generating 10,000 MCMC samples followed by a burn-in of 1,000 samples for two parallel chains using WinBUGS.

Year	S	Z
2000	0.90	0.11
	(0.010)	(0.011)
2001	0.90	0.11
	(0.010)	(0.011)
2002	0.91	0.09
	(0.012)	(0.013)
2003	0.91	0.09
	(0.011)	(0.012)
2004	0.92	0.08
	(0.011)	(0.011)
2005	0.91	0.10
	(0.009)	(0.0102)
DIC = 49.2, p_D = 6.0		
Combined	0.90	0.11
	(0.005)	(0.005)
DIC = 43.6, p_D = 1.0		

Table 2: Posterior mean (posterior standard deviation) estimates of survival and instantaneous mortality rates from random effects model for catch curve data from Apostle Island populations of lake trout in Lake Superior, 2000-2005. Estimates based on generating 15,000 MCMC samples following a burn-in of 5,000 samples for two parallel chains using WinBUGS.

Parameter	Mean	s.d.
S_{2000}	0.89	0.009
S_{2001}	0.89	0.009
S_{2002}	0.90	0.010
S_{2003}	0.90	0.010
S_{2004}	0.91	0.010
S_{2005}	0.90	0.008
S^*	0.90	0.008
$\sqrt{\frac{S^*(1-S^*)}{\tau+1}}$	0.02	0.007
Z^*	0.11	0.009
DIC = 46.3, $p_D = 4$		
fDIC = 75.3, $p_D = 2$		

Table 3: Survival rate and standard deviation estimates from simulation study on random effects. The standard errors for all estimated parameters are less than 0.001, and the point estimates are based on 10,000 MCMC runs following a burn-in time of 1,000 runs for each of two chains.

S	σ		UMVUE Estimates	Bayesian Estimates with Jeffreys Prior
0.6	0.00	$\hat{S}$	0.59 (0.049)	0.59 (0.049)
		$\hat{\sigma}_{\hat{S}}$	0.026	0.026
	0.05	$\hat{S}$	0.59 (0.049)	0.59 (0.049)
		$\hat{\sigma}_{\hat{S}}$	0.026	0.026
	0.25	$\hat{S}$	0.59 (0.055)	0.59 (0.055)
		$\hat{\sigma}_{\hat{S}}$	0.026	0.026
0.75	0.00	$\hat{S}$	0.74 (0.032)	0.74 (0.032)
		$\hat{\sigma}_{\hat{S}}$	0.013	0.013
	0.05	$\hat{S}$	0.74 (0.032)	0.74 (0.033)
		$\hat{\sigma}_{\hat{S}}$	0.013	0.013
	0.25	$\hat{S}$	0.74 (0.036)	0.74 (0.036)
		$\hat{\sigma}_{\hat{S}}$	0.013	0.013
0.9	0.00	$\hat{S}$	0.90 (0.012)	0.90 (0.012)
		$\hat{\sigma}_{\hat{S}}$	0.003	0.003
	0.05	$\hat{S}$	0.90 (0.012)	0.90 (0.012)
		$\hat{\sigma}_{\hat{S}}$	0.003	0.003
	0.25	$\hat{S}$	0.89 (0.013)	0.89 (0.013)
		$\hat{\sigma}_{\hat{S}}$	0.004	0.004

Table 4: Model selection using fDIC with and without the presence of a random effect. Asterisks (**) indicate a significant difference ($p < .0001$) between the fDICs for the random model and the DICs for the fixed model. The estimates of p_D were all equal to 1.0 with standard errors of less than 0.002, while the estimates of p_D^f were all equal to 2.3 with standard errors of less than 0.002.

Size of the Random Effect, σ	Mean DIC, fixed effect model		Mean fDIC, random effect model			Proportion selecting fixed effect model	
						fDIC	DIC
0.00	100.4	(0.88)	93.5	(0.11)	**	0.48	0.04
$S = 0.8$ 0.05	103.5	(0.96)	93.6	(0.12)	**	0.44	0.03
0.10	111.7	(1.24)	94.1	(0.15)	**	0.38	0.03
0.20	165.8	(3.54)	95.9	(0.21)	**	0.21	0.01
0.00	71.8	(0.39)	72.2	(0.12)		0.63	0.18
$S = 0.6$ 0.05	75.1	(0.46)	72.9	(0.12)	**	0.54	0.12
0.10	82.6	(0.68)	74.1	(0.14)	**	0.39	0.08
0.20	124.3	(2.06)	77.4	(0.20)	**	0.15	0.03