Ecology, 90(2), 2009, pp. 356–362

Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning

José Miguel Ponciano,[1,5] Mark L. Taper,[2] Brian Dennis,[3] and Subhash R. Lele[4]

[1]*Centro de Investigación en Matemáticas, CIMAT A.C. Calle Jalisco s/n, Col. Valenciana, A.P. 402, C.P. 36240 Guanajuato, Guanajuato, México*
[2]*Montana State University, Department of Ecology, 301 Lewis Hall, Bozeman, Montana 59717-3460 USA*
[3]*Department of Fish and Wildlife Resources and Department of Statistics, University of Idaho, Moscow, Idaho 83844-1136 USA*
[4]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1 Canada*

Abstract. Hierarchical statistical models are increasingly being used to describe complex ecological processes. The data cloning (DC) method is a new general technique that uses Markov chain Monte Carlo (MCMC) algorithms to compute maximum likelihood (ML) estimates along with their asymptotic variance estimates for hierarchical models. Despite its generality, the method has two inferential limitations. First, it only provides Wald-type confidence intervals, known to be inaccurate in small samples. Second, it only yields ML parameter estimates, but not the maximized likelihood values used for profile likelihood intervals, likelihood ratio hypothesis tests, and information-theoretic model selection. Here we describe how to overcome these inferential limitations with a computationally efficient method for calculating likelihood ratios via data cloning. The ability to calculate likelihood ratios allows one to do hypothesis tests, construct accurate confidence intervals and undertake information-based model selection with hierarchical models in a frequentist context. To demonstrate the use of these tools with complex ecological models, we reanalyze part of Gause's classic *Paramecium* data with state–space population models containing both environmental noise and sampling error. The analysis results include improved confidence intervals for parameters, a hypothesis test of laboratory replication, and a comparison of the Beverton-Holt and the Ricker growth forms based on a model selection index.

Key words: AIC; Bayesian statistics; data cloning; frequentist statistics; hierarchical models; likelihood ratio; Markov chain Monte Carlo; maximum likelihood; model selection; profile likelihood; state–space models; stochastic population models.

Reports

Introduction

Reliable understanding of complex ecological data depends on the formulation of proper statistical models of the underlying processes. Hierarchical statistical models have proved highly useful for achieving such understanding in many ecological systems (see Table 1 in Lele et al. 2007). Such models allow researchers to incorporate variability in parameters that otherwise might be unrealistically treated as fixed. In addition, these models allow the incorporation of multiple layers of process and observation uncertainty. Stochastic population models with added observation error (De Valpine and Hastings 2002, Clark and Bjørnstad 2004, Staples et al. 2004, Dennis et al. 2006, Lele 2006, Newman et al. 2006, Sæther et al. 2007), stochastic models of species abundance distributions (Etienne and Olff 2005), and capture–recapture models with uncertain capture probabilities (George and Robert 1992) are just a few examples of this broad class of random effects models.

Until recently, computational difficulties rendered many frequentist statistical inferences for hierarchical models unfeasible. For all but the simplest models, computing the likelihood function needed for such inferences requires computing an intractable, high-dimensional integral. Inferences using computer intensive Bayesian methods side step this difficulty by simulating observations from a posterior distribution using one of the various Markov chain Monte Carlo (MCMC) algorithms (Robert and Casella 2005). Although other approaches are possible, the new data cloning (DC) algorithm by Lele et al. (2007) provides convenient tools to carry out frequentist estimation of the parameters in general hierarchical models. An often-repeated justification of the Bayesian approach is the fact that as sample size increases the Bayesian solution approaches the maximum likelihood solution (Walker 1969). The trick in data cloning is to apply a Bayesian methodology to a data set constructed by duplicating the original data set enough times that the Walker theorems apply.

Manuscript received 23 May 2008; revised 6 October 2008; accepted 21 October 2008. Corresponding Editor: A. M. de Roos.
[5] E-mail: ponciano@cimat.mx

DC applies Bayesian prior distributions and MCMC simulations to k copies (clones) of the data. If the number of clones is large, the sample mean vector of the resulting simulated posterior distribution corresponds to the maximum likelihood (ML) estimates of the parameters. As well, the sample variance-covariance matrix of the posterior, multiplied by k, provides estimates of variances and covariances of the ML estimates (inverse of the observed information matrix in the theory of ML estimation). The estimated variances can be used to obtain Wald-type confidence intervals (see Lele et al. [2007] for a full discussion).

Despite its general applicability, the DC method as proposed by Lele et al. (2007) has two important limitations. First, Wald-type symmetric confidence intervals often have less than nominal coverage rates (see Meeker and Escobar [1995] for an accessible discussion), and hence Wald-type DC confidence intervals share this defect. Improved confidence intervals exist; for example, it is well known that profile likelihood (PL) based confidence intervals tend to have better statistical coverage properties than the Wald-type intervals (Meeker and Escobar 1995, Pawitan 2001). Second, while DC provides ML estimates of the model parameters, it does not provide the value of the likelihood function evaluated at those estimates. This is a drawback because the maximized likelihood value is used to carry out model selection using information criteria (Burnham and Anderson 1998), to compute PL confidence intervals and to compute the likelihood ratio (LR) statistic for hypothesis testing (Pawitan 2001).

In this paper, we improve and extend the frequentist inferences for hierarchical models. First, we provide a straightforward algorithm for calculating LRs for hierarchical models using DC. We call it the data cloned likelihood ratio (DCLR) algorithm. This algorithm exploits the fact that just the LR and not the maximized likelihood value is needed for LR hypothesis tests, PL confidence intervals and for model selection (Thompson 1994, Pawitan 2001). Second, we describe the use of the DCLR algorithm for conducting LR hypothesis testing and for calculating the differences in values of information-theoretic criteria needed for model selection. For PL confidence intervals we extend the DCLR algorithm to accomplish PL calculations in a computationally efficient manner. We illustrate these algorithms by conducting various statistical inferences for state-space models fitted to data from Gause (1934) on the growth of laboratory *Paramecium* sp. populations.

Methods

The DC method described by Lele et al. (2007) is a general technique to compute ML estimates along with estimates of their asymptotic variances for hierarchical models. The DC method is applicable to any hierarchical model of the following form:

$$\begin{aligned} \mathbf{Y} &\sim f(\mathbf{y} \mid \mathbf{X} = \mathbf{x}, \boldsymbol{\phi}) \\ \mathbf{X} &\sim g(\mathbf{x} \mid \boldsymbol{\theta}) \end{aligned} \tag{1}$$

where $\mathbf{Y}$ is a vector of observations, and $\mathbf{X}$ is a vector of unobserved random quantities (often called latent variables or random effects) on which the observations depend. Through this article, random variables like the multivariate random vectors $\mathbf{Y}$ and $\mathbf{X}$ will be denoted with capital letters. Realizations of the random variables will be denoted with lower case letters. For instance, $\mathbf{x}$ and $\mathbf{y}$ in Eqs. 1 and 2 below, denote realizations (fixed constants, samples already drawn) from the random vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively. The probability density functions for $\mathbf{Y}$ and $\mathbf{X}$ are denoted by f and g, respectively. In addition, $\boldsymbol{\theta}$ is a vector of unknown fixed parameters affecting the latent variables in $\mathbf{X}$, and $\boldsymbol{\phi}$ a vector of unknown fixed parameters related to the observation model. In some cases, the latent variable model g is, in turn, formulated with hierarchical layers of random effects, but such cases do not need to be notationally distinguished in the methods we present. For the class of models represented by Eq. 1, the likelihood function is given by a multidimensional integral:

$$\begin{aligned} L(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{y}) &= \int\int \cdots \int f(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\phi}) g(\mathbf{x} \mid \boldsymbol{\theta}) dx_1 dx_2 \cdots dx_p \\ &= \int f(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\phi}) g(\mathbf{x} \mid \boldsymbol{\theta}) d\mathbf{x}. \end{aligned} \tag{2}$$

Notice that the dimension of the integral is the same as the number of components (p) in the vector $\mathbf{X}$, potentially quite high. For hierarchical models that are not based on the normal distribution, the integral is usually too difficult to evaluate, even numerically.

The DC method commandeers the Bayesian computational approach for frequentist purposes. In the standard Monte Carlo Bayesian approach, one generates samples from the posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{y})$ that is proportional to product of the likelihood function $L(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y})$ and a specified proper prior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\phi})$, without actually having to evaluate the likelihood function. In DC, one generates samples from the posterior distribution $\pi^{(k)}(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{y})$ that is proportional to the kth power $[L(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y})]^k$ of the likelihood and a specified proper prior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\phi})$, without having to evaluate the likelihood (Eq. 2). The expression $[L(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{y})]^k$ is the likelihood for k copies of the original data. Lele et al. (2007) show that, for k large enough, $\pi^{(k)}(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{y})$ converges to a multivariate normal distribution with mean equal to the ML estimate of the model parameters and variance–covariance matrix equal to $1/k$ times the inverse of the Fisher information matrix for the ML estimates (Appendix in Lele et al. 2007). This factor of $1/k$ adjusts for the fact that the cloned data set has k times more information than the original data set. DC proceeds by generating a large number of random numbers from the DC posterior distribution using MCMC; the sample mean vector of the generated random numbers provides the

ML estimates of the model parameters, and k times their sample variance–covariance matrix is an estimate of the asymptotic variance–covariance matrix for the ML estimates.

LR and PL computation

In what follows we show how DC, combined with the result in Appendix A, can be used to compute the ratio of likelihoods evaluated at two different sets of parameter values.

Let $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\phi}^{(0)})$ be a particular set of parameter values (i.e., a point in the parameter space) and let $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\phi}^{(1)})$ be another such point. The DCLR algorithm computes the ratio $L(\boldsymbol{\theta}^{(0)}, \boldsymbol{\phi}^{(0)})/L(\boldsymbol{\theta}^{(1)}, \boldsymbol{\phi}^{(1)})$. Such a ratio is needed, for instance, for a LR test between two nested models. In that case, $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\phi}^{(0)})$ are the data cloned ML estimates under the constrained model and $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\phi}^{(1)})$ are the data cloned ML estimates under the full model. The steps in the DCLR algorithm are the following:

1) Generate m random data samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(m)}$ from the conditional distribution

$$h\left(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\phi}^{(1)}\right) \propto f\left(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\phi}^{(1)}\right) g\left(\mathbf{x} \mid \boldsymbol{\theta}^{(1)}\right)$$

of the latent variables. These samples are obtained with a straightforward MCMC algorithm using $g(\mathbf{x} \mid \boldsymbol{\theta}^{(1)})$ as the prior distribution and $f(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\phi}^{(1)})$ as the likelihood in a typical Bayesian calculation.

2) Estimate the desired LR as

$$\frac{L\left(\boldsymbol{\theta}^{(0)}, \boldsymbol{\phi}^{(0)}\right)}{L\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\phi}^{(1)}\right)} \approx \frac{1}{m} \sum_{j=1}^{m} \frac{f\left(\mathbf{y} \mid \mathbf{x}^{(j)}, \boldsymbol{\phi}^{(0)}\right) g\left(\mathbf{x}^{(j)} \mid \boldsymbol{\theta}^{(0)}\right)}{f\left(\mathbf{y} \mid \mathbf{x}^{(j)}, \boldsymbol{\phi}^{(1)}\right) g\left(\mathbf{x}^{(j)} \mid \boldsymbol{\theta}^{(1)}\right)}.$$

See Thompson (1994) and Appendix A for a derivation of this approximation.

The above algorithm extends to the calculation of profile likelihoods. Suppose we are interested in the profile likelihood for a subset $\boldsymbol{\theta}_S$ of the components of $\boldsymbol{\theta}$. The remaining components of $\boldsymbol{\theta}$ are denoted by $\boldsymbol{\theta}_C$ so that $\boldsymbol{\theta} = [\boldsymbol{\theta}_S, \boldsymbol{\theta}_C]$. One could also single out components of $\boldsymbol{\phi}$ for study, but for simplicity of exposition we describe the algorithm in terms of $\boldsymbol{\theta}_S$ only. Then, the steps to calculate the profile likelihood are

1) Calculate ML estimates $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ using DC.

2) For the parameter(s) of interest $\boldsymbol{\theta}_S$ select an array of dimension J of fixed values/vectors $\boldsymbol{\theta}_S^{(1)}, \boldsymbol{\theta}_S^{(2)}, \cdots, \boldsymbol{\theta}_S^{(J)}$ bracketing the ML estimate(s) sufficiently broadly for the desired confidence levels to be contained.

3) For each value $\boldsymbol{\theta}_S^{(1)}, \boldsymbol{\theta}_S^{(2)}, \cdots, \boldsymbol{\theta}_S^{(J)}$ in turn, conduct data cloning to maximize the likelihood for all other parameters $\boldsymbol{\theta}_C$. Then, for each array element, this produces a vector of constrained ML estimates $(\{\hat{\boldsymbol{\theta}}_C^{(1)}, \hat{\boldsymbol{\phi}}^{(1)}\}, \{\hat{\boldsymbol{\theta}}_C^{(2)}, \hat{\boldsymbol{\phi}}^{(2)}\}, \cdots, \{\hat{\boldsymbol{\theta}}_C^{(J)}, \hat{\boldsymbol{\phi}}^{(J)}\})$.

4) Generate m random samples $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(m)}$ from $h(x \mid y, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$.

5) Then, for each $\boldsymbol{\theta}_S^{(i)}$, $i = 1, 2, \cdots, J$ calculate the following sample average:

$$\frac{L\left(\boldsymbol{\theta}_S^{(i)}, \hat{\boldsymbol{\theta}}_C^{(i)}, \hat{\boldsymbol{\phi}}^{(i)}\right)}{L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})} \approx \frac{1}{m} \sum_{j=1}^{m} \frac{f\left(\mathbf{y} \mid \mathbf{x}^{(j)}, \hat{\boldsymbol{\phi}}^{(i)}\right) g\left(\mathbf{x}^{(j)} \mid \boldsymbol{\theta}_S^{(i)}, \hat{\boldsymbol{\theta}}_C^{(i)}\right)}{f\left(\mathbf{y} \mid \mathbf{x}^{(j)}, \hat{\boldsymbol{\phi}}\right) g\left(\mathbf{x}^{(j)} \mid \hat{\boldsymbol{\theta}}\right)}.$$

Note that the above sample averages can be evaluated simultaneously for the whole array of parameter values using the single extra MCMC chain generated from $h(x \mid y, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$, along with vectorized calculations.

6) The plot of the surface of sample averages versus the fixed parameter values $\boldsymbol{\theta}_S^{(1)}, \boldsymbol{\theta}_S^{(2)}, \cdots, \boldsymbol{\theta}_S^{(J)}$ is the relative profile likelihood for the parameter of interest, $\boldsymbol{\theta}_S$.

A $100(1 - \alpha)\%$ confidence region for $\boldsymbol{\theta}_S$ based on the chi-square approximation to the LR consists of the region of values of $\boldsymbol{\theta}_S$ for which -2 times the log-profile LR is less than the $100(1 - \alpha)$th percentile from a chi-square distribution. The degrees of freedom of the chi-square distribution is equal to the number of parameters in $\boldsymbol{\theta}_S$. For one parameter, the 95% critical region occurs when the relative profile likelihood is above 0.1464. We emphasize that the entire profile likelihood requires only one MCMC simulation (the values $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(m)}$ simulated using $h(x \mid y, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ in step 4).

Model selection with information criteria

The differences of Akaike Information Criteria (AIC) needed for model comparison depend only on the ratio of the maximized model likelihoods and the difference in the number of parameters between the two models. (Burnham and Anderson 2002):

$$\mathrm{AIC}_1 - \mathrm{AIC}_2 = -2\ln\left(\frac{\hat{L}_1}{\hat{L}_2}\right) + 2(d_1 - d_2) \qquad (3)$$

where d_1 and d_2 are the number of parameters estimated under models 1 and 2, respectively. Differences of other common information criteria such as the AIC_c (Hurvich and Tsai 1989) and the SIC (Schwarz 1978) also depend on the likelihood only through the likelihood ratio and thus can be calculated using the DCLR algorithm. For a collection of more than two models, the LRs need to be calculated for all pairs of models. However, LRs for all pairs can be generated from the smaller set of LRs resulting from pairing each model with a single reference model by taking appropriate quotients (Burnham and Anderson 2002). As with the profile likelihood, all LRs in the smaller set can be evaluated using a single set of latent variables (random effects) simulated from their posterior distribution under the reference model.

Examples Using Gause's Experiments

Data description

To illustrate the methods proposed in this paper, we revisited the single-species population growth data from Gause's (1934) laboratory experiments with *Paramecium aurelia*. Lele et al. (2007) used the same data set to exemplify the DC method for calculating ML parameter estimates and standard errors, and here we extend their analyses to include LR hypothesis testing, information

criterion based model identification, and PL confidence intervals.

Gause initiated three independent replicate liquid cultures on day 0 at a concentration of two individuals per 0.5 cm^3 of culture media. Then, on days 2–19, he took daily 0.5-cm^3 samples of the microbe cultures and counted the number of cells in each sample. Although Lele et al. (2007) fitted a hierarchical state–space model to Gause's data set taking into account both process noise and observation error, many inferential questions remain unanswered. For instance, do the three replicates represent realizations of a single process or do they represent three different processes? Which model, among a suite of stochastic population growth models, best describes the data? Are the Wald-type confidence intervals proposed by Lele et al. adequate given that Gause's time series consist of just 20 time steps? In what follows we illustrate how LR tests, PL confidence intervals, and AIC-based model selection all implemented using the DC algorithms described above may serve to answer these questions.

Description of alternative models

We used two discrete time stochastic models of population dynamics to describe Gause's data. The two models were state-space versions of the Ricker and the Beverton-Holt models. Each model incorporates a latent variable component $\mathbf{X}$ to describe an unobserved time series of actual population abundance (numbers of individuals per 0.5 cm^3 culture media). The elements of $\mathbf{X}$ are denoted as X_t, where the time index t goes from 0 to q. The latent variable component contains density dependence (Ricker or Beverton-Holt) and stochastic process noise. Each model incorporates an observation component $\mathbf{Y}$ as well, to account for variability caused by sampling.

The latent variable component is formulated in terms of the underlying log-abundance $X_t = \ln(N_t)$ and has the form: $X_t = m(X_{t-1}) + \sigma Z_t$, where $Z_t \sim \mathcal{N}(0, 1)$ and σ is a positive constant. The function $m(\cdot)$ specifies the deterministic skeleton of the population dynamics model (that is, the type of density dependence) as a function of the log abundance x:

$$m(x) = \begin{cases} a + x - be^x & \text{(Ricker)} \\ \ln\lambda + x - \ln[1 + \beta\exp(x)] & \text{(Beverton-Holt).} \end{cases} \quad (4)$$

In this equation, b, λ, and β are positive constants. The joint distribution for a single time series of population log-abundances, $\mathbf{x} = [x_1, x_2, \cdots, x_q]^\mathsf{T}$ (where T denotes the transpose), is following Dennis and Taper (1994):

$$g(\mathbf{x}|\boldsymbol{\theta}) = \prod_{t=1}^{q} (\sigma^2 2\pi)^{-(1/2)} \exp\left\{ -\frac{[x_t - m(x_{t-1})]^2}{2\sigma^2} \right\}. \quad (5)$$

For the Ricker model, $\boldsymbol{\theta} = [a, b, \sigma^2]^\mathsf{T}$, and for the Beverton-Holt model, $\boldsymbol{\theta} = [\lambda, \beta, \sigma^2]^\mathsf{T}$. The initial log-abundance is x_0, and for Gause's data $x_0 = \ln 2$. The distribution $g(\mathbf{x} \mid \boldsymbol{\theta})$ will serve as the latent variable component in the hierarchical model (Eq. 1).

The *Parameceium* samples were pipetted from well-mixed cultures, and therefore a Poisson sampling model is appropriate. Let $\mathbf{y} = [y_2, y_3, \cdots, y_q]^\mathsf{T}$ denote the time series of sample counts from a single population culture. The joint distribution for $\mathbf{y}$ given the underlying population trajectory is $\mathbf{x}$, is given by a product of Poisson probabilities:

$$f(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\phi}) = \prod_{t=2}^{q} \frac{e^{-n_t} n_t^{y_t}}{y_t!} \quad (6)$$

where $n_t = \exp(x_t)$. The distribution $f(\mathbf{y} \mid \mathbf{x}, \boldsymbol{\phi})$ will serve as the observation component of the hierarchical model (Eq. 1). Note, for this model, the parameter $\boldsymbol{\phi}$ does not exist because the Poisson sampling process is specified by its expectation, $n_t = \exp(x_t)$ in this case, with no additional unknown parameters.

Likelihood function

The data consists of trajectories from three population cultures. Assuming the cultures are independent from each other the likelihood function is

$$L(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3) = \prod_{j=1}^{3} \int f(\mathbf{y}_j \mid \mathbf{x}_j) g(\mathbf{x}_j \mid \boldsymbol{\theta}_j) d\mathbf{x}_j. \quad (7)$$

The parameter vector for the jth culture is denoted by $\boldsymbol{\theta}_j$, $j = 1, 2, 3$.

LR test.—Do the three replicates represent realizations of a single process or do they represent three different processes? In ecological data, heterogeneities abound. One would hope that in the controlled laboratory setting the biological processes would be nearly homogeneous, but one should test. Homogeneous populations will have identical parameters in their growth models, while parameters for heterogeneous populations will be distinct. The corresponding statistical hypotheses are:

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_3 = \boldsymbol{\theta}$$
$$H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \neq \boldsymbol{\theta}_3.$$

The vector $\boldsymbol{\theta}$ contains the parameter values in common among the populations. The LR test statistic is

$$\Lambda = \frac{L_0(\hat{\boldsymbol{\theta}})}{L_1(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \hat{\boldsymbol{\theta}}_3)}.$$

Here, $\hat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$, $L_0(\hat{\boldsymbol{\theta}}) = L(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}})$ (see Eq. 6) and $\hat{\boldsymbol{\theta}}_1$, $\hat{\boldsymbol{\theta}}_2$, $\hat{\boldsymbol{\theta}}_3$ are the ML estimates under the alternative hypothesis. For illustrative purposes we fitted the Beverton-Holt model to *P. aurelia* data. The DC ML estimates for the parameters under the null hypothesis were $\hat{\lambda} = 2.274$, $\hat{\beta} = 0.00235$, $\hat{\sigma} = 0.1274$. Under the alternative the DC ML estimates for the parameters were $\hat{\lambda}_1 = 2.173$, $\hat{\beta}_1 = 0.00213$, $\hat{\sigma}_1 = 0.1187$, $\hat{\lambda}_2 = 2.2610$, $\hat{\beta}_2 = 0.00247$, $\hat{\sigma}_2 = 0.1390$, and $\hat{\lambda}_3 = 2.4160$, $\hat{\beta}_3 = 0.00251$, $\hat{\sigma}_3 = 0.1062$.

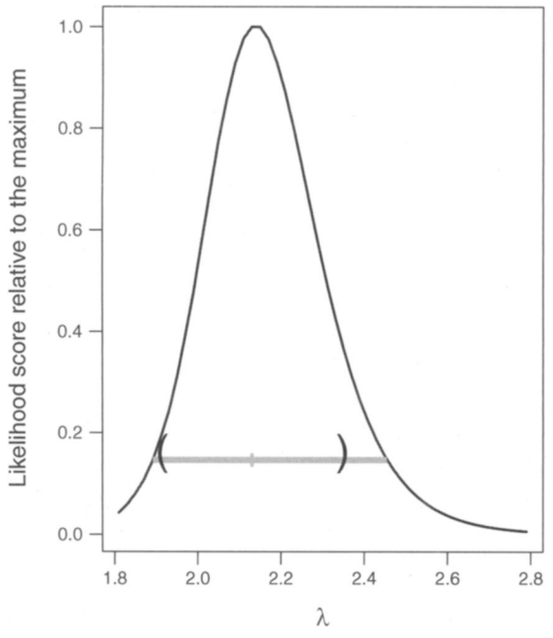

FIG. 1. Profile likelihood for the parameter λ for the first of the three replicated time series for *Paramecium aurelia*. The profile likelihood (solid, black curve) is given relative to the maximum, i.e., the likelihood evaluated at the maximum likelihood (ML) estimates. The gray horizontal bar denotes the 95% profile likelihood confidence, and the gray short vertical dash denotes the ML estimate. The black parentheses mark the Wald-type 95% confidence limits. The ML estimates used in this plot were $\hat{\lambda} = 2.13$, $\hat{\beta} = 0.2047$, and $\hat{\sigma} = 0.1203$; these were calculated by passing preliminary estimates from data cloning through the likelihood ratio algorithm (see Supplement for details).

Reports

The LR test failed to reject the null hypothesis of common parameters $-2 \ln \Lambda = 3.87$ df $= 9 - 3 = 6$, $P = 0.70$.

Model selection

In addition to the Beverton-Holt model we fitted the Ricker model to the *P. aurelia* data under the assumption of common parameters among replicates. The parameters estimates under the Ricker model were $\hat{a} = 0.7742$, $\hat{b} = -0.001429$, $\hat{\sigma} = 0.1375$. Hypothesis testing does not lend itself easily to comparing these two models because the models are not nested. Therefore we used the AIC for model comparison. Taking model 1 to be the Ricker model ($d_1 = 3$) and model 2 to be the Beverton-Holt model ($d_2 = 3$) the difference in AIC values was 3.738764. Because this difference is positive we conclude that the Beverton-Holt model provides a better description of the data than does the Ricker model. AIC differences greater than 2 are generally thought to be significant, and differences greater than 3 very significant (Burnham and Anderson 2002, Taper 2004).

Profile likelihood and confidence intervals

Now we illustrate the use of the profile likelihood algorithm to compute confidence intervals for the parameter λ in the Beverton-Holt model. To emphasize the difference between Wald-type and profile likelihood-type confidence intervals for small samples we performed calculations for just one of the three *P. aurelia* cultures. The resulting profile likelihood for λ was somewhat asymmetric, and the corresponding 95% confidence limits were asymmetric around the ML point estimate (Fig. 1). Furthermore the Wald-type confidence interval is liberal relative to profile likelihood confidence interval with the Wald interval being contained inside the profile likelihood interval. The main difference between the two is a lower upper limit for the Wald interval than for the profile likelihood interval (Fig. 1).

DISCUSSION

The DC method produces ML estimates and estimates of standard errors, from which Wald-type confidence intervals can be constructed based on asymptotic ML theory.

Profile likelihood intervals, however, often have better coverage than Wald-type confidence intervals (Meeker and Escobar 1995, Pawitan 2001; and see Appendix B). Also, profile likelihood intervals are invariant to transformation of the parameters, whereas Wald-type intervals are not. Wald-type intervals vary in coverage depending on the transformation used.

In addition to providing confidence intervals, an entire likelihood profile is a convenient graphical summary about the information contained in the data about a particular parameter. In particular, likelihoods for hierarchical models often have multiple local maxima (Searle et al. 1992, Dennis et al. 2006), and the profile likelihood can serve as a diagnostic tool to reveal such problems.

Confidence intervals and hypothesis tests can, of course, be done with parametric bootstrapping. The necessary computational effort to obtain parametric bootstrap confidence intervals (see Appendix B) is greater than that needed for profile likelihood, but not substantially. Parametric bootstrap confidence intervals share the transformation invariance property with profile likelihood intervals. Furthermore, confidence intervals for any functions of parameters are often easier to calculate via parametric bootstrap.

However the information portrayed in the bootstrap sampling distribution of the parameter estimates is somewhat different than the information portrayed in the profile likelihood: the profile likelihood depicts the evidence contained in the data regarding different values of a given parameter (Royall 2000), whereas the bootstrap sampling distribution depicts the reliability of the parameter estimate (Efron 2003). Here, reliability is measured in terms of the variation of the parameter estimate under repeated hypothetical realizations of the data generation process.

In our experience, constructing the profile likelihood using the LR algorithm helps improve the numerical accuracy of the ML estimates under DC. In DC, the numerical accuracy increases as the number data clones used increases; however, the calculation time increases as well. Once an approximate ML estimate is found with DC, then using that estimate to calculate the profile likelihood can reveal the location of the likelihood peak more accurately. As a matter of routine we now calculate profile likelihoods in two passes: first, with the approximate ML estimate found with a preliminary DC analysis, and second, with the improved ML estimate found from the first profile likelihood. Explicit details of how the profile likelihood was calculated are given in the R code shown in the Supplement.

Concluding remarks

A suite of classical frequentist and evidential inferences is now available for complex biological problems that are modeled using hierarchical statistical models. In addition to the maximum likelihood parameter estimates and Wald confidence intervals made feasible with the DC algorithm (Lele et al. 2007), profile likelihood confidence intervals, hypotheses tests and information theoretic model selection can be accomplished with the DCLR algorithm as described in this paper. Until now, hierarchical models have mostly been analyzed in a Bayesian framework because frequentist inferences were computationally intractable. The choice between Bayesian and frequentist approaches is no longer a matter of feasibility or convenience but rather can and should be based on the philosophical foundations of statistical inference preferred by the investigator. This choice, while philosophical, has profound practical implications. In this paper, we have focused on the tools that we are making available for frequentist/likelihood inference in hierarchical models and not on the philosophical question. We urge the reader to explore the vast and controversial literature that is developing surrounding the philosophical question (see Barnett 1999, Taper and Lele 2004, Cox 2006, and Thompson 2007 for comparative discussions). Dennis (2004) and Lele and Dennis (2009) express some of our concerns with using the Bayesian approach in science.

Acknowledgments

The authors thank two anonymous reviewers for their valuable suggestions and comments. This work was supported in part by Montana Fish, Wildlife, and Parks contract 060327 to M. L. Taper, NSERC funding to S. R. Lele, and the Center for Research in Mathematics (CIMAT A.C.).

Literature Cited

Barnett, V. 1999. Comparative statistical inference. Third edition. John Wiley and Sons, Chichester, UK.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

Clark, J. S., and O. N. Bjørnstad. 2004. Population time series: process variability, observation errors, missing values, lags and hidden states. Ecology 85:3140–3150.

Cox, D. R. 2006. Principles of statistical inference. Cambridge University Press, Cambridge, UK.

Dennis, B. 2004. Statistics and the scientific method in ecology. Pages 327–378 *in* M. L. Taper and S. R. Lele, editors. The nature of scientific evidence: statistical, empirical and philosophical considerations. University of Chicago Press, Chicago, Illinois, USA.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. 2006. Estimating density dependence, process noise and observation error. Ecological Monographs 76:323–341.

Dennis, B., and M. L. Taper. 1994. Density dependence in time series observations of natural populations: estimation and testing. Ecological Monographs 64:205–224.

De Valpine, P., and A. Hastings. 2002. Fitting population models incorporating process noise and observation error. Ecological Monographs 72:57–76.

Efron, B. 2003. Second thoughts on the bootstrap. Statistical Science 18:135–140.

Etienne, R. S., and H. Olff. 2005. Confronting different models of community structure to species abundance data: a Bayesian model comparison. Ecology Letters 8:493–504.

Gause, G. F. 1934. The struggle for existence. Williams and Wilkins, Baltimore, Maryland, USA.

George, E. I., and C. P. Robert. 1992. Capture–recapture estimation via Gibbs sampling. Biometrika 79:677–683.

Hurvich, C. M., and C. L. Tsai. 1989. Regression and time-series model selection in small samples. Biometrika 76:297–307.

Lele, S. R. 2006. Sampling variability and estimates of density dependence: a composite- likelihood approach. Ecology 87: 189–202.

Lele, S. R., and B. Dennis. 2009. Bayesian methods for hierarchical models: Are ecologists making a Faustian bargain? Ecological Applications 19, *in press*.

Lele, S. R., B. Dennis, and F. Lutscher. 2007. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov Chain Monte Carlo methods. Ecology Letters 10:551–563.

Meeker, W., and L. A. Escobar. 1995. Teaching about approximate confidence regions based on maximum likelihood estimation. American Statistician 49:48–53.

Newman, K. B., S. T. Buckland, S. T. Lindley, L. Thomas, and C. Fernández. 2006. Hidden process models for animal population dynamics. Ecological Applications 16:74–86.

Pawitan, Y. 2001. In all likelihood: statistical modeling and inference using likelihood. Oxford University Press, Oxford, UK.

Robert, C. P., and G. Casella. 2005. Monte Carlo Statistical methods. Second edition. Springer-Verlag, New York, New York, USA.

Royall, R. 2000. On the probability of observing misleading statistical evidence. Journal of the American Statistical Association 95:760–768.

Sæther, B.-E., M. Lillegård, V. Grøtan, F. Filli, and S. Engen. 2007. Predicting fluctuations of reintroduced ibex populations: the importance of density dependence, environmental stochasticity and uncertain population estimates. Journal of Animal Ecology 76:326–336.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. John Wiley and Sons, New York, New York, USA.

Staples, D. F., M. L. Taper, and B. Dennis. 2004. Estimating population trend and process variation for PVA in the presence of sampling error. Ecology 85:923–929.

Taper, M. L. 2004. Model identification from many candidates. Pages 448–524 *in* M. L. Taper and S. R. Lele, editors. The

nature of scientific evidence: statistical, philosophical and empirical considerations. The University of Chicago Press, Chicago, Illinois, USA.
Taper, M. L., and S. R. Lele, editors. 2004. The nature of scientific evidence: statistical, philosophical and empirical considerations. The University of Chicago Press, Chicago, Illinois, USA.
Thompson, B. 2007. The nature of statistical evidence. Springer, New York, New York, USA.
Thompson, E. A. 1994. Monte Carlo likelihoods in genetic mapping. Statistical Science 9:355–366.
Walker, A. M. 1969. On asymptotic behaviour of posterior distributions. Journal of the Royal Statistical Society Series B 31:80–88.

APPENDIX A

A proof showing that the ratio of two likelihoods initially written as the ratio of two integrals can be written as the integral of a ratio, as shown in steps 2 and 5 of the data cloned likelihood ratio algorithm and of the profile likelihood algorithm, respectively (*Ecological Archives* E090-027-A1).

APPENDIX B

Simulation study (*Ecological Archives* E090-027-A2).

SUPPLEMENT

R program illustrating the calculations needed to construct a profile likelihood for one of the stochastic Beverton-Holt state space model parameters (*Ecological Archives* E090-027-S1).