BULLETIN OF MARINE SCIENCE, 70(2): 455–471, 2002 MOTE SYMPOSIUM INVITED PAPER

REVIEW OF METHODS FOR FITTING TIME-SERIES MODELS WITH PROCESS AND OBSERVATION ERROR AND LIKELIHOOD CALCULATIONS FOR NONLINEAR, NON-GAUSSIAN STATE-SPACE MODELS

Perry de Valpine

ABSTRACT

A key challenge for analyzing fisheries time-series data has been to incorporate sources of uncertainty such as process error, observation error, and model-structure uncertainty. Recent years have seen promising advances in methods for handling the first two together in a state-space framework, but likelihood calculations for state-space models require high-dimensional integrals, which make their use computationally challenging. The first section of this paper reviews model-fitting methods that use a state-space model structure, including errors-in-variables methods, Bayesian methods that do and do not use the state-space likelihood, and the possibility of classical likelihood analysis with nonlinear, non-Gaussian state-space models. It also discusses the relationship between true likelihood calculations and errors-in-variables likelihoods, as well as the role of Monte Carlo methods in implementing Bayesian and/or state-space model analyses. The second section introduces a numerical method for calculating state-space likelihoods without Monte Carlo methods and gives examples in a classical maximum-likelihood framework. The method is applicable when the dimension of the state space at each time step is low. Although recent advances in model-fitting and analysis methods are promising, inferences from noisy data and complex processes will continue to be variable and uncertain.

Some types of inference about fisheries dynamics rely on fitting time-series data to mathematical models of population dynamics and harvesting. Not long ago, a wide disjunction separated roughly realistic models and statistically usable models; one could not take just any model and estimate parameters and their uncertainty without taking more or less awkward steps. One reason was that any realistic model must include variability in how populations change—process error—as well as variability in estimates of the true state of the population—observation error. Together, these two greatly complicate the statistical relationship between model and data. Recent years have seen wholesale improvements in methods for incorporating both process and observation error into statistical analysis of fisheries data, but these methods have also brought to the fore several complex issues.

The simplest method for handling process and observation error is to assume that one or the other is absent (see, e.g., Hilborn, 1979; Polacheck et al., 1993; Raftery et al., 1995; Kinas, 1996). Later methods include both but require the assumption that a ratio involving their variances is known (Ludwig and Walters, 1981, 1989; Walters and Ludwig, 1981; Collie and Sissenwine, 1983; Ludwig et al., 1988; Schnute, 1994; Schnute and Richards, 1995; Richards et al., 1997; Fargo and Richards, 1998). What these methods have in common is that the fit of model to data is measured by only a single estimated population trajectory that might have produced the data. The latest methods incorporate the full range of population trajectories that might have produced the data to calculate the fit of model to data, and they do not require assumptions about relative magnitudes of

process and observation error (Mendelssohn, 1988; Sullivan, 1992; Pella, 1993; Gudmundsson, 1994; McAllister et al., 1994; Schnute, 1994; Freeman and Kirkwood, 1995; Kimura et al., 1996; McAllister and Ianelli, 1997; Meyer and Millar, 1999; Millar and Meyer, 2000a,b; de Valpine and Hastings, 2002). Mathematically speaking, these latest methods use likelihood calculations that integrate over unknown process errors or population states. The framework for these methods is called a 'state-space model', but some other methods also use a state-space model structure without the integrated likelihood calculations as part of the fitting method.

The earliest applications of state-space model likelihoods used linear models with additive Gaussian process and observation error, in some cases to approximate nonlinear and/or non-Gaussian models (Mendelssohn, 1988; Sullivan, 1992; Pella, 1993; Gudmundsson, 1994; Schnute, 1994; Freeman and Kirkwood, 1995; Kimura et al., 1996). In these cases the likelihood calculation is given or approximated by the Kalman filter or extended Kalman filter. These analyses used maximum-likelihood parameter estimates with a classical (as opposed to a Bayesian) view of parameters. Later work combined two important developments: use of state-space models with a Bayesian view of parameters and use of Monte Carlo numerical methods for the analysis, which facilitates both the Bayesian treatment of parameters and the state-space treatment of likelihoods (McAllister et al., 1994; McAllister and Ianelli, 1997; Meyer and Millar, 1999; Millar and Meyer, 2000a,b). (Please note that the present paper is not intended to express any opinion on the relative merits of the Bayesian and classical views.)

Here, I discuss a non-Monte Carlo numerical method for calculating likelihoods for nonlinear, non-Gaussian state-space models. This method is practical for models such as biomass-dynamics models that have low-dimensional state spaces at each time step, but it would be impractical for models such as explicitly age-structured models with high-dimensional state spaces at each time step. I illustrate the method with the work of de Valpine and Hastings (2002), using classical maximum-likelihood parameter estimation. Before discussing this approach, I review model-fitting methods to highlight some of their relationships. I introduce the state-space framework; introduce the way Monte Carlo samples are used for Bayesian parameter analysis, state-space likelihood integration, or both together; and classify several methods according to whether they calculate state-space likelihoods and whether they use a Bayesian or a classical treatment of parameters.

Throughout, I use the term 'parameter' to refer only to model parameters such as intrinsic population growth rate or survival rates and not to unknown process errors or population states, as it is sometimes used (correctly) in Bayesian analysis. This distinction matters because parameters and states (or process errors) play different roles with respect to likelihood calculations and asymptotic likelihood theory (both discussed below), even in Bayesian analysis.

State-space Models

State-space models (Harvey, 1989, 1993; Schnute, 1994) include three fundamental types of quantities: population states, observations, and parameters (ignoring control variables for simplicity). Population states describe the true state of the population, such as its total size and/or stage, age, or spatial structure. Denote the population state at time t as vector X_t, and consider $X = (X_1, \ldots, X_T)$ to be one big vector containing the population states at all times over some study period. Observations are inaccurate estimates of some

or all of the population states. For example, observations may include estimates of total size but not of stage structure, or vice versa, or both. Denote the observation from time t as vector Y_t, and consider $Y = (Y_1, \ldots, Y_T)$ to be one big vector containing all of the Y_t values for the study period. Finally, denote parameters in a vector Θ.

A simple example for illustration is a Ricker model with stochastic population growth rate and observations from independent surveys with error. Suppose the true state of a population at time t is the biomass, and the next population state is determined by

$$X_t = \left(X_{t-1} - C_{t-1}\right) e^{r\left(1-(X_{t-1}-C_{t-1})/K\right)+v_{t-1}} \qquad \text{Eq. 1}$$

where r is intrinsic population growth rate, K is carrying capacity, v_t is normally distributed with mean 0 and variance σ_v^2, and C_t is catch in year t, assumed to be accurately measured. The randomness of v_t provides the process noise of this model.

The true state of the system can never be known, but suppose we have estimates from independent surveys,

$$Y_t = q_t X_t \qquad \text{Eq. 2}$$

where q_t is a random variable with some distribution, such as a beta distribution, whose mean and variance, μ_q and σ_q^2, need to be estimated. The randomness in q_t provides observation error. For example, if one knew every q_t exactly, then one could calculate the state X_t exactly from Y_t, but because q_t is random, so is any estimate of X. Given a series of data, Y, the goal of analysis is to estimate parameters $\Theta = (r, K, \sigma_v^2, \mu_q, \sigma_q^2)$ and states, X, or at least the most recent state, X_T, and to describe in some sound way our uncertainty about those estimates.

These models are called 'state-space models' because they have separate equations for underlying state dynamics and for observations of the states. The 'state space' of their name refers to the entire range of possible trajectories through time of the population states. What distinguishes state-space models from other models is not the biology that is represented—one could use any biological model for the state dynamics. Rather it is the potential to handle properly the probability structure created by states and observations.

Likelihoods

The sequences of states and observations together are a jointly distributed random variable with complicated, nonlinear, non-Gaussian lack of independence. Although this random variable has many dimensions, one can think of it conceptually as like a two-dimensional random variable with probability density $f_\Theta(X, Y)$. The subscript Θ on the probability density indicates that it depends on the parameters, Θ. It is easy to write down (but hard to evaluate) the likelihood of the observations, which is the marginal probability:

$$L(\Theta \mid Y) = \int f_\Theta(X, Y) dX \qquad \text{Eq. 3}$$

where dX indicates integration over the many dimensions of true states over the study period. This integral is the heart of the matter of incorporating process and observation

error together into any likelihood-based inference, whether Bayesian or classical; I refer to it here as the 'true likelihood'. One way to think conceptually about the true likelihood is that it accounts for all possible underlying trajectories of state values that might have produced the data.

Schnute (1994) used the state-space framework to discuss the relationship between errors-in-variables (EV) and Kalman Filter (KF) methods. For linear, Gaussian models, the true likelihood is given by the KF. That is, if the state-space model is

$$X_t = AX_{t-1} + v_t \qquad \text{Eq. 4}$$

$$Y_t = BX_t + \varepsilon_t \qquad \text{Eq. 5}$$

where v_t and ε_t are normally distributed, then the KF provides (and can be derived from) a calculation of the true likelihood, Eq. 3. The term 'true likelihood' is intended to distinguish Eq. 3 from other uses of the term 'likelihood' in the fisheries literature, notably for EV 'likelihoods' (see, e.g., Ludwig and Walters, 1981; Schnute, 1994; Schnute and Richards, 1995), and to include nonlinear, non-Gaussian models rather than just the linear, Gaussian models of the KF.

It is important to realize that we could alternatively view the sequences of $v = (v_1,\ldots, v_T)$ and Y as the jointly distributed random variables of interest. That is, we might consider working with v instead of X. For fixed parameters, either (v,Y) or (X,Y) define everything about a specific population trajectory. Either case involves two groups of random variables, those whose values are known and those whose values are not known, i.e., observations and either states or process errors, respectively.

The likelihood integration gives the same answer whether we integrate over states or process errors. That is, as explained in the appendix,

$$\int f_\Theta(X,Y)dX = \int f_\Theta(v,Y)dv\,. \qquad \text{Eq. 6}$$

Readers who wish to skip the brief math in the appendix can note that, intuitively, the sum of probabilities over the full range of v (the integral over v) is the same as the sum of probabilities of a function of v (i.e., the integral over X).

The Bayesian and classical views of the unknown population states are the same. In both views, unknown population states are random variables, and the likelihood integral, Eq. 3, is valid. To place this comment in perspective, recall that a fundamental difference between Bayesian and classical analysis is that Bayesian analysis treats parameters as random variables, whereas classical analysis does not. In the Bayesian view then, states and parameters are the same type of quantity—random variables—and are often both referred to as 'parameters' or 'unknowns'. Because states are random variables under either framework, however, the likelihood integration, Eq. 3, is inherently neither Bayesian nor classical. The difference between the roles of states and true parameters in true likelihood calculations and in asymptotic likelihood theory is discussed in the next section.

Stochastic Harvest Example.—The above framework is valid even if process and observation errors are not as distinct as in Eqs. 1–2. For example, suppose that we do not have independent survey estimates of stock size but instead wish to use catches as esti-

mates of stock size. We might do so by supposing that harvest rate is a random variable with a mean and variance that must be estimated (this scenario might be simplistic in a real management situation, but it illustrates the point here). The Ricker example would then be

$$X_t = \left(X_{t-1} - Y_{t-1}\right) e^{r\left(1-\left(X_{t-1}-Y_{t-1}\right)/K\right)+v_{t-1}} \quad \text{Eq. 7}$$

$$Y_t = h_t X_t \quad \text{Eq. 8}$$

where Y_t is now catch and h_t is a beta-distributed harvest rate. In this case the h_t values cause both observation and process error, but (X, Y), or (v, Y), are still jointly distributed random variables, and Eqs. 3 and 6 are still valid. In fact, if we have both catch and independent survey data, this information can all be combined in the same framework, where states (or process errors), catches, and independent survey estimates are all jointly distributed random variables, and the joint likelihood of catches and survey estimates still requires integration over the state or process error dimensions.

Why Do True Likelihoods Matter?

The term 'likelihood' has come to be used for two types of calculations with different theoretical properties. In much of statistics, the likelihood of a set of parameters is defined as the probability (more technically, the value of a probability density function) that the model with those parameters produced particular data. This value is what I call the 'true likelihood'. In fisheries, the term 'likelihood' is also used in an EV model-fitting framework. The EV likelihood is the probability (or value of a probability density function) of an estimated series of population states as well as the observations from those states (Ludwig and Walters, 1981; Schnute, 1994; Schnute and Richards, 1995). In a Bayesian framework states are parameters, so the EV likelihood is also the probability of the data given the parameters, but I am distinguishing it from the true liklihood because it has different theoretical properties. Millar and Meyer (2000a) call EV likelihoods 'penalized likelihoods' to distinguish them from true likelihoods. In the setting of missing-data models, Robert and Casella (1999) call the same type of quantity a 'complete-data likelihood'.

Likelihoods have a strong claim as a good foundation for statistical inference from either a classical or a Bayesian perspective (see, e.g., Stuart and Ord, 1991; Edwards, 1992; O'Hagan, 1994; Dennis et al., 1995). From a classical perspective, asymptotically, maximum-likelihood estimators are unbiased with decreasing variance (i.e., consistent), are normally distributed, and have minimum variance (Stuart and Ord 1991). 'Asymptotically' here means 'as the amount of data becomes infinite', so, on average, as the amount of data increases, consistent parameter estimates will converge to the truth, and an asymptotically normal estimator permits increasingly accurate assessments of the uncertainty of those estimates (e.g., confidence intervals). If the estimator has asymptotically minimum variance, then as the amount of data increases, one cannot find an estimation algorithm that has lower variance than maximum-likelihood estimation. These properties apply to independent-sampling models (Stuart and Ord, 1991); linear, Gaussian

time-series models (Harvey, 1989); and many nonlinear, non-Gaussian state-space models (Bickel et al., 1998; Jensen and Petersen, 1999), which include fisheries models.

Many of the desirable properties of Bayesian inference also rely on asymptotic likelihood theory. For example, when O'Hagan (1994) said that "Bayesian inference is inherently 'consistent'," he meant that, as the amount of data increases, the posterior will eventually put all weight on the true parameters (provided they had nonzero prior probability). As O'Hagan pointed out (section 3.21), it will do so for the same reason that maximum-likelihood estimators are consistent. As with classical likelihood inference, the likelihood-based asymptotic properties of Bayesian inference apply only for true likelihoods.

The parameters to which these asymptotic properties apply are actual model parameters, not unknown states or process errors. Evaluating methods based on asymptotic properties may at first seem curious, because data are always limited, but good asymptotic properties have at least three virtues. First, they imply that, as more data accumulate, the answer should, on average, get better. Second, methods with good asymptotic properties are often fairly good when only limited data are available. Third, a likelihood framework often provides the theoretical structure to estimate limited-data corrections for deviations from the asymptotic theory. Asymptotic likelihood theory forms the primary foundation for the popularity of likelihood-based inference.

Situations like fisheries model-fitting, however, pose a legitimate difficulty for reliance on likelihood asymptotics as sound justifications for using likelihood methods. If there is considerable uncertainty about the model structure itself, i.e., in how to represent fisheries biology with simple equations, then the idea that more data on average result in better parameter estimates loses meaning. In this paper, I attempt to discuss the issues but not to determine which methods are best in all circumstances.

Monte Carlo Numerical Methods

Some of the calculations involved in Bayesian model-fitting approaches (to be described below) have been implemented with Monte Carlo numerical methods. I introduce Monte Carlo methods to distinguish their roles for true likelihood integrations and Bayesian parameter inference. The method for calculating true likelihoods discussed later in this paper is not Bayesian and does not use Monte Carlo methods, but it is related to methods that do.

Monte Carlo methods involve using a sample of points from a probability distribution to approximate properties of the distribution, such as the mean, variance, or maximum. Monte Carlo analysis of state-space models can use samples of Bayesian parameter distributions, samples of state distributions, or samples of states and parameters together. Throughout this section, X could be replaced with ν, and 'state' with 'process error', and the formulas and discussion would remain valid. Gilks et al. (1996) and Robert and Casella (1999) discuss Monte Carlo statistical methods in depth.

The central calculation of Bayesian analysis, not necessarily for a state-space model, is

$$\pi(\Theta \mid Y) = \frac{L(Y \mid \Theta) P(\Theta)}{P(Y)} \qquad \text{Eq. 9}$$

where $P(\Theta)$ and $\pi(\Theta|Y)$ are the prior and posterior parameter distributions, respectively; $L(Y|\Theta)$ is the likelihood; and $P(Y) = \int L(Y|\Theta)P(\Theta)d\Theta$ is a normalizing constant that is often hard to calculate. The Monte Carlo approach to Bayesian analysis is to obtain a sample of multivariate points, $\{\Theta_1, \ldots, \Theta_m\}$, from $\pi(\Theta|Y)$. The average posterior parameter,

$$\overline{\Theta}\,|\,Y = \int \Theta\pi(\Theta\,|\,Y)d\Theta \qquad \text{Eq. 10}$$

can then be approximated as the average of the Monte Carlo sample:

$$\overline{\Theta}\,|\,Y \approx \frac{1}{m}\sum_{i=1}^{m}\Theta_i \qquad \text{Eq. 11}$$

Similarly, the variance could be approximated by the variance of the Monte Carlo sample, and the density $\pi(\Theta|Y)$ can be estimated by any method for estimating densities from samples, such as a kernel density-estimator. I do not go into methods for obtaining the sample from $\pi(\Theta|Y)$ (see Gilks et al., 1996; Robert and Casella, 1999). A variety of methods exists for doing so, the most general of which is a wide class of algorithms known as Markov chain Monte Carlo (MCMC) algorithms. A critical feature of MCMC algorithms is that they can sample from distributions that are known only up to a constant. Therefore, in terms of Eq. 9, even though the constant $P(Y)$ is unknown, one can sample from $\pi(\Theta|Y)$ if $L(Y|\Theta)P(\Theta)$ is easy to calculate for any value of Θ. The Gibbs sampler is a popular MCMC algorithm for relatively simple situations.

To use a Bayesian analysis for a state-space model, one must use likelihoods like Eq. 3 that are integrals over state (or process error) dimensions:

$$\pi(\Theta\,|\,Y) = \frac{\left[\int P(Y\,|\,X,\Theta)P(X\,|\,\Theta)dX\right]P(\Theta)}{P(Y)} \qquad \text{Eq. 12}$$

where the unknown normalizing constant is $P(Y) = \int P(Y|X,\Theta)P(X|\Theta)P(\Theta)dXd\Theta$. At first obtaining a sample from the posterior seems difficult because each calculation of $L(Y|\Theta)$ requires a difficult integral.

Fortunately this problem has a simple resolution. For a Monte Carlo sample, $\{(A_1,B_1), \ldots, (A_m,B_m)\}$, from a joint distribution, $P(A,B)$, a sample from the marginal distribution of A is just $\{A_1, \ldots, A_m\}$. In other words, to obtain a sample from

$$P(A) = \int P(A,B)dB \qquad \text{Eq. 13}$$

which is the marginal distribution of A, we can just use a sample from $P(A,B)$ and drop the B values.

To apply this principle to the above problem, we can write the Bayesian posterior as

$$\pi(\Theta\,|\,Y) = \int P(X,\Theta\,|\,Y)dX \qquad \text{Eq. 14}$$

where the relation $P(Y,X,\Theta) = P(X, \Theta|Y)P(Y)$ has been used. An MCMC algorithm can produce a sample from $P(X,\Theta|Y)$. Then if we drop the X values, we have a sample from $\pi(\Theta|Y)$. Doing so does not mean we are not interested in the X values—for fisheries we probably would be. Rather, the point is that the Θ values of a Monte Carlo sample from $P(X,\Theta|Y)$ represent a Bayesian posterior distribution using the true likelihood, i.e., integrating over unknown states. For the state distributions, a Bayesian might want to calculate the distribution of states given observations marginalized over the posterior distribution of parameters:

$$P(X \mid Y) = \int P(X \mid Y,\Theta)\pi(\Theta \mid Y)d\Theta \qquad \text{Eq. 15}$$

For a Monte Carlo approximation of this calculation, one could sample from the joint distribution of Θ and X defined by $P(X|Y,\Theta)\pi(\Theta|Y)$ and then drop the Θ values. Fortunately this procedure turns out to be the same as sampling from $P(X,\Theta|Y)$, which one would have done for Eq. 14 anyway. It is the same because

$$P(X \mid Y,\Theta)\pi(\Theta \mid Y) = P(X \mid Y,\Theta)\frac{P(Y \mid \Theta)P(\Theta)}{P(Y)} = P(X,\Theta \mid Y) \qquad \text{Eq. 16}$$

As I discuss in the next section, a number of recent fisheries analyses take this approach of Monte Carlo sampling from $P(X,\Theta|Y)$ (e.g., McAllister et al., 1994; McAllister and Ianelli, 1997; Quinn and Deriso, 1999; Bjornstad et al., 1999; Meyer and Millar, 1999; Millar and Meyer, 2000a). All of these studies explain the approach (correctly) as a Bayesian approach where the parameters are X and Θ, the likelihood is $P(Y|X,\Theta)$, and the Monte Carlo posterior is a sample representing $P(X,\Theta|Y)$. I have presented this approach in a different way to emphasize that it uses true likelihoods for the model parameters (i.e., not the unknown states), so asymptotic likelihood theory applies, and also gives a distribution of states conditioned on the posterior parameters, which may be useful for decision analysis.

Overview of Some Model-Fitting Methods

Here, I classify methods into four categories, according to (a) whether they calculate true likelihoods or calculate some other likelihood (such as an EV likelihood) and (b) whether they take a Bayesian or a classical view of parameters. I list examples of each category, but this review is not exhaustive.

Classical Treatment of Parameters without Integration Over States.—A common method that uses classical parameters and a nonintegrated 'likelihood' is the EV method. It estimates specific values of parameters and states, Θ and X, that maximize the joint probability of those particular states along with the observations. In other words, it maximizes

$$P(Y \mid X,\Theta)P(X \mid \Theta)$$

which is a particular value of the integrand of the true likelihood (Eq. 3), but not the true likelihood itself. This method played an important historical role in facing the problem of process and observation error together (Ludwig and Walters, 1981; Ludwig et al., 1988; Quinn and Deriso, 1999) and continues to be recommended (e.g., by Fargo and Richards, 1998).

Two points about EV are important. First, it requires a constraint on the relationship between variances of process and observation error, which may be difficult to estimate. Second, the asymptotics of likelihood-based inference do not necessarily apply because EV likelihoods are not the true likelihoods of statistical theory. That EV methods differ importantly from true-likelihood methods has been recognized for a long time (e.g., by Mendelssohn, 1988; Schnute, 1994; Millar and Meyer, 2000b).

Bayesian Treatment of Parameters without Integration over States.—In this approach, the Bayesian parameter analysis (Eq. 9), or a Monte Carlo representation of it, is used but with a likelihood that does not involve integration over unknown states. For example, Raftery et al. (1995) and Kinas (1996) used this approach where the population models were defined without process error. These investigators used Monte Carlo numerical tools to handle the Bayesian posterior (Eq. 9), but their models did not involve the likelihood integral (Eq. 3) or its Monte Carlo approximation. If process error was present in reality, its omission from their models would in many cases lead to biases in parameter estimates (Ludwig and Walters, 1981; Polacheck et al., 1993; McAllister et al., 1994; Schnute and Richards, 1995; de Valpine and Hastings, 2002). This type of bias occurs in the likelihood function, so it would affect either a Bayesian or a classical analysis.

Bayesian Treatment of Parameters with Integration Over States.—Neither of the above approaches uses the potential of a state-space model structure for calculating true likelihoods. More recently a few lines of work have taken this important step. McAllister et al. (1994) analyzed a model with process error from stochastic recruitment, and the dimensions for these random variables were included along with the dimensions of the parameter distributions in their Monte Carlo sampling. In the notation of the present paper, they worked with $P(\nu,Y)$, not $P(X,Y)$, where ν are random variables for recruitment. They obtained a Monte Carlo sample representing (ν,Θ) given Y, which is like working with (Eq. 14) but with the probability densities transformed to be in terms of ν instead of X. McAllister and Ianelli (1997) further developed and applied this approach. Meyer and Millar (1999) and Millar and Meyer (2000a) also studied state-space models with a Bayesian treatment of parameters. They worked with $P(X,Y)$, not $P(\nu,Y)$, and used a Gibbs sampler to obtain Monte Carlo samples from (X,Θ) given Y. See also Millar and Meyer (2000b), Bjornstad et al. (1999), Quinn and Deriso (1999, example 5.4), and Punt and Hilborn (1997).

Analyses That Use Non-Bayesian Treatment of Parameters but Still Integrate over States.—Finally, numerical methods can be used to deal with the state-space integral (Eq. 3) while parameters are treated in a classical framework. One such method is discussed in the next section. I am not aware of examples of this approach in the fisheries literature, other than with the KF for linear models or linear approximations of nonlinear models. For discussions of Monte Carlo methods in classical inference for missing-data models (including state-space models), see Geyer and Thompson (1992), Geyer (1996), McCulloch (1997), Booth and Hobert (1999), and Robert and Casella (1999).

SUMMARY OF OVERVIEW.—The combined effects of process and observation error can be described with a state-space model structure. Calculating true likelihoods for state-space models involves a difficult integration problem. Some approaches avoid the true-likelihood calculation either by using an EV likelihood, in which specific states are estimated along with parameters, or by ignoring process or observation error. These approaches have been used both with and without Bayesian treatment of parameters. In contrast, true-likelihood approaches attempt to incorporate process error and observation error together by means of the likelihood integration (Eq. 3). All such analyses of which I am aware for nonlinear models in the fisheries literature use a Bayesian view of parameters, in which case a Monte Carlo Bayesian parameter analysis and a Monte Carlo state-space integration can be naturally combined by means of a joint Monte Carlo sample of states and parameters. I am not aware of approaches in the fisheries literature that calculate the true-likelihood integral (Eq. 3) for nonlinear, non-Gaussian models with a classical treatment of parameters. I present one here.

THE DIRECT APPROACH TO STATE-SPACE CALCULATIONS

Here I describe what might be called a 'direct' method for making state-space model integrations, to distinguish it from Monte Carlo methods. At first glance, the integral (Eq. 3) seems difficult to implement numerically. For example, for a one-dimensional state variable (e.g., population size) and a time series of length 20, say, each likelihood requires a 20-dimensional integral. Numerically calculating a 20-dimensional integral by a grid method would be impractical because a 20-dimensional grid is enormous—it suffers from the "curse of dimensionality" (Robert and Casella, 1999). This is the main reason that Monte Carlo numerical methods have become so popular: they are more efficient than grid methods for high-dimensional problems.

Many state-space likelihood integrals contain more structure, however, than is evident when they are viewed, to use the above example, simply as 20-dimensional integrals. In particular, the Markov property—that each state depends only on the previous state—allows us to write the integral in simpler form. Instead of a 20-dimensional integral, we can obtain 20 1-dimensional integrals and 20 2-dimensional integrals (actually not 2-dimensional integrals, but their computational equivalent), and in some cases the 2-dimensional integrals permit fast numerical tricks. The more general case of an n-dimensional state space at each time step (e.g., for 3 stage classes, n = 3) for T time steps involves T n-dimensional integrals and T 2n-dimensional integrals (or the close equivalent). The result is that, for fisheries models with low-dimensional state spaces at each time step, a good alternative to Monte Carlo methods is available.

The approach is to use the Markov structure of Eq. 3 to split the full integral into sequential, lower-dimensional integrals for each time step and to approximate the probability density functions very closely by splitting up the possible state and observation values very finely. Once the state and observation domains are split into small bits, the integrals and other calculations can be closely approximated. This is essentially a grid method, similar to Gaussian quadrature, for the lower-dimensional integrals. The approach was introduced by Kitagawa (1987) and explored for Ricker and Beverton-Holt models by de Valpine and Hastings (2002). Here I summarize the method and a few results of de Valpine and Hastings (2002). To describe the dimensions involved, I will suppose that each X_t value has n dimensions.

The likelihood equation (Eq. 3) can be factored into

$$L(\Theta) = P(Y_1)\prod_{i=2}^{i=T} P\left(Y_i \mid Y^{(i-1)}\right) \qquad \text{Eq. 17}$$

A parenthetical superscript indicates an entire sequence up to and including the time in the superscript (cf. Schnute, 1994). For example, $Y^{(i-1)} = (Y_1, Y_2, \ldots, Y_{i-1})$. Each probability is also a function of θ, but that is not written explicitly.

The factors in the product (Eq. 17) can be calculated iteratively as follows. Suppose we have the distribution of states conditioned on all observations up to and including time t, which will be written as $P(X_t|Y^{(t)})$. Then the projected distribution of states at time $t+1$ is

$$P\left(X_{t+1} \mid Y^{(t)}\right) = \int P\left(X_{t+1} \mid X_t\right) P\left(X_t \mid Y^{(t)}\right) dX_t \qquad \text{Eq. 18}$$

Because this n-dimensional integral must be evaluated for each value of X_{t+1}, it is computationally similar to a 2n-dimensional integral. The projected distribution $P(X_{t+1}|Y^{(t)})$ can be used to calculate the probability of the $(t+1)$th observation, which is the $(t+1)$th factor in the likelihood (Eq. 17):

$$P\left(Y_{t+1} \mid Y^{(t)}\right) = \int P\left(Y_{t+1} \mid X_{t+1}\right) P\left(X_{t+1} \mid Y^{(t)}\right) dX_{t+1} \qquad \text{Eq. 19}$$

Because this probability must be calculated only for the value of Y_{t+1} that was actually observed, it is an n-dimensional integral. So far, for this time step we have used an n-dimensional integral (Eq. 19) and something similar to a 2n-dimensional integral (Eq. 18). Finally the distribution of states at time $t+1$ can be updated to account for the information in the $(t+1)$th observation by the conditional probability or 'filtering' equation:

$$P\left(X_{t+1} \mid Y^{(t+1)}\right) = \frac{P\left(Y_{t+1} \mid X_{t+1}\right) P\left(X_{t+1} \mid Y^{(t)}\right)}{P\left(Y_{t+1} \mid Y^{(t)}\right)} \qquad \text{Eq. 20}$$

This process does not require much computational trouble because all of the numerical pieces for the numerator will have been produced for the integral (Eq. 19). (Although Eq. 20 is a statement of Bayes' law, its use here is separate from the issue of Bayesian or classical treatment of parameters because states are random variables in both cases.)

The distribution $P(X_{t+1}|Y^{(t+1)})$ can be used in the right hand side of Eq. 18 to calculate $P(X_{t+2}|Y^{(t+1)})$ and so on. It remains to specify the starting distribution of states before any observations have been made, $P(X_1)$. De Valpine and Hastings (2002) used the stationary distribution of the stochastic model for state dynamics as an objective choice for $P(X_1)$.

If n is small, the density functions of X and Y can be approximated by means of a fine grid for the range of values of X. De Valpine and Hastings (2002) used this approach to compare the state-space method with the least-squares methods that ignore either process noise or observation error. The numerical integrations can be implemented in efficient

ways; full implementation details are given by de Valpine and Hastings (2002) and are not reproduced here. De Valpine and Hastings (2002) referred to this approach as the 'numerically integrated state-space approach' (NISS) and referred to the least-squares approaches as 'least squares with process noise only' (LSPN) and 'least squares with observation error only' (LSOE). ('Noise' and 'error' are interchangeable in this context.) The above formulas (Eqs. 17–20) are just the nonlinear, non-Gaussian extension of the mathematics behind the Kalman filter.

A few of the results from de Valpine and Hastings (2002) are reproduced in Figure 1. They used a stochastic Ricker model with lognormal errors:

$$X_{t+1} = X_t e^{r\left(1-\frac{X_t}{K}\right)} e^{v_t}$$

and a stochastic Beverton-Holt model with lognormal errors:

$$X_{t+1} = X_t \frac{\lambda}{1+\gamma X_t} e^{v_t}$$

where in both cases v_t is a normally distributed process noise. The numerically integrated state-space approach can include any form of noise distribution, but lognormal noises facilitated comparison with least-squares methods that ignore either process noise or observation error. For the Ricker model, they used $r = 1.5$, 2.4, or 2.6 and $K = 100$. These values of r produce a stable equilibrium, a stable two-cycle, and a stable four-cycle, respectively, in the deterministic model. For the Beverton-Holt model, they used $\lambda = 4.48$ and $\gamma = 0.0348$, which have the same equilibrium and growth rate at small population sizes as the $r = 1.5$, $K = 100$ Ricker model. For an observation model, de Valpine and Hastings (2002) used

$$Y_t = X_t e^{\varepsilon_t}$$

where ε_t is normally distributed, again to facilitate comparison with least squares methods. In the results reproduced here, the standard deviations of v_t and ε_t are 0.2. No assumptions of variance ratios were made for maximum-likelihood parameter estimation. Distributions of maximum-likelihood estimates from 300 time series of length 20 show that the numerically integrated state-space method produced lower-bias, often lower-variance estimators than least-squares methods that ignore either process noise or observation error (Fig. 1). More results were presented by de Valpine and Hastings (2002). Among other questions, they asked whether the distribution of differences of the maximum log-likelihood from the log-likelihood of the true parameters was close to the theoretical distribution (chi-squared for –2 times the log-likelihood ratio) to which it converges asymptotically (i.e., as the amount of data increases). If convergence is good, then asymptotic distributions provide good approximations for assessing uncertainty in parameter estimates, i.e., estimating confidence intervals. They found fairly good convergence for the Ricker model and less good convergence for the Beverton-Holt model.

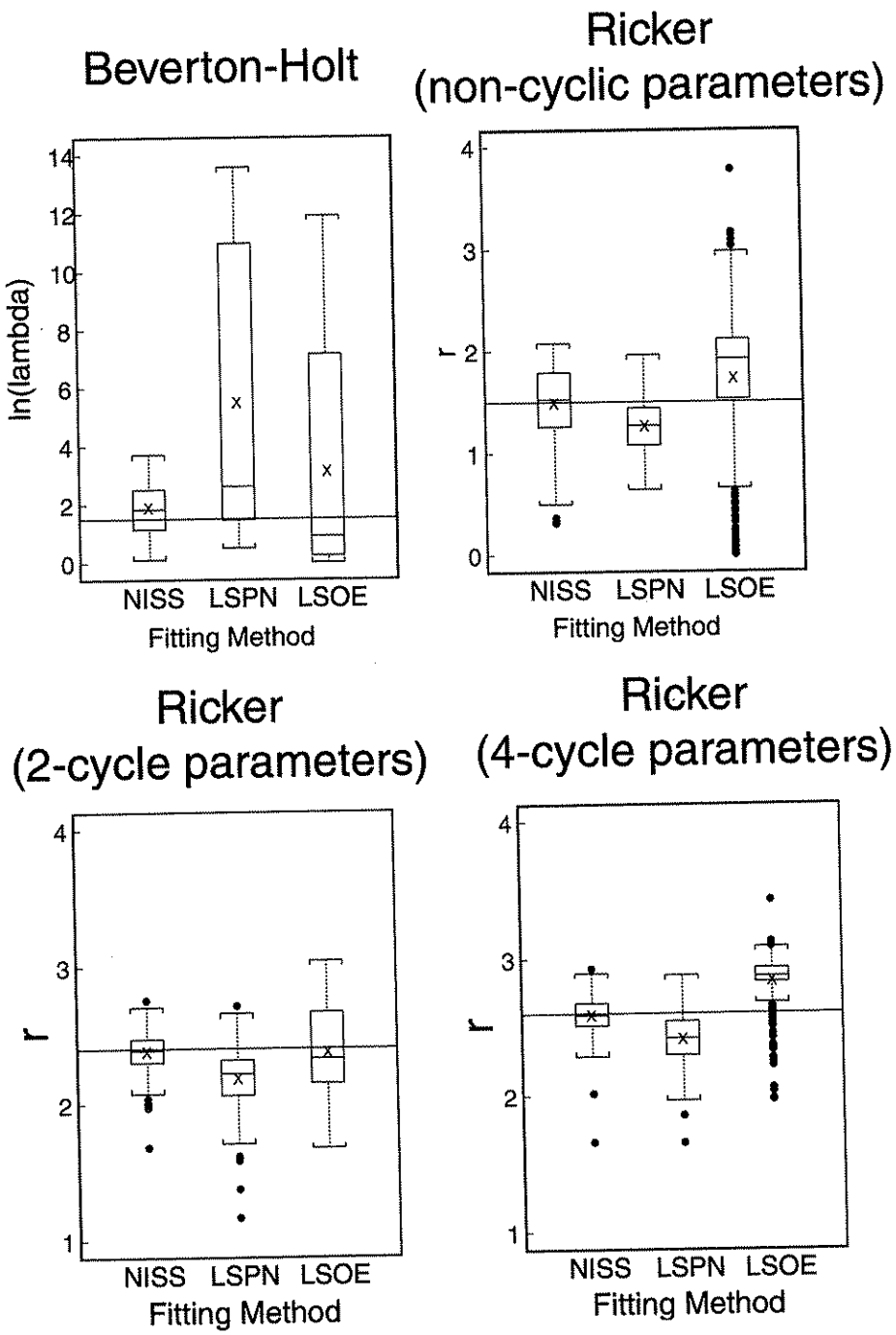

Figure 1. (A) Distribution of maximum-likelihood parameter estimates for Beverton-Holt and Ricker ($r = 1.5$) models for 300 time series of length 20. (B) Distribution of maximum-likelihood parameter estimates for Ricker models with $r = 2.4$ and $r = 2.6$. Adapted from de Valpine and Hastings (2002) with permission. Fitting methods are 'numerically integrated state-space' (NISS), 'least-squares with process noise only' (LSPN), and 'least-squares with observation error only' (LSOE).

Conclusion

Recent years have seen great progress in methods for incorporating a wide range of model and error structures fully into model-fitting methods using true likelihoods. The examples of de Valpine and Hastings (2002) show that a non-Monte Carlo method for numerically integrating likelihoods is feasible when the state variable at each time step has low dimension. This type of calculation could be used for either classical or Bayesian parameter analysis. In simulated examples, maximum-likelihood estimators performed better than estimators that ignored process or observation error.

Considering robustness of model-fitting methods raises two difficult questions. First, how should one handle uncertainty in model structure itself (see, e.g., Chatfield, 1995)? The problem of model structure is in some ways outside of the boundaries in which much statistical theory has been developed. For example, likelihood theory takes a well-defined model structure as its starting point. Indeed, Ludwig and Walters (1989) and Ludwig et al. (1988) framed the EV approach not in terms of likelihood asymptotics but in terms of the need for useful estimation properties even with incorrect models.

Second, given that fisheries systems are inherently very noisy and difficult to model, is it reasonable to expect decent inferences and assessments of uncertainty from time-series data? Even under simple model structures, noisy data naturally yield variable and uncertain parameter estimates. Advances in model-fitting methods may provide a more exact picture of how very uncertain our inferences will remain. For this reason, technical statistical advances—although an important component of fisheries science—are unlikely to solve fisheries management problems (Ludwig et al., 1993).

Acknowledgments

I thank the Mote conference for the opportunity to write this paper, two anonymous reviewers for helpful suggestions, and A. B. Thistle for editing. One reviewer in particular offered numerous questions that led me to expand the review section.

Literature Cited

Bickel, P. J., Y. Ritov and T. Rydèn. 1998. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. Annals Stat. 26: 1614–1635.

Bjornstad, O. N., J.-M. Fromentin, N. C. Stenseth and J. Gjosaeter. 1999. Cycles and trends in cod populations. Proc. Natl. Acad. Sci. 96: 5066–5071.

Booth, J. G. and J. P. Hobert. 1999. Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm. J. R. Stat. Soc. B 61: 265–285.

Chatfield, C. 1995. Model uncertainty, data mining and statistical inference. J. R. Stat. Soc. A 158: 419–466.

Collie, J. S. and M. P. Sissenwine. 1983. Estimating population size from relative abundance data measured with error. Can. J. Fish. Aquat. Sci. 40: 1871–1879.

de Valpine, P. and A. Hastings. 2002. Fitting population models with process noise and observation error. Ecol. Monogr. 72: 57–76.

Dennis, B., R. A. Desharnais, J. M. Cushing and R. Costantino. 1995. Non-linear demographic dynamics: mathematical models, statistical methods, and biological experiments. Ecol. Monogr. 65: 261–281.

Edwards, A. W. F. 1992. Likelihood. Expanded edition, The Johns Hopkins Univ. Press, Baltimore, Maryland. 275 p.

Fargo, J. and L. J. Richards. 1998. A modern approach to catch-age analysis for Hecate Strait rock sole (*Pleuronectes bilineatus*). J. Sea Res. 39: 57–67.

Freeman, S. and G. Kirkwood. 1995. On a structural time series method for estimating stock biomass and recruitment from catch and effort data. Fish. Res. (Amst.) 22: 77–98.

Geyer, C. J. 1996. Estimation and optimization of functions. Pages 241–258 *in* W. R. Gilks, S. Richardson and D. J. Spiegelhalter, eds. Markov chain Monte Carlo in practice. Chapman & Hall, New York, New York.

_____ and E. A. Thompson. 1992. Constrained Monte Carlo maximum likelihood for dependent data. J. R. Stat. Soc. B 54: 657–699.

Gilks, W. R., S. Richardson and D. J. Spiegelhalter, eds. 1996. Markov chain Monte Carlo in practice. Chapman & Hall, New York, New York. 486 p.

Gudmundsson, G. 1994. Time series analysis of catch-at-age observations. Appl. Stat. 43: 117–126.

Harvey, A. C. 1989. Forecasting, structural time series models, and the Kalman filter. Cambridge Univ. Press, Cambridge, U.K. 554 p.

_____. 1993. Time series models. 2nd ed., MIT Press, Cambridge, Massachusetts. 308 p.

Hilborn, R. 1979. Comparison of fisheries control systems that utilize catch and effort data. J. Fish. Res. Bd. Can. 36: 1477–1489.

Jensen, J. L. and N. V. Petersen. 1999. Asymptotic normality of the maximum likelihood estimator in state space models. Annals Stat. 27: 514–535.

Kimura, D. K., J. W. Balsiger and D. H. Ito. 1996. Kalman filtering the delay-difference equation: practical approaches and simulations. Fish. Bull., U.S. 94: 678–691.

Kinas, P. G. 1996. Bayesian fishery stock assessment and decision making using adaptive importance sampling. Can. J. Fish. Aquat. Sci. 53: 414–423.

Kitagawa, G. 1987. Non-Gaussian state-space modeling of nonstationary time series (with discussion). J. Amer. Stat. Assoc. 82: 1032–1063.

Ludwig, D. and C. J. Walters. 1981. Measurement errors and uncertainty in parameter estimates for stock and recruitment. Can. J. Fish. Aquat. Sci. 38: 711–720.

_____ and _____. 1989. A robust method for parameter estimation from catch and effort data. Can. J. Fish. Aquat. Sci. 46: 137–144.

_____, _____ and J. Cooke. 1988. Comparison of two models and two estimation methods for catch and effort data. Nat. Resour. Model. 2: 457–498.

_____, R. Hilborn and C. Walters. 1993. Uncertainty, resource exploitation, and conservation: lessons from history. Science 260: 17–18.

McAllister, M. K. and J. N. Ianelli. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54: 284–300.

_____, E. K. Pikitch, A. E. Punt and R. Hilborn. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 51: 2673–2687.

McCulloch, C. E. 1997. Maximum likelihood algorithms for generalized linear models. J. Amer. Stat. Assoc. 92: 162–170.

Mendelssohn, R. 1988. Some problems in estimating population sizes from catch-at-age data. Fish. Bull., U.S. 86:617–630.

Meyer, R. and R. B. Millar. 1999. Bayesian stock assessment using a state-space implementation of the delay difference model. Can. J. Fish. Aquat. Sci. 56: 37–52.

Millar, R. B. and R. Meyer. 2000a. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. Can. J. Fish. Aquat. Sci. 57: 43–50.

_____ and _____. 2000b. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. Appl. Stat. 49: 327–342.

O'Hagan, A. 1994. Kendall's advanced theory of statistics, vol. 2B, Bayesian inference. Arnold, London, U.K. 330 p.

Pella, J. J. 1993. Utility of structural time series models and the Kalman Filter for predicting consequences of fishery actions. Pages 571–593 *in* G. Kruse, D. Eggers, R. Marasco, C. Pautzke and T. Q. Quinn II, eds. Proc. Int'l. Symp. Management Strategies for Exploited Fish Population. Alaska Sea Grant College Program Rpt. no. 93-02. Alaska Sea Grant College Program, Univ. Alaska, Fairbanks, Alaska.

Polacheck, T., R. Hilborn and A. E. Punt. 1993. Fitting surplus production models: comparing methods and measuring uncertainty. Can. J. Fish. Aquat. Sci. 50: 2597–2607.

Punt, A. E. and R. Hilborn. 1997. Fisheries stock assessment and decision analysis: the Bayesian approach. Rev. Fish Biol. Fish. 7: 35–63.

Quinn, T. J., II and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford Univ. Press, New York, New York. 542 p.

Raftery, A. E., G. H. Givens and J. E. Zeh. 1995. Inference from a deterministic population dynamics model for bowhead whales. J. Amer. Stat. Assoc. 90: 402–416.

Rice, J. A. 1988. Mathematical statistics and data analysis. Wadsworth & Brooks/Cole, Pacific Grove, California.

Richards, L. J., J. T. Schnute and N. Olsen. 1997. Visualizing catch-age analysis: a case study. Can. J. Fish. Aquat. Sci. 54: 1646–1658.

Robert, C. P. and G. Casella. 1999. Monte Carlo statistical methods. Springer, New York, New York. 507 p.

Schnute, J. T. 1994. A general framework for developing sequential fisheries models. Can. J. Fish. Aquat. Sci. 51: 1676–1688.

_____ and L. J. Richards. 1995. The influence of error on population estimates from catch-age models. Can. J. Fish. Aquat. Sci. 52: 2063–2077.

Stuart, A. and J. K. Ord. 1991. Kendall's advanced theory of statistics, vol. 2, Classical inference and relationship, 5th ed., Oxford Univ. Press, New York, New York.

_____ and _____. 1994. Kendall's advanced theory of statistics, vol. 1, Distribution theory, 6th ed., Edward Arnold, London, U.K. 676 p.

Sullivan, P. J. 1992. A Kalman filter approach to catch-at-length analysis. Biometrics 48: 237–257.

Walters, C. J. and D. Ludwig. 1981. Effect of measurement errors on the assessment of stock-recruit relationships. Can. J. Fish. Aquat. Sci. 38: 704–710.

ADDRESS: *National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101. E-mail: <devalpin@nceas.ucsb.edu>.*

APPENDIX

To consider the change of coordinates $(X,Y) \rightarrow (v,Y)$ and to determine whether the likelihood integral stays the same, denote this change of variables by $v = g_1(X,Y)$, $Y = g_2(X,Y)$. Of course g_2—the identity function—seems a bit silly, but it is useful to write it this way to stay general. Denote the inverse transformation $X = h_1(v,Y)$, $Y = h_2(v,Y)$.

We now use a subscript notation on probability densities to indicate which coordinates the density uses. For example $f_{(X,Y)}(a,b)$ is the probability density in (X,Y) coordinates of $(X = a,\ Y = b)$. The relationship between the probability density functions of (X,Y) and (v,Y) is

$$f_{(v,Y)}(v,Y) = f_{(X,Y)}\big(X = h_1(v,Y), Y = h_2(v,Y)\big)\,|J_h| \qquad \text{Eq. 21}$$

where $|J_h|$ is the determinant of the Jacobian of the inverse transformation, $(v,Y) \rightarrow (X,Y)$. In this case, $|J_h| = |\delta h_1/\delta v|$. Derivation of this rule for transforming probability distribu-

tions can be found in many probability texts, such as Rice (1988) and Stuart and Ord (1994).

Now write the likelihood integral using the (v, Y) coordinates:

$$\int f_{(v,Y)}(v,Y)dv = \int f_{(X,Y)}\big(X = h_1(v,Y), Y = h_2(v,Y)\big) \mid J_h \mid dv$$

If we change coordinates for integration from v to X, we will need to use $dX = dv|J_h|$, so the integral is just $\int f_{(X,Y)}(X,Y)dX$, or the likelihood integrated in the (X, Y) coordinates. In effect we have transformed into (v, Y) coordinates and then back out for integration, with no side effects. Actually, this is the reason that Eq. 21 is true: the cumulative distributions (i.e., integrals) over any subset of outcomes (whether in v or X) must match. The backward presentation here emphasizes that we can switch between v and X without changing the likelihood.